BAYESIAN APPROACHES FOR MAPPING FOREST SOIL ORGANIC CARBON: ADDRESSING SPATIAL AND MODEL UNCERTAINTY

By

BRIAN J. CLOUGH

A dissertation submitted to the

Graduate School- New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Ecology and Evolution

Written under the direction of

Dr. Edwin J. Green

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2014

ABSTRACT OF THE DISSERTATION

Bayesian Approaches for Mapping Forest Soil Organic Carbon: Addressing Spatial Uncertainty

by BRIAN J. CLOUGH

Dissertation Director:

Edwin J. Green

Forest soil organic carbon (SOC) is the largest terrestrial pool of carbon, and its management plays a significant role in global efforts to mitigate atmospheric carbon concentrations. Despite its importance, much of the world is still lacking good baseline data of forest soil carbon stocks. In the past, broad scale stocks of forest SOC have been derived from soil surveys based on a small number of sampling units, and the resulting estimates are highly uncertain. More recently, predictive statistical models have received attention as an approach for mapping soil carbon at scales relevant to climate change policy and research. However, in order for these models to be useful they must provide full and accurate accounting of uncertainty, in addition to accurate predictions.

This dissertation aims to improve prediction of forest SOC by incorporating two potentially important sources of uncertainty into the modeling process: (1) spatial dependence in soil inventory data; and (2) error associated with assuming a single model to be "true". In order to address these issues, we turn to well established techniques in the Bayesian statistics literature. Our primary focus is on exploring the application of spatial Bayesian hierarchical regression models for improving estimates of forest carbon. This line of research involves both characterizing the spatial dependence in forest SOC

inventories at regional, national, and continental scales (the focus of chapters 1 and 3), and applying spatial hierarchical models for mapping SOC and validating this method against non-spatial approaches (chapter 4). Additionally, in chapter 2 we compare methods for model selection and weighting, as well as the effect of model averaging to account for model uncertainty, through rigorous predictive checks. This work is conducted with both forest SOC data as well as other ecological datasets.

Taken together, these studies highlight the need for a consistent statistical framework in order to generate reproducible estimates of forest SOC stocks across the globe. Our results argue for hierarchical models, and especially spatial hierarchical models, as a reasonable way forward for predictive mapping of SOC. However, they also highlight significant methodological development that will be necessary in order to obtain predictively accurate models.

Acknowledgements

I wish to thank my advisor, Dr. Ed Green and my dissertation committee; Dr. Richard Lathrop, Dr. Peter Smouse, and Dr. Andrew Finley. Additionally, I would also like to thank Dr. Harold Burkhart and Dr. Peter Morin for furnishing some of the datasets used in this dissertation. I would also like to acknowledge my family, and particularly my wife Amy Manning, for their everlasting support as I pursued my degree. Lastly, special thanks are owed to my first dissertation advisor, the late Dr. Joan Ehrenfeld. Dr. Ehrenfeld has been a major influence on me, and without her tutelage I would not be where I am today.

Table of contents

List of Tables

List of figures

Introduction

Globally, the forestry sector accounts for approximately 17 percent of annual greenhouse gas emissions, primarily through deforestation and forest degradation in tropical nations (Houghton 2003, Le Quere et al. 2009, Edwards et al. 2010). In 2010, the United Nations' "Reducing Emissions from Deforestation and Forest Degradation in Developing Countries" (REDD+) program estimated that an international investment of approximately $22-38 billion could lead to a 25 percent reduction in forest sector emissions by 2015 (Angelsen 2009, Corbera and Schroeder 2011). Since then, approximately $2.72 billion has been pledged by more than ten nations to support activities that meet the goals outlined by the REDD+ strategy. While this is far off from the target to achieve a 25 percent reduction, it is clear that forest sector emissions represents a 'low hanging fruit' in global efforts to mitigate the effects of climate change, and the value of these efforts is widely recognized by policymakers in developed nations.

The establishment of a Monitoring, Verification, and Reporting (MVR) system for assessing forest C stocks has been identified as a core work area of the REDD+ program, and this has resulted in significant methodological development by the forest science community (Gibbs et al. 2007, Asner 2011, Herold and Skutsch 2011). Much of this is focused on the design of optimal sampling networks for newly established national forest inventories (Maniatis and Mollicone 2010, Kohl et al. 2011), and on integrating these data with suitable remote sensing datasets capable of supplying 'wall to wall' coverage for developing regional forest C baseline estimates (Achard et al. 2007, Goetz and Dubayah 2011, De Sy et al. 2012). In the latter case, this work is primarily conducted on existing national forest inventories in developed nations, such as the US Forest

Service's Forest Inventory and Analysis (FIA) database (Birdsey et al. 2013). In general forest scientists agree on a 'three tiered' approach, where fine-scale measurements (e.g. ecological monitoring stations) taken at few locations may be related to inventory datasets with many observations covering large spatial ranges, as well as related remote sensing data (Hoover 2008).

Public availability of these datasets, coupled with increasing computing power and methodological developments in computational statistics, have motivated many studies aimed at developing forest carbon stocks via model-based prediction (Potter 1999, Purves and Pacala 2008, Saatchi et al. 2011). Generally this involves identifying covariates from remote sensing datasets, establishing relationships between these and ground measured observations of forest carbon stocks or fluxes from inventory data, and then employing guided interpolation across a sampling grid. A variety of methodological approaches have been put to this purpose within the literature, including multiple regression (Brown 1997, Chave et al. 2004), "data model fusion" techniques (Luo et al. 2011), and random forest or regression tree models (Grimm et al. 2008, Powell et al. 2010).

Thus far, most of these efforts have been aimed at aboveground forest carbon stocks and fluxes. This is natural, given that traditionally forest inventories have been designed to sample trees. National forest biomass equations such as those developed by Jenkins et al. (2003) allow for the estimation of aboveground forest carbon stocks from typical forest inventory measurements. Further, there is a significant body of literature relating aboveground biomass to a variety of remote sensing datasets, including spectral data from satellite platforms such as Landsat or MODIS, and light detection and ranging

(LiDAR) surveys (Dubayah and Drake 2000, Lu 2006, Goetz and Dubayah 2011). These covariates, along with climatic, topographic, and other datasets, have been incorporated into predictive models for aboveground carbon storage at a variety of spatial scales.

While there has been significant development in quantifying aboveground carbon, the forest soil carbon pool remains poorly understood. This is in spite of the fact that forest soil carbon is the largest global terrestrial carbon pool (Goodale et al. 2002, Pan et al. 2011), and that soil carbon sequestration has been cited as an important strategy for offsetting atmospheric C emissions (Heath et al. 2003, Lal 2005). This is particularly true in tropical regions, the principle targets of the REDD+ program, where reforestation and forest restoration may lead to increased rates of soil carbon accretion in degraded and/or desertified soils (Post and Kwon 2000, Lal 2004). Given the long residence time of carbon stored in soils (Chapin et al. 2002), soil C sequestration seems like a vital component of any policy or management plan looking to engineer permanent increases in terrestrial carbon stocks.

In order for these gains to be verifiable, good baseline estimates, along with reasonable estimates of uncertainty, are necessary. Global estimates of soil C storage have traditionally been the domain of soil survey, such as the USDA's national soils database (Bredja et al. 2001). However, these are at a coarse resolution, derived by assigning average values of soil carbon from a small number of samples to taxonomic map units (McBratney et al. 2003). As a result, significant sources of error go ignored and the resulting estimates are highly uncertain. Further, since these stocks are derived using deterministic 'map algebra' approaches (i.e. assigning average values to taxonomic

map units and summing these to gain regional C stocks), this uncertainty is difficult to quantify.

These shortcomings of soil survey are widely recognized within the soil science community, and have given rise to a large literature on 'digital soil mapping' (DSM) (McBratney et al. 2003, Grunwald 2009). As with predictive models for aboveground carbon, DSM involves established statistical relationships between soil attributes and remote sensing covariates in order to carry out 'rasterized' prediction across the study region. The earliest literature on guided prediction of soil attributes focused on terrain analysis (Moore et al. 1991, Odeh et al. 1994, Mueller and Pierce 2001), but many data sources have been used since, including: land cover classifications (Tóth et al. 2013), satellite remote sensing data (McGrath and Zhang 2003, Simbahan et al. 2006, Wang et al. 2013), and other soil attributes (Guhaniyogi et al. 2013).

Not surprisingly soil carbon has been a popular focus of digital soil mapping studies (Minasny et al. 2013), and several recent efforts have been aimed at mapping soil carbon across broad spatial scales. Much of this work has focused on agricultural landscapes. For example, Tóth et al. (2013) used multiple regression to map soil properties for croplands within the member states of the European Union, using the recently published LUCAS topsoil database. Similar studies are available for karst landscapes in China (Liu et al. 2006), as well as grassland ecosystems in Ireland (McGrath and Zhang 2003). A variety of studies have also sought to map soil C across mixed-used landscapes (Bernoux et al. 2006, Minasny et al. 2006, Malone et al. 2009, Mishra et al. 2010, Vasques et al. 2010a).

There are few examples of regional DSM for forest soil carbon. Until recently, this has primarily been due to a lack of data. Surveys conducted by soil scientists and agronomists are typically biased towards croplands, and the additional expense of incorporating soil sampling into forest survey led it to be excluded from early national forest inventories, where the primary aim was to estimate standing stock of timber resources rather than carbon. The exception is the National Forest Inventory of Finland, which began monitoring forest soil C in 1985 (Liski and Westman 1997). The Forest Inventory and Analysis (FIA) program of the United States Forest Service included soil attributes among the "forest health measurements" taken as part of their phase 3 program beginning in 2000 (Heath et al. 2003). More recently, the LUCAS topsoil database sought to establish a large network of point-referenced samples of soil attributes across all landcover types within the European Union, including forests (Tóth et al. 2013b).

The availability of these and similar datasets provides an excellent opportunity to develop predictive models for broad scale modeling of forest soil carbon distribution. Such an effort not only supplies important baseline information for the countries already possessing forest soil C inventory data, but also contributes positively to the goals of the REDD+ program. In order for participating nations to fully monitor changes in forest C stocks, an approach for establishing baseline estimates of forest soil C will be a necessary component of the MVR system. Developing reasonable protocols and predictive models for generating these estimates, based on existing data, represents a significant contribution to the monitoring of forest carbon pools.

In designing predictive models for forest soil carbon, the primary goals should be two-fold. First, models should possess mean structures that are capable of generating

accurate predictions anywhere within the study region. Achieving this goal requires the careful selection of model covariates, as well as considerations for the relationships between these variables and soil C. A large literature on designing functions for DSM already exists (*see* McBratney et al. 2003 *and* Minasny et al. 2013 *for thorough reviews*). While much of this literature focuses on agricultural and mixed-use landscapes, it provides a reasonable basis for the design of predictive functions for forest soil C as well.

The second goal is a model that provides reasonable estimates of uncertainty surrounding predictive estimates. This is an equally important quality in models, yet a large portion of DSM mapping studies make no attempt to formally quantify prediction error (Grunwald 2009). If estimates of forest soil C derived from predictive models are to be used as a basis for policy making and management decisions, or if they are to be fed into global climate models, an accurate estimate of uncertainty is necessary (Huntingford et al. 2009, Shvidenko et al. 2010). Failing to provide such an estimate, or providing inaccurate error bounds that suggest a higher degree of precision than the model actually provides, may have serious implications.

Fully quantifying uncertainty in predictive models of forest C means more than providing error bars around aggregate stock estimates. If our goal is to design the 'best' predictive model possible, it may be useful to understand where and how uncertainty arises. Error in model predictions may come from multiple sources, but can generally be grouped into three categories: (1) error associated with data collection and sample analysis; (2) "model error" stemming from parameter estimation; and (3) error arising from spatial and temporal variability underlying the data (Banerjee et al. 2004, Finley et al. 2010). Each of these features requires specific consideration, but one or more is

frequently ignored in predictive modeling studies. This is in large part due to the limitations of the methodological techniques traditionally used to fit predictive models for DSM and forest C mapping, though advances in modern computational statistics offer some appealing solutions.

Uncertainty arising from data measurement error and parameter estimation has been addressed in the forest carbon literature. Forest inventories usually include rigorous quality assurance quality control (QAQC) procedures to help minimize error within the data themselves, and widely used analytical techniques such as linear regression attempt to capture this variation with "white noise" error terms (Miehle et al. 2006). Over the last decade, forest scientists have also come to recognize the importance of capturing uncertainty in estimates of model parameters (Radtke et al. 2002, Williams et al. 2005). This has contributed to the prevalence of Bayesian analytical techniques in the forest sciences, including hierarchical models (Green and Valentine 1998, Cressie et al. 2012), Bayesian synthesis (Green et al. 1999, 2000, MacFarlane et al. 2000), and so-called "data model fusion" techniques (Keenan et al. 2011, 2012, Luo et al. 2011).

Other sources of uncertainty have only received attention more recently. Model error associated with selecting a single best fitted model and assuming it to be "true" has been cited in the statistics literature as an important consideration in model selection problems (Raftery et al. 1997, Wintle et al. 2003) . This has given rise to a large literature on model averaging, either based on weighting information criteria used for model selection such as Akaike's information criterion (AIC) or the deviance information criterion (DIC) (Hoeting et al. 1999, Burnham and Anderson 2002), or through Bayesian measures of model fit such as posterior model probability (Key et al. 1999, Clyde et al.

2011). In fact, given the large variation in typical natural resources datasets, some authors recommend model averaging is adopted as standard practice (Johnson and Omland 2004, Burnham et al. 2011). Forest carbon models are no exception. Recently, Green and Bollock (2013) observed that ignoring model error in a biomass model for Loblolly Pine resulted in credible intervals that underestimated total variance by 10-40 percent.

Failure to explicitly model spatial processes is another key source of error in predictive models of forest SOC. Modeling spatial dependence in soil inventory data is not a recent development. In fact, applied geostatistics has roots in geology and natural resource science, and soil scientists and agronomists were early adopters of spatial interpolation methods such as kriging, primarily for within field soil nutrition mapping (Odeh et al. 1994, McBratney et al. 2003, Hengl et al. 2004). While kriging methods represent a step forward in that they explicitly model spatial dependence via spatial covariance functions (Isaaks and Srivastava 1989), they do not allow for uncertainty in the parameters of these functions. This can be problematic since spatial covariance models are derived by fitting an empirical semivariogram, and significant residual error in the spatial pattern is ignored in this process (Banerjee et al. 2004). Theoretical variogram models rarely fit the empirical variogram very well, and in fact "cleaning up" the variogram by removing outliers is often recommended as standard practice (Goovaerts 1997). The very large spatial datasets that have become commonplace in forestry (i.e., large forest inventories and remote sensing datasets) demand a more flexible approach.

In the Bayesian paradigm, hierarchical regression models that include spatial random effects provide the kind of versatile method necessary for addressing macroscale

issues in forest science. These models describe spatial dependence as a stochastic Gaussian process (Gelfand 2012), introduced as a "spatial random effects" term in a regression model, where our model of interest is described as:

$$y(s) = \mu(s) + w(s) + \varepsilon(s)$$

Where $y(s)$ are the observed data at locations s, $\mu(s)$ is some mean function (i.e. a linear relationship with a chosen set of covariates), $w(s)$ are the spatial random effects, and $\varepsilon(s)$ is a "white noise" error term representing both data/measurement error as well as spatial variance at finer scales than those considered by the analysis (Banerjee et al. 2004). In this sense, we partition the total variance around the mean structure into two parts: the spatial and non-spatial components. Since we are working within a hierarchical framework, it is possible to place prior specifications on the parameters controlling the spatial process (i.e. the spatial range and sill parameters), as well as that in $\varepsilon(s)$ (equivalent to the "nugget effect" in geostatistics). In this way, we may fully capture uncertainty within the spatial dependence underlying forest soil carbon.

Spatial hierarchical models have been applied to map aboveground forest attributes at broad spatial scales, including forest biomass (Finley et al. 2010, Johnson et al. 2014), forest community composition (Finley et al. 2009), deforestation rates (Agarwal et al. 2005), and individual tree structure variables (Babcock et al. 2013). Additionally, Guhaniyogi et al. (2013) used multivariate spatial hierarchical models to map a suite of forest soil nutrients, including SOC, at a fine scale within a tropical forest in Costa Rica. Thus far, these models have not been applied to map soil carbon at scales relevant to international climate change research and policy, though their use in other macroscale forest science studies indicates significant potential. Similarly, to the best of

our knowledge techniques for addressing model uncertainty have not been applied to digital soil mapping, though research on model averaging for prediction throughout ecology and natural resources argues for their consideration here.

The aim of this dissertation is to contribute to the development of a consistent modeling approach for forest soil carbon at regional, national, and continental scales. To accomplish this, we explore the development of predictively accurate functions, but place particular emphasis on the quantification of uncertainty. More specifically, we focus on the roles model error and spatial dependence may play in forest soil C models, given that these sources of error are frequently overlooked by DSM studies. Here, we present a collection of papers focusing on computational methods, primarily from the Bayesian statistics literature, that are useful tools for forest and soil scientists interested in mapping soil C or other forest attributes at broad spatial scales.

Chapter one addresses the issue of modeling regional scale spatial dependence with sparse inventory data. For economic reasons, forest inventory plots are often only available at low densities, and frequently sampling strategies are not specifically designed with spatial analysis in mind. This study uses a low density set of forest soil carbon inventory data for the Coastal Plain of New Jersey, USA to ask whether, for this case, modeling spatial dependence offers any improvement in prediction accuracy relative to comparable non-spatial models. Two non-Bayesian methods from conventional geostatistics, universal kriging and co-kriging, are compared to non-Bayesian, non-spatial multiple regression models. We find that spatial dependence is unimportant for this dataset, though it is not clear if this is a result particular to our region, or one that is generalizable across other temperate forest ecosystems.

Chapter two provides a departure from the primary focus on spatial dynamics of forest soil C to consider the role of 'model error' in natural resources and ecological datasets. Here we directly compare the performance of several widely used criteria for model selection: Akaike's information criterion (AIC) (Akaike 1973), the deviance information criterion (DIC) (Spiegelhalter et al. 2002), and Bayesian posterior model probability (Kass and Raftery 1995, Bayarri et al. 2012). We use both 'best' model and model averaged predictors with these metrics, and compare their performance for three datasets based on posterior predictive checks (Gelman 2003). The conclusions drawn in this study do not advance one method for model selection over others, but instead suggest that any can be a reasonable approach, providing rigorous evaluation using predictive diagnostics is subsequently performed.

Chapter 3 is the first of two on applying spatial Bayesian hierarchical models to the prediction of forest soil carbon at broad spatial scales. Here we explore whether spatial dependence exists in forest inventory datasets for three different spatial scales: the Coastal Plain physiographic province of New Jersey, forested lands within with the nation of Germany, and forests within the member states of the European Union. We use variograms and surface plots of model residuals and spatial random effects to examine the spatial structure in soil carbon, soil nitrogen, and the residuals of these variables. The results of this study suggest that spatial dependence in forest soil C and N is primarily present at broader spatial scales, perhaps driven by increasing environmental heterogeneity associated with crossing large spatial extents. However, we also find that the soil C and N residuals exhibit spatial structure in all of the datasets, suggesting that these variables co-vary across space at all of the scales we considered.

Lastly in chapter 4 we turn to the comparison of spatial and non-spatial models for these three study regions. We construct linear functions based on the 'scorpan' model of McBratney (2003), and compare models for each dataset via *k*-folds cross validation. We find that spatial models offer better predictive performance, even for the datasets that did not exhibit much spatial dependence. We find an interpolated surface of soil N (from our inventory data) to be a good predictor of soil C, but do not find strong linear relationships with other remote sensing covariates. Our results highlight spatial hierarchical models as an ideal framework for predictive mapping of soil C, but also suggest that linear functions are not optimal. In general, we find the construction of good mean structures for forest soil C to be both scale and context dependent.

Taken together, these four studies contribute to the development of predictive models for forest soil C, and highlight spatial hierarchical models as an appropriate framework for broad scale mapping of forest soil carbon. However, there is much more to be done. Our work raises questions about controls over spatial dynamics in SOC, and how they vary across multiple scales. Further, the results of our model comparison studies argue for new lines of research on the development of optimal predictive models of forest soil carbon. These issues will be highlighted in the discussion sections of the proceeding chapters.

Digital mapping of soil carbon, particularly forest SOC, is a nascent field. Undoubtedly in the coming decades the increasing availability of both soil inventory and remote sensing data, as well as the development of new efficient methods for analyzing massive datasets, will lead to significant improvements over the current literature on predictive modeling of soil attributes. However, if these models are to be widely used

within the management and policy community, and are to lead to verifiable estimates of soil organic carbon stocks, a consistent modeling framework must be employed. It is our hope that, if nothing else, the work presented here highlights a methodological way forward that future work may build upon.

Overview of the datasets.

This dissertation relies on several soil carbon datasets, collected at different scales both within New Jersey, United States and across Europe. Additionally, chapter two employs two unrelated published datasets representing statistical variable selection problems. Before proceeding with the presentation of the research chapters, here we present an overview of these data. Most importantly, we outline the survey and sampling protocol for the soil carbon inventory collected by the authors in 2011 and 2012, as part of this dissertation research.

The New Jersey soil carbon data is the only dataset used in all four research chapters. Chapter one relies solely on this dataset, collected in forests across New Jersey's physiographic plain. Chapter two again utilizes these data, but also considers data on individual tree volume in Loblolly Pine from Burkhart et al. (1984) as well as ecosystem productivity in laboratory microcosm experiments from McGrady-Steed et al. (1997). In chapters 3 and 4, we also make use of the European Union Joint Research Centre's LUCAS topsoil dataset, which is a publicly available set of geo-referenced observations of a variety of soil properties. Details on these data, and the associated sampling protocols, may be found in (Toth et al. 2013a, 2013b). What follows is a full description of the NJ data used within our studies.

Sampling design.

500 plots were established throughout New Jersey's coastal plain, using a stratified random sampling design based on dominant forest type and drainage class. Forest cover classes were extracted from the 2010 LULC classification of New Jersey performed by the Center for Remote Sensing and Spatial Analysis (CRSSA), Department

of Ecology & Evolution, Rutgers University (Hasse and Lathrop 2010). A drainage class map was created from the SSURGO data set (Soil Survey Staff, NRCS) using the soil mapper function. Four vegetation classes (coniferous forest, deciduous forest, mixed forest, wetland forests) and three drainage classes (excessively drained, well drained, poorly drained) were used in a factorial design, resulting in ten sampling strata. Manipulation of source vegetation and soil data to create individual files for each sampling stratum was accomplished using ArcMap 10.0 (ESRI, Redlands, CA).The 500 sampling plots were then randomly generated, proportional to stratum area, using the Geospatial Modeling Environment module (Spatial Ecology, 2010).

Several other variables were considered for sampling stratification, including elevation, soil texture, and fire history. All of these, along with vegetation and soil moisture, can greatly influence soil organic matter distribution (Chapin et al., 2002). However after exploring the available data, the latter two were decided upon as the two dominant factors that would work the best for dividing the entire region into a reasonable number of sampling strata. Other variables will be incorporated into the model, and the distribution of our sampling regime should incorporate enough spatial variation to make use of these factors for predicting SOC distribution.

The sampling model was parameterized to establish plots at a minimum distance of 80 meters, and an 80 meter buffer around all paved roads and other land use types was also established. 80 meters was selected as an appropriate lag distance based on a pilot study we undertook in the fall of 2010.

Plot sampling

A subset of 182 plots was sampled during the course of the study. Many more plots than we intended to sample were generated in order to account for the fact that many would be located on private land, and present access issues. The decision was made to not constrain the sample distribution to public land for two reasons. First, the proportion of each stratum type on public land may not be the same as it is across the entire coastal plain. Second, while good spatial data of all state land boundaries was available, we were lacking land held by NGOs or other non-governmental agencies that would permit us access. In selecting which plots to sample care has been taken to maintain roughly the same proportions in each strata, and to spread sampling out across the entire geographic area.

Sampling points were located to +/- 5 meters using a Magellan mobile mapper handheld GPS unit. Soil sampling was conducted at the center of each plot. One 625 cm^2 litter sample was collected, along with three soil samples in the following depth increments: 0-10 cm, 10-20 cm, and 20-30 cm. Each increment sample consisted of a bulk density core and a second sample of approximately equal volume for lab analysis. Bulk density cores were carefully inserted into the soil matrix, using a sharp knife to cut around the outside of the core, in order to minimize soil compaction. The whole cores were extracted, carefully cleaned of excess soil, and wrapped in aluminum foil to be brought back to the lab. Analytical samples were collected with a small trowel and transferred to brown vials for transport.

Laboratory analysis

Four laboratory analyses were completed on the field collected samples: bulk density, loss-on-ignition (LOI) analysis for soil organic matter content, elemental

spectroscopy (CHN) analysis for soil carbon and nitrogen content, and CHN analysis for leaf litter content. The decision was made to analyze soils for both SOM and SOC to help correct for the very small volume of material (well less than 1 cm^3) typically used for CHN analysis. The extrapolation of carbon content for a sample > 100 g in size based on a very small analytical subsample may be prone to considerable measurement error.

Bulk density samples were dried for 48 hours at 105 ^oC and passed through a 2 mm sieve. Both the less than 2 mm and greater than 2 mm fractions were weighed, and bulk density was calculated as:

$$BD = FM/(V-(CM/PD)$$

Where BD = bulk density, FM = <2 mm fraction, V = volume of the sampling core, CM = >2 mm fraction, and PD = particle density, a constant of 2.65 $g*cm^{-3}$. Bulk density values were calculated for each of the depth increments.

The analytical samples were air dried for at least 72 hours, and then sieved to 2 mm. The fine fraction material was ground into a fine powder with a mortar and pestle, and homogenized for a period of 15 minutes by shaking and rolling the container. Two sub-samples were then drawn for the LOI and CHN analyses.

Soil organic matter content was estimated using loss-on-ignition (LOI). Samples were placed in a Lindberg muffle furnace (General Signal, Watertown, WI, USA) at 400 ^oC for 24 hours. This temperature allowed for all soil organic matter to fully evolve, while minimizing mass loss from matrix water and inorganic compounds (Ben Dor and Banin 1989). Soils in New Jersey's Coastal Plain are low in carbonates (Tedrow, 1986), so little measurement error is expected from evolution of inorganic carbon. Final mass was taken while the samples were still hot in order to limit measurement error due to

condensation on the soil material. Gravimetric soil organic matter content (cm^3*cm^{-3}) is then calculated as:

$$(IM_c\text{-}FM)/IM_c$$

Where IM_c = initial soil mass, corrected for residual water volume and FM = mass of remaining material after 24 hours at 400 ^oC.

Percent carbon and nitrogen by weight were determined using CHN analysis. Analyses were performed by the Ecosystems Lab of the University of Nebraska. Litter samples were dried to 70 ^oC, weighed, and ground to fine a powder using a Wiley mill. This material homogenized and also submitted for CHN analysis.

Calculation of SOC & SOM stocks

Prior to analysis, both soil organic carbon and organic matter stocks for each depth increment were converted to metric tons per hectare, a standard areal unit for expressing soil carbon stocks, according to the following equation:

$$SC = C_c * BD * V$$

where:

SC = soil organic carbon stock (t/ha)

C_c = soil carbon concentration ($kg*kg^{-1}$)

BD = soil bulk density ($kg*m^{-3}$) of the < 2 mm fraction

V = the volume of a rectangle 1 hectare in area and 10 cm in depth

Litter carbon stock is also expressed on an area basis, in units of metric tons/hectare, as:

$$LC = LC_c * A$$

where:

LC = litter carbon stock (t/ha)

LC_c = litter carbon content of the 625 cm^2 sampling frame

A = the area of a 1 hectare rectangle

Total soil carbon stock is then calculated as:

$$SC_{total} = LC + SC_{0\text{-}10cm} + SC_{10\text{-}20cm} + SC_{20\text{-}30cm}$$

Data archiving & storage

The final datasets, which include raw bulk density and gravimetric contents of soil organic carbon (SOC), soil organic matter (SOM), and soil nitrogen (SN), as well as derived areal estimates in terms of megagrams/hectare, will be archived in several locations. First, these data will remain as part of the permanent databank owned by the Center for Remote Sensing and Spatial Analysis (CRSSA) at Rutgers University. A special folder will be created for all relevant files, as well as associated metadata. This folder will also include all post-processed raster layers used as model predictors (see the methods outlined in Chapter 4 and it's appendix for details). In addition, the datasets will be made publicly available via the International Soil Carbon Network (ISCN) database.

Chapter 1: Comparing spatial and non-spatial approaches for predicting forest soil organic carbon at unsampled locations.

Abstract

Prediction of soil organic carbon (SOC) at unsampled locations is central to statistical modeling of regional SOC stocks. This is often accomplished by applying geostatistical techniques to plot inventory data. However, in many cases inventory data is sparsely sampled (<0.1 plots/km^2) relative to the region of interest, and it is unknown if geostatistics provides any advantage. Our objective was to test whether modeling spatial autocorrelation, in multivariate and univariate predictive models, improved estimates of SOC at prediction locations based on sparsely-sampled inventory data. We conducted our study using a dataset sampled across all forested land in the Coastal Plain physiographic province of New Jersey, USA. We considered five models for predicting SOC, two linear regression models (intercept only and multiple regression with predictor variables), ordinary kriging (a univariate spatial approach), and two multivariate spatial methods (regression kriging and co-kriging). We conducted a simulation study in which we compared the predictive performance (in terms of root mean squared error) of all five models. Our results suggest that our sparsely-sampled SOC data exhibits no spatial structure, though several of the covariates are spatially autocorrelated. Multiple linear regression had the best performance in the simulation study, while co-kriging performed the worst. Our results suggest that when inventory data is dispersed across the region of interest, modeling spatial autocorrelation does not provide significant advantage for predicting SOC at unsampled locations. However, it is unknown whether this autocorrelation does not exist at broad scales, or if sparse sampling strategies are unable

to detect it. We conclude that in these situations, multiple regression provides a straightforward alternative to predicting SOC for mapping studies, but that more work on the spatial structure of soil carbon across multiple scales is needed.

Introduction

Globally, forests are thought to store approximately 861 Pg of carbon, with about 44% of this mass found in forest soils (Pan et al., 2011). The large capacity of the forest soil pool to sequester carbon makes its management a viable option for mitigating the effects of atmospheric carbon emissions (Goodale et al., 2002; Lal, 2008). Naturally, there is considerable interest in the quantification of forest SOC pools for carbon monitoring projects, and the development of market-based carbon accounting schemes. There is a need for methodologies that produce consistent results with a degree of accuracy acceptable to policymakers (Chen et al., 2000a; Houghton, 2003; Shvidenko et al., 2010).

Forest carbon stocks are typically measured using forest inventories, and areal estimates are gained by 'scaling up' these measurements across the region of interest (Birdsey, 1992; Goodale et al., 2002). However, these data are often sparse relative to the extent of the stock estimate. In the case of forests, soil carbon sampling is often excluded from large inventory efforts due to the additional time and cost needed to collect and process samples relative to aboveground forest measurements. As a result, regional estimates of forest soil carbon storage are often highly uncertain, leading to wide disparity among the literature. For example, estimates of carbon stocks for European forest soils have ranged from 3 Pg C to as high as 79 Pg C; a difference that constitutes

approximately 9% of the global forest soil carbon stock (Cannell et al., 1992; Goodale et al., 2002; Liski et al., 2002; Jones et al., 2005).

Developing a regional soil carbon stock from inventory data involves prediction of the response variable at many unsampled locations (i.e. all squares of a grid covering the region of interest). Spatial autocorrelation, where nearby points are on average more similar than points that are further apart, is a common property in environmental datasets and, when present in inventory data, may be leveraged to increase prediction accuracy (Simbahan et al., 2006). Spatial autocorrelation may be summarized by computing the semivariance, a measure of spatial similarity, and plotting these values for all pair-wise combinations of the sampling points as a function of distance (Goovaerts, 1997). These plots, typically referred to as the empirical semivariogram, may be fitted with 'theoretical' semivariogram models, such as Matérn class or spherical functions. Kriging methods, a widely used class of spatial interpolators, incorporate such theoretical semivariogram models to weight predictions at unsampled locations (Isaaks and Srivastava 1989). This feature, combined with the fact that these methods may be extended to model spatial covariance between the response and predictor variables, makes kriging a logical approach for the prediction of soil carbon.

Geostatistical techniques have been successfully applied to predict soil organic carbon at unsampled locations, based on plot inventory data, at a variety of spatial scales. Several studies are available for agricultural fields, where very dense sampling regimes (>400 plots/km^2) can be achieved, and clear patterns of spatial variation are often elucidated (Chen et al., 2000a; Lark, 2000; Mueller and Pierce, 2001; Simbahan et al., 2006). In such situations, geostatistical models have been shown to offer considerable

improvement in prediction results when compared to non-spatial regression models (Simbahan et al., 2006).

Fewer examples are available for regional soil carbon mapping, where reduced sampling density may make spatial autocorrelation more difficult to detect. Still, several studies have shown an increase in prediction accuracy when incorporating geostatistical techniques. Liski and Westman (1997) used block kriging to interpolate measurements of soil organic carbon taken as part of the national forest inventory (NFI) in Finland, and detected significant spatial structure in these clustered, but densely sampled (~5 plots/km^2), data. More recently Mishra et al. (2010) compared the performance of several geostatistical methods, including geographically weighted regression and regression kriging, to multiple regression models for predicting SOC across a heterogeneous region in the northern Midwestern United States. Their results suggest a significant increase in prediction accuracy (~22% relative improvement) when incorporating spatial error structure into the model. Other examples where significant spatial structure was detected and used to model SOC are available for grasslands in Ireland (McGrath and Zhang 2003, Zhang et al. 2011) and agricultural landscapes in the karst region of China (Liu et al. 2006, Zhang et al. 2012).

While the aforementioned studies modeled soil organic carbon across large spatial extents, most took advantage of reasonably dense plot inventory data ($\geq$ 0.1 plots/ km^2), and in the case of Liski and Westman (1997) approximately 5 plots/km^2. The exception is the study by Mishra et al., which utilized sparsely sampled data (approx. 0.003 plots/ km^2), but modeled SOC across a heterogeneous landscape with several major cover types and a pronounced latitudinal gradient (from the upper Peninsula of Michigan south to

Kentucky, USA); both of which may exert strong controls on soil organic carbon distribution. When the region of interest comprises a single cover class, as it would in forestry applications, or does not span many degrees of latitude, it is less clear that modeling spatial autocorrelation presents any advantage for predicting forest soil organic carbon. In fact, a few studies provide evidence suggesting this is the case. Studies in tropical forests that examined forest SOC across multiple scales, in tropical dry forests in the West Indies (Gonzalez and Zak, 1994) and in the Brazilian Amazon (Cerri et al., 2000; Bernoux et al., 2006), suggest that the spatial structure is limited to fine scales only.

In this study, our primary objective was to assess whether incorporating spatial autocorrelation into models for predicting forest soil organic carbon at unsampled locations improved results for sparsely sampled (<0.1 plots km^2) inventory data. To meet our objective, we compared the performance of both univariate and multivariate spatial models to similar linear regression models. We predicted that the spatial models would perform the best when predicting forest SOC at independent validation locations, despite the sparsity of our sampling locations relative to the region of interest. To test this prediction, we developed a simple simulation experiment to directly compare the predictive accuracy of all models considered in the experiment.

Methods

2.1 Study region

This study was conducted on the Coastal Plain physiographic province of New Jersey, USA (Fig. 1.1). This region is largely forested, and the remaining landcover mainly consists of residential and agricultural development. Three major upland forest

communities dominate the region: (1) Pitch Pine (*Pinus rigida*) forest, (2) Oak (*Quercus* sp.) forest, and (3) mixed communities that span a gradient between these two classes (Hasse and Lathrop 2010). On the inner coastal plain, these communities mix with other hardwood species such as American Beech (*Fagus grandifolia*) and Hickory (*Carya*) species. Forested wetlands are common along river courses or in low areas. Most of these are hardwood swamps dominated by Red Maple (*Acer rubrum*), Sweet Gum (*Liquidambar styraciflua*), and Black Gum (*Nyssa sylvatica*). However, forested peat bogs with pure stands of Atlantic White Cedar (*Chamaecyparis thyoides*) are also present across the landscape. Soils in the region are largely typic Hapludults and Quartzisappamments of marine or alluvial origin (Tedrow, 1986). Soils range from very poorly to excessively drained, and are primarily sandy in texture. However, clayey and mucky soils are frequent in wet areas. The total area of the study region (i.e. all forested land in New Jersey's Coastal Plain) is approximately 4,522 km^2.

2.2 The datasets

We considered two plot inventory datasets for this study. The primary dataset, hereafter referred to as the 'small' dataset consists of 62 plots, and possesses measurements for forest soil organic carbon and all of the covariates used in the model experiments. This corresponds to a sampling density of approximately 0.013 plots/km^2. The 'large' dataset consists of 120 plots and contains measurements of the model covariates only, and was used for the co-kriging analysis. The small dataset is a subset of the large dataset, so those 62 plots are co-located and present in each. The plots were sampled in a stratified random design across the landscape, based on both forest community type and soil drainage class (Fig. 1.1).

At each sampling location, soil was collected from three depth intervals: 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. Bulk density samples were dried for 24 hours at 105 ^{o}C and passed through a 2 mm sieve to remove the coarse fragments (i.e. gravel and litter material) that are not a component of the soil organic matter pool. The analytical samples were air dried for at least 48 hours, sieved to 2 mm, then ground into powder with a mortar and pestle and homogenized.

Percent soil organic carbon was estimated by elemental ('CHN') analysis on a subsample of the air-dried analytical sample. A second subsample was used to measure percent soil organic matter (SOM) by loss-on-ignition (LOI). These SOM measurements were recorded for all 120 plots, and used as a covariate in the multivariate models. Soil organic matter typically has a significant relationship with SOC, and has been used as a predictor for soil organic carbon in several studies (Konen et al., 2002; De Vos et al., 2005). Samples were placed in a Lindberg muffle furnace (General Signal, Watertown, WI, USA) at 400 ^{o}C for 24 hours. Both percent SOC and percent SOM were converted to stock estimates using the following formula (Hoover 2008):

$$S = P * BD * V * g \quad (1)$$

Where S is the stock estimate (Mg*ha^{-1}), P is a percent measurement of SOC or SOM, BD is soil bulk density (g·cm^{-3}), V is the volume of a 1 ha^{2} rectangle with a depth of 30 cm, and g is a unit scaling constant.

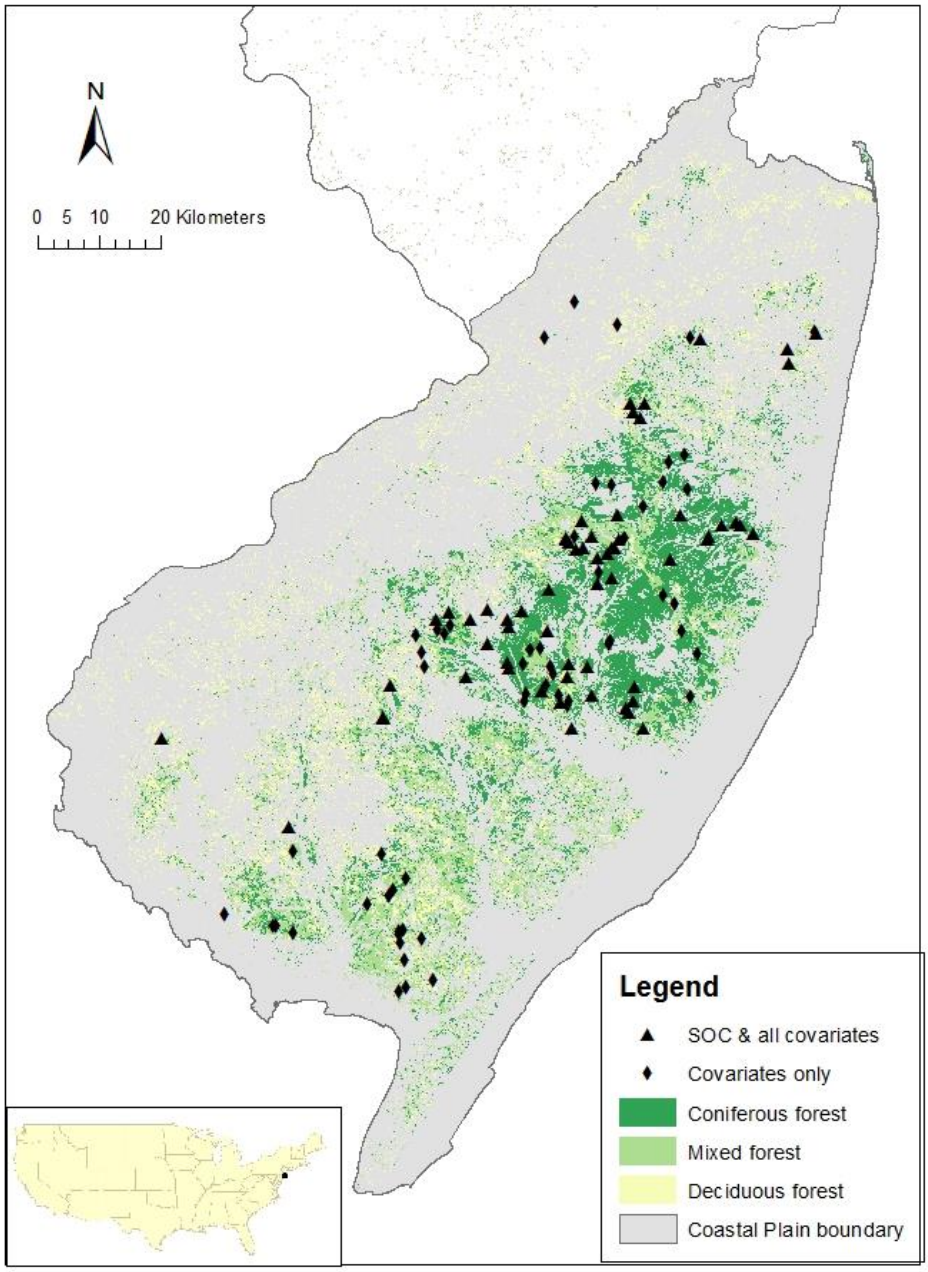

Figure 1.1: Distribution of sampling locations and primary forest cover types across the study region.

2.3 Model covariates

In addition to the plot measured soil organic matter data, we utilized four covariates extracted from remote sensing and GIS datasets: normalized difference vegetation index (NDVI), band 2 of the 'tasseled cap' transform (TC2), compound topographic index (CTI), and elevation (ELEV). These variables represent a reasonable set of potential predictors for soil organic carbon, and are similar to covariates incorporated by several recent regional SOC mapping studies (McGrath and Zhang, 2003; Mishra et al., 2010; Zhang et al., 2012; Vasques et al., 2012). NDVI and TC2, the "greenness" band of the tasseled cap transform, are both related to net photosynthetic output, which has a theoretical relationship to inputs into the soil organic carbon pool (Chapin et al., 2002). Terrain position can have a strong influence on soil organic carbon storage, so we included two related variables: elevation and estimates of the compound topographic index (CTI). CTI is a steady state wetness index designed to model soil water content based on values of slope and flow direction extracted from a digital elevation model (Moore et al., 1991). It has been shown to correlate with soil moisture content, which may exert influence over soil organic carbon formation and storage (Barling et al., 1994).

To extract the NDVI and TC2 measurements for our sampling locations, cloud-free Landsat TM scenes (glovis.usgs.gov) were downloaded for a single date during the study, July 14th 2011, and tiled into a mosaic of the study region. We used a Level 1 data product from the Landsat 5 thematic mapper instrument that had been previously corrected for radiometric error and terrain variability, geo-referenced, and converted to

Universal Transverse Mercator (UTM) projection. The Erdas Imagine software package (Leica Geosystems, Atlanta, GA, USA) and ArcGIS (ESRI, Redlands, CA) were used to separately generate rasters for both variables with a grid cell size of 30 m^2 for all forested land within the study region, and to extract values of NDVI and TC2 for our sampling locations. Both the elevation and CTI data were derived from a 10m digital elevation model (DEM) provided by the Center of Remote Sensing and Spatial Analysis (CRSSA), Rutgers University. Compound topographic index was calculated for all cells in the DEM using ArcGIS.

2.3 Modeling approaches

Our objective in this study was to test whether explicitly modeling spatial autocorrelation improved prediction accuracy for our sparsely sampled forest SOC data. To accomplish this we considered five models that represent spatial and non-spatial approaches for both univariate (SOC data only) and multivariate (incorporating the predictor variables) cases (Table 1.1). This design allowed us to both examine the effect of modeling spatial autocorrelation only, in the case of the univariate spatial model (OK), as well as the influence of the spatial variance term for two different multivariate approaches.

Table 1.1: Dimensions and spatial variance assumptions for the five predictive models considered in this study.

Model	Dimensions	Spatial variance term
Intercept only regression (IR)	univariate	no
Multiple linear regression (MLR)	multivariate	no
Ordinary kriging (OK)	univariate	yes
Regression kriging (RK)	multivariate	yes
Co-kriging (COK)	multivariate	yes

Our non-spatial approach was linear regression models of the general form:

$$Y_i = \alpha + \sum_{j=1}^{p} X_{ij}\beta_j, \quad (2)$$

where Y_i represents the observations of soil organic carbon, α is the intercept, β_j is a slope parameter associated with each of the p covariates, and $i = 1, \ldots, n$ where n is the number of observations.

All three of the spatial models we incorporated are variations on the kriging algorithm, where spatial prediction is accomplished as a function of the theoretical semivariogram; a model fitted to a plot of semivariance values against distance for each pair-wise combination of sampling locations in the dataset (i.e. the empirical variogram)

(Goovaerts 1997). For the univariate model (OK), the linear estimator used to predict new values of the response variable, for some set of locations *u*, takes the form:

$$Z^*(u) = \sum_{\alpha=1}^{n(u)} \lambda_i(u)[(Z(u_i) - m(u)] \quad (3)$$

Where $Z^*(u)$ is the vector of predicted values of the response variable at new locations, $Z(u_i)$ is the vector of known values of the response at sampled locations, $m(u)$ is the mean response, and the λ_i is the vector of 'kriging weights' for each sampled location, that are determined by the semivariogram model, and *n*(*u*) is the neighborhood size (Goovaerts, 1997; Simbahan et al., 2006). In the case of ordinary kriging, note that the mean is taken to be a function of the locations *u* so that it is allowed to vary across the region (Isaaks and Srivastava, 1989). In this way, we are removing the mean function, commonly referred to as the 'drift' in geostatistics, and kriging the residuals.

In addition to the univariate ordinary kriging model, we considered two different approaches for incorporating covariates into spatial models. The first of these is regression kriging (RK), which is very similar to OK in principle. The difference is that the residuals of the response and predictor variables are interpolated, and in this way co-varying spatial patterns are indirectly incorporated into the analysis (Odeh et al., 1994; Hengl et al., 2004; Simbahan et al., 2006). For prediction at new locations, the spatially predicted residuals must be added back on to the mean trend, resulting in the following linear estimator for $Z^*(u)$:

$$Z^*(u) = \beta_0 + \sum_{k=0}^{p} \beta_k q_k(u) + \sum_{\alpha=1}^{n(u)} \lambda_i e(u) \quad (4)$$

Where β_k are the regression parameters associated with the predictors q_k, p is the number of predictors, and $e(u)$ are the residuals between the response and covariables (Hengl et al., 2003, 2004). The rest of the terms in the model are as defined above. We wish to note that the technique we outline here is one of several closely related approaches that have all variously been termed 'regression kriging', 'kriging with external drift', and 'kriging with a trend' (Goovaerts, 1997; Wackernagel, 1998; Chiles and Delfiner, 1999). We follow Hengl et al. (2004) in describing the method outlined above, where the non-spatial trend and the spatial interpolation of the residuals are accomplished separately, as regression kriging.

The second multivariate method considered here is co-kriging, which represents a particularly flexible approach to modeling multivariate geostatistical data. Rather than interpolating residuals between the response and predictor variables, co-kriging starts with the fitting of both direct and cross variograms for all variables in the model, typically with a linear model of coregionalization (Gelfand et al., 2004). Essentially, this means we model the spatial covariance of the response and predictors individually, but also model the spatial cross-covariance among the model variables. This variogram system is employed to weight predictions of the response variable at new locations, according to the following linear estimator:

$$Z^*(u)_i = \sum_{i_1=1}^{n_1(u)} \lambda_{i_1}(u)\left[Z_1(u_{i_1}) - m_1(u_{i_1})\right] + \sum_{i=2}^{N_v} \sum_{i_i=1}^{n_i(u)} \lambda_{i_j}(u)\left[Z_i\left(u_{i_j}\right) - m_i\left(u_{i_j}\right)\right] \quad (5)$$

Where $Z^*(u)_i$ are the predicted values of the response variable at new locations, λ_{i_1} is the weight assigned to the response variable Z_1 and λ_{i_j} represents the weights for the covariates Z_i (Goovaerts, 1997; Simbahan et al., 2006). In this model, the expected values m_i are subtracted from the data, indicating we consider the spatial association between the response and predictor variables to be a multivariate non-stationary process.

Co-kriging is appropriate for situations in which a response variable that is expensive to measure is sampled sparsely, while several 'cheap' covariates have been sampled in the same as well as additional locations. In our case, we have the 'large' dataset available, which contains 120 measurements of all of the model covariates. These additional values are used to fit the direct and cross variograms during co-kriging, along with the 62 observations which also contain measurements of the response variable. This situation lends itself well to the co-kriging approach.

2.4 Model comparison simulation

To compare the performance of the four models for predicting soil organic carbon, we devised a simulation that compared predicted vs. actual results for independent validation data. We first randomly divided the 'small' dataset into fitting and validation datasets. We reserved 25% of the data for validation (N=15) and used the remaining 47 plots to fit the models. We split the data this way, rather than using an even split, because initial runs of several kriging models resulted in undefined covariance functions when n=31 for the model fitting data. A fitting set of 47 plots translates to a

density of approximately 0.01 plots/km^2 across the study region. In the case of the co-kriging model, the additional covariate observations in the 'large' dataset were included in the model fitting, as the structure of the coregionalization model permits this design. To increase normality, all variables were log-transformed prior to fitting the models. Each model was used to predict soil organic carbon for the validation dataset, and we computed root mean squared error (RMSE) to assess model performance. Prior to computing RMSE, predicted values of log(SOC) were back-transformed into their original units. To avoid biasing results by selecting a single, favorable fitting dataset we ran this simulation for 10,000 iterations and tracked mean RMSE for the entire study. This is especially relevant for the geostatistical models, as relatively sparse datasets such as ours may possess spatial autocorrelation with some configurations but not with others.

To initialize the OK and COK models, we supplied values for the sill, range, and nugget parameters derived by fitting a Matérn class covariance function to the empirical variograms for soil organic carbon in the full dataset. In the co-kriging model, these values were used to initialize the parameters for all direct and cross variograms. In the case of regression kriging, we supplied initial parameter values from a theoretical variogram fitted to the residuals of SOC and the model covariates. All model fitting was accomplished with ordinary least squares. A Matérn covariance function was selected because it is a particularly flexible model for spatial autocorrelation, and is a popular choice in current geostatistical research (Stein, 1999; Finley et al., 2010). The simulation was conducted using the R statistical computing environment. Variogram fitting, ordinary kriging, and regression kriging were conducted using the geoR package (Ribeiro and Diggle, 2001), and co-kriging was accomplished in the gstat package (Pebesma, 2004).

Results

3.1 Exploratory analysis

Table 1.2 presents the mean and standard deviation for all variables, as well as the regression parameters for the MLR model and the correlation coefficients between log(SOC) and each of the covariates for the full dataset (n=62). Soil organic matter is highly correlated with SOC (ρ=0.708), while the remaining variables are not notably correlated (ρ<0.2 for each). For the intercept only model, α = 3.59.

Table 1.2: Mean (μ), standard deviation (σ^2), slope parameters (β_j) and correlation coefficients (ρ) for the five covariates and SOC.

	M	σ^2	βj	ρ
SOC (Mg/ha^2)	65.93	65.67	**	**
SOM (Mg/ha^2)	113.17	153.66	0.678	0.708
NDVI	0.61	0.05	0.395	0.103
TC2	29.54	9.13	-0.507	0.046
CTI	9.99	2.48	0.332	0.106
ELEV (m)	26.35	12.25	0.157	0.098

Examination of the spatial structure in the SOC dataset does not reveal any spatial autocorrelation among the 62 sample sites (Moran's *I* = -0.05, p = 0.39). However, slight positive spatial autocorrelation was noted for the following covariates: TC2 (Moran's *I* = 0.06, p = 0.04), CTI (Moran's *I* = 0.05, p = 0.09), and ELEV (Moran's *I* = 0.036, p <

0.001). Both SOM and NDVI do not possess significant spatial autocorrelation ($p > 0.10$). The empirical variograms, as well as the fitted Matérn covariance functions (i.e. the theoretical variograms), agree with these results (Fig. 1.2). TC2, CTI, and ELEV show an increase in semivariance across distance, each with an asymptotic range > 120,000 m. However, note that there is considerable residual error between the empirical semivariance values and the fitted covariance model. NDVI suggests an increase in semivariance, but the scale of the y-axis for this plot indicates a minute change across the effective range. Both SOC and SOM do not show spatial structure in either the empirical or theoretical semivariograms.

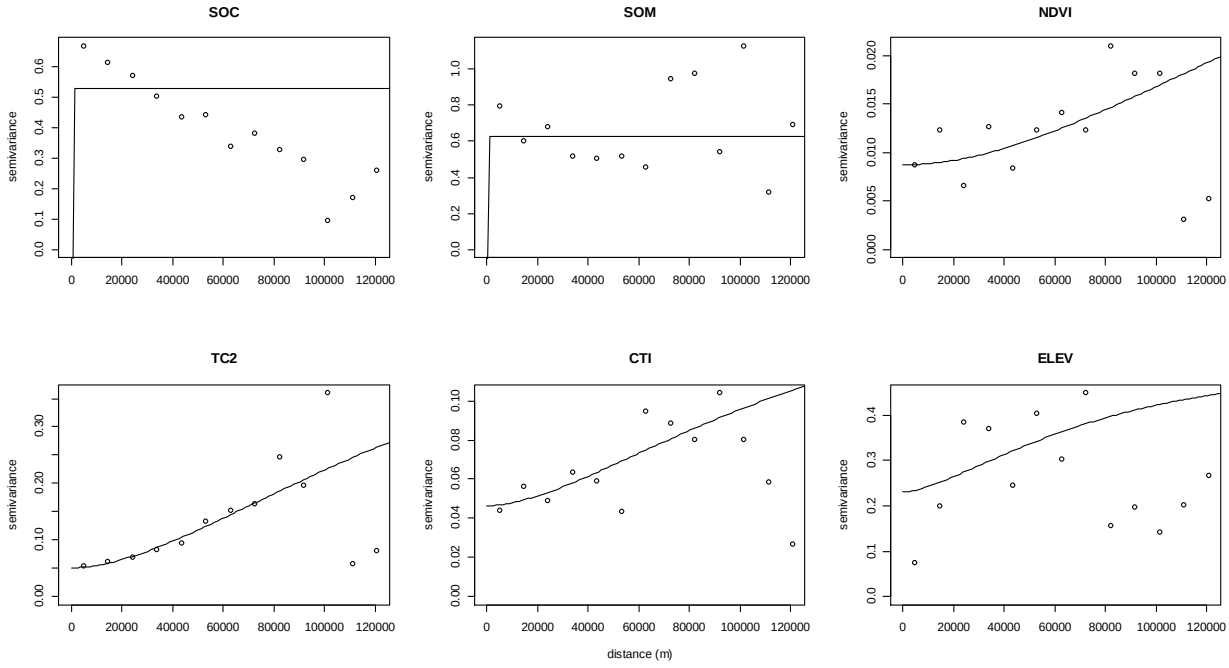

Figure 1.2: Empirical (open circles) and theoretical (solid lines) variograms for the response (SOC) and the five covariates.

3.2 Model comparison experiment

The results of the simulation experiment show that the multiple regression model provided the most accurate predictions for the validation data, in terms of mean RMSE over the 10,000 trials in the simulation (Table 1.3). Ordinary kriging was the worst performing model, followed by co-kriging. These results correspond with the general lack of structure in the SOC data described above. Regression kriging had a similar, though slightly inferior, performance relative to multiple regression. This is not surprising, given that the regression kriging estimator is simply an extension of that of MLR. Comparing the two univariate methods also suggests a disadvantage to modeling spatial autocorrelation, as the intercept only regression model reduced error when compared to the ordinary kriging model.

Table 1.3: Results of the simulation experiment. Note that this table presents back-transformed values of forest SOC. Mean RMSE refers to the mean root mean squared error over the 10,000 trials in the simulation experiment. RI refers to the relative improvement in predictive performance of each model, when compared to the worst performing method (Ordinary Kriging).

Model	**mean RMSE($Mg \cdot ha^{-1}$)**	**RI (%)**
Intercept only regression (IR)	59.65	12.1
Multiple regression (MLR)	51.52	24.1
Ordinary kriging (OK)	67.9	--
Regression kriging (RK)	53.43	21.3
Co-kriging (CK)	61.01	10.1

Discussion

In contrast to studies where prediction of SOC is accomplished with relatively dense plot inventories (>0.1 plots/km^2) (e.g. Liski and Westman, 1997; Lark, 2000; McGrath and Zhang, 2003; Simbahan et al., 2006; Zhang et al,. 2012), we found that modeling spatial autocorrelation did not improve prediction accuracy at unsampled locations for our sparse inventory data. Both variogram analysis and Moran's *I* statistics suggest a lack of spatial autocorrelation in our soil carbon data. While spatial structure was noted in some of the covariates, the lack of spatial structure in SOC resulted in inferior performance of the spatial models relative to multiple regression. However, note that the RMSE of all models is large relative to the mean of soil carbon for the whole dataset (65.9 Mg·a^{-1}), suggesting that there is a high degree of uncertainty in all five models.

Generally, these results highlight the difficulties of spatial prediction of forest soil carbon. A number of studies have identified spatial structure at local scales in a variety of forest types, with variogram range parameters from 4-500 meters (for example, Robertson et al.,, *n.d.*; Lister et al., 2000; Wang et al., 2002; Garten Jr. et al., 2007; Worsham et al., 2010). It is not known, however, if these fine-scale spatial dynamics are meaningful to predictions for regional datasets, where distances between plots may range from one to hundreds of kilometers. Further, it remains unclear whether spatial autocorrelation at broad scales is an important factor in understanding regional forest carbon dynamics. Our results suggest otherwise, as do those of the few other studies that have looked at this question (Liski and Westman, 1997; Cerri et al., 2000; Bernoux et al., 2006).

In our data, determining whether regional forest SOC data truly exhibits no spatial structure, or if this is the result of a detectability issue caused by low sampling densities, remains unclear. Assuming spatial structure exists at the regional scale, describing it may require a large number of observations relative to the region of interest. While national forest inventories, such as the US Forest Service's Forest Inventory and Analysis (FIA) program, may achieve the requisite densities for aboveground measurements (Finley et al., 2007), collection of data on soil variables is often only completed at a fraction of these plots. Further, the ability to detect spatial autocorrelation is influenced by the sampling design (Fortin et al., 1989). Thus, surveys may need to be specifically designed to detect broad-scale spatial structure in forest soils.

Those studies which have detected regional spatial autocorrelation in the soil organic pool have typically done so over heterogeneous landscapes, spanning multiple cover classes (McGrath and Zhang, 2003; Mishra et al., 2010; Vasques et al., 2010; Zhang et al., 2011). In these contexts, the spatial structure of soil carbon is influenced by other spatially-explicit dynamics, such as patterns in land use and land cover, which may make regional patterns easier to define (Vasques et al., 2012). Modeling soil carbon over very large regions, such as the northern portion of the Midwestern United States (Mishra et al., 2010), also incorporates the effect of latitudinal climate gradients which are well known to influence soil organic carbon (Chapin et al., 2002). In this way, Mishra et al. detected an advantage to spatial approaches (geographically weighted regression and regression kriging) over multiple regression despite a very low plot density (<0.001 plots/km^2). In forestry applications, particularly across comparatively small regions such as the Coastal Plain of New Jersey, there may be fewer influences on the regional spatial

structure of soil organic carbon. However, additional studies in different forest types will be necessary to determine if this is in fact the case.

Incorporating covariates of soil carbon into predictive models is a typical strategy, and one employed by almost all of the studies outlined here. In our case, four of the five covariates we considered were not strongly correlated with forest SOC. These patterns may be unique to our region in some ways. For instance, one would expect a strong relationship between soil organic carbon and elevation. However, the Coastal Plain of New Jersey is a fairly low-relief landscape, and fully capturing the covariance between SOC and elevation in an inventory dataset may be particularly challenging.

Field measured soil organic matter was the one covariate that was reasonably correlated with SOC, but given that this variable was also sampled as part of our forest inventory it has limited usefulness for predicting soil carbon at unsampled locations. For regression models, it is generally necessary to have values for the covariates at the prediction locations (i.e., for all cells of a sampling grid in mapping applications). Methods based on fitting coregionalization models, such as co-kriging, are attractive in that they do not share this prerequisite (Goovaerts, 1997; Banerjee et al., 2004; Gelfand et al., 2004). However, in the absence of spatial structure in the response variable, these methods will likely yield poor results, as was the case with our data. An alternative strategy is to model the spatial dynamics of the covariates themselves. For example, soil organic matter may be interpolated based on ancillary variables in order to inform a sampling grid for soil carbon. However, this introduces additional sources of uncertainty which may propagate through to the final estimate of the response variable.

The results of our study have applications for forest SOC mapping projects, particularly where new inventories are being established to accomplish these goals. This may be especially relevant in developing countries, where international funding mechanisms such as the United Nations' Reducing Emissions from Deforestation and Degradation (REDD+) program has motivated increased interest in managing forests to offset carbon emissions (Edwards et al., 2010). Newly established forest inventories will be important for both gathering baseline data on forest carbon stocks in these regions, and for verifying gains in carbon sequestration (Maniatis and Mollicone, 2010). Our results suggest that when plot inventories are sparsely distributed (<0.1 plots/km^2), there is no spatial autocorrelation present in forest SOC data, and as a result modeling spatial structure does not result in increased prediction accuracy. In these cases, multiple linear regression presents a straightforward alternative, providing a set of reasonable covariates can be identified for all prediction locations.

Taken in context with the existing literature on the spatial dynamics of forest SOC, our work highlights the need for more studies that explicitly model soil carbon across a range of spatial scales. Without these data, it remains unknown whether regional spatial autocorrelation does not exist or requires more dense sampling schemes to detect. Further, our results are from but one forest type, and it is not clear that the dynamics we describe are generalizable to other forest ecosystems. That all of our models provide a fairly poor fit to our SOC data demonstrates just how challenging characterizing uncertainty in regional soil carbon stocks can be. Advanced statistical modeling techniques such as geostatistics present many appealing methods for the prediction of forest soil carbon, but their utility is premised on a set of assumptions that the available

data may not meet. We advocate that these methods are considered for the prediction of forest carbon, and for SOC mapping studies, but only when their use is warranted by the data.

Conclusions

When predicting soil organic carbon at unsampled locations based on sparse inventory datasets, it may be difficult to detect a significant degree of spatial autocorrelation. This is especially true on homogenous landscapes, or for studies that only consider one cover type, as there may be spatial structure associated with the correlation between SOC density and land cover type. In such cases, geostatistical models may be inappropriate, and multiple linear regression offers an appealing and straightforward alternative. Including covariates can increase the predictive performance of statistical models. The best predictors will not only be closely correlated with soil organic carbon, but will be available for the full extent of the study region. The results of our study have implications for SOC mapping approaches using existing inventories, where analytical efforts are constrained by data quality and availability, as well as for new sampling efforts where resources are limited. Future work should look to model spatial autocorrelation of soil carbon across multiple scales, to fully characterize the relationship of well-described local spatial structure to broad-scale, regional patterns.

Chapter 2: Model selection and evaluation: lessons learned from ecological data

Abstract.

Model selection is important in ecology, typically with the goal of selecting the 'best' model from some set of candidate models. However, the choice of model selection criteria may influence results. This creates a lack of consistency which is exacerbated by the fact that evaluation of models is rarely performed, and has led to confusion in the literature over 'best practices' for model selection. Here we present three case studies with the goal of evaluating the performance of different model selection strategies. We selected models using methods that span both Information Theoretic and Bayesian approaches. We evaluated performance by generating simulated datasets from the posterior predictive distribution and performing posterior predictive checks. Our results show that each criterion provided quite different results both for model selection and evaluation. Given these inconsistences, evaluation based on prediction revealed important features of the models that would have otherwise been overlooked. Posterior predictive checks are a useful, general framework for evaluating models based on their predictive performance, and we will conclude with some practical advice for incorporating this approach into the model selection process.

Keywords: Model selection, Bayesian posterior model probabilities, AIC, DIC, posterior predictive checks

Introduction.

The model selection problem, where the goal is to find the 'best' model or models among a candidate set $M_1, \ldots, M_k$, is one that most ecologists and natural resource

scientists entertain on a regular basis. Model selection typically takes one of several forms: including the selection of suitable covariates for some response variable(s) of interest; determination of an appropriate probability distribution for a given dataset; or 'curve fitting' to describe the relationship between two or more variables. The last decade has seen many ecologists moving away from classic hypothesis testing, where one 'alternative' model is compared to a 'null' model, and toward multi-model inferential frameworks where multiple alternative models may be entertained simultaneously (Johnson and Omland 2004). As a result, statistical methods for evaluating and ranking models are now commonplace in the literature.

In modern statistics, model selection is accomplished by using a test statistic designed to estimate the 'fit' of the model to the data available to the analyst. Two principle paradigms for model selection have emerged within the literature: (1) information theoretic (IT) approaches, such as Akaike's information criterion (AIC) (Akaike 1973, Burnham and Anderson 2002); and (2) Bayesian model selection, including Bayes factors (Kass and Raftery 1995), model posterior probability (Berger and Barbieri 2004, Bayarri et al. 2012), and the Bayesian information criterion (BIC) (Schwarz 1978). While these paradigms all share a common goal, they differ in approach. AIC and various derivatives compare models by minimizing the Kullback-Leibler divergence between the estimated model and some "true" distribution (Burnham and Anderson 2002), while the Bayesian approaches attempt to select the model with the highest posterior model probability (Yang 2005). A very large literature exists on the implications of these differences, and we refer readers to those sources for more background on this debate (Shao 1997, McQuarrie and Tsai 1998, Burnham et al. 2011).

Regardless of paradigm, most analysts employ a similar approach to model selection:

1. Establish a set of candidate models, based on *a priori* hypotheses or candidate predictor variables;
2. Evaluate each model using the criterion of choice; and
3. Proceed with the analysis; whether it is prediction, interpretation, or both; using the "best" fitted model or a model-average of all reasonable candidates.

This is a deceptively simple outline for a process that is quite complex. Most, if not all, scientists agree that it is necessary to think carefully about the models that will be compared, even when this process is aided by computational tools such as statistical variable selectors or data-mining strategies. When using model averaging, which is recommended as routine practice to address model uncertainty (Raftery et al. 1997, Burnham and Anderson 2002), choosing which models to include requires consideration, and a variety of computational and heuristic methods have been recommended to address this question (Hoeting et al. 1999). In short, standard practice calls for careful thought at both the beginning and the end of the model selection process.

However the second step, where model fit is actually evaluated, is more often than not left to a single test statistic. This is despite the fact that several comparative studies have demonstrated that the relative performance of IT and Bayesian model selection criteria is often problem specific and context dependent (Kuha 2004, Wang and Liu 2006, Ward 2008). As Aho et al. (2014) recently observed, the search for a

'consistent and efficient' model selection strategy seems impossible. The question then becomes how best to proceed when no optimal model selection criterion exists.

A logical strategy is to consider how well models under consideration perform when applied to a common analytical goal: prediction. Posterior predictive checking, where replicate datasets are generated from simulated model parameters and used as the basis for checking model fit (Rubin 1984, Gelman et al. 1996), is an approach that has gained traction in the statistics literature. In posterior predictive checking, models are evaluated based on their ability to replicate the distribution of the data of interest. Many faux datasets are generated, using values sampled from the joint posterior distribution of the model parameters, to account for uncertainty in parameter estimates (Gelman 2003). Comparisons may be made quantitatively, through computation of predefined test metrics such as the posterior predictive *p*-value (Meng 1994), or simply by qualitative (visual) assessment (Gelman 2007). Posterior predictive checking is appealing in its simplicity; it shows us how well a model under consideration does at reproducing the data of interest, without relying on derived quantities that lack such a straightforward interpretation. This makes posterior predictive checking a highly interpretable method for evaluation in model selection problems.

The apparent lack of consistency among criteria for evaluating models implies a need for more rigorous approaches to the second step of the model selection process we outline above; statistical evaluation. We envision a more iterative approach to the model selection problem, where IT or Bayesian methods are coupled with stringent predictive checks to first cull poor models from consideration, and then to provide direct evidence (i.e. ability to generate replicate data resembling the data at hand) for model fit. Here, we

examine this approach by considering three case studies that represent a common, and particularly vexing, analytical problem for ecological data: variable selection for linear regression. Our objectives here are twofold: (1) compare the performance of several IT and Bayesian model selection criteria for a given set of models; and (2) consider how the results from posterior predictive checks compare to those of the model selection criteria. After presenting the results of this analysis, we conclude with practical recommendations for undertaking robust evaluation of models.

Background

Variable selection in linear regression

The variable selection problem arises when we have a large number of potential covariates $\{X_1, \ldots, X_q\}$ for some response variable *Y*, and we are interested in using the linear regression model $y = x_1 * \beta_1 + \cdots + x_p * \beta_p + \varepsilon$, where $\{x_1, \ldots, x\}$ is some subset of the full covariate set that best fits or predicts *Y* (Berger and Barbieri 2004, Liang et al. 2008, O'Hara and Sillanpää 2009). The primary motivations for choosing a 'best subset' are typically to develop a reduced predictive model for *Y*, or for developing hypotheses/drawing inference on how *Y* relates to the selected covariates. In either case, we are interested in the most parsimonious model that will meet our criterion (i.e. minimum squared error in the case of prediction). Scientists are often interested in both prediction and inference, and suitable analyses will look to foster both goals.

In both Bayesian statistics and maximum likelihood analysis, it is common to treat variable selection as a model comparison problem. In this case, we consider 2^q possible models, which contain all possible sub-models that may be constructed from the full covariate set (George and McCulloch 1993). From this, a candidate model or set of

models may be evaluated by the test criteria of the analyst's choice. This perspective is useful, but the number of models under consideration quickly becomes unwieldy with increasing *q*. For example, a variable selection system with 30 covariates can generate > one billion potential models. For model sets this large, performing even a simple AIC calcuation for each model is not feasible. Variable selection has attracted the interest of many statisticians in recent years, particularly in Bayesian statistics, and an array of simulation-based approaches now exist to address this problem.

Bayesian approaches to model selection

Model selection within the information theoretic framework, based on AIC and similar criteria, is widely used within the ecological literature and we refer readers to authoritative sources on this approach for additional background information (i.e., Burnham and Anderson 2002). Bayesian model selection approaches, particularly selection via posterior model probability and posterior predictive checking, remain somewhat underutilized in the ecological sciences. Johnson and Omland (2004) suggested that this is due to the computational requirements of these methods, as well as a lack of software that facilitates their easy implementation. These methods sometimes require specialized Markov chain Monte Carlo (MCMC) algorithms that are difficult for non-statisticians to 'tune' and use reliably. Over the last decade, however, there has been a significant increase in the tools available for performing Bayesian model selection. Here we include a brief introduction to the approaches we employ in our simulation study.

Bayesian model comparison is based on posterior model probability. Given data Y, the posterior probability of model M_i, that is contained within some set of models under consideration $\boldsymbol{M} = \{M_i\}$ can be defined as:

$$\rho_i = P(M_i|Y) = \frac{P(Y|M_i)*P(M_i)}{\sum_j P(Y|M_j)*P(M_j)} = \frac{P(Y,M_i)}{\sum_j P(Y,M_j)}$$

Where $P(Y|M_i)$ is the marginal likelihood of model M_i and $P(M_i)$ represents the prior information about M_i. From the above, it can be observed that the posterior model probability is simply the ratio of the joint probability of Y and M_i to the sum of the joint probabilities of the data and models over all possible models. The simplicity of this derivation lends a high degree of interpretability to PMP as a model evaluation criterion; it provides support for M_i by computing the probability of selecting it as a reasonable model given data Y and the other models in $\boldsymbol{M}$.

Classic Bayesian hypothesis testing compares two models M_1 and M_2, given data *D*, using the Bayes factor (Jeffreys 1961, Kass and Raftery 1995):

$$BF(M_1, M_2) = \frac{P(M_1|D)/P(M_2|D)}{P(M_1)/P(M_2)}$$

Note that we are simply looking at the ratio of the ratio of the posterior odds of M_1 to the prior odds of M_1(Ando 2010). A larger Bayes factor indicates a higher marginal likelihood for M_1 given the data *D*, with BF > 3 typically seen as providing substantial evidence for preferring this model over the other (Kass and Raftery 1995). Applying Bayes' theorem to each term in the numerator of (2) reveals:

$$BF(M_1, M_2) = \frac{\left[\frac{P(D|M_1)P(M_1)}{P(D)}\right] \Big/ \left[\frac{P(D|M_2)P(M_2)}{P(D)}\right]}{P(M_1)/P(M_2)} = \frac{P(D|M_1)}{P(D|M_2)}$$

This is the ratio of the probability of *D* given M_1 over the probability of *D* given M_2. Further, we can see that if the prior probabilities of the two models are equal, then:

$$BF(M_1, M_2) = \frac{P(M_1|D)}{P(M_2|D)}$$

which is a ratio of the posterior probabilities of the two models. This is a useful property of the Bayes factor in "objective" Bayesian analysis, where it is customary to assume that all models under consideration have equal, weakly informative priors.

The Bayes factor is a conceptually straightforward comparative metric of the evidence contained within two models, but it is often difficult to quantify in practice, and this seems to have limited its application for model selection problems. The difficulty arises from the need to compute an exact marginal likelihood from each model, which requires integration over the prior distribution (Kadane and Lazar 2004). Various methods for exact computation or asymptotic calculation of the marginal likelihood have been proposed (Gelfand and Dey 1994, O'Hagan 1995), though these approaches are computationally expensive. They require proper priors if one hopes to obtain closed form solutions, and calculation of the marginal likelihood can be problematic with the vague 'proper' priors that are routinely used in practical Bayesian statistical modeling (Bayarri et al. 2012). The Bayes factor is further limited in being a pair-wise comparison by definition, and it is cumbersome for multimodel inference with three or more models.

Recent advances in Markov Chain Monte Carlo (MCMC) sampling methods have made it possible to compute the model posterior probabilities directly, thus circumventing the need to calculate the marginal likelihoods (George and McCulloch 1993, Casella et

al. 2009, Ando 2010). If we include model selection as a step in the MCMC sampler, then posterior model probabilities are calculated from MCMC output as:

$$P(M_i|Y) = \frac{1}{L}\sum_{j=1}^{L}(\gamma_j = \gamma_i)$$

where L is the number of MCMC samples, and γ_i is a 'model indicator' related to M_i (Dellaportas et al. 2002, Ntzoufras 2009). Thus, we use a unique indicator related to each model under consideration and then track which model is selected in each iteration of the sample. From that output, the posterior model probability is calculated as the number of iterations in which γ_i is selected over the total number of samples.

While computation of the model posterior probability is easy, ensuring the MCMC algorithm will move among candidate models in order to obtain a representative sample can be challenging. This is especially true in variable selection problems, where hundreds or even thousands of models may be entertained (Kuo and Mallick 1998). Proper exploration of the model space is influenced by prior specification on the regression parameters, and there is an extensive literature on prior choice for linear regression models which we will not elaborate on here (but see Zellner and Siow 1984, Zellner 1986, Liang et al. 2008, Bayarri et al. 2012 and the references therein). The design of the algorithm can also affect the efficiency of the sampler, and many methods are available in the literature. O'Hara and Silänpää (2009) present an extensive review and performance comparison of these differing methods.

The principle drawback of these approaches is that a large number of samples must be collected in order to ensure that the algorithm has fully explored the model space. As a result, such algorithms can be computationally intensive. Green (1995)

introduced reversible jump MCMC (RJMCMC), which is a particularly efficient algorithm. RJMCMC 'jumps' between models from iteration to iteration, by drawing moves (i.e. addition or subtraction of predictor variables) from a proposal distribution, and then evaluating whether those moves should be made or the current model should be retained. This approach, coupled with modern computing technology, makes it easy to draw a sufficient number of samples even when the model space is very large. While practical Bayesian model selection was previously reliant on ad-hoc estimates of posterior probability, such as the Bayesian information criterion (BIC), these advances render the direction calculation of PMP convenient. This is advantageous for many problems, as the computation of BIC for large model spaces is impractical and it has been shown to be sensitive to sample size (Posada and Buckley 2004).

Model evaluation & posterior predictive checking

We now turn our attention to the question of model evaluation, which we approach using posterior predictive checks. Unlike model selection, which assesses the evidence for model fit based on the marginal likelihood of each model under consideration, posterior predictive checking evaluates models based on their ability to replicate the data at hand (Gelman et al. 1996). These checks are based on the posterior predictive distribution, which is defined according to:

$$P\left(Y_{pred}\middle|Y_{obs}\right) = P\left(Y_{pred}\middle|M_i\right)P(Y_{obs}|M_i)$$

Here, Y_{pred} represents the new data to be predicted, Y_{obs} are the data at hand, and M_i is the model under consideration. The Bayesian posterior predictive distribution is unique in that it uses the entire posterior distribution of the model, given an existing set of data to generate a probability distribution for each new observation, rather than a point estimate

(Ntzoufras 2009). Model fit can then be assessed based on some test statistic T by plotting Y_{pred} against Y_{obs} and estimating the tail area probability:

$$p_b(Y_{obs}) = P[T(Y_{pred}) \geq T(Y_{obs})|M, Y_{obs}]$$

This is the classical *p*-value, averaged over the posterior distribution of the model parameters in M, and corresponds to the 'posterior predictive *p*-value' first described by Rubin (1984).

The choice of T is important, and a method that detects discrepancies between the observed and replicated datasets should be selected (Meng 1994). Gelman (2000) reviews a number of potential choices for T, including simple measures of central tendency and variance (i.e. mean and standard deviation), as well as checks on the residuals between the observed and replicated data. The decision of which discrepancy measures to use is somewhat problem specific, and requires careful thought on the part of the analyst prior to undertaking the model checking endeavor. This challenge, as well as potential calibration issues related to the lack of a uniform distribution of the posterior predictive *p*-value under the null hypothesis (Bayarri and Berger 2004, Bayarri and Castellanos 2007), has led some authors to suggest that graphical diagnostics are more informative for model checking than are test statistics *per se* (Gelman 2003, 2007, Gelman et al. 2003). These checks may comprise the full range of standard diagnostic plots that are routinely used in the preliminary stages of data analysis; including comparison of the frequency distributions of the observed and replicated datasets via histograms, residual plots, and q-q plots.

Posterior predictive checking offers some major advantages as a method for model evaluation. Comparing how well a model reproduces a given set of data via

graphical diagnostics allows the analyst to consider the range of potential discrepancies that might exist between the observed and simulated datasets. These checks, as well as quantitative metrics such as posterior predictive *p*-values, are closely related to methods that most scientists find intuitive and easy to understand (Anderson et al. 2001). However, posterior predictive checking is impractical as a model selection strategy when many models are under consideration. Instead, it is best used for evaluating a narrowed list of models selected via other approaches. Posterior predictive checking has also been criticized for 'using the data twice' (Bayarri and Castellanos 2007), and thus it is recommended that it is combined with cross-validation procedures. We note in passing that empirical Bayes estimation, which today seems to be a reasonably well accepted method, was initially also subject to the criticism of using the data twice.

Methods

We considered three datasets, hereafter referred to as the 'protist data', 'forest data', and 'soils data'. All of these examples represent 'variable selection' problems in the linear regression context. Our goal is to identify datasets where a single response variable (Y) is considered as a function of a reasonably large ($1,\ldots,Q$ where $Q > 10$) set of predictor variables (X_q). It is important to note that these datasets typify the so-called "*M*-open" model selection problem (Key et al. 1999), where many models are considered and none are believed to be the 'true' model underlying the data. For these type of problems, the information theoretic approach has traditionally been viewed as preferable (Link and Barker 2006).

The protist data are that of McGrady-Steed et al. (1997), and it consists of 85 observations of cumulative CO_2 flux in a community microcosm experiment. The dataset

has 30 covariates, which are species richness and the abundance of each species included in the microcosms. The original study aimed to understand how diversity metrics, as well as presence/absence of member organisms in model communities, influenced ecosystem productivity (estimated as CO_2 flux). The forest data consists of 189 observations of total outside bark cubic foot volume of unthinned loblolly pine (*Pinus taeda*) from old field sites in Virginia. The data contains 11 covariates, which include age, mean diameter, tree basal area, and several other typical forest mensuration measurements. These data were collected specifically for the development of forest yield equations. The soils data contains 172 observations of forest soil organic carbon from plots on the Coastal Plain physiographic province of New Jersey, USA. The dataset has 18 covariates, including field measured soil organic matter content, vegetation indices extracted from Landsat data, and soil physical properties extracted from the USDA's SSURGO database. These are the same data used elsewhere in this dissertation for developing predictive models of forest SOC. A full list of the covariates in each dataset may found in appendix A, table A.1.

For each of the three datasets, we defined a model set M that consisted of the top five models, according to their posterior model probability, computed from a reversible jump Markov Chain Monte Carlo (RJMCMC) procedure using the 'Jump' extension for WinBUGS (Lunn et al. 2006). We retained the intercept term and employed statistical variable selection to determine the additional covariates that would be included in each model. We acknowledge that the choice to keep the top 5 models selected by RJMCMC is arbitrary. Here the goal is to compare the relative merits of several model selection approaches for a given set of models *M*, so we prefer to select *M* by an automated

approach. In a variable selection problem, *M* will consist of all possible models that may be generated from the chosen set of covariates. We can explore this model space within an MCMC by attaching an indicator parameter $\gamma_i$to each covariate in the full model, and tracking which variables were included at each iteration of the sampler (Ntzoufras et al. 2000, Dellaportas et al. 2002). Here, we used equal prior probabilities on the model indicator parameters ($\gamma_i \sim Bern(0.2)$), so that each model in *M* would have the same prior likelihood of being retained at each step in the MCMC.

For each dataset, we considered the following predictors: (1) 'best' model selected by posterior model probability (PMP); (2) model averaging with PMP; (3) the 'median probability model' based on PMP (Berger and Barbieri 2004); (4) 'best' model selected by the deviance information criterion (DIC) (Spiegelhalter et al. 2002); (5) model averaging with DIC; (6) 'best' model selected by Akaike's information criterion (AIC); and (7) model averaging with AIC. PMP for each model in the set was obtained via RJMCMC when defining the model sets. Model averaging was implemented by scaling the posterior probabilities of the selected models to sum to one. The median probability model, as proposed by Berger and Barbieri (2004), is constructed by including all covariates with a posterior inclusion probability > 0.5. This model can in some sense be viewed as providing a "model averaged" result, and it has been shown to provide superior predictive performance in certain cases (Berger and Barbieri 2004, Bayarri et al. 2012).

We used the corrected form of AIC that accounts for small samples, which is recommended by Burnham and Anderson (2002) as standard practice. We also included DIC in our analysis as it is widely used for model selection in the Bayesian framework in

ecology and evolution. While DIC is a Bayesian procedure it is based on a decision theoretic argument, which differs from methods based on posterior model probability (Spiegelhalter et al. 2002). Model averaging for DIC and AIC is based on 'AIC weights', using the approach advocated by Burnham and Anderson (2002). Table 2 displays the posterior model probabilities, AIC, and DIC scores for all models within each model set. PMP and DIC were calculated by fitting the full dataset as a hierarchical model in the BUGS program, while AIC was derived by fitting the models via ordinary least squares in the R programming language.

We now describe the procedure we devised to compare and evaluate the performance of each predictor via posterior predictive checks. First, the datasets were randomly divided into 'fitting' and 'validation' sets, where approximately 66% of the observations were used to fit the models and the remaining 33% were retained for validation. For each model, the fitting data were used to draw 10,000 parameter samples from the posterior predictive distribution. All models were fit as hierarchical models, with vague priors and hyper-priors on the regression coefficients ($\beta_j \sim N(0, 1000)$) and model variance term ($\tau^2 \sim Gamma(0.001, 0.001)$). Model selection was performed on the fitting data for each dataset, using posterior model probability, DIC, and AIC in order to construct the seven predictors described above.

In the case of AIC, it is important to note that we used it to select and rank models in a non-Bayesian context, using the selected models fit via ordinary least squares (OLS), then we proceeded to make predictions with the AIC-selected models via the posterior predictive distributions obtained from fitting the same models in a Bayesian framework. This differs from the normal use of AIC within the literature, but is appropriate given our

goal of comparing the performance of different model selection criteria based on posterior predictive checks. Naturally we wished to remain consistent in how these tests were performed. Burnham and Anderson have argued that AIC has a Bayesian derivation (2004), so we do not see an inherent conflict in using it in conjunction with Bayesian prediction.

For posterior predictive checking, we adopted a 'hold-out predictive' approach, where the data are randomly divided into fitting and validation sets. In our study, we used approximately 66% of each dataset for model fitting, and retained one third for validation. Parameter samples drawn from the joint posterior distribution of the fitting data were applied to predict the response variable, based on the observations of the covariates in each model from the validation data (Vehtari and Ojanen 2012). This approach differs from typical model checks using the posterior predictive distribution, where fitting and validation are accomplished using the full joint posterior distribution of all the data available. In our case, we prefer using an independent hold-out set to provide a more stringent test, and to ameliorate the issues associated with posterior predictive checks related to "using the data twice" (i.e. for parameter estimation and validation) (Bayarri and Castellanos 2007, Gelman 2007). We are confident that models which pass these independent posterior predictive checks will be reasonable choices for prediction with each dataset.

For each set of parameter samples (n=10,000), simulated values of the response variable, based on the validation data, were retained for the seven predictors we considered (i.e. 'best' model and model averaged predictors with PMP, AIC, and DIC, as well as the median probability model according to PMP). Posterior predictive checks

were performed using the 2.5 percent and 97.5 percent credible interval bounds of the simulated datasets. We computed a posterior predictive '*p*-value' to check whether the CI bounds of the observed data were within the 95 percent credible interval of our simulated values. Additionally, the predictors were evaluated with root mean squared error (RMSE) and mean absolute error (MAE) by comparing the average value of the 10,000 observations we retained from the simulation. Model selection was then conducted using these posterior predictive checks, as well as information supplied by considering PMP, DIC, and AIC for the model sets.

Results

In general, the three model selection criteria were not consistent in the models they selected (Table 2.1). The soils data were the only case where PMP, AIC, and DIC agreed on a 'best' fitted model, and even here there were differences in relative support for the candidate models among each predictor. PMP has strong support for the model containing only soil organic matter (SOM) and the intercept, while both AIC and DIC consider one or more additional models as reasonable competitors. A similar effect is also noted in both the soil and forest data. The forest data provides the most striking example of disagreement among each method, with the three criteria each selecting a different 'best' model. These results show that proceeding with both top model selection and model averaging with each method may result in considerably different predictors within each dataset.

Table 2.1: Model structure, posterior model probability (PMP), deviance information criterion (DIC), and Akaike's information criterion (AIC) for the five models considered for each dataset. Asterisks indicate the top model according to each method.

a. Protist Data

Model	PMP	DIC	AIC
x5 + x16	7.0E-5*	863.2	864.2*
x5	5.44E-5	860.4*	865.2
x5 + x10	4.66E-5	865.6	866.9
x5 + x12 + x16	4.55E-5	863.5	866.3
x5 + x7 + x16	4.08E-5	864.1	866.6

b. Forest Data

Model	PMP	DIC	AIC
x5 + x6 + x7 + x8	0.218*	-475.1	-251.2
x1 + x2 + x5 + x6 + x7 + x10	0.114	-498.9*	-251.1
x5 + x6 + x8	0.11	-455.9	-244.7
x5 + x6 + x8 + x11	0.099	-469.8	-250.1
x6 + x7 + x8	0.092	-461.2	-251.7*

c. Soil Data

Model	PMP	DIC	AIC
x3	0.825*	53.8*	41*
x3 + x9	0.066	55.87	44.1
x3 + x4	0.036	55.11	43.9
x3 + x8	0.019	55.68	41.9
x3 + x11	0.012	55.41	44

Table 2.2 presents the results of the posterior predictive checks we performed on each predictor. In general the results of two checks, squared error terms and the 95% quantiles, differed in whether they preferred 'best' model or model averaged predictors. The model averaged predictors reduced error better than single models, with PMP model

averaging performing the best for the protist and forest data, and DIC model averaging performing the best for the soils data. For all three datasets, AIC model averaging performed the worst among the model averaged predictors.

Conversely, posterior predictive checks on the 2.5 and 97.5 percent quantiles reveal that model averaging tends to pull prediction towards the mean, overestimating the lower credible interval bound and underestimating the upper. For the protist data, both CI bounds for all three model averaged predictors are significantly different than the observed CIs ('p' < 0.05). For the forest and soil data, the effect is still present but less apparent. In fact, model averaging with PMP is a reasonable predictor according to this diagnostic. Histograms of the simulated 2.5 and 97.5 credible intervals for each dataset can be found in appendix A (Figs. A.1-A.3).

Table 2.2: Results of the posterior predictive checks: mean root mean squared error (RMSE) and mean absolute error (MAE), as well as posterior predictive *p*-values for the lower and upper credible interval bounds. * Indicates a significant difference at the α = 0.1 level (i.e. the 90% credible level), while ** indicates significance at α = 0.05. Graphical diagnostics for these checks may be found in appendix A (figures 1-3).

a. Protist Data

Predictor	RMSE	MAE	lower	upper
PMP, top model	845.9	702.3		
PMP, model av.	**683.8**	**544.6**	**	**
DIC, top model	863.6	717.7		
DIC, model av.	738.4	596.3	**	**
AIC, top model	845.26	702.33		
AIC, model av.	697.32	557.15	**	**
MPM	871.8	725.2		

b. Forest Data

Predictor	RMSE	MAE	lower	upper
PMP, top model	0.048	0.04		
PMP, model av.	**0.04**	**0.033**		*
DIC, top model	0.052	0.043		
DIC, model av.	0.046	0.038	*	**
AIC, top model	0.05	0.041		
AIC, model av.	0.044	0.037		**
MPM	0.049	0.04		

c. Soil Data

Predictor	RMSE	MAE	lower	upper
PMP, top model	0.81	0.63		
PMP, model av.	0.78	0.59		
DIC, top model	0.81	0.63		
DIC, model av.	**0.71**	**0.52**		**
AIC, top model	0.81	0.63		
AIC, model av.	0.72	0.53		**
MPM	0.81	0.63		

Discussion

Our results highlight a need for careful evaluation of models following selection. In our study, conducting model selection with IT and Bayesian criteria led to divergent results. In each case, proceeding based on only one criterion presents a risk of ignoring predictors with superior performance. This is best exemplified in the Soils Data, where PMP assigned very low marginal likelihoods to models 2-5, but DIC retained three of these as competitors to model 1. Model averaging with DIC turned out to be a good predictor for these data, and considering PMP only would have led us to dismiss reasonable models off-hand.

These differences in performance among the datasets may be in part explained by features of the data themselves. The protist data is a case where none of the models fit the data well, as evidenced by their low posterior probabilities (Table 2.1). By contrast, the forest data model set possesses several reasonable alternatives, with reasonably competitive PMPs and AIC scores, while one 'best' model was preferred by all three statistics in the soils data. Examining posterior inclusion probabilities for the covariates in these datasets (appendix A, table A.1) shows a large number of variables in the protist data that do not have strong relationships with the response variable. The forest data is somewhat the opposite case, with many covariates possessing high posterior inclusion probabilities, while one variable (soil organic matter content) is strongly favored for the soils data. Thus, in the case of the protist and forest data, it is perhaps not surprising that PMP, DIC, and AIC would to different results, given that they available models compete closely with one another. Of course, the difference is that none of the models are meaningful for the protist dataset, while all provide good predictions in the forest data.

For the soils data, where there was less parity among candidate predictor variables, all three statistics converged on the same ‘best’ model.

In these case studies, subsequent evaluation of models based on simulated data was important for identifying which predictors were the best at reproducing observed values. Posterior predictive checks using both squared and absolute error, as well as the 2.5 and 97.5 percent quantiles, revealed that regardless of the criterion used model averaging tended to misestimate the CI bounds. On the other hand, model averaging often led to significant reductions in bias according to root mean squared error (RMSE) and mean absolute error (MAE). These features were only made apparent by examining multiple tests for model support.

Considering these results, the decision on whether or not to model average is dependent on the goals of the analysis. If the aim is to generate an aggregate sample or population mean, we may be more interested in reducing bias at the expense of missing more extreme values in the dataset. Conversely, if we are more interested in model-based inference, such extremes may be important to consider. Note however that it was the posterior predictive checks that revealed these trends and that failing to perform robust model evaluation following selection with IT and/or Bayesian criteria would have led us to ignore potentially important features of our models.

It is interesting to note that when comparing ‘best’ model and model averaged predictors, posterior model probability did as well or better than AIC and DIC for both the protist and forest data; two datasets where our results suggest multiple models should compete closely with one another. This is despite the expectation for AIC to perform better than Bayesian criteria in “*M*-open” problems such as variable selection (Burnham

and Anderson 2004, Yang 2005, Aho et al. 2014). Many authors have suggested that the class of Bayesian model selection criteria related to the Bayes factor may be superior to IT approaches when we do believe the 'true' model is present in *M* (i.e., "*M*-closed" cases) (Kass and Raftery 1995, Berger and Pericchi 1996, Johnson and Omland 2004, Link and Barker 2006), but it is not clear that this implies inferior performance in "*M*-open" problems. Further, model averaged predictors with DIC and AIC did better than PMP for the soils data, where all three statistics agreed on the 'best' model, suggesting that in this case PMP over-penalized reasonable models relative to DIC and AIC.

The frequentist properties of posterior predictive checking make it a method with a high degree of interpretability for most scientists. Anderson et al. (2001) have previously promoted predictive checks as a straight-forward method for displaying the results of complex Bayesian models in the ecological literature. We certainly agree with this sentiment, but argue that analysts using the maximum likelihood and frequentist paradigms should utilize these methods as well. Information theoretic measures of model fit have no absolute meaning, and though they can identify the "best" among a group of candidate models, they do not indicate how well the "best" model actually fits the data. Here, posterior predictive checking provides a simple method for evaluating models that have been selected using IT model choice. Frequentist statistical tests are typically premised on the computation of *p*-values, though in standard applications the distribution of 'future trials' upon which the *p*-value is based never exists (Spiegelhalter et al. 2002). By basing frequentist checks on many simulated datasets, posterior predictive methods avoid this significant pitfall (Rubin 1984). While our checks were performed based on the posterior predictive distribution, an inherently Bayesian concept, similar approaches

could be devised based on alternative sampling strategies such as bootstrapping or simulated annealing.

Taken together, the three analytical scenarios we present here provide a strong case for the need to carefully evaluate models following selection. We have noted that the model selection paradigm used affects predictive performance, and that in our examples model averaging requires a tradeoff between complete coverage of the credible interval and minimizing predictive bias. The analytical scenario we present; comparison of several nested linear models; is a common problem for many ecologists, and it follows that these issues would occur in other linear model selection problems. To address this, we offer the following practical suggestions for carrying out model selection and evaluation:

1. Carefully consider the models, or covariates, of interest *a priori*: As others have observed, all models under consideration should represent reasonable hypotheses for the question at hand (Hoeting et al. 1999, Burnham and Anderson 2002). In the case of variable selection problems, it is necessary to only include covariates for which there is a reasonable expectation that they could be a good predictor for the response variable. Automatic variable selection can be very useful when there are many potential covariates and little is known about their relationships to the response variable prior to the analysis (Murtaugh 2009, Guan and Stephens 2011, Garcia-Donato and Martinez-beneito 2013). However, the resulting models should be carefully considered before they are submitted for model selection.

2. Consider multiple methods when selecting among candidate models. In these case studies we found posterior model probability, DIC, and AIC to all be

reasonable methods for selecting models, even though they provided divergent results. It may in fact be reasonable to consider both information theoretic and Bayesian approaches when undertaking model selection. PMP provides a very different kind of evidence in support of one model over another when compared to AIC or DIC, as it is based on a different concept for evaluating model fit (Ando 2010). Even if one's particular statistical philosophy prefers the use of one model selection procedure over another, we believe it is good practice to see how the results of an alternative method compare prior to proceeding with the analysis.

Determining how many models to include from the initial set also requires careful consideration. In our analysis we simply selected the top five models according to posterior model probability, since our goal was only to compare the relative performance of several model selection methods with a given set of models. However, we acknowledge that this is a suboptimal approach when the analysis aims to make meaningful inference on or prediction of ecological data. There is a large literature on this question (*see* Hoeting et al. 1999, *and* Burnham and Anderson 2002 *as well as their references)*, and we refer readers to authoritative sources reviewing different methods for approaching this problem.

3. Evaluate both single and model averaged predictors: While model averaging is often suggested as standard practice, the results of our model evaluation suggest that model averaging may sacrifice accurate coverage of the confidence interval in order to minimize uncertainty in prediction. In many cases, such as the Forest and Soils data, this may be inconsequential, and in these situations model averaging is still preferred in order to account for model uncertainty in the analysis (Raftery et al. 1997).

However, our evaluation revealed significant bias towards the mean from model averaging within the Protist dataset. For situations such as this, where there are many weakly correlated predictor variables but no information is available to dismiss them *a priori*, using a single model constructed from the best covariates may be logical. The median probability model of Berger and Barbieri (2004) represents one formal attempt to deal with this issue, but this approach generally performed poorly in our analyses.

4. Conduct model evaluation with simulation based methods, such as posterior predictive checks: Ultimately, the best way to evaluate a model is to see how well it reproduces independent observations of the response variable. Posterior predictive checking is a rich and flexible approach for accomplishing this, and we recommend that ecologists adopt it as standard practice. In our analysis, we were able to simulate a large sample (N = 10,000) of replicate datasets using 'coda' output from the BUGS software and a simple simulation model. Posterior predictive checks on the 2.5 and 97.5 percent CI bounds revealed important features of the predictors we considered that would have been overlooked by standard cross-validation procedures. This general approach can be adapted to compare the distribution of model parameters and other aspects of the data, and explicit quantitative '*p*-values' are easy to compute.

The 'hold-out predictive' approach that we adopted, where the posterior distribution of model parameters is derived from a fitting dataset, and prediction is accomplished using a separate 'validation' set of covariate observations, provides a stringent test for model predictive performance. We recommend that this approach is employed when possible (i.e., when there is a suitably large number of observations to facilitate data-splitting). When smaller datasets are available, it is possible to combine

posterior predictive evaluation into simulation-based cross validation procedures, where one or several observations are held out for prediction at each iteration.

Conclusion

Thanks to rapid advances in statistical methodology for model selection, scientists now have more options available to them than ever before. While there is significant theoretical literature devoted to debating the merits of the major model selection paradigms, particularly Bayesian model selection and Information Theoretics, their relative performance in applied contexts remains an open question. We have presented three case studies in which we compared several different model selection approaches, and submitted the resulting predictors to formal evaluation via posterior predictive checking. While our results confirm our general preference for posterior model probability, these data demonstrate that the relative performance of these model selection paradigms is context dependent. We believe it is logical to examine both, and that the model or models with which the analysis proceeds, as well as the decision to model average or not, should be guided by the data and the goals of the analysis. Regardless of how model selection is accomplished, we advocate that model evaluation with simulation-based methods is adopted as standard practice.

Chapter 3: Quantifying spatial dependence in forest soil carbon and nitrogen pools across multiple spatial scales.

Abstract

Forest soils contain a globally significant terrestrial carbon stock, so accurate accounting of forest soil carbon pools is necessary for any efforts to mitigate the effects of climate change. National forest inventories play an important role in forest C accounting, including forest soils, but predictive models are necessary to "scale up" plot-level observations to broad spatial scales. Specifically accounting for spatial autocorrelation within these models will be necessary in calibrating soil C stocks from inventory data, but the spatial dependence of soil C at broad spatial scales is poorly understood. Here we conduct exploratory analyses aimed at characterizing the spatial structure of soil chemical inventories over regional to continental spatial scales, considering both soil carbon data and its spatial covariance with soil nitrogen. We demonstrate that spatial autocorrelation is not reliably estimable over micro and meso scales (within a single physiographic province), but is detectable and substantial at macro (both national and continental) scales. We show, however, that spatial covariance between soil C and N is detectable at all scales. Our results suggest that allowing for spatial autocorrelation will be necessary for improving the accuracy of predictive models and for fully quantifying uncertainty in national and continental forest soil C stocks, and that co-modeling soil nitrogen with carbon may be beneficial across all three scales we consider. In addition to discussing the implications for current forest soil monitoring

schemes, we conclude by discussing future research needed to formulate general expectations about spatial structure in soil C data.

Keywords: Forest soil carbon, geostatistics, macroscale, variograms, spatial hierarchical modeling

Introduction

Soil contains the largest pool of terrestrial carbon, so coordinated international efforts to mitigate the effects of climate change require reasonable baselines of terrestrial C stocks. Forest soils contain approximately 48% of that soil inventory (Pan et al. 2011). Given the large contribution of forested lands to carbon sequestration, full and accurate estimates of forest soil organic carbon (SOC) are important (Goodale et al. 2002, Shvidenko et al. 2010). The United Nations initiative on "Reducing Emissions from Deforestation and forest Degradation" (REDD+) funds the next generation of forest inventories in developing nations (Edwards et al. 2010, Maniatis and Mollicone 2010), and understanding how best to incorporate soil sampling into broad scale forest inventories will contribute to improving global forest carbon stock baselines.

Improving predictive models for forest soil inventory data constitutes an important step towards improving global forest soil C stocks. Traditional broad scale approaches for forest soil carbon accounting rely on either inventory-based analyses (Eswaran et al. 1993) or on 'mass balance' methods, where the SOC pool is estimated based on other forest C pool and flux estimates (Dixon et al. 1994, Goodale et al. 2002, Pan et al. 2011). Digital soil mapping (DSM), where prediction of soil attributes is

propagated across a rasterized grid that covers the study region, has received attention as an alternative method (McBratney et al. 2003, Grunwald 2009, Minasny et al. 2013). This results in high resolution maps of soil attributes, generated by predictive statistical models, which offers significant potential for improving predictive accuracy and uncertainty quantification. Developing a good predictive model requires a thorough understanding of the factors that introduce uncertainty into model estimates. When spatial autocorrelation is not included in predictive models, error arising from spatial processes underlying soil attribute data may mistakenly be attributed to model error, thereby reducing the predictive accuracy and precision of soil nutrient models. (Grunwald 2009, Minasny et al. 2013). By leveraging analytical techniques capable of explicitly modeling spatial dependence, it may be possible to improve the performance of predictive models for forest soil carbon and other nutrients substantially.

Further improvements may be achieved by coupled spatial modeling of forest SOC with related nutrients, such as soil Nitrogen. Given the close relationship between soil carbon and nitrogen (Chapin et al. 2002), spatial dependence in the residual error of forest soil C and N may offer improved predictive performance when these variables are modeled jointly, even if there is little spatial structure in the soil C data. However, whether considering spatial dependence within or among variables, these improvements are dependent on a modeling approach that captures apparent spatial autocorrelation in soil inventory data at scales relevant to national and international forest inventories.

Geostatistical practice encompasses a broad suite of techniques from both Bayesian and frequentist statistical paradigms, united in attempting to explicitly model

spatial autocorrelation among geo-referenced samples (Goovaerts 1997). These methods typically facilitate interpolation by fitting a model, such as an exponential or Matérn-class function, to a spatial covariance matrix for all pairwise points within the sample (Isaaks and Srivastava 1989, Banerjee et al. 2004). Changes in variance among samples, as a function of distance, can be used to guide predictions at unsampled locations. A substantial body of literature exists demonstrating that geostatistical models can lead to improved predictions when interpolating environmental variables at the landscape scale, including soil nutrients (Finley et al. 2010, Mishra et al. 2010, Guhaniyogi et al. 2013).

Geostatistics offers powerful tools for interpolating point data across smooth surfaces, but the utility of these techniques for mapping forest soil carbon and nitrogen remains poorly explored. These methods can improve the prediction accuracy, relative to non-spatial regression models, provided that there is sufficient spatial structure with the point data themselves. The ability to detect spatial dependence can depend on several factors, including sampling design - the distribution and density of samples across the study region (Fortin et al. 1989), the scale of measurement error (Cressie 1993, Goovaerts 1997), and the overall heterogeneity of factors influencing soil formation (Ettema and Wardle 2002). Many existing soil databases were not generated with geostatistical applications in mind, and in these cases the utility of spatial methods in digital soil mapping remains unclear (Minasny et al. 2013). Moreover, soil formation is a complex process, and many environmental and anthropogenic factors influence the accretion of soil carbon and nitrogen (Jenny et al. 1968, Chapin et al. 2002). Spatial patterns within these processes may either contribute to spatial dependence within soil nutrient data, or may make it more difficult to detect (Vasques et al. 2012, Clough and Green 2013).

Detecting and modeling spatial dependence may be particularly challenging when the spatial scale is restricted. While several studies have shown spatial dependence of soil carbon data across large, heterogeneous landscapes (Malone et al. 2009, Marchetti et al. 2010, Mishra et al. 2010, Zhang et al. 2012, Vasques et al. 2012, Toth et al. 2013a, de Brogniez et al. 2014), most studies of spatial dependence in forest soils have focused on small spatial scales (McBratney et al. 2003, Minasny et al. 2013). Spatial variation at fine scales is typically assumed to be unimportant for either regional or national estimates, and is incorporated into the estimated 'nugget effect' (Banerjee et al. 2004, Finley et al. 2010).

There is some evidence that spatial dependence exists in forest soil attributes at broader scales, and McKenzie and Ryan (1999) used geostatistical methods to improve forest soil carbon predictions for a 50,000 ha national forest in Australia. Spatial dependence has also been noted for both soil organic carbon and soil pH within a large forested watershed in the Medicine Bow Mountains, Wyoming, USA (Rahman et al. 1996). At larger spatial scales, Liski and Westman (1997) used a block (based on 16 km^2 sampling units) kriging design to interpolate soil carbon for all forested lands within Finland, based on that country's National Forest Inventory data. Bernoux et al (2006) also found kriging preferable to non-spatial models for a large forested region (approximately 334,000 km^2) in the Amazon basin of Brazil. However, in the latter study a large geostatistical 'nugget effect' suggests that much of the spatial dependence was limited to scales smaller than those the authors considered.

As with soil carbon, spatial dependence in soil N has mainly been studied at fine scales. Examples are available from agricultural systems (Sutherland et al. 1993, Ruffo et al. 2005), grasslands and scrub shrub communities (Gallardo and Parama 2007), and forests (Lister et al. 2000, Wang et al. 2007). Several studies have considered spatial dependence in soil N for mixed use landscapes, and have applied geostatistical models to improve predictions of this and other soil nutrients (Qu et al. 2012, Wang et al. 2013). However, as with carbon, the spatial patterns of soil N at the regional scale have not been widely described. Similarly, spatial co-dependence in soil C and N at broad scales has not been explicitly explored, though Watt and Palmer (2012) found significant spatial structure in soil Carbon:Nitrogen ratio from an inventory dataset collected across New Zealand.

Geostatistical modeling of forest nutrients at larger scales (> 100,000 ha) has not been common, because the data accommodating such efforts have only recently become available (McBratney et al. 2003, Poggio and Gimona 2014). Traditional soil surveys are designed to map soil taxonomy and related properties for aggregated areal units, and typically do not possess a sufficient density of sampling points to facilitate geostatistical modeling within a single land cover type. More recently, the European Union has established the LUCAS topsoil database (Tóth et al. 2013b), providing quality controlled, point-referenced observations (approximately 20,000 plots total) of soil C and related variables for the 23 member states of the European Union. Additionally, soil carbon is now part of many national forest inventory programs, including the US Forest Inventory and Analysis (FIA) database, and the Finnish National Forest Inventory (NFI), providing opportunities to evaluate spatial dependence of forest SOC at national inventory scales.

The objective of this study is to explore how spatial dependence in soil carbon and soil nitrogen, as well as spatial covariance between them, changes across multiple spatial scales. We characterized the spatial structure of forest soil nutrient data at three different spatial scales: within a single physiographic region (Coastal Plain of New Jersey, USA), across a national scale landscape (national forests of Germany), and across continental scales (forested lands within the European Union). To guide our study, we proceed with the following hypotheses: (1) Spatial autocorrelation will exist at all three scales for both soil carbon and nitrogen; (2) The distributions of soil carbon and nitrogen will be coupled across space at all scales; and (3) Spatial dependence will be better defined for more heterogeneous landscapes, where sampling regions cut across multiple forest communities, geologies, and broad environmental gradients.

Methods

Datasets

We consider two principle datasets for this study: (1) a forest soil inventory conducted by the authors in 2011 and 2012 for the Coastal Plain physiographic province in New Jersey, United States, and (2) the LUCAS topsoil database, which covers the 23 member states of the European Union and is made publicly available by the European Commission's Joint Research Centre. Both datasets possess geo-referenced observations of soil organic carbon and soil nitrogen, scaled by field measured soil bulk density and expressed in units of Mg/ha^2. The NJ data represent 172 point observations, sampled only within forests in a stratified random design, based on forest community type and soil drainage class. The LUCAS database consists of approximately 20,000 point

observations, covering a range of land use land cover (LULC) classes. Of these the 4,818 sample points were located within forests, as determined by rasterized LULC datasets available as part of the LUCAS survey, were retained for our study. The NJ data cores were sampled to a depth of 30 cm, while those for the LUCAS dataset were sampled to 20 cm. Thorough descriptions of these datasets can be found in (Clough and Green, 2013) and (Toth et al. 2013a, 2013b), respectively.

Using these datasets, we examined spatial dependence in soil carbon and soil nitrogen at regional, national, and continental scales. Table 3.1 summarizes the number of observations, study area, minimum distance between points, and observation density (plots/km^2) in each dataset. Note that the plot density is higher for our regional dataset and approximately equal for the national and continental samples, and that at all scales the sampling densities are small relative to the region of interest. This is typical of regional soil inventories, given the effort and expense required to collected point referenced soils data. As a result, we focus our investigation only on spatial dependence at regional (NJ), national (DE), or continental (EU) scales. This is important to note, because spatial dependence may also exist at finer scales than our inquiry. Maps of each study region, including the distribution of the sampling locations, are shown in Figure 3.1.

Table 3.1: Location and sampling distribution of the regional, national, and continental datasets

Study Region	Total #/observations	Study area (km^2)	Sampling density (plots/km2)
New Jersey, USA (NJ)	**172**	**130,140**	**0.1**
Germany (DE)	**382**	**357,790**	**0.001**
European Union (EU)	**4,818**	**4,055,580**	**0.001**

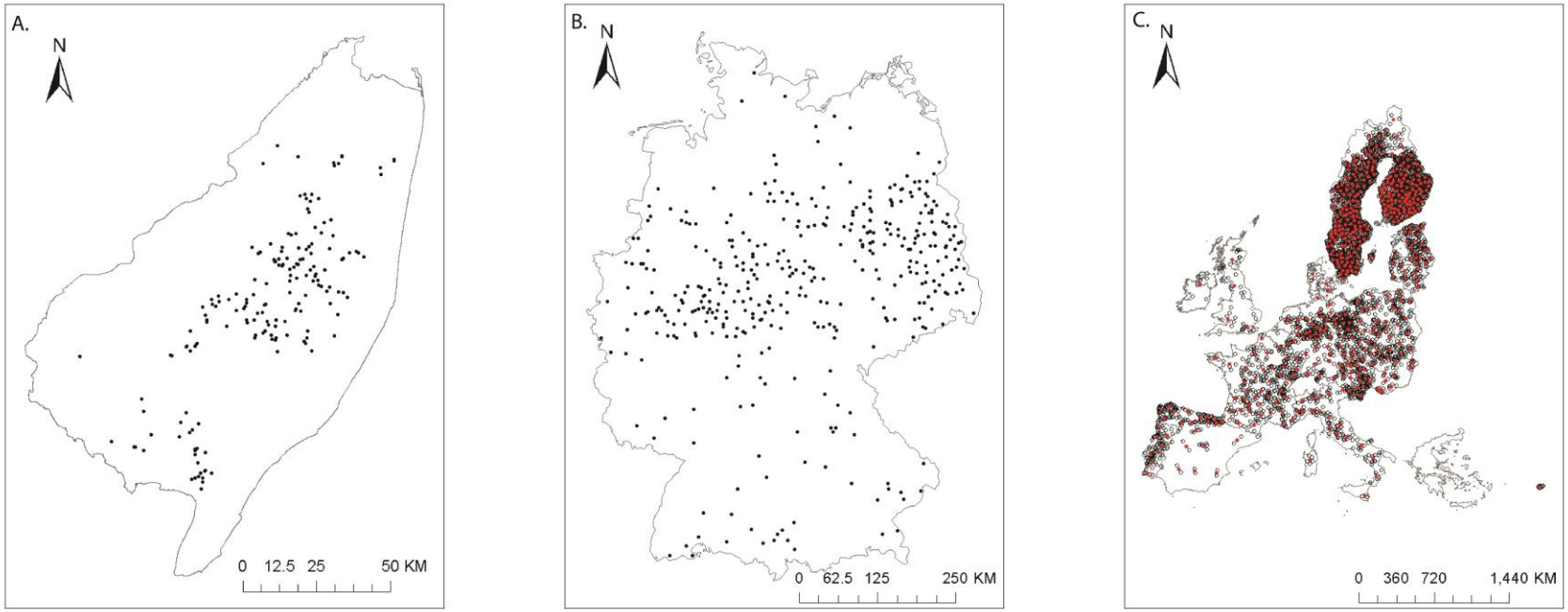

Figure 3.1: Sampling locations within the three study regions: the Coastal Plain of New Jersey (A.), Germany (B.), and the member states of the European Union (C.). Note that in the EU map, the sampling locations are indicated by open black circles, while the "knots" used to model the spatial random effects re marked in red.

Data analysis

Our analysis consisted of two steps: (1) exploratory analysis using semivariograms; and (2) mapping of spatial random effects from univariate spatial Bayesian hierarchical models (Banerjee et al. 2004), with direct comparison to non-spatial residuals. Both analyses were applied to soil carbon and nitrogen to understand the spatial structure within each variable, as well as to the residuals between carbon and nitrogen components. Here we use spatial structure within residuals to characterize the joint spatial distribution between SOC and soil nitrogen.

Variogram analysis

Semivariance, a measure of pair-wise divergence between all combinations of sampling locations, can be related to spatial separation by computing an empirical semivariogram (Isaaks and Srivastava 1989):

$$\gamma(h) = \frac{1}{2|N(h)|}\sum_{N(h)}(z_i - z_j)^2$$

Where $N(h)$ is the set of all pairwise distances such that $|x_i - x_j| = h$, $|N(h)|$ is the number of unique pairs in $N(h)$, and the z's are the measurements of the variable of interest at locations i and j respectively. As indicated above, we computed empirical variograms for carbon, nitrogen, and carbon/nitrogen residuals at each of the three scales we considered. In our analysis, we assume any spatial dependence within our data will be isotropic (i.e., h only represents distance, not direction), and stationary (Cressie 1993).

We fit each empirical variogram with an exponential covariance function (Goovaerts 1997, Ribeiro and Diggle 2001):

$$\gamma(h;\ \phi, \sigma^2, \tau^2) = \ \tau^2 + (\sigma^2)\left(1 - e^{-\frac{3h}{\phi}}\right) \ \ , h > 0$$

Here ϕ is the effective range of the spatial dependence, σ^2 is the *partial sill* and τ^2 is the geostatistical nugget effect, representing both "white noise" error and spatial dependence below our scale of inquiry (Banerjee et al. 2004, Finley et al. 2013). This procedure produces a 'theoretical variogram', which can be plotted against the empirical variogram. Inspection of these variogram models can reveal important information about the spatial dependence within the dataset.

Hierarchical modeling

In order to characterize spatial dependence across the study domain, we fit univariate spatial Bayesian hierarchical models to our variables of interest, within each region. These models are of the form:

$$y(s) = \ \mu(s) + w(s) + \ \varepsilon(s) \ ,$$

where $\mu(s) = \ \beta_0$ is the mean at location s, $w(s)$ is a spatial Gaussian process, and $\varepsilon(s) \sim N(0, \tau^2)$ is a typical error term, representing measurement error and other 'random effects', as well as microscale spatial dependence; i.e., the geostatistical "nugget effect" (Cressie 1993, Banerjee et al. 2004).

The $w(s)$'s, frequently termed the spatial 'random effects', represent the portion of the model error that is explained by a spatial process model; in our case an exponential covariance function. They are of particular interest for characterizing spatial dependence within point-referenced datasets (Finley et al. 2007a). Within a hierarchical framework, the spatial random effects are specified as:

$$w|\sigma^2, \phi \sim N(0, \sigma^2 R(\phi))$$

Where R is a valid spatial autocorrelation model (e.g., an exponential function), and $\sigma^2$and ϕ are the sill and spatial range parameters as previously described (Banerjee et al. 2004, Finley et al. 2013).

These models are simply an extension of a standard, non-spatial linear model. The key difference is that the model variance is decomposed into a spatial ($w(s)$) and non-spatial ($\varepsilon(s)$) component. This allows us to explicitly model the spatial random effects as described above, and facilitates direct comparison with the non-spatial residuals. In this way, we can examine how much variance is explained by the spatial information, and how much remains as residuals (i.e., measurement error and microscale spatial variation). Surface plots, interpolations of both non-spatial residuals and the spatial random effects across the study region, are a convenient way to visualize and understand these results.

Spatial dynamics of soil carbon and nitrogen

Both of these analyses were applied to observations of soil C and N for each of the three datasets described above (regional, national, and continental data). In the first stage of the analysis, we computed empirical variograms and fitted theoretical variogram

models to these results. The empirical variograms were fitted with exponential covariance functions via nonlinear least squares, and initial values for the sill (σ^2) and range (ϕ) were derived from visual inspection of the empirical variogram plots. For this analysis, we assume an isotropic spatial pattern for both soil C and N. All variogram analyses were conducted using the geoR package for the R statistical computing environment (Ribeiro and Diggle 2001).

Following variogram analysis, we fit univariate spatial Bayesian hierarchical models to both the soil C and N data. For this analysis, we placed a flat prior on the model intercept (β_0), and 'vague' inverse gamma priors on $\sigma^2$and τ^2 (e.g., $\sim IG(0.001,0.001)$). For the range parameter ϕ, we used an weakly informative uniform prior with a restricted domain around the estimated spatial range from fitting the theoretical variograms. These specifications are as follows: regional, $\phi \sim Unif(\frac{3}{40{,}000}, \frac{3}{80{,}000})$; national, $\phi \sim Unif\left(\frac{3}{175{,}000}, \frac{3}{225{,}000}\right)$; continental, $\phi \sim Unif(\frac{3}{375000}, \frac{3}{425000})$. The bounds on these priors are derived by considering an effective spatial range D_0 as the distance where the correlation between locations equals 0.05. Assuming an exponential correlation function, solving the equation $0.05 = \exp(-\phi * D_0)$ reveals that $D_0 = -\log(0.05)/\phi$, or approximately $3/\phi$. The model was fit via a Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm, and the MCMC chains were initialized using estimated values of the spatial parameters from the variogram fitting. A posterior sample of the spatial parameters was obtained from 2,500 MCMC iterations, following a burn-in period of 7,500 iterations.

In the case of the EU data, fitting a "full rank" spatial hierarchical containing all 4,818 observations of soil C is not computationally feasible. Sampling the spatial random effects for an n x n spatial covariance matrix, where n = 4,818, requires a huge computational expense; well beyond what standard modern computers are capable of providing. To avoid this issue, we fit the EU data with a 'low rank' predictive process model (Banerjee et al. 2008, Wikle 2010, Finley et al. 2012), where the spatial random effects were modeled using a set of 500 "knots" distributed across the region via a *k*-means clustering algorithm (Finley et al. 2013). We define these knots as a set of locations s^* such that $s^* \ll s$, where observed values of the response variable within the 'neighborhood' defined by each knot have been summarized. Spatial hierarchical models with low rank predictive processes have been shown to be effective at capturing spatial dependencies elsewhere in the natural resources literature (Finley et al. 2009, Latimer et al. 2009, Guhaniyogi et al. 2013). However, the number and spatial configuration of knots on which the predictive process is based will have a large impact on the quality of the resulting model (Gelfand et al.2013). Using *k*-means clustering weights the distribution of knots along with the distribution of observations, and we tested both lower and higher numbers of knots to establish that 500 provided a reasonable balance between computational efficiency and spatial coverage.

From this sample, we drew posterior estimates of the spatial surface $w(s)|y(s)$. In addition, we fitted a non-spatial, 'intercept only' regression ($y = \mu + \varepsilon$) and recovered the residuals (ε) for each observation in the sample. To assess the degree to which $w(s)$ explains model variance, these residuals and the spatial random effects were

interpolated for the sampling area, using multilevel B-splines (Lee et al. 1997). Multilevel B-splines are an efficient non-statistical interpolation method that is frequently used in computer science and image analysis. For our purposes, they provide a convenient way to smooth the residuals and spatial random effects across our study regions, to better visualize how these correspond with one another. Spatial regression models were fitted in the spBayes package for R (Finley et al. 2007, 2013), and the interpolation was accomplished using the MBA package for R.

Spatial co-dependence between soil C and N

In order to understand how the distribution of soil C and N are coupled across space, we first examined variograms for the residuals between carbon and nitrogen. These residuals were computed by fitting $C = \beta_0 + \beta_1 N + \varepsilon$ via ordinary least squares. Empirical and theoretical variograms were fitted as described above. In addition, we fitted the following spatial regression model:

$$C(s) = \beta_0 + \beta_1 N(s) + w(s) + \varepsilon(s)$$

We again interpolated both the non-spatial residuals and spatial random effects via multilevel B-splines, and examined these plots to assess the degree of spatial covariance within each dataset.

Results

Figure 3.2 shows the empirical variograms, along with the fitted variogram models, for both soil C and N at the three scales we considered. For both carbon and nitrogen, the national scale data exhibits strong spatial dependence, while no clear spatial

pattern exists for the regional dataset. At the continental scale, soil carbon exhibits some spatial dependence, while soil nitrogen does not. Note that the scale of these variograms is not commensurate with the study regions. After inspecting initial empirical variograms with a maximum distance equal to the longest pairwise distance in each dataset, a reasonable maximum distance was selected and the variograms were recomputed for display. These distances are: regional, 60 km; national, 200 km; and continental, 4,000 km. Beyond these distances, the initial variograms revealed no spatial autocorrelation.

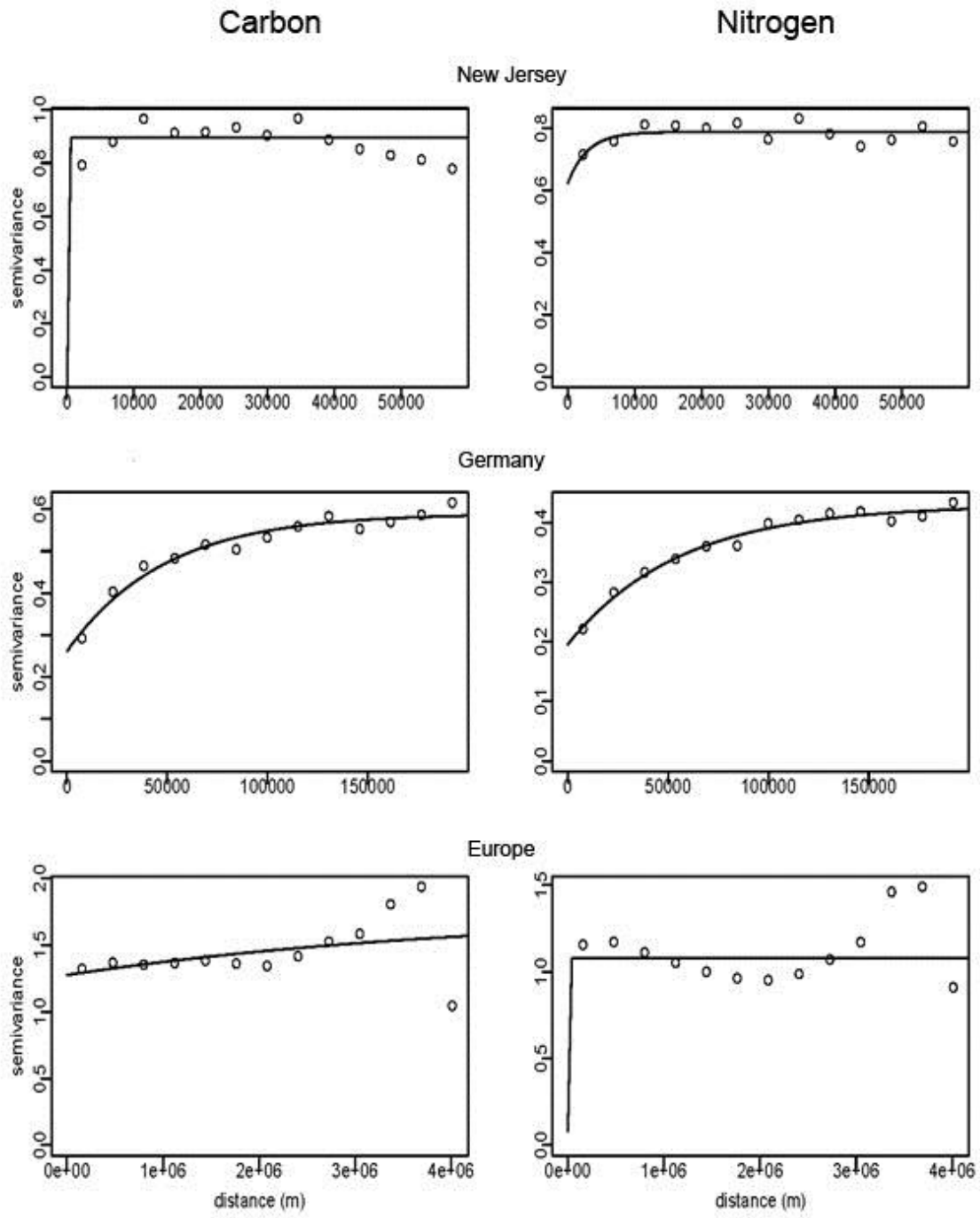

Figure 3.2: Empirical and fitted theoretical variograms for soil carbon and nitrogen at three spatial scales.

The theoretical variograms (exponential covariance functions) fit both the regional and national datasets well, generally confirming a lack of spatial dependence within the former and clear spatial patterns in both soil C and N in the latter. In the case of regional soil N, the theoretical variogram does suggest a slight spatial pattern at local scales (3-5 km). For the continental dataset, however, there is no discernable spatial pattern in the soil N data. In the case of C, the empirical variogram suggests some degree of spatial dependence out to approximately 3,750 km, though a sharp drop in semivariance at a slightly greater distance (4,000 km) influences the fit of the exponential function. This may be caused by a small number of extreme values in that spatial range. In the case of soil N, the empirical variogram exhibits a general trend of increasing semivariance from fine to broad scales, but it dips at approximately 2,000 km. In this case, an exponential covariance function may be a suboptimal choice for a semivariogram model.

Variograms for C/N residuals suggest some spatial dependence at all three spatial scales (Figure 3.3). In the case of the regional dataset, the pattern is fairly weak and the theoretical variogram model provides a poor fit to the empirical variogram. Both the national and continental variograms exhibit reasonably strong spatial dependence. This is interesting in the latter case, given that there was no spatial dependence observed in Nitrogen at the continental scale (Figure 3.2).

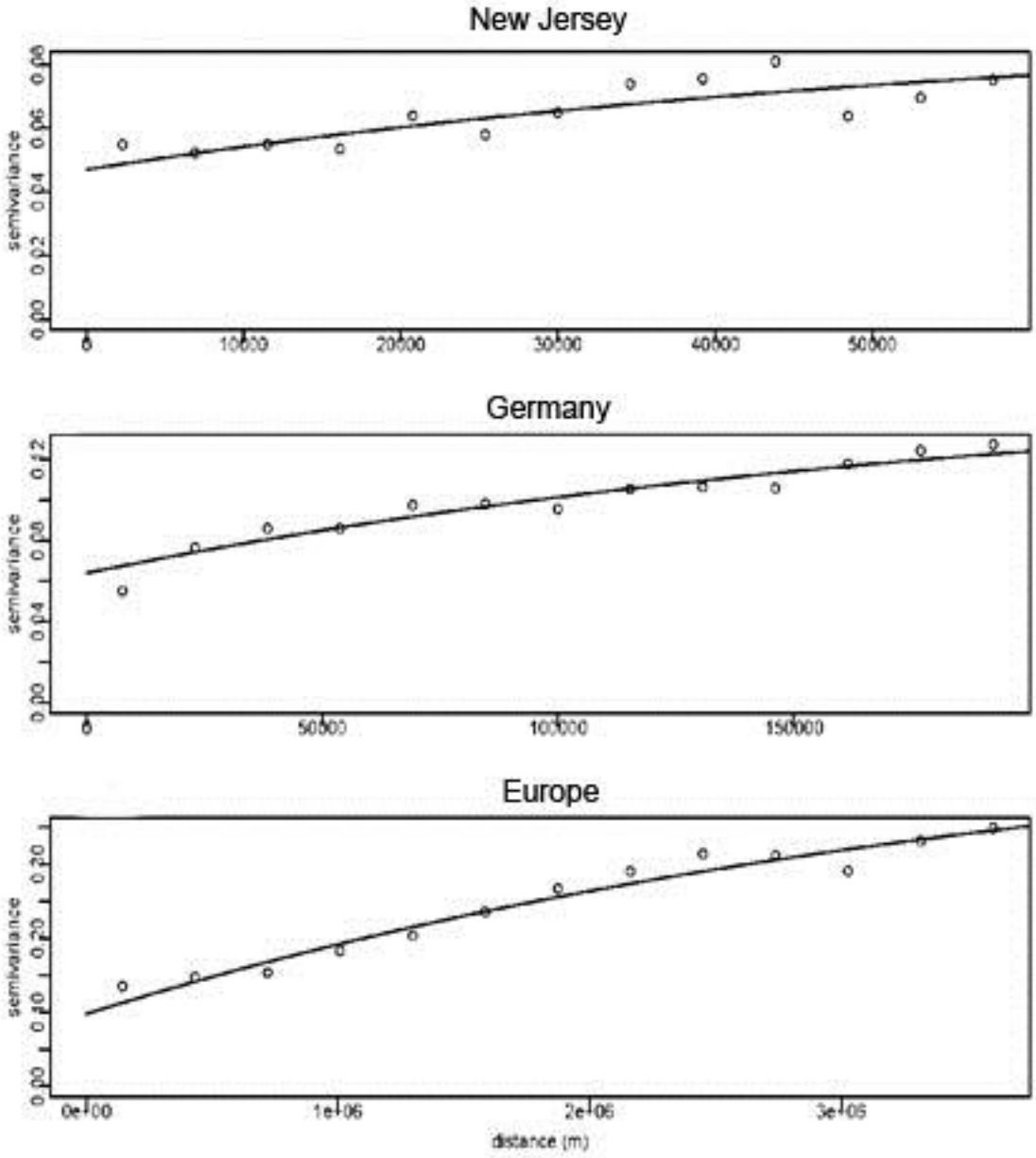

Figure 3.3: Empirical and theoretical variograms of the carbon/nitrogen residuals at three spatial scales

While variograms are capable of capturing general spatial patterns across the study area, they do not reveal any information about how patterns of spatial dependence change across the region, or how this spatial error is related to the overall variance within the model. To address such issues, we examined surface interpolations of non-spatial model residuals, and the spatial random effects ($w(s)$) from the univariate spatial Bayesian hierarchical models. These plots reveal how much of the total variance across the study region is explained by the spatial random effects. Where the surface plots of the model residuals and spatial random effects exhibit similar patterns, $w(s)$ accounts for a larger proportion of the total variance. In this way, comparing these plots is useful for describing and visualizing spatial dependence in point-referenced data.

In the case of soil carbon (Figure 3.4), the spatial random effects are not a large component of the total variance for the regional dataset, but demonstrate a clear pattern of spatial dependence over larger national and continental scales. In the spatial random effects of the regional data, we see the entire study area is dominated by a flat expanse of low to moderate values, even where the residuals reveal significant variation. There are some peaks that are correlated with high values in the residuals, but they are moderate in comparison, suggesting the spatial term in the model provides a poor fit for these regions. This contrasts with both the national and continental data, where the spatial random effects do a much better job of capturing extreme values; for instance in the northern and northwestern portions of the continental dataset. Note that these results differ from the theoretical variogram model, which suggested only limited spatial dependence in these data.

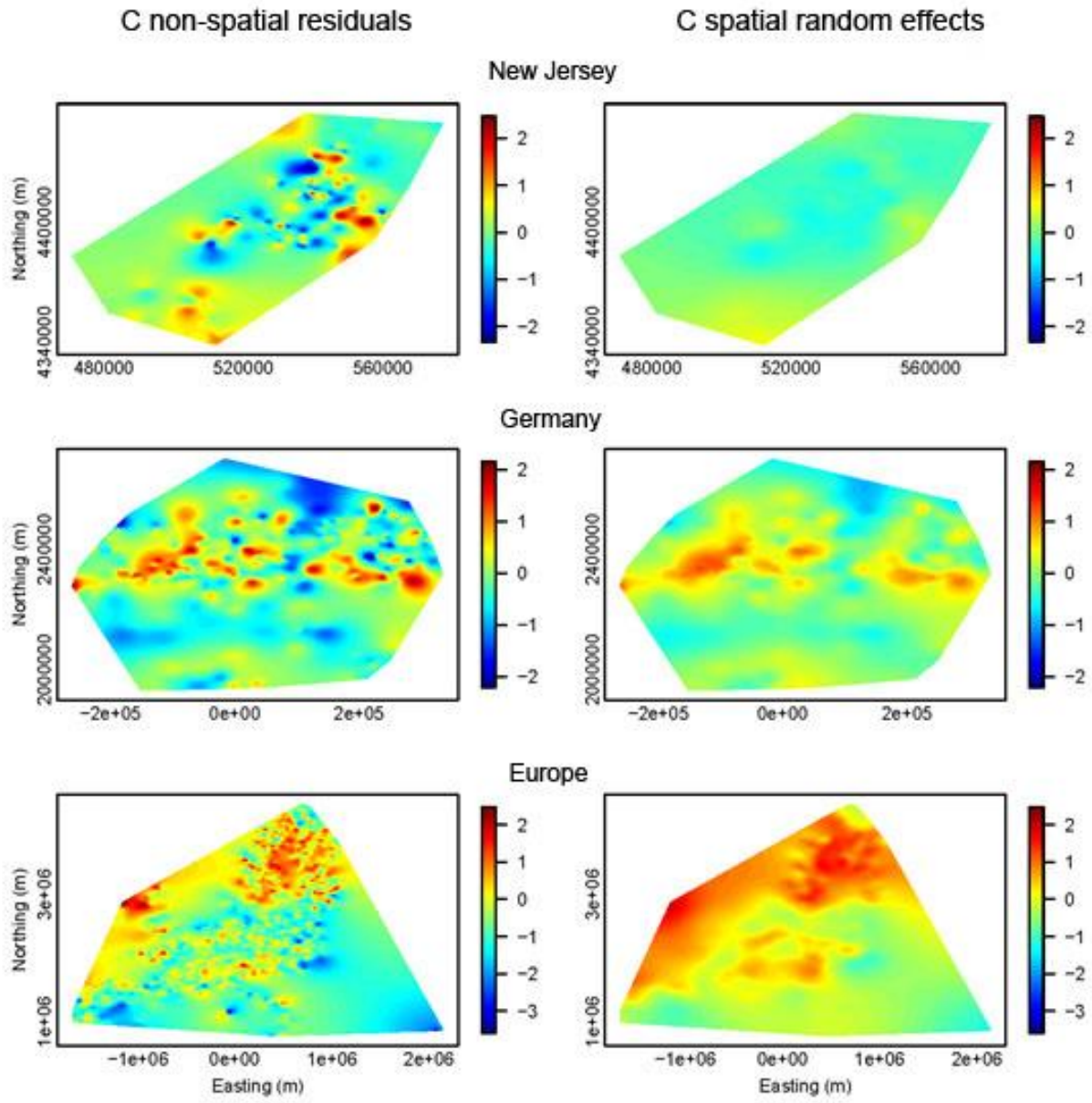

Figure 3.4: Non-spatial residuals and spatial random effects of soil C at three scales

The nitrogen plots (Figure 3.5) reveal fairly similar trends, with spatial dependence being strongest for the national dataset, somewhat weaker at the continental scale, but weak in the regional data. Unlike carbon, spatial random effects do describe some of the total variance for the regional dataset. This is most apparent for the very low values (the blue regions in the center of the plot). Note that the presence of some spatial dependence in these plots is consistent with the results of the variogram analysis, which suggested some spatial structure in soil N at fine scales (< 5 km). For the national dataset, as with carbon, the spatial random effects have a very similar distribution to the residuals, suggesting they explain a large portion of the total variance.

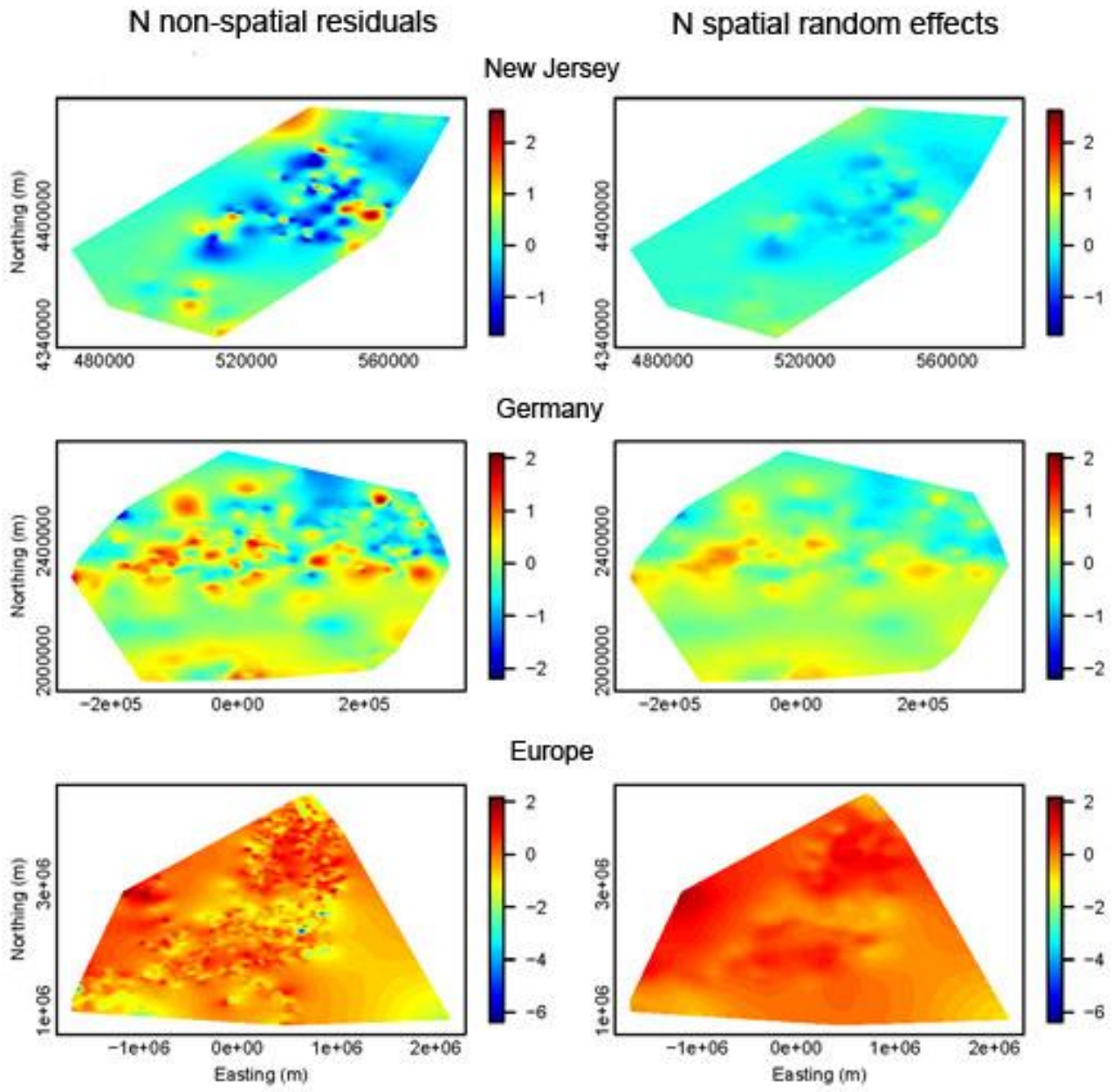

Figure 3.5: Non-spatial residuals and spatial random effects of soil nitrogen at three spatial scales.

The continental data exhibit a similar general trend to carbon, with generally higher N stocks in the northern and northwestern portions of the region. Note, however, that there is a high degree of local heterogeneity in the interpolated residuals, much of which is smoothed over by the spatial random effects. Given that we restricted the prior on the range parameter ϕ to 3,500-4,500 km, it is not surprising that the model did a poor job of capturing the local dynamics.

The interpolated surfaces for soil C and N residuals (Figure 3.6) generally confirm the results of the variogram analysis. The regional spatial random effects do exhibit some spatial dependence, largely driven by several extreme values in the dataset. At the national scale, the spatial random effects of the C/N residuals exhibit a similar spatial pattern to that of effect variable individually, while the continental dataset shows the same general north-south gradient that was observed in soil C.

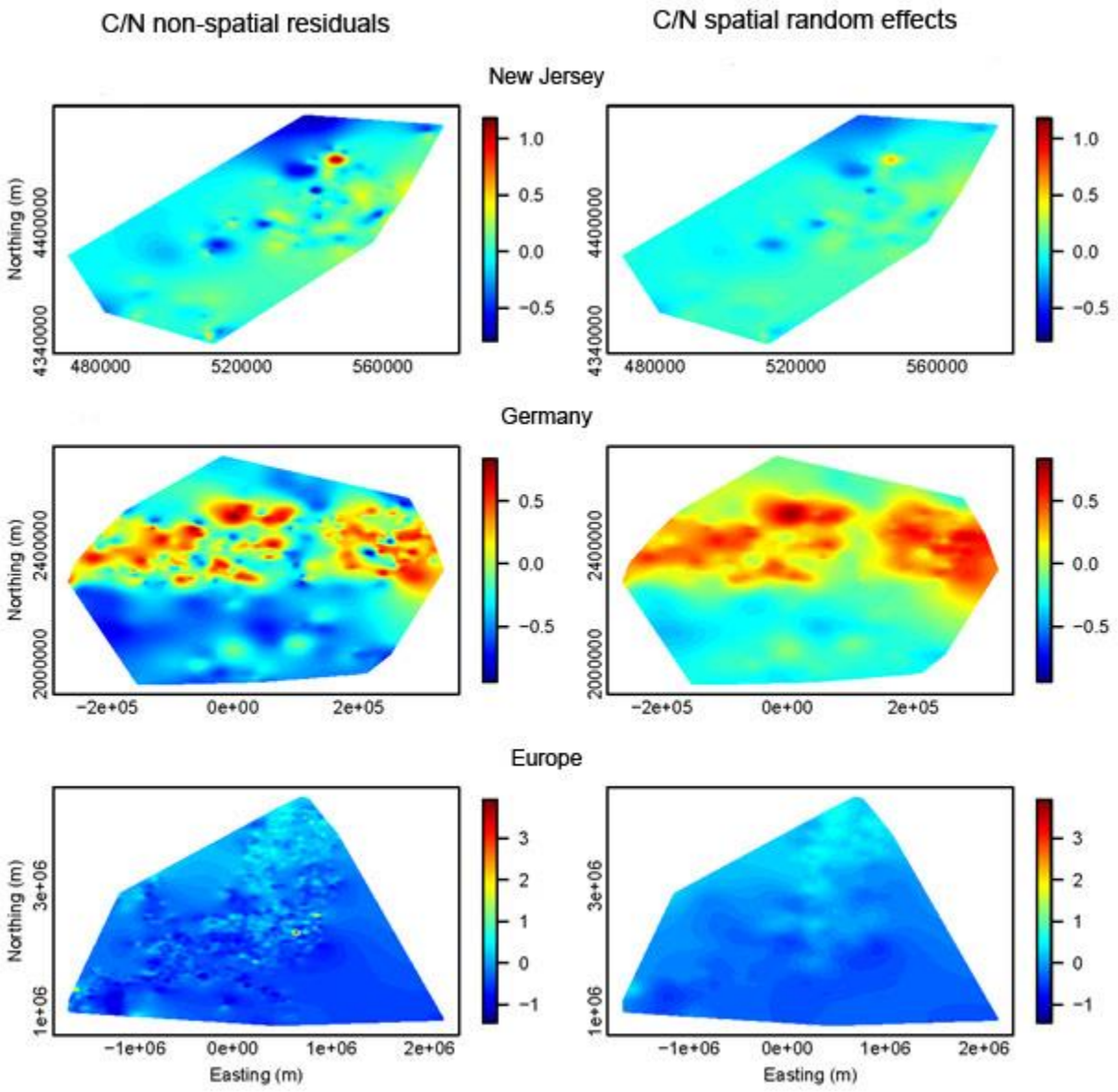

Figure 3.6: Non-spatial residuals and spatial random effects of the C/N residuals at three spatial scales.

Discussion

Our results suggest that spatial structure in forest soil C and N varies across spatial scales. Both variograms and surface plots of spatial random effects show that the strongest spatial pattern was apparent at the intermediate scale (Germany), and weakest at the finest scale (Coastal Plain, New Jersey). At the continental (European) scale, our results suggest moderate spatial dependence in soil C, represented as a general north to south gradient in C stocks, and no spatial dependence for soil Nitrogen.

Taken together, these results confirm some of our hypotheses, but do not agree with others. We expected to find spatial autocorrelation at all scales of our analysis. Many studies have shown spatial dependence in forest soil nutrients at fine scales (McBratney et al. 2003, Minasny et al. 2013), and our results confirm that it exists at very broad scales, in accord with other work (Liski and Westman 1997, Vasques et al. 2012). The failure to detect regional scale spatial autocorrelation is intriguing, particularly in view of the fact that the regional database exhibits 100 times the sample density as either the national or continental datasets (Table 3.1).

This result may be explained in part by the generally homogenous nature of this study region. New Jersey's Coastal Plain is typical of this physiographic province, with mostly very well drained sandy soils, originating from marine and alluvial deposits (Tedrow 1998). Forest communities are mainly pure and mixed Pine forests (in this area, *Pinus rigida*), and the landscape generally lacks steep changes in elevation. By restricting our analysis within a single physiographic province, and not crossing multiple geologic formations and associated forest communities, it is possible we did not introduce enough

broad scale variance into the data for any spatial structure to exist at the regional scale. At the same time factors that control spatial distribution of soil nutrients at local scales, such as variability in microclimate and microtopography, are too localized to be detected by even the relatively dense sampling density of our study, and are subsumed by an inflated nugget effect (τ^2); in fact, the variogram for regional soil N (Figure 3.2) suggested possible spatial dependence with a range (ϕ) of approximately 4 km.

It is important to note that these results may be specific to the Coastal Plain ecosystem, and not typical of all temperate forests. It is possible that Coastal Plain soils are particularly homogenous in concentrations of soil nutrients, and Lister et al. (2000) have found only limited spatial dependence, even at the stand level, in Coastal Plain forest soils. Similar efforts should be repeated for other temperate forest ecosystems in both the United States and Europe, but in practice, forest soil datasets collected at the regional scale are rare.

At the broadest extent, the continental scale, spatial dependence in soil C and soil N was apparent but not profound. The variograms suggested moderate spatial structure in soil C and none in soil N (figure 3.2), and the surface plots of $w(s)$ generally confirm this (figure 3.3). In the case of soil C, the general pattern revealed by these figures conforms to expectations, with the highest concentrations of soil C observed in northerly regions with cold climates, such as Scandanavia and the northern United Kingdom, and a gentler latitudinal gradient moving southward through continental Europe. The non-spatial residuals of soil N show significant local variation, but the spatial random effects fail to

capture a general continental trend. This suggests that soil N stocks are more strongly influenced by local factors than by a continental gradient.

While we observed strong spatial dependence in both soil C and soil N at the national scale, weaker spatial structure at the continental scale does not accord with our second hypothesis that spatial dependence should generally be easier to detect when crossing heterogeneous regions. However, in the same fashion that the regional dataset was collected at too broad a scale to capture the microscale spatial dependence driven by local factors, the continental dataset may have been too coarse to capture the finer details of geological, forest ecological or elevational patterns. Moreover, the spatial patterns we detect in soil C may arise from 'global' factors such as climate and the distribution of net primary productivity (NPP). More work is necessary in order to fully untangle the factors influencing soil C and N stocks at multiple scales. Here we only point to these possible factors to illustrate the major message that arises from these results; that spatial dependence in soil nutrient stocks is highly scale dependent.

At all three scales, we observed spatial dependence in the residuals between carbon and nitrogen, suggesting that these nutrients co-vary across space. Predictably, this was most prominent in the national dataset, where both C and N exhibit their own clear spatial patterns. However, even in the regional and continental data, where variogram analysis suggested one or both of these variables were not spatially dependent, spatial structure was observed within the residuals. Given the close linkages between the soil carbon and nitrogen cycles, it is not surprising that these variables would be jointly distributed across space, but the fact that their residuals display their own spatial

dependence is an interesting result. It may suggest that other spatially structured environmental factors are influencing the interaction of soil carbon and nitrogen across the landscape. Given the role aboveground-belowground interactions play in linking C and N cycling, forest community structure and composition could be a potential explanatory factor. These observations merit further study.

Taken together, our results have important implications for efforts to monitor forest soil carbon stocks. Coordinated international efforts such as the UN REDD+ program, as well as international agreements such as the Kyoto accords, require forest carbon accounting at the national level, and monitoring forest soil carbon stocks is an important component of these efforts (Maniatis and Mollicone 2010). When spatial autocorrelation is present in soil inventory data, utilizing spatial models can offer improved prediction accuracy and provide a more accurate representation of predictive uncertainty across the study region (Hengl et al. 2004, Simbahan et al. 2006). Here, we observed spatial dependence in soil C across Germany as well as across the entire European Union. These are the primary scales of interest for broad scale soil carbon accounting, and we suggest that a spatial modeling approach merits consideration for national efforts to estimate forest soil C stocks.

Of course, prior to applying a geostatistical model to soil inventory data, careful exploratory analyses must be made to characterize the spatial structure. We used two approaches to examine spatial dependence in our datasets: variograms and direct comparison of non-spatial residuals and spatial random effects via interpolated surface plots. Variograms are a foundational tool in the geostatistical literature, and have been

employed by most efforts aimed at characterizing spatial dependence in soil nutrient data. Variograms provide information about the general tendencies of the spatial dependence, but they do not provide any direct information on how the spatial pattern is arranged across the study. Variograms are also quite sensitive to outliers, which may lead to misrepresentations of the data's spatial structure (Goovaerts 1997). This may explain the disagreement between the variogram and hierarchical model analyses for the continental data, as the general spatial pattern was sensitive to a few extreme observations.

The hierarchical Bayesian approach allowed for direct examination of the spatial random effects, which helps to address these issues. By comparing the residuals from a non-spatial model to those of spatially random effects, we were able to assess how much of the total model variance was explained by spatial pattern. This analysis also provided inference on the distribution of soil C and N, based on an interpolated surface of the spatial random effects ($w(s)$).

It is important to note that our analysis of C and N alone makes two key assumptions about the spatial dependence in soil nutrients, i.e. that the distributions are both stationary and isotropic, whereas our analysis of C-N residuals assumes nonstationaity of C These are common assumptions for environmental geostatistics, though for many natural resources datasets, they may be poor (Isaaks and Srivastava 1989). By contrast, our results indicate that soil C patterns across continental Europe may be anisotropic models. Further, the significant local variation in our residual plots suggests that it is very possible the mean and variance of soil C and N shift across the region. Modeling anisotropy and/or nonstationarity introduces significant challenges to

model design and development, but future studies should seek to implement these models in order to compare results.

In this study, we consider only one example from each of the representative scales, and the results may be particular to the datasets we analyzed. Our general observation that spatial dependence in forest C and N is a function of increasing scale and landscape heterogeneity is logical, however, given current understanding of the controls over soil nutrient pools. As more nations work to incorporate soil sampling into forest inventories, additional broad scale datasets will become available from across the globe. Our collective understanding of the spatial distribution of forest soil attributes is very much "still in development", and further inquiry will be important in extending our understanding. Future work should seek to characterize spatial dependence in soil C for new regions, particularly in tropical and boreal forests, and also employ analyses that consider more complex spatial dynamics.

Conclusion

Spatial dependence of soil carbon and nitrogen is both scale and context dependent. In our study, spatial structure was an important feature of soil carbon datasets, particularly at the scales most relevant to national and international forest inventories. For this reason, when using point inventory data to develop broad scale estimates of soil nutrient stocks, spatial models should be considered as a potential approach. Exploratory analyses of the type we performed provide a rich picture of the spatial structure of a point-referenced dataset, and results should be examined carefully before deciding on a modeling approach. More advanced modeling strategies that account for additional

features of spatial datasets like anisotropy or nonstationarity may be preferable to the models we fit here. Finally, we have shown that soil carbon and soil nitrogen co-vary across space at all three scales we considered. Coupled modeling of these and other soil attributes may also lead to improved estimates of soil nutrient stocks.

Chapter 4: Comparing spatial and non-spatial Bayesian hierarchical models for mapping forest soil organic carbon at regional, national, and continental scales

Abstract

Model based prediction has become a popular alternative to soil survey for mapping soil attributes such as forest soil organic carbon (SOC). However, the methods generally used in the literature present significant shortcomings, particularly in their failure to fully quantify prediction uncertainty. In this study, we explore the use of spatial Bayesian hierarchical regression models for mapping forest SOC at three scales: within a single physiographic province (the Coastal Plain of New Jersey), the national scale (Germany), and across continental Europe. To validate the spatial approach, we compare prediction results to non-spatial hierarchical models of the same structure. Our results suggest that including a spatial process improves predictions of forest SOC at all three scales we studied. However, the linear models we utilized are highly uncertain, despite stringent efforts to identify suitable predictor variables. Additional work is needed to improve the performance of predictive models for forest SOC, particularly in identifying suitable nonlinear functional forms to fully capture relationships between SOC and environmental covariates. Spatial hierarchical models provide a convenient framework for accomplishing these goals, while allowing for fully accounting of prediction uncertainty.

Introduction

The forest soil organic carbon (SOC) pool constitutes one of the largest terrestrial pools of carbon worldwide (Goodale et al. 2002, Heath et al. 2003, Pan et al. 2011). Accretion of atmospheric carbon into forest soils, particularly in young forests, has been promoted as an important strategy for mitigating atmospheric carbon emissions (Lal 2008), and losses of forest soil carbon related to forest degradation and climate change have been a major focus of international cooperatives such as the United Nations' "Reducing Emissions from Deforestation and Forest Degradation" (REDD+) program (Birdsey et al. 2013). Despite the importance of the forest soil carbon pool to global carbon dynamics, good baseline data on forest SOC stocks is still lacking for much of the world.

Rapid advances in computational statistics, coupled with increasing availability of large soil attribute datasets, have led to improvements in spatial approaches for modeling the distribution of soil nutrients (McBratney et al. 2003). Increasingly soil scientists are focusing on point-referenced observations of soil attributes, coupled with 'rasterized' remote sensing and GIS datasets, to develop predictive models for mapping soil nutrient distributions over large spatial scales (Minasny et al. 2013). These methods present many advantages over traditional survey-based approaches, from which areal estimation of soil properties is accomplished by extrapolating observations from a small number of samples across soil survey map units, including: (1) higher spatial resolution; (2) incorporation of spatial and/or temporal processes to improve model-based forecasting; (3) leveraging statistical relationships between soil variables and related predictor variables; and (4) probabilistic estimation of prediction uncertainty (Grunwald 2009). The development of a model-based framework for forest SOC that is capable of cheaply and efficiently

producing high resolution maps at broad scales would enhance global research and policy related to carbon management.

When designing a predictive model for forest soil carbon, we seek an approach that possesses several characteristics:

- **Predictive accuracy:** Model design, particularly selection of predictor variables, should be both theoretically sound and informed by analyses aimed at maximizing predictive utility.
- **Scalability:** An ideal framework should be applicable at any spatial scale, and should accommodate the "scaling up" of patterns observed at finer scales to broad scale estimates.
- **Uncertainty quantification:** Predictions should reflect uncertainty of the data, model parameters, and other potentially important sources of variation (i.e. spatial and temporal processes) where applicable.
- **Computational efficiency:** The above characteristics must be balanced against the limits of computational resources.

Designing an approach with these features in mind will lead to a model that is generally applicable, adaptable to different ecological contexts, and is capable of producing predictions with reasonable estimates of uncertainty. Achieving this goal requires careful consideration of both the mean structure of the model and the selected computational method.

A large literature on selection of suitable covariates for predicting soil attributes already exists (*see reviews by* McBratney et al. 2003, Grunwald 2009, *and* Minasny et al. 2013). Most of this work focuses attention on environmental factors that are considered to be important influences of the spatial distribution of soil properties, and then seeks to identify available sources of remote sensing and GIS data related to those factors. This is perhaps best formalized by the 'scorpan' model (McBratney et al. 2003), which is based on the famous soil forming factors of Jenny (1941). A general presentation of the 'scorpan' model can be written as:

$$S_a = f(s, c, o, r, p, a, n) ,$$

where S_a is the soil attribute(s) of interest, s refers to other soil properties, c is climate, o is the influence of organisms (plants, animals, human activity), r signifies topographic variables, p refers to parent material and underlying geology, a stands for soil age, and n refers to spatial position. The 'scorpan' approach provides a theoretically sound framework from which precise functions for the prediction of soil attributes may be constructed. The set of chosen covariates, as well as how they should be scaled to the attribute of interest, must be established via careful analytical consideration.

Following the 'scorpan' approach addresses two of our target qualities for a predictive model of forest SOC: predictive accuracy and scalability. By providing a general set of guidelines, rather than a rigid set of model variables, we are free to construct functions for available data that are both context- and scale-dependent. We can reasonably expect that the role of various covariates will change, depending on the scale of inquiry. For example, mean annual temperature may not be a good predictor of SOC

within small regions, but is probably quite important for continental scale analyses. Similarly, we may be interested in a host of within-plot effects that will manifest as 'noise' over broader spatial scales. The 'scorpan' approach tailors the variables we use to the spatial scale of examination, but selection of covariates within each factor must be conducted carefully to avoid over-fitting models. A logical approach for this may incorporate both *a priori* expert knowledge and computational approaches such as statistical variable selection.

Once a reasonable predictive function (or set of functions for multi-scale studies) has been established, we turn our attention to computational approaches for model fitting. As outlined above, we seek a procedure that will provide reasonable estimates of uncertainty, while still remaining computationally efficient. Policy-makers rely on scientists' error estimates to inform decision-making processes (Shvidenko et al. 2010), and propagating uncertainty in SOC stock baselines into climate projections is a principle goal of ensemble climate models (Le Quere et al. 2009, Beer et al. 2010). Despite the importance of accounting for uncertainty, Grunwald (2009) found that approximately 36% of DSM studies surveyed made no attempt to quantify uncertainty in predicted values.

Fully characterizing uncertainty in digital maps of forest SOC requires more than error bars around aggregate estimates. We may be interested in determining how much of our uncertainty stems from measurement error in the data, versus that from model parameter estimates or from spatial variability (Banerjee et al. 2004). We may also wish to map uncertainty in order to plan future sampling efforts. Finally, it would also be

useful to consider the uncertainty of model choice, rather than simply adopting a particular model. In practice, 'model uncertainty' is not widely considered in digital soil mapping, but it can be described and dealt with via model averaging procedures (Raftery et al. 1997, Hoeting et al. 1999, Burnham and Anderson 2002).

Both statistical and machine learning approaches have been proposed for rasterized prediction of SOC. In a recent review, Grundwald (2009) found that least squares regression, decision/classification schemes, tree-based modeling, and univariate kriging are all common methods for digital soil mapping. Simulation models, where the SOC pool is determined as the balance between soil carbon inputs and outputs, have also been used to map soil nutrients, though less frequently than other methods (Bricklemyer et al. 2007, Causarano et al. 2007, Bauer et al. 2008). In general, multiple linear regression seems to be the most popular approach, with kriging and related approaches nearly as common.

Each of these approaches addresses some of our criteria for a good predictive model, but fails to cover all. Deterministic models provide detailed description of the mean structure, resulting in good predictive accuracy when properly specified, but frequently make no attempt to quantify uncertainty (Keith et al. 2009). Machine learning methods are highly efficient, but the model structure is usually predetermined by the methodological approach, and quantifying uncertainty from all sources may prove difficult (Grunwald 2009). Regression methods offer flexibility in specifying the functional form and selecting covariates, and provide a probabilistic framework for quantifying uncertainty. However, fitting methods such as least squares ignore parameter

uncertainty, and standard regression models do not explicitly incorporate spatial error. Geostatistical approaches, such as kriging models (Goovaerts 1997, Stein 1999), provide the ability to specifically model error arising from spatial dependence, but univariate kriging does not use model covariates to improve prediction (Hengl et al. 2003).

Geostatistical methods that incorporate predictor variables, such as regression kriging (Hengl et al. 2004), geographically weighted regression (Mishra et al. 2010, Mishra and Riley 2012) and co-kriging (Odeh et al. 1994, Goovaerts 1997, Wang et al. 2013), come closest to addressing our predictive model criteria. These methods are fit in a statistical framework, allow for a good degree of flexibility in specifying mean functions, specifically incorporate spatial processes, and are reasonably efficient with moderately sized datasets. However, when the spatial covariance (semivariogram) function of a geostatistical model is fit with least squares, which is typically the case, model parameter error is again ignored. Further, fitting a function to the spatial covariance matrix is computationally expensive (Finley et al. 2007a, Banerjee et al. 2008), and standard geostatistical methods may be impractical for very large datasets.

In order to avoid these shortcomings and to obtain a well-behaved predictive model of forest SOC, we turn to an alternative, spatial Bayesian hierarchical models (Banerjee et al. 2004). These models are extremely flexible, accommodate a wide array of mean structures, and facilitate modeling of spatial dependence via explicit spatial covariance models (i.e., exponential or Matérn functions). Further, by working in the Bayesian inferential framework, inclusion of appropriate prior and ‘hyper-prior’ specifications of model parameters yields a full accounting of uncertainty arising from

data, model parameters, and spatial processes (Congdon 2006). By generating a Bayesian posterior predictive distribution that can be marginalized for each new estimate (Ntzoufras 2009, Vehtari and Ojanen 2012), we can propagate uncertainty into the end result in seamless fashion. Finally, advances in statistical computing, particularly the development of 'low rank' predictive process models (Banerjee et al. 2004, 2008, Wikle 2010, Finley et al. 2012), renders the computation of spatial Bayesian hierarchical models feasible, even for very large datasets.

Spatial Bayesian hierarchical models have been used for other natural resource applications, including spatial mapping of forest attributes (Finley et al. 2007b, 2010), crop production (Finley et al. 2011), regional-scale climate modeling (Salazar et al. 2011), and mapping of species range distributions and abundance (Thogmartin et al. 2004). Guhaniyogi et al. (2013) have used multivariate spatial hierarchical models to study the distribution of several soil nutrients at local scales in tropical hardwood forests in central Costa Rica, but there are no examples of their application for regional mapping of forest soil attributes. Bayesian hierarchical models have also been used to scale soil properties across a heterogeneous urban landscape surrounding the city of Phoenix, USA (Oleson et al. 2006, Kaye et al. 2008), but these studies did not attempt to model a spatial process explicitly.

In this study, we applied spatial Bayesian hierarchical models to map regional forest SOC at three different scales: the Coastal Plain physiographic province of New Jersey, USA (NJ), the nation of Germany (DE); and the member states of the European Union (EU). Our objectives were to: (1) establish a suitable set of covariates, based on

the ‘scorpan’ framework, for modeling forest SOC within each study region; (2) develop maps of forest SOC distribution, along with corresponding maps of prediction uncertainty; and (3) compare the performance of the spatial approach with non-spatial hierarchical models. This final objective is aimed at determining whether the extra computational demands of a spatial process provide enough predictive improvement to warrant the effort.

Methods

Datasets

For this study, we used the same three datasets we analyzed in chapter 3: the dataset collected by the authors on the Coastal Plain physiographic region of New Jersey (NJ), and the LUCAS topsoil database of the European Union (Tóth et al. 2013). As in chapter 3, we consider only forested land for the European dataset at two scales: national (DE), and continental (EU). Complete descriptions of these datasets may be found in Clough and Green (2013), and Tóth et al. (2013). Refer to Chapter 3, Figure 1 for a map of the three study regions.

Covariate selection

Selection of predictor variables for SOC was carried out in two stages. In the first stage, we assembled a broad range of datasets that characterize the major soil forming factors considered by the ‘scorpan’ model discussed above (McBratney et al. 2003). Once the relevant datasets were in hand, we conducted both statistical variable selection and exploratory analyses aimed at revealing which covariates would serve as useful

predictors of SOC. Our goal was to derive models containing variables that represented all of the major soil forming factors, but to utilize computational approaches to refine the list, rather than simply choosing a set of covariates *a priori*. Table 4.1 lists the initial covariates considered for each dataset as well as the source of these data.

All of the covariate sets were first converted to a standard pixel size (30 m for NJ, 250 m for Germany and Europe). These were selected because they are the same as the finest resolution datasets for each analysis. All NJ soil-data layers were geo-referenced with the plot observations and projected in Universal Transverse Mercator (UTM), while the German and European datasets were converted to an Alber's equal area projection. Once all datasets were in a common resolution and projection, pixel values that coregistered with the SOC observation points were extracted from each layer for variable selection, model construction, and validation. Initial processing of covariate data was accomplished with the ARC and ERDAS Imagine programs.

Since our aim was to use the 'scorpan' framework as a general guide, we selected covariates that were reasonable for our separate scales of inquiry, leading to somewhat different initial datasets for each study region. Table 4.1 lists the initial set of predictor variables for the NJ dataset. We used an interpolated surface of total soil nitrogen, generated using multilevel B-splines (Lee et al. 1997) from field measured observations that were co-located with the SOC measurements in the dataset. For topographic variables, we included elevation and slope as potential predictors, as initial exploration of other terrain attributes were not promising on this low-relief landscape. A 10 m digital elevation model of the study region was used, and slope was calculated from these data.

The climate variables we include are interpolated surfaces from the PRISM climate dataset. We considered two Landsat derived metrics that are known to relate to aboveground biomass, the normalized difference vegetation index (NDVI) (Steltzer and Walker 2006) and the so-called "Greenness" band of the tasseled cap transformation (Goetz and Dubayah 2011). We also included the "brightness" and "wetness" bands of the tasseled cap transform. Landsat data are available at a fine resolution (30 m pixels), so these data, rather than coarser satellite datasets such as MODIS, were a good choice for our regional analysis. Since we do not expect large variations in the underlying geology across New Jersey's coastal plain, we instead include two predictors related to soil texture and chemistry from the SSURGO dataset: available water capacity and soil pH.

Table 4.1: Selected covariates for the NJ data.

Data layer	*Source*
Total soil nitrogen (N)	Field-measured
Elevation (ELEV) Slope (SLOPE)	10 cm digital elevation model (DEM)
Precipitation (PRECIP) Maximum temperature (MAXTEMP) Minimum temperature (MINTEMP)	PRISM climate data, 30 year normals
TC (tasseled cap(Brightness (TC1) TC Greeness (TC2) TC Wetness (TC3) Normalized difference vegetation index (NDVI)	Landsat ETM 5
Available water capacity (AWC) Soil pH (pH)	USDA SSURGO database

For Germany and Europe, we again used interpolated soil N surfaces derived with the same method as the NJ data, and considered the same primary terrain attributes (elevation and slope). Given the coarser scale, the elevation data were derived from NASA's shuttle radar mission (SRM) global topography dataset. For climatic effects, we used the 19 "BIOCLIM" variables from the WORLDCLIM database. These variables are derived from monthly estimates of rainfall and temperature, and are designed to represent biologically meaningful climatic effects, including annual averages, seasonality, and extremes in temperature and precipitation (Hijmans et al. 2005). We expected greater variation in underlying lithology for Germany and Europe, so we used geologic classification data as more general measures of the parent material effects. Similarly, we considered only a coarser-scale estimate of net primary production (NPP), derived from MODIS spectral data. The covariates used for the German and Europe analyses are summarized in table 4.2. With the exception of co-occurring observations of soil nitrogen within each dataset, we used only publicly available data.

Table 4.2: Initial covariates for the Germany and Europe analyses

Data layer	*Source*
Total soil nitrogen (N)	Field-measured
Elevation (ELEV) Slope (SLOPE)	NASA SRTM
Mean annual temperature (BIO1) Mean diurnal range (BIO2) Temperature isothermality (BIO3) Temperature seasonality (BIO4) Maximum temperature of warmest month (BIO5) Minimum temperature of coldest month (BIO6) Temperature annual range (BIO7) Mean temp. of wettest quarter (BIO8) Mean temp. of driest quarter (BIO9) Mean temp. of warmest quarter (BIO10) Mean temp. of coldest quarter (BIO11) Annual precipitation (BIO12) Precip. of driest month (BIO13) Precip. Seasonality (BIO15) Precip. of wettest quarter (BIO16) Precip. of driest quarter (BIO17) Precip. of warmest quarter (BIO18) Precip. of coldest quarter (BIO19)	WORLDCLIM database
Soil parent material (GEO)	BGR geologic map of Europe
Annual Net Primary Production (NPP)	MODIS derived 30 year average

Computational evaluation of model covariates consisted of both automated variable selection and subsequent exploratory analysis. In order to gain an idea of which covariates tended to be more important for predicting SOC, we first submitted the full linear models containing all covariates to an automated Bayesian variable selection procedure. Bayesian variable selection is typically evaluated via posterior model

probability (Clyde et al. 2011). Consider an initial model space $M = 2^p$, where p is the number of potential covariates under consideration. Thus, *M* contains all possible models that may be constructed with the available predictors, and an efficient Markov Chain Monte Carlo (MCMC) algorithm is used to search this model space and identify the 'best' models based on their posterior probability (George and McCulloch 1993, Casella et al. 2009) We evaluated the following linear regression model $Y = I\beta X$, where Y is a *n* x *1* vector of the response variables, *X* is an *n* x *p* design matrix, and β is a *p* x *1* vector of model coefficients. The diagonal element I is a binary "indicator parameter" that tracks whether a variable is 'in' or 'out' at each iteration of the MCMC sampler.

From this output, posterior inclusion probabilities (PIPs) can be computed for each covariate as the number of iterations where $I_{ii} = 1$ over the total number of iterations in the MCMC sample (Ntzoufras et al. 2000, Dellaportas et al. 2002). This analysis was conducted with the "Robust" reference priors of Bayarri et al. (2012) on the regression coefficients and an equal prior inclusion probability (i.e., $I_{ii} \sim Binom(0.5)$) on their associated index parameters. Robust priors are specifically designed as 'uninformative' and are widely used in "objective" Bayesian analysis (Berger 2006). These, combined with a prior on the indicator parameters that allows for an equal probability of inclusion at each iteration, allowed us to conduct variable selection that is only informed by the data. Given the large number of initial covariates in each dataset, a Gibbs sampler was used to draw 100,000 posterior samples of the model parameters, and inference was conducted following a burn-in of 50,000 iterations. We retained variables for further consideration if they had a PIP > 0.5, which is a common cut off in the

Bayesian statistics literature (Berger and Barbieri 2004).. Variable selection was conducted using the 'BayesVarSel' package for R (Garcia-Donato and Forte 2012).

We examined basic diagnostics for each covariate selected by the variable selection procedure, including histograms and scatterplots with the response variable, as well as correlation coefficients and R^2 values. Since our variable selection was not conducted in spatial context, we also examined variograms of the residuals between the covariate and response. Final selection of covariates for each dataset was then made based on the results of both these exploratory analyses and our variable selection procedure.

While we employed statistical methods to select predictor variables, we preserved the general structure of the 'scorpan' model. For instance, climate related variables were included in each model, regardless of the results of our analysis, because there is a strong theoretical basis for the effect of climate on SOC distribution. However, given that multiple climate variables were initially considered (3 for NJ, 19 for DE and EU), computational evaluation is useful for identifying which of these are most likely to explain the distribution of SOC. By using this general approach, we retain a theoretically valid model for SOC prediction, while avoiding the need to make decisions on precise covariates *a priori*, as well as potential colinearity issues related to over-fitting linear models with many correlated variables.

Hierarchical modeling

Following covariate selection, the final spatial Bayesian hierarchical models were all of the general form:

$$y(s) = \mu(s) + w(s) + \varepsilon(s),$$

where $y(s)$ is the SOC observation at location s, $\mu(s)$ is the mean structure, $w(s)$ is a univariate spatial process modeled by an exponential covariance function, and $\varepsilon(s)$ is the residual "white noise" error term (Banerjee et al. 2004). Model fitting was conducted using the same procedures as in chapter 3, where both the NJ and DE datasets were fit as full rank spatial models, and the EU data were fit with a modified predictive process (Finley et al. 2012) to reduce computational burden. The same initial values, drawn from the variograms for SOC, and prior specifications for both the regression coefficients and spatial parameters that were used in our exploratory analyses were used again here. Please refer to chapter 3 for a detailed description of this approach.

Comparison of Spatial and non-Spatial Models

To compare the performance of spatial and non-spatial models, we fit both to each of the three datasets, containing all of the covariates selected in the first phase of our analysis. The spatial models were given vague priors on the regression coefficients, partial sill (σ^2), and nugget (τ^2) and weakly informative priors on the spatial range (ϕ), resulting from preliminary variogram analysis. The latter priors are weakly informative because the bounds were selected based on the observed emprircal semivariograms, but the priors were considered to be flat within those bounds. The non-spatial models were fit with vague prior specifications on all parameters. Model fitting, prediction, and checking were carried out using the spBayes package available for R (Finley et al. 2007a).

We applied *k*-fold cross validation (Vehtari and Ojanen 2012) to test predictive performance. Each dataset was randomly divided into ten approximately equal size sets, and each of these was successively held out for validation, while the model was fit to the remaining nine partitions. This process was repeated until all 10 data sets had been used for model validation. Prediction was performed using the covariates from each validation location, via draws from the posterior predictive distribution:

$$p(y_{new}|y) = \int p(y_{new}|y,\theta,\beta)p(\theta,\beta|y)d\theta d\beta \,,$$

where y_{new} are the prediction locations, y are the fitting data, θ contains the spatial process parameters, and β represents the regression coefficients. We computed root mean squared error (RMSE) and mean absolute error (MAE) for each validation, and compared predictive performance based on the mean of these results. We also compared models based on the deviance information criterion (DIC) (Spiegelhalter et al. 2002).

Mapping forest SOC at large spatial scales

Following model evaluation and selection, the 'best' fitted model (either spatial or non-spatial) was used to produce distribution maps of SOC (Mg/ha), as well as the 95% credible interval, for all forested land in each study region. Ideally, this would involve prediction at each pixel of a grid covering the region. Such an approach can become computationally expensive for large raster datasets, particularly when Bayesian prediction has been used to generate a posterior predictive distribution for each pixel, rather than a single point estimate. Given the high resolution of our prediction grids, full Bayesian prediction at every pixel was not computationally feasible.

Balancing computational efficiency against the need to produce a spatially-registered 'uncertainty surface' for SOC, we combine sparse Bayesian prediction with a highly efficient nonstatistical interpolation algorithm; multilevel B-splines (Lee et al. 1997). This is a method from computer science and image visualization that utilizes a hierarchical lattice framework to approximate smooth surfaces with small computational investment. We produced a dense prediction grid consisting of 40,000 observations, distributed at regular intervals across each study region. Initial exploration suggested this resolution was a reasonable compromise, particularly for the very large Europe dataset. Values of the model covariates were drawn from the raster datasets for these points, and independent predictions were made via the posterior predictive distribution for each location, following a burn-in period of 9,000 MCMC iterations. A sample of 100 posterior predictive observations, thinned from 1,000 MCMC samples, was drawn for each of the 40,000 locations. In initial tests, 100 observations were sufficient to produce the posterior predictive distribution.

Following prediction, the samples were summarized by a mean value of SOC at each location, as well as a 95% credible interval estimated from the posterior predictive distribution. We thus produced a regular grid of estimated mean forest SOC, along with an associated credible envelope. These values were then passed to the spline interpolator to create smooth surfaces of mean SOC and the 95% CIs for all forested land in the study region, at a resolution matching the original prediction grids. As with model validation, prediction was accomplished in the spBayes package, while the spline interpolation was performed using the MBA package for R.

Results

Table 4.3 shows the structure of the models used to predict SOC for each dataset, based on the results of our variable selection procedure. For the New Jersey and German datasets, log-transformed soil N was the best single covariate, and the remaining covariates had somewhat weaker relationships with SOC. The New Jersey data showed the largest R^2 for the selected model, but this is mostly explained by soil N, which is highly correlated with soil C in these data. In the German dataset, soil N is also the only covariate with a strong relationship to soil C. Soil N was not a strong predictor in the EU data, and in fact none of the selected covariates had strong relationships with SOC, leading to a low R^2 when compared to the New Jersey and German models. In both cases, the remaining covariates allow us to consider all of the major soil forming factors. The full results of the variable selection for each dataset can be found in Appendix B.

Table 4.3: Results of the covariate selection procedure. Note that the non-spatial models are shown here, but the same structure is also utilized for the spatial models.

	Model	R^2
New Jersey	$\log(soil\ C) = \beta_0 + \beta_1 * SLOPE + \beta_2 * PRECIP + \beta_3 * TC1 + \beta_4 * NDVI + \beta_5 * AWC + \beta_6 * pH + \beta_7 * N + \varepsilon$	0.93
Germany	$\log(soil\ C) = \beta_0 + \beta_1 * \log(SLOPE) + \beta_2 * BIO8 + \beta_3 * BIO14 + \beta_4 * BIO15 + \beta_5 * GEO + \beta_6 * \log(NPP) + \beta_7 * N + \varepsilon$	0.51
Europe	$\log(soil\ C) = \beta_0 + \beta_1 * ELEV + \beta_2 * BIO1 + \beta_3 * BIO2 + \beta_4 * BIO14 + \beta_5 * BIO19 + \beta_6 * GEO + \beta_7 * NPP + \beta_8 * N + \varepsilon$	0.18

For all three regions, spatial models yielded lower values for both root mean squared error (RMSE) and mean absolutely error (MAE), based on ten-fold cross validation, but the relative improvement varied (Table 4.4). For the New Jersey data, the spatial hierarchical model reduced RMSE by approximately 30%, while the gains were more modest for the German (9.3%) and EU (1.2%) datasets. Recall that the observations in the German and EU datasets were less dense (0.001 ha^{-1}) than they were for the New Jersey dataset (0.1 ha^{-1}), and further that spatial random effects in the EU model were fit with a predictive process based on 500 knots. This may in part explain why the gains offered by modeling spatial dependence were smaller for these datasets. Deviance information criterion (DIC) also supports spatial models for Germany and the EU, but actually supports the non-spatial model for the New Jersey data. Given the results of the cross validation analyses, however, this result should not weigh heavily in our model selection, because DIC uses point estimates rather than the full posterior predictive distribution for prediction (Spiegelhalter et al. 2002). This fact has led several authors to

observe that DIC may over-fit in situations where none of the competing models provide a strong fit to the data (Plummer 2008, Spiegelhalter et al. 2014).

Table 4.4: Deviance information criterion (DIC), and root mean squared error (RMSE) and mean absolute error (MAE) from *k*-folds cross validation. The DIC scores result from fitting models to the complete datasets.

		DIC	RMSE	MAE
New Jersey	Spatial	99.01	80.91	47.39
	Non-spatial	96.59	116.71	83.18
Germany	Spatial	-162.46	23.51	13.95
	Non-spatial	-95.66	25.26	15.76
Europe	Spatial	-171.33	134.4	71.98
	Non-spatial	-156.56	136.3	73.31

Based on our model selection results, we proceeded with spatial hierarchical models to map forest SOC, including 95% credible intervals, for our three study regions (Figure 4.1). For the NJ data, predicted SOC ranged from 5 Mg/ha – 300 Mg/ha. The lowest values are predicted for the central region of New Jersey's Coastal Plain. This corresponds to a soil deposit known as the Beacon Hill formation, which is comprised of excessively well drained and nutrient poor soils (Tedrow 1998). The model also predicts generally higher concentrations of SOC for forests on the inner Coastal Plain, an expanse along the northwestern edge of the study region, where soils tend to be somewhat higher in silt concentrations and as a result more nutrient rich (Tedrow 1986). Note, however,

that the accompanying credible intervals reveal that the predictions supplied by this model are highly uncertain. In large portions of the study area, the variance estimates are an order of magnitude larger than the predicted estimates. This is consistent with the results of our model selection procedure, which suggested that the final model for the NJ data provided a poor fit. While we gained predictive accuracy by modeling a spatial process, as suggested by RMSE and MAE from the cross validation exercise (table 4.4), both models were poorly fitting.

Of the three study regions, we had the greatest success predicting forest SOC for Germany. However, results here were still highly uncertain. The majority of the sampling observations were concentrated in several large forested areas in the northern central portion of the nation (Chapter 3, figure 1), particularly in the states of Rhineland-Palatinate and Hesse. These are some of the most densely forested regions in Germany, so this spatial bias is not surprising. Still, the distribution of our predictive results are in accordance with other studies on spatial distribution of forest SOC in Germany, showing generally lower concentrations in the Scots Pine (*Pinus sylvestris*) dominated regions in the northeast (Wellbrock and Bolte 2008), and higher values in the humic-rich soils in the northwest as well as the Alpine soils in the southern portion of the country (Wiesmeier et al. 2013, Gruneberg et al. 2014). Wellbrock and Bolte (2008) also found significant local variation in forest SOC concentrations across the nation, though much of that is smoothed out by our spline interpolation approach.

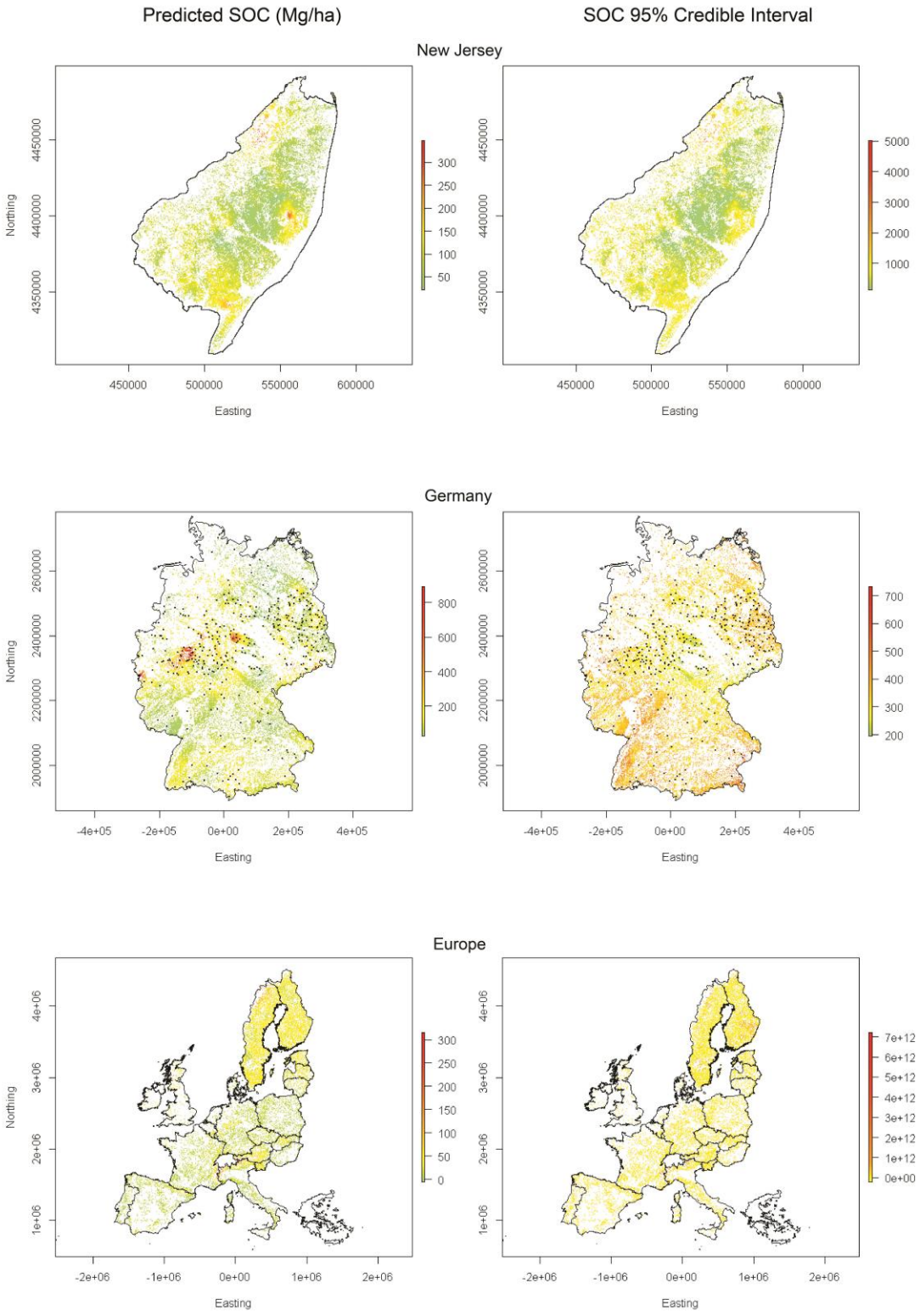

Figure 4.1: Predicted forest soil organic carbon, and accompanying 95% credible intervals, for the three study regions.

The SOC map of Europe likewise conforms to expectations, with higher forest soil carbon concentrations in the high latitude boreal forests as well as the alpine regions in central Europe. However, these results had the highest uncertainty of the three regions we studied. Note that while the scale suggests some extremely large uncertainty values, these are mainly outliers and through the majority of the region uncertainty, while still large, is more concordant with the mean predictions. Given the very low density of the spatial process knots, as well as the grid we used for statistical prediction, the EU results are more smoothed than those from New Jersey or Germany. Still, it is interesting to observe that the model generally captured expected trends in SOC distribution along latitudinal and elevational gradients.

Discussion

In each of our three case studies, spatial Bayesian hierarchical models provided a reasonable alternative to a non-spatial predictive model for forest SOC, though in all three cases both models revealed significant uncertainty in prediction estimates. In the New Jersey and German datasets, the gains in predictive accuracy were significant, while for the European dataset, they were slight at best. Despite the uncertainty, the spatial predictive models did a reasonable job capturing the spatial distribution of forest SOC, conforming to our understanding of the study system as well as the work of others. Bayesian prediction allowed for the creation of an ‘uncertainty surface’ via interpolation of the range of the posterior predictive distribution of each new observation. In our study, this was important for evaluating model performance, and for understanding how model uncertainty was influenced by environmental processes and the spatial distribution of the sampling locations.

The gains provided by fitting a spatial hierarchical model must be weighed against the added computational burden. The New Jersey and German datasets were relatively small (< 500 observations), so generation of adequate posterior samples of the model parameters could be accomplished in minutes. The European dataset, on the other hand, required approximately 10 hours on a computing system possessing a quad-core Intel i5 processor and 32 Gb of memory to draw the 10,000 posterior samples we needed. This was even after fitting $w(s)$ with a modified predictive process, based on 500 knots, a very sparse framework from which to model continental-scale spatial dependence in forest SOC. Fitting the "full rank" model using all observations would not have been possible with available computational resources. In this case, given the limited benefits to be obtained from modeling spatial dependence, a non-spatial model might be sensible.

Regardless of whether predictive models include a spatial term, we advocate usage of a Bayesian hierarchical framework to predict forest SOC. By generating a posterior predictive distribution, conditional on the joint distributions of the observed data and the model parameters, we are able to obtain a credible uncertainty estimate around each predicted value (Gelman et al. 2003, Ntzoufras 2009). This is quite valuable in mapping of forest SOC stocks, where uncertainty estimates are of particular interest for both policy makers and other scientists (Shvidenko et al. 2010). Generating 'uncertainty surfaces' allows analysts to 'track' how model error is distributed across space, and to incorporate this information into future sampling efforts. Further, adopting a Bayesian approach frees us from the need to assume that model parameters will be fixed at any single value, particularly important for spatial models, where the spatial covariance

functions used in traditional kriging often provide a poor fit to empirical variograms (Goovaerts 1997). Traditional kriging ignores potentially important sources of uncertainty, and as a result may provide biased mean and variance estimates. Bayesian hierarchical models allow us to capture the full range of parameter uncertainty, seamlessly propagating it into predictions.

As noted, Bayesian prediction is computationally more intensive, and this may become an issue when the goal is to predict variables at high resolution across very broad spatial scales. We addressed this issue by predicting on a lower resolution grid, and then interpolating the resulting means and 95% credible intervals with highly efficient multilevel B-splines. The resulting maps have inevitably been somewhat smoothed, and their usefulness is dependent on the goals of the analysis. If the aim were simply to characterize the general distribution of forest SOC, as well as model uncertainty across space, our results may suffice. In general, our models did a good job of producing forest SOC distributions that conform to reality, despite the uncertainty in the predictive estimates. However, if the goal is to develop a highly credible map of the regional forest SOC stock, then it is ideal to use the predictive model to generate a posterior predictive distribution for every pixel within the study region. Future work should pursue alternative strategies for easing the computational burden of prediction on high resolution grids that will avoid the need to smooth results as widely as we do here.

In attempting to construct valid mean structures for predicting regional forest soil C, we found the ‘scorpan’ approach to be a useful tool in guiding the selection of covariates (McBratney et al. 2003, Grunwald 2009, Minasny et al. 2013). Drawing an initial set of predictor variables based on the major categorical factors influencing the

formation and distribution of soils is sensible, but this exercise would have led us to greatly over-fit our datasets if we had included all of the initial covariates we considered. Combining a philosophical model such as 'scorpan' with computational variable selection provided a productive 'middle ground'. Using this approach, we were able to construct models that were theoretically sound but also informed by observed relationships between forest SOC and the covariates within the dataset. If our work on spatial dynamics of forest SOC, both here and within Chapter 3, teaches us nothing else, it is that controls over SOC distribution will be both scale and context dependent. Approaching the design of predictive models with this reality in mind is the best way to achieve reasonable results.

Even with a guided approach, we had difficultly establishing strong relationships between forest SOC and potential covariates. The New Jersey and German models provided a good fit, which was almost entirely driven by the inclusion of soil nitrogen, predicted from co-located measurements, in the models. The remaining covariates we considered had highly variable relationships with SOC, and there were no clear trends to discern in some cases. Incorporating remote sensing data for the prediction of forest SOC will probably be more difficult than for agricultural landscapes, for which one can use spectral information about the soil itself (Hengl et al. 2004, Simbahan et al. 2006, Tóth et al. 2013b), while heavy forest vegetation cover leaves us reliant on relationships mediated by aboveground carbon dynamics. We found little evidence of Landsat or MODIS derived products that were useful for mapping forest SOC.

Topographic and climate variables were somewhat important in our analyses, in agreement with the work of other authors (Odeh et al. 1994, Mueller and Pierce 2001,

Zhang et al. 2012, Vasques et al. 2012). Even so, our models did a poor job of capturing the variance in such relationships. Moreover, we modeled log-transformed SOC with linear functions of selected covariates, but linearity may not provide proper scaling. Other authors have suggested that models used for digital soil mapping should be able to capture nonlinear relationships between soil attributes and covariates (McBratney et al. 2003, Malone et al. 2009). This is possible to accommodate within a spatial hierarchical model, and we suggest that future work should involve rigorous model comparison between linear and nonlinear functional forms.

In ecosystem science, the SOC pool is often estimated via complex process models that treat it as a balance between C inputs (litter deposition and decomposition) and outputs (soil respiration) (Chapin et al. 2002). Given that relationships between these carbon fluxes and several variables we considered are well established in the literature; for instance, the influence of climatic gradients of litter decomposition rates have been well reported (Couteaux et al. 1995); biologically realistic process models may provide an ideal approach for developing accurate mean structures. However, in order to address our need for valid uncertainty estimates in model predictions, it is necessary to use these models within a probabilistic framework. Bayesian melding, where complex model functions are implemented in the context of a statistical hierarchical model (Raftery et al. 1995, Poole and Raftery 2000), can be used to accomplish this goal. This technique, and closely related methods such as Bayesian synthesis and ‘data assimilation’, have been used to quantify uncertainty in a variety of forest science contexts, including stand growth models (Green et al. 1999, MacFarlane et al. 2000), and forest carbon models (Schwalm et al. 2010, Keenan et al. 2012, LeBauer et al. 2013). Adopting this approach

for modeling SOC may be logical, given our difficulty in establishing linear relationships between forest SOC and other biophysical variables.

Bayesian melding can also be carried out in the context of spatial hierarchical models. Finley et al. (2011) used this technique to improve yield estimates within agricultural fields, using a complex crop production model. This method is especially attractive for digital soil mapping, since it allows for the direct modeling of spatial dependence for the input data, as well as the model parameters. This is potentially important, since it is reasonable to expect that rates of carbon fluxes related to SOC accretion will vary significantly across large regions. Of course, the use of even more complex models within a hierarchical framework will add to the computational difficulty we encountered when producing high resolution maps via Bayesian prediction. To address this issue, it may be necessary to explore computational alternatives to MCMC sampling for deriving posterior distributions, such as integrated nested Laplacian approximation (INLA) (Rue et al. 2009). INLA is an analytical technique that offers significant computational gains, and has already been used for reducing the burden presented by high dimensional hierarchical models (Eidsvik et al. 2012). Approximate Bayesian Computation (ABC) is another promising method, which unlike INLA may also be applied to fit non-Gaussian models.

In addition to accommodating complex mean structures, a hierarchical framework allows us to address other simplifying assumptions we made here, and that are frequently made by other analysts mapping SOC distributions. The datasets we used contained only surface soil samples, so resulting estimates ignore subsoil C storage. This is particularly problematic for humic and wetland soils, where significant long-term carbon storage may

occur at depths greater than 30 cm (Grunwald 2009). Several authors have proposed the use of soil profile depth functions, developed from sparse profile datasets, for estimating SOC storage across a wider depth interval. Malone et al. (2009) used geostatistical techniques to interpolate parameters of an exponential depth function for mapping soil carbon and available water capacity in the Lower Namoi Valley of southern Australia, which offers a significant improvement over models that only predict for surface soil layers. However, their approach only allowed for point estimates of PTF parameters at each pixel. As discussed, a fully Bayesian generalization would allow us to account for uncertainty in these models.

Regardless of the particular functional forms employed to map forest SOC, or the spatial scope of these studies, it is clear that establishing reproducible baselines of forest soil carbon stocks at the national scale requires a consistent framework. Though there is much further to go, our results argue for Bayesian hierarchical models as a central approach. Spatial hierarchical models in particular seem promising, as we have identified spatial structure in soil carbon data at the scales most relevant to international policy mechanisms such as REDD+. Using this approach, we were able to obtain reasonable maps of the distribution of soil organic carbon, even if the results were highly uncertain. Perhaps more importantly given the magnitude of uncertainty, a Hierarchical Bayesian approach provides an honest accounting of all levels of uncertainty. As more forest soil inventory datasets become available, it will become possible to test a variety of functional forms across multiple study regions. Spatial hierarchical models will accommodate such searches for improved predictive accuracy, while remaining scalable and allowing for full accounting of predictive uncertainty.

Conclusions

Here, we have identified four qualities that we seek in a predictive model for forest SOC: (1) predictive accuracy; (2) scalability; (3) quantification of uncertainty; and (4) computational efficiency. Our results suggest that Bayesian hierarchical spatial models provide a reasonable framework for addressing these goals. That said, our linear model versions of the 'scorpan' functions did not provide accurate predictions, despite considerable efforts to identify useful covariates. Further, we found that full prediction on high resolution grids is too computationally burdensome when using MCMC. More complex functions may allow us to account for the nonlinear relationships between SOC and environmental predictor variables, but relationships between soil inventory data and remote sensing covariates will probably still be highly variable. Given this reality, accounting for uncertainty in model-based predictions will be important for both policy and scientific applications. There is a clear need for a consistent framework that allows for full accounting of model error, and Bayesian hierarchical models accomplish this.

While we argue that hierarchical models are ideal for digital soil mapping of forest SOC, there are still some computational challenges that arise from this approach, particularly for high resolution prediction. More efficient estimation of the posterior predictive distributions will be required before the method can be widely applied. We have attempted here to circumvent this difficulty by interpolating predictions from a coarse grid via multilevel B splines, but the resulting maps ignore much of the local detail. Efficient parameter estimation in high dimensional models is an area of active research in the statistics literature (Rue et al. 2009, Eidsvik et al. 2012), and future work should look to apply these methods to high dimensional spatial prediction.

Model-based prediction offers significant improvement over traditional soil survey for mapping soil attributes. Advances in computational statistics have facilitated the development of the next generation of predictive models in forest and soil science. Despite these developments, we still have a long way to go, before we can fully utilize the massive remote sensing datasets that have attracted significant interest over the last decade. Future collaborations between forest scientists, soil scientists, and applied statisticians will contribute to the development of new practical tools for soil nutrient mapping. While we suggest that Bayesian hierarchical models are an appropriate method for fitting models, and that more complex nonlinear functions will be necessary to maximize predictive accuracy, it is important to remember that there is no "one size fits all" approach for mapping soil attributes. Each model will be both context and scale dependent, and model design should be tailored to each particular study. With this in mind, developing more general rules for predictive models such as those we suggest here may be more useful than selecting a model *a priori* for achieving consistency and reproducibility among soil mapping studies.

Appendix A: Additional tables and figures for chapter 2

Table A.1: Covariates of each dataset, with posterior inclusion probabilities (PIP) from the reversible jump MCMC procedure. PIP tracks the number of iterations in which each covariate was included in the selected model, and can be regarded as a level of support for that variable as a predictor.

Soil data	PIP
Northing (X)	0
Easting (Y)	0
field-measured soil organic matter (SOM)	1
Normalized difference vegetation index (NDVI)	0.04
Tassled cap, greenness band (TC2)	0.009
Compound topographic index (CTI)	0.0006
Elevation (ELEV)	0.0003
Leaf off NDVI (LO NDVI)	0.023
Available soil water content (AWC)	0.074
Soil bulk density (BD)	0.003
Cation exchange capacity (CEC)	0.015
Electrical Conductivity (ECEC)	0.001
# frost free days (FROST)	0.002
Organic matter content, SSURGO (OrgMat)	0.0004
Percent Clay (CLAY)	0.0005
pH	0.003
Water content at 15 bars (WC15bar)	0.012
Protist data	
Species richness (RICHNESS)	0.182
Ankistrodesmus (ANKISTRO)	0.378
Chlamydomonas (CHLAMYS)	0.186
Microflagellates (MICROFLA)	0.191
Halteria (HALTERIA)	0.835
Monostyla (MONOSTY)	0.247
Stylonychia(STYLON)	0.266
Amoeba (AMOEBAL)	0.273
Amoeba (AMOEBAS)	0.239
EYESPOTS	0.272
Rotaria (ROTARIA)	0.234
Brachionus (BRACHION)	0.292
Frontonia (FRONTON)	0.302

Scendesmus (SCENDES)	0.176
Diatom sp. (DIATOM)	0.274
Ankistrodesmus (ANKISTRS)	0.469
Staurastrum (STAURST)	0.2
Stentor (STENTI)	0.3
Colpidium (COLP)	0.335
Heliozoa (HELIO)	0.277
Paramecium (PARAMECIUM)	0.304
Euglena (EUGLENA)	0.303
Gastrotrich sp. (GASTRO)	0.344
Oxytricha (OXYTRICH)	0.301
P. bursarii (PBURSARI)	0.281
Stentor (STENTO)	0.286
Spriostomum (SPIROSTO)	0.301
Netrium (NETRIUM)	0.3
Amoeba (AMOEBA)	0.297
Aspidisca (ASPIDISC)	0.295
Forest data	
Age (A)	0.39
Mean diameter (MD)	0.957
Basal area (BA)	0.0009
# trees/acre (N)	0.022
1/age (INVA)	0.999
Log basal area (LB)	1
Log mean diameter (LMD)	0.785
Height of dominants (HD)	0.614
HD/age (SQ)*	0.585
Log HD/age (LSQ)*	0.588
Log # trees/acre (LN)	0.246

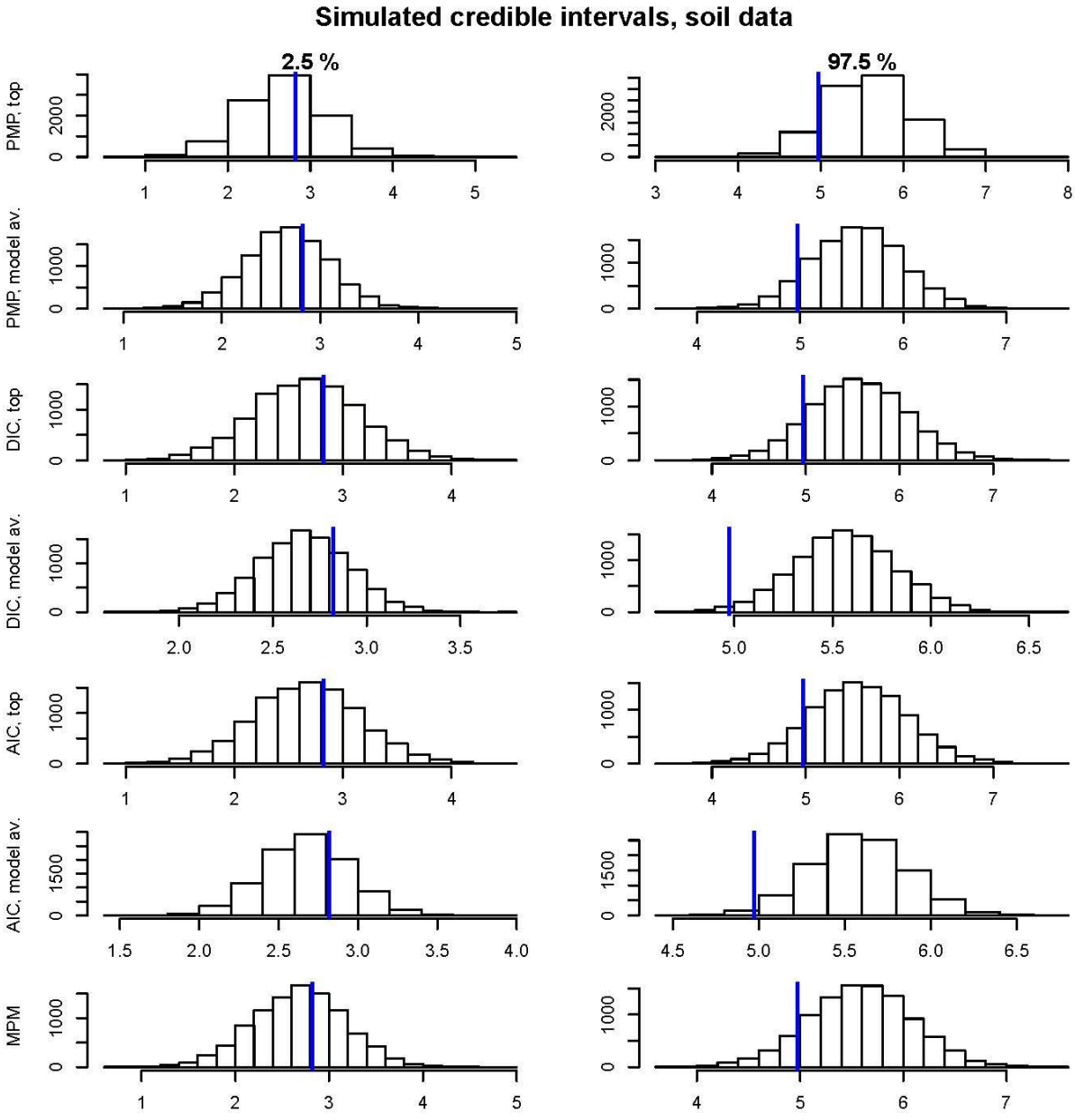

Figure A.1: Histograms of simulated 2.5 and 97.5 percent credible intervals for the soils dataset. The blue line indicates the observed CI for the validation data.

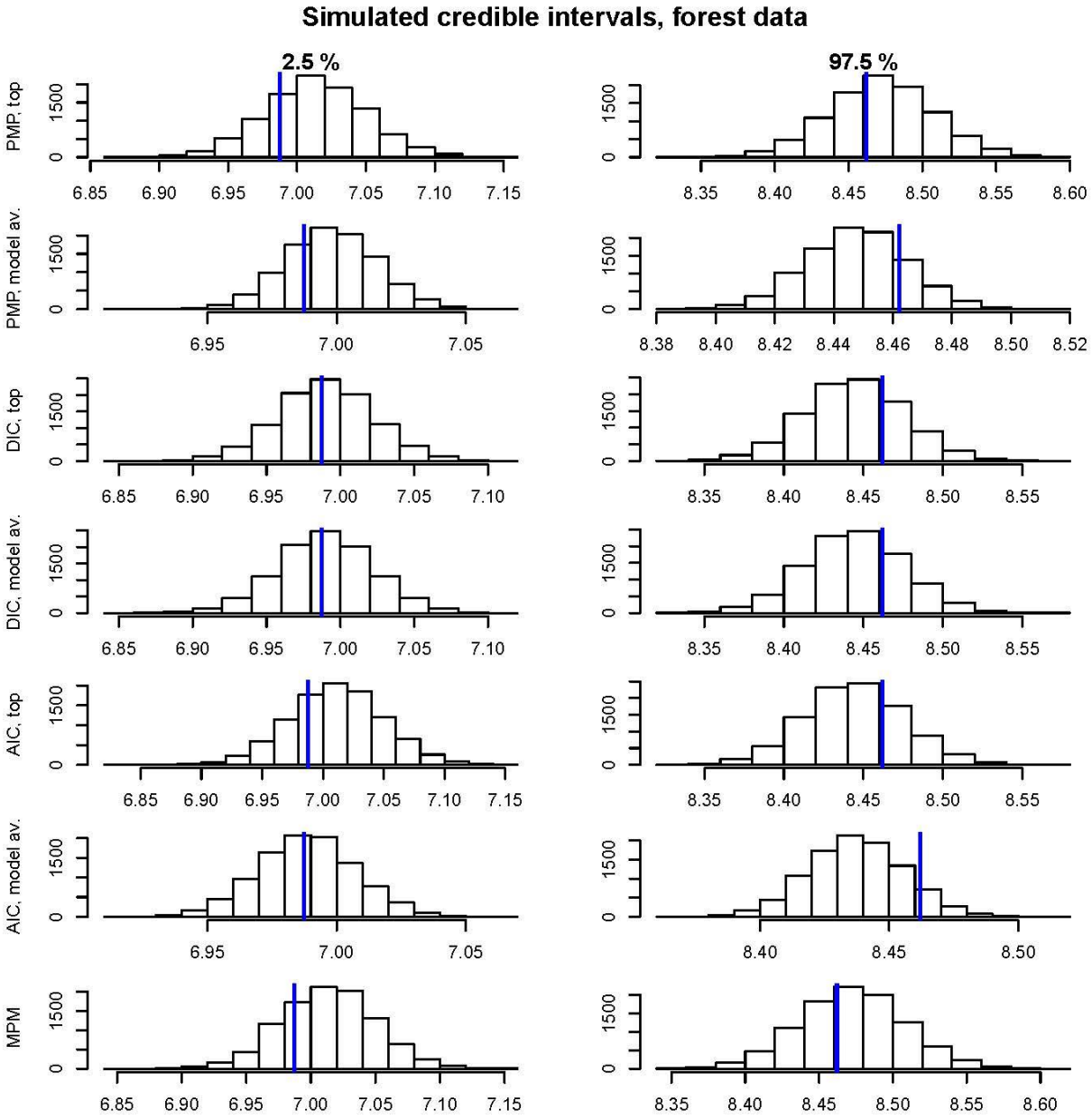

Figure A.2: Histograms of simulated 2.5 and 97.5 percent credible intervals for the forest data. The blue line indicates the observed CIs of the validation dataset.

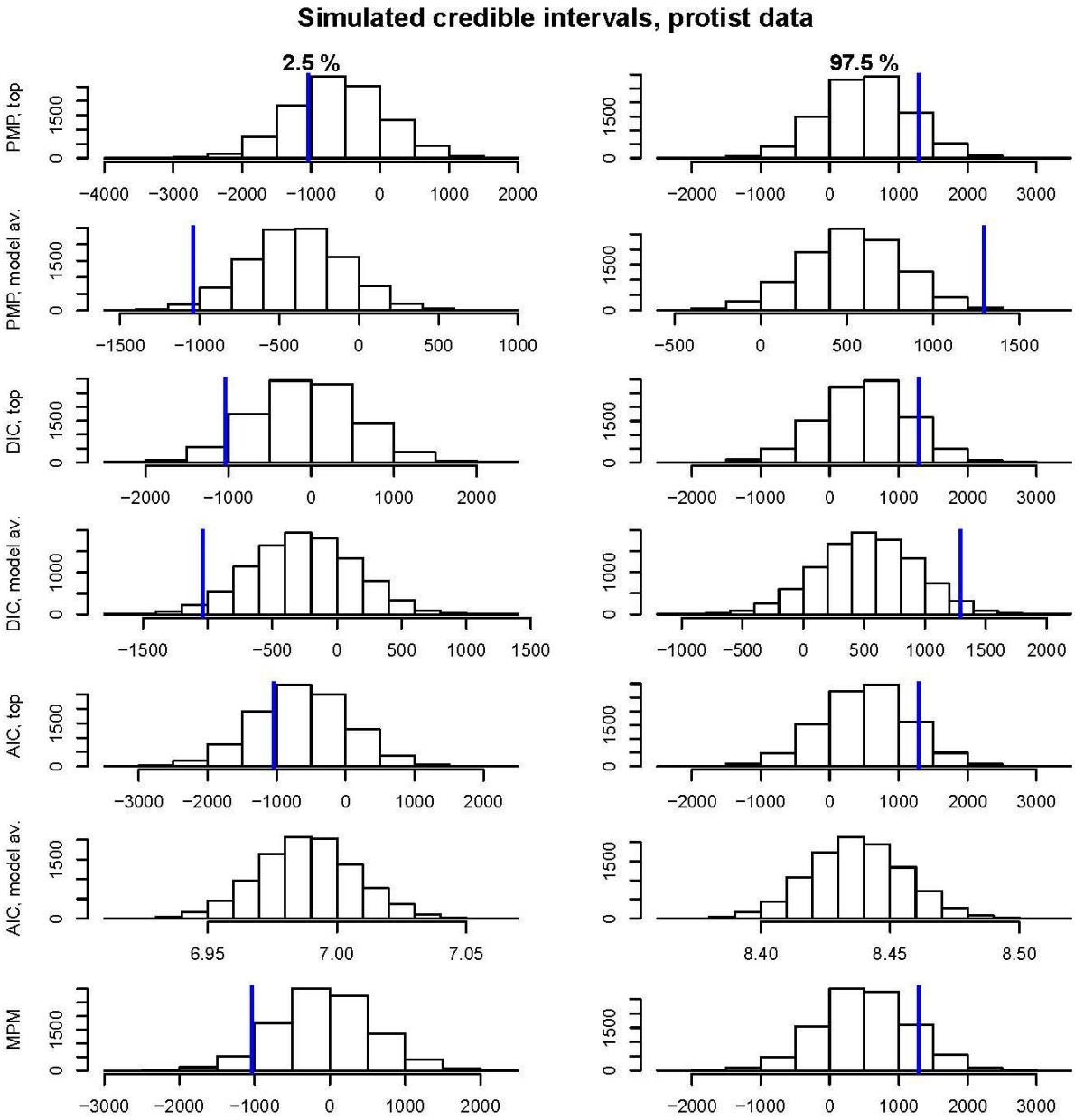

Figure A.3: Histograms of simulated credible intervals for the protist data. The blue line indicates the observed 2.5 and 97.5 CI of the hold-out validation dataset.

Appendix B: Complete description of the variable selection analysis for identifying predictive models for forest SOC

Table B.1 presents the results of the Bayesian variable selection procedure for each dataset. Note that for all three datasets, most of the covariates we initially considered had low posterior inclusion probabilities (PIPs). These values correspond to the proportion of MCMC iterations in which each covariate was included in the candidate model by the variable selector, and are frequently used in Bayesian inference as an indicator of the relationship between a covariate and the response (Berger and Barbieri 2004). Looking at the highest probability models (HPMs) according to variable selection, the NJ data resulted in the sparsest model with 4 of 11 covariates. The model for Germany contained 6 of 23 potential covariates, while the model for Europe included 12 of 23 covariates. However, note that in each case, several variables that may exhibit collinearity have all been included in the HPM. For example, the HPM for Europe contains mean diurnal range (BIO2), maximum temperature of the warmest month (BIO5), temperature annual range (BIO7), mean temperature of the driest quarter (BIO9), and mean temperature of the coldest quarter (BIO11). Further analysis is merited to establish which of these climatic predictors provides the best fit to soil carbon and to refine predictive models.

Note also that the results outlined in table 1 do not include soil Nitrogen. For each dataset total N was easily the best predictor of soil carbon, with a PIP = 1.0 in initial runs of the variable selector. The variable selection routine we employed is sensitive to this situation, where one variable provides a much better fit than any of the others, and it is

possible for the MCMC routine to become ‘fixed’ and provide a poor posterior sample for the other covariates. Since our goal is produce a good predictive model based on the ‘scorpan’ function, we excluded soil N from the variable selection analysis to avoid this issue. Nitrogen is included in all models, and summary plots of its relationship to soil C in each dataset are shown below.

Table B.1: Posterior inclusion probabilities from the Bayesian variable selection procedure for all potential covariates. An asterisk indicates that a variable was retained in the highest probability model retained by the variable selector.

New Jersey	PIP	HPM
ELEV	0.2	
SLOPE	0.51	
PRECIP	0.81	*
MAXTEMP	0.23	
MINTEMP	0.26	
TC1	0.78	*
TC2	0.22	
TC3	0.36	
NDVI	0.554	*
AWC	0.428	
pH	0.94	*

Germany	PIP	HPM	**Europe**	PIP	HPM
ELEV	0.19		ELEV	0.99	*
SLOPE	0.56	*	SLOPE	0.83	*
BIO1	0.3		BIO1	1	*
BIO2	0.24		BIO2	0.99	*
BIO3	0.22		BIO3	0.14	
BIO4	0.24		BIO4	0.15	
BIO5	0.25		BIO5	0.55	*
BIO6	0.3		BIO6	0.54	
BIO7	0.23		BIO7	0.73	*
BIO8	0.63	*	BIO8	0.55	
BIO9	0.19		BIO9	0.59	*
BIO10	0.24		BIO10	0.16	

BIO11	0.32		BIO11	1	*
BIO12	0.36		BIO12	0.15	
BIO13	0.32		BIO13	0.4	*
BIO14	0.98	*	BIO14	0.99	*
BIO15	0.76	*	BIO15	0.26	
BIO16	0.49	*	BIO16	0.3	
BIO17	0.23		BIO17	0.42	
BIO18	0.32		BIO18	0.16	
BIO19	0.24		BIO19	0.3	*
GEO	0.63	*	GEO	1	*
NPP	0.17		NPP	0.09	

As discussed in the methods, our goal is to preserve the general model structure of the ‘scorpan’ approach, since it is based on the major factors that influence soil formation. As a result, in each dataset we included covariates with low posterior inclusion probabilities in our further analysis. For the NJ data, these were slope and available water capacity (AWC). We include the former because it is a topographic variable and the latter because it is closely related to soil texture.

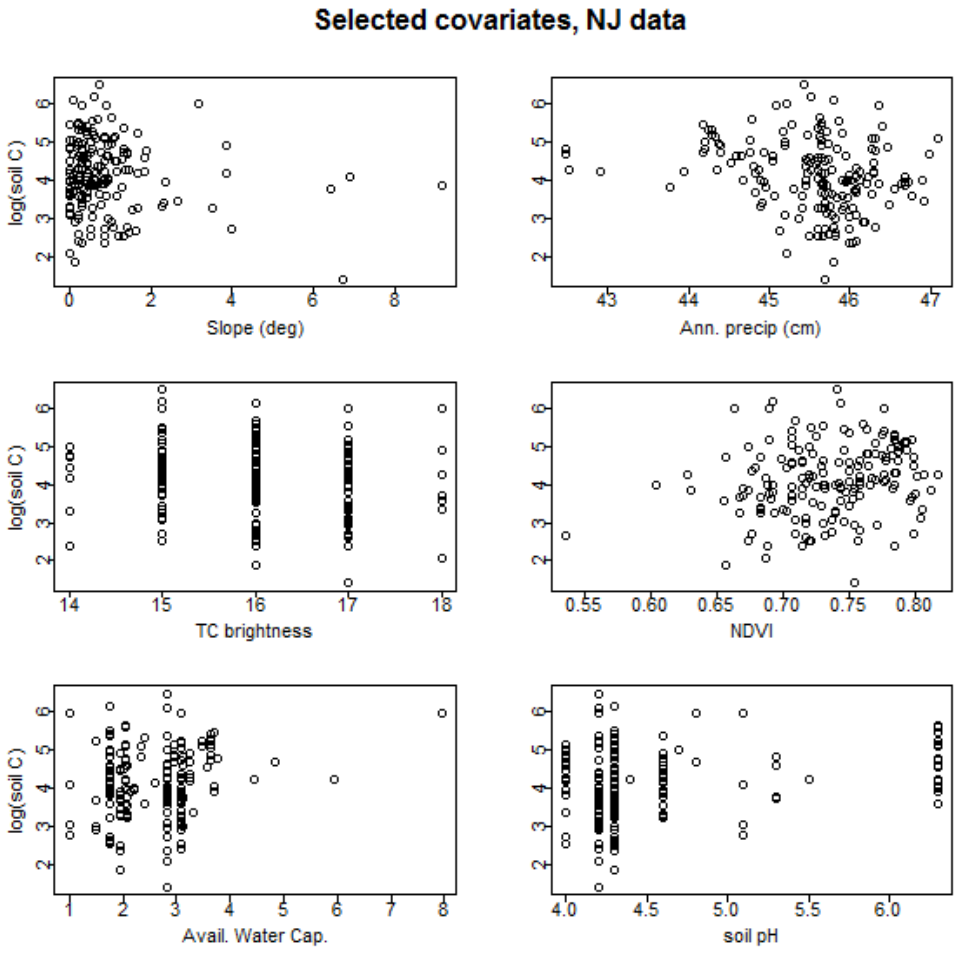

Figure B.1: Scatterplots of log-transformed soil organic carbon with selected covariates for the NJ dataset.

In general, none of these variables are strongly correlated with soil carbon. The strongest relationships are with annual precipitation, available water capacity, and soil pH, each with $\rho = \sim 0.2$ (figure B.1). The residuals between soil C and each of these covariates do not exhibit significant spatial dependence (figure B.2). Given that we have previously established that there is no spatial structure in soil carbon for our NJ data, this is not a surprising result.

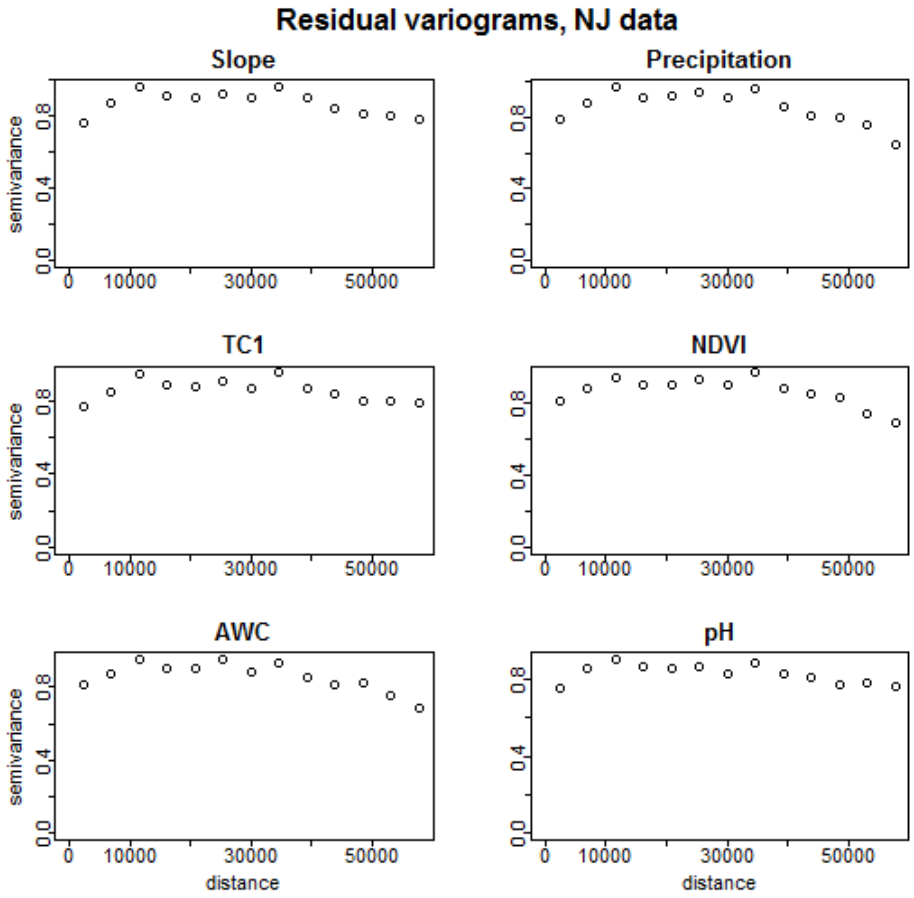

Figure B.2: Empirical variograms of the residuals between log-transformed soil C and each covariate.

For the Germany dataset, four climatic variables had the strongest relationship with soil C; mean temperature of the wettest quarter (BIO8), mean precipitation of the driest month (BIO14), precipitation seasonality (BIO15), and precipitation of the wettest quarter (BIO16) (figure B.3). We also consider slope, bedrock geology (GEO), and mean annual net primary production (NPP) in order to include variables related to all soil forming factors. While the four climate variables are related to one another, exploratory

analyses (e.g. R^2, AIC) suggested that the model including all four provides the best fit for the data. In general, the residuals between these covariates and soil C are spatially autocorrelated (figure B.4).

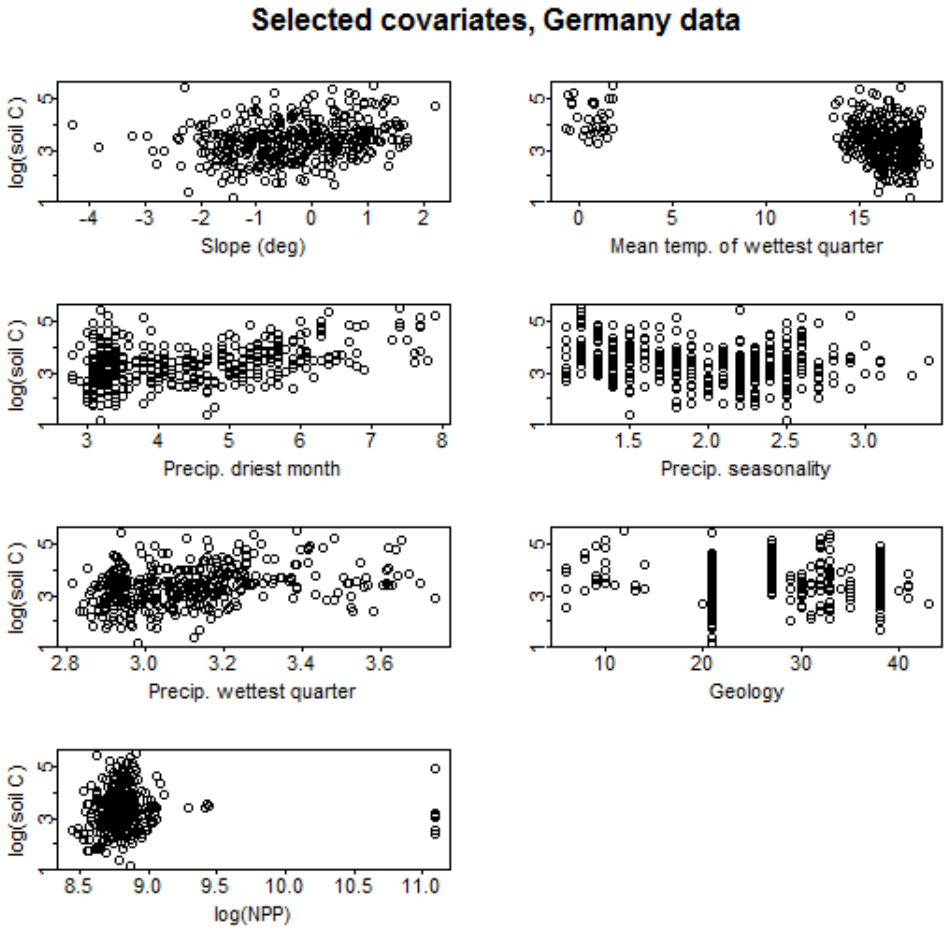

Figure B.3: Scatterplots between log-transformed soil carbon and selected covariates for the Germany dataset.

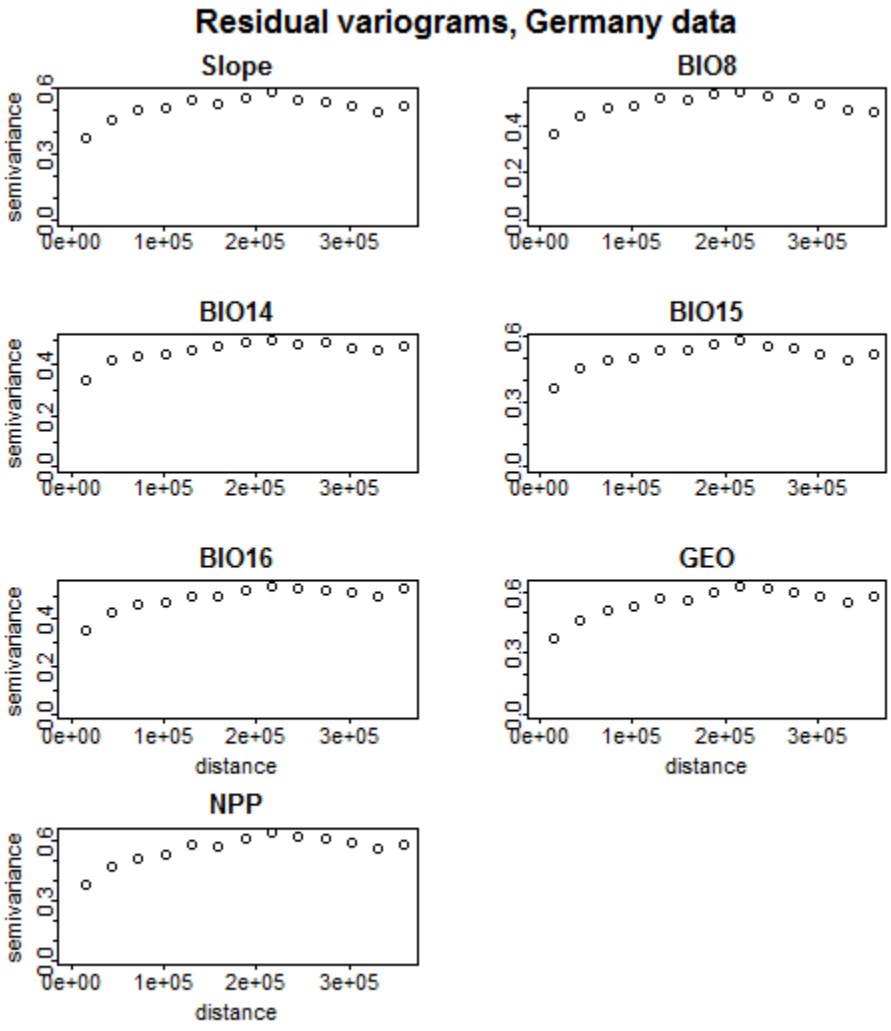

Figure B.4: Variograms of the residuals between log-transformed soil C and the selected covariates for the Germany dataset.

Selected covariates for the Europe dataset include elevation (ELEV), mean annual temperature (BIO1), mean diurnal range (BIO2), precipitation of the driest month (BIO14), precipitation of the coldest quarter (BIO19), bedrock geology (GEO) and net primary production (NPP) (figure B.5). As with the preceding two datasets, the

relationship between these covariates and forest soil C are rather variable. Several do exhibit general linear trends, most notably BIO1 and Bio14, though there is significant residual error in this relationship. Geology and NPP are generally poor predictors, but we again include them in the final model to specify a complete 'scorpan' function.

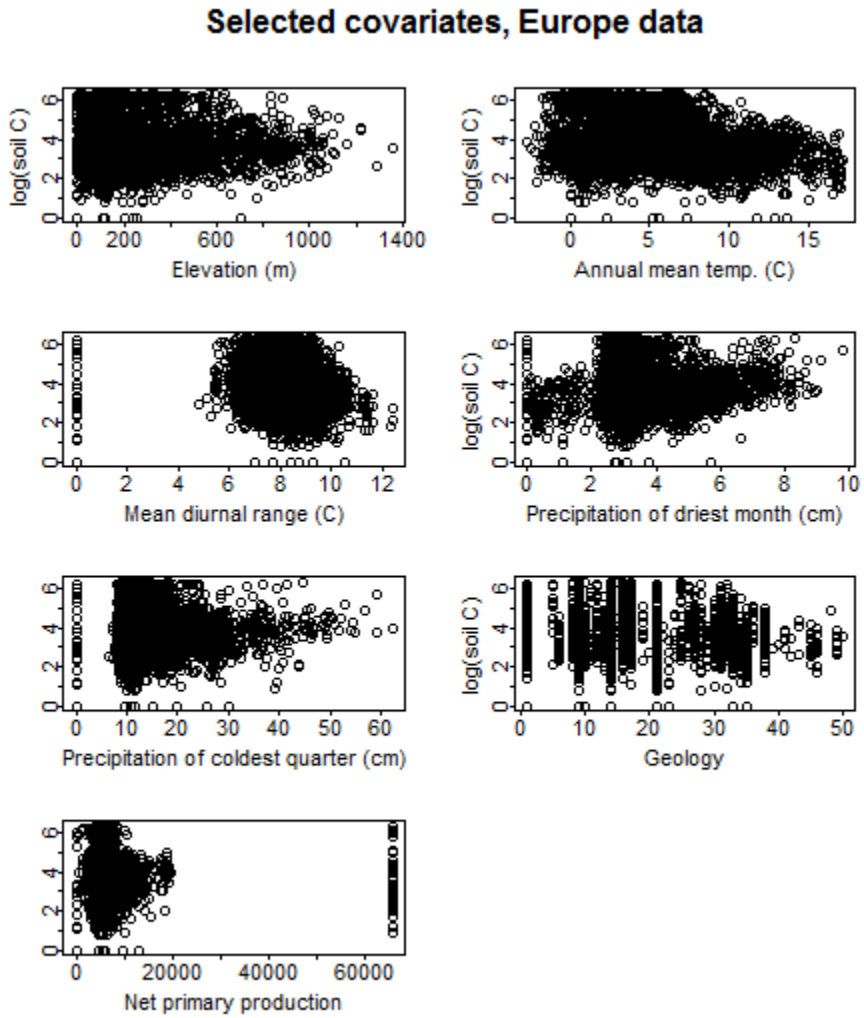

Figure B.5: Selected covariates for the Europe dataset.

References

Achard, F., R. DeFries, H. Eva, M. Hansen, P. Mayaux, and H. J. Stibig. 2007. Pan-tropical monitoring of deforestation. Environmental Research Letters 2.

Agarwal, D. K., J. a. Silander, A. E. Gelfand, R. E. Dewar, and J. G. Mickelson. 2005. Tropical deforestation in Madagascar: analysis using hierarchical, spatially explicit, Bayesian regression models. Ecological Modelling 185:105-131.

Aho, K., D. Derryberry, and T. Peterson. 2014. Model selection for ecologists: the worldviews of AIC and BIC. Ecology 95:631-636.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267-281 *in* B. N. Petrox and F. Caski, editors. 2nd International Symposium on Information Theory. Budapest.

Anderson, D. R., W. A. Link, D. H. Johnson, and K. P. Burnham. 2001. Suggestions for presenting the results of data analyses. The Journal of Wildlife Management 65:373-378.

Ando, T. 2010. Bayesian Model Selection and Statistical Modeling. CRC Press, Boca Raton, FL.

Angelsen, A. 2009. Realising REDD+: National Strategy and Policy Options. Center for International Forestry Research, Bogor.

Asner, G. P. 2011. Painting the world REDD: addressing scientific barriers to monitoring emissions from tropical forests. Environmental Research Letters 6.

Babcock, C., J. Matney, A. O. Finley, A. Weiskittel, and B. D. Cook. 2013. Multivariate spatial regression models for predicting individual tree structure variables using LiDAR data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6:6-14.

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall, Boca Raton, FL.

Banerjee, S., A. E. Gelfand, A. O. Finley, and H. Sang. 2008. Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70:825-848.

Barling, R., I. D. Moore, and R. B. Grayson. 1994. A quasi-dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content. Water Resources Research 30:1029-1044.

Bauer, J., M. Herbst, J. Huisman, L. Weihermuller, and H. Vereecken. 2008. Sensitivity of simulated soil heterotrophic respiration to temperature and moisture reduction functions. Geoderma 145:17-27.

Bayarri, M. J., and J. O. Berger. 2004. The Interplay of Bayesian and frequentist analysis. Statistical Science 19:58-80.

Bayarri, M. J., J. O. Berger, A. Forte, and G. Garcia-Donato. 2012. Criteria for Bayesian model choice with application to variable selection. The Annals of Statistics 40:1550-1577.

Bayarri, M. J., and M. E. Castellanos. 2007. Bayesian checking of the second levels of hierarchical models (with discussion). Statistical Science 56:322-367.

Beer, C., M. Reichstein, E. Tomelleri, P. Ciais, M. Jung, N. Carvalhais, C. Rödenbeck, M. A. Arain, D. Baldocchi, G. B. Bonan, A. Bondeau, A. Cescatti, G. Lasslop, A. Lindroth, M. Lomas, S. Luyssaert, H. Margolis, K. W. Oleson, O. Roupsard, E. Veenendaal, N. Viovy, C. Williams, F. I. Woodward, and D. Papale. 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science (New York, N.Y.) 329:834-8.

Ben Dor, E., and A. Banin. 1989. Determination of organic matter content in arid-zone soils using a simple "loss-on-ignition" method. Communications in Soil Science and Plant Analysis 20:1675-1682.

Berger, J. 2006. The case for objective Bayesian analysis. Bayesian Analysis 1:385-402.

Berger, J. O., and M. M. Barbieri. 2004. Optimal predictive model selection. The Annals of Statistics 32:870-897.

Berger, J. O., and L. R. Pericchi. 1996. The intrinsic Bayes factor for model selection and prediction. Journal of the American Statistical Association 91:109-122.

Bernoux, M., D. Arrouays, C. E. P. Cerri, and C. C. Cerri. 2006. Regional organic carbon storage maps of the western Brazilian Amazon based on prior soil maps and geostatistical interpolation. Developments in Soil Science 31:497-506.

Birdsey, R. A. 1992. Carbon storage and accumulation in United States forest ecosystems. USDA Forest Service General Technical Report WO-59.

Birdsey, R., G. Angeles-Perez, W. A. Kurz, A. J. Lister, M. Olguin, Y. Pan, C. Wayson, B. Wilson, and K. Johnson. 2013. Approaches to monitoring changes in carbon stocks for REDD+. Carbon Management 4:519-537.

Blake, G. R., and K. H. Hartge. 1986. Bulk Density. Pages 377-382 *in* A. Klute, editor. Methods of Soil Analysis Part 1. Physical and Mineralogical Methods. . American Society of Agronomy, Inc.; Soil Science Society of America Inc., Madison, WI, USA.

Bredja, J. J., M. J. Mausbach, J. J. Goebel, D. L. Allan, T. H. Dao, D. L. Karlen, T. B. Moorman, and J. L. Smith. 2001. Estimating surface soil organic carbon content at a regional scale using the national resource inventory. Soil Science Society of America Journal 65:842-849.

Bricklemyer, R., P. Miller, P. Turk, K. Paustian, T. Keck, and G. Nielsen. 2007. Sensitivity of the Century model to scale-related soil texture variability. Soil Science Society of America Journal 71:784-792.

de Brogniez, D., C. Ballabio, B. van Wesemael, R. J. Jones, A. Stevens, and L. Montanarella. 2014. Topsoil Organic Carbon Map of Europe. Pages 393-405 *in* A. E. Hartemink and K. McSweeney, editors. Soil Carbon. . Springer.

Brown, S. 1997. Estimating biomass and biomass change of tropical forests: a primary. . Rome.

Burkhart, H., and P. Sprinz. 1984. Compatible cubic volume and basal area projection equations for thinned old-field Loblolly Pine plantations. Forest Science 30:86-93.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Edition. . Springer, New York, NY.

Burnham, K. P., and D. R. Anderson. 2004. Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods Research 33:261.

Burnham, K. P., D. R. Anderson, and K. P. Huyvaert. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral Ecology and Sociobiology 65:23-35.

Cannell, M. G. R., R. C. Dewar, and J. H. M. Thornley. 1992. Carbon flux and storage in European forest. Pages 256-271 *in* A. Teller, P. Mathy, and J. N. R. Jeffers, editors. Responses of Forest Ecosystems to Environmental Changes. . Elsevier, London.

Casella, G. R., F. J. Giron, M. L. Martinez, and E. Moreno. 2009. Consistency of Bayesian procedures for variable selection. The Annals of Statistics 27:1207-1228.

Causarano, H., J. Shaw, A. Fransluebbers, D. Reeves, R. Raper, K. Balkcom, M. Norfleet, and R. Isaurralde. 2007. Simulating field-scale soil organic carbon dynamics using EPIC. Soil Science Society of America Journal 71:1174-1185.

Cerri, C. C., M. Bernoux, D. Arrouays, B. J. Feigl, and M. C. Piccolo. 2000. Carbon Stocks in Soils of the Brazilian Amazon. Page *in* R. Lal, M. Kimble, and B. A. Stewart, editors. Global climate change and tropical ecosystems. . CRC Press.

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

Chave, J., R. Condit, S. Aguilar, A. Hernandez, S. Lao, and R. Perez. 2004. Error propagation and scaling for tropical forest biomass estimates. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 359:409-420.

Chen, F., D. E. Kissel, L. T. West, and W. Adkins. 2000a. Field-scale mapping of surface soil organic carbon using remotely sensed imagery. Soil Science Society of America Journal2 64.

Chen, J., and Z. Chen. 2008. Extended Bayesian information criterion for model selection with large model spaces. Biometrika 95:759-771.

Chen, W., J. Chen, J. Liu, and J. Cihlar. 2000b. Approaches for reducing uncertainties in regional forest carbon balance. Global Biogeochemical Cycles 14:827-838.

Chilles, J., and P. Delfiner. 1999. Geostatistics: Modeling Spatial Uncertainty. . Wiley, New York.

Clough, B., and E. J. Green. 2013. Comparing spatial and non-spatial approaches for predicting forest soil organic carbon at unsampled locations. Mathematical and Computational Forestry and Natural Resource Science 5:115-125.

Clyde, M. A., J. Ghosh, and M. L. Littman. 2011. Bayesian adaptive sampling for variable selection and model averaging. Journal of Computational and Graphical Statistics 20:80-101.

Congdon, P. 2006. Bayesian Statisical Modeling, 2nd Edition. . Wiley.

Corbera, E., and H. Schroeder. 2011. Governing and implementing REDD+. Environmental Science and Policy 14:89-99.

Couteaux, M., P. Bottner, and B. Berg. 1995. Litter decomposition, climate and litter quality. Trends in ecology & evolution 10:63-66.

Cressie, N. A. C. 1993. Statistics for Spatial Data. . Springer, New York, NY.

Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, C. K. Wikle, E. Applications, S. Clark, M. Jay, K. Wikle, and N. Carolina. 2012. Accounting for Uncertainty in

Ecological Analysis : the Strengths and Limitations of Hierarchical Statistical Modeling Reviewed work (s): Published by : Ecological Society of America content in a trusted digital archive . We use information technology a.

Dellaportas, P., J. J. Forster, and I. Ntzoufras. 2002. On Bayesian model and variable selection using MCMC. Statistics and Computing 12:27-36.

Dixon, R. K., A. M. Solomon, S. Brown, R. A. Houghton, M. C. Trexier, and J. Wisniewski. 1994. Carbon pools and flux of global forest ecosystems. Science 263:185-190.

Dubayah, R., and J. Drake. 2000. Lidar remote sensing for forestry. Journal of Forestry 98:44-46.

Edwards, D. P., B. Fisher, and E. Boyd. 2010. Protecting degraded rainforests: enhancement of forest carbon stocks under REDD+. Conservation Letters 3:313-316.

Eidsvik, J., A. O. Finley, S. Banerjee, and H. Rue. 2012. Approximate Bayesian inference for large spatial datasets using predictive process models. Computational Statistics & Data Analysis 56:1362-1380.

Eswaran, H., E. Van Den Berg, and P. Reich. 1993. Organic carbon in soils of the world. Soil Science Society of America Journal 57:192-194.

Ettema, C. H., and D. A. Wardle. 2002. Spatial soil ecology. Trends in ecology & evolution 17:177-183.

Finley, A. O., S. Banerjee, and B. Basso. 2011. Improving Crop Model Inference Through Bayesian Melding With Spatially Varying Parameters. Journal of Agricultural, Biological, and Environmental Statistics 16:453-474.

Finley, A. O., S. Banerjee, and B. P. Carlin. 2007a. spBayes: An R package for univariate and multivariate hierarchical point-referenced spatial models. Journal of Statistical Software 19:1-24.

Finley, A. O., S. Banerjee, and A. E. Gelfand. (n.d.). spBayes for large univariate and multivariate point-referenced spatio-temporal data models. Journal of Statistical Software.

Finley, A. O., S. Banerjee, and A. E. Gelfand. 2012. Bayesian dynamic modeling for large space-time datasets using Gaussian predictive process models. Journal of Geographical Systems:29-47.

Finley, A. O., S. Banerjee, and D. W. Macfarlane. 2010. A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas. Public Health.

Finley, A. O., S. Banerjee, and R. E. McRoberts. 2007b. A Bayesian approach to multi-source forest area estimation. Environmental and Ecological Statistics 15:241-258.

Finley, A. O., S. Banerjee, and R. E. McRoberts. 2009. Hierarchical spatial models for predicting tree species assemblages across large domains. Annals of Applied Statistics 3:1052-1079.

Fortin, M.-J., P. Drapeau, and P. Legendre. 1989. Spatial autocorrelation and sampling design in plant ecology. Vegetatio 83:209-222.

Gallardo, A., and R. Parama. 2007. Spatial variability of soil elements in two plant communities of NW Spain. Geoderma 139:199-208.

Garcia-Donato, G., and A. Forte. 2012. Package "BayesVarSel."

Garcia-Donato, G., and M. A. Martinez-beneito. 2013. On sampling strategies in Bayesian variable selection problems with large model spaces. Journal of the American Statistical Association 108:340-352.

Garten Jr., C. T., S. Kang, D. J. Brice, C. W. Schadt, and J. Zhou. 2007. Variability in soil properties at different spatial scales (1m-1km) in a deciduous forest ecosystem. Soil Biology and Biochemistry 39:2621-2627.

Gelfand, A. E. 2012. Hierarchical modeling for spatial data problems. Spatial Statistics 1:30-39.

Gelfand, A. E., S. Banerjee, and A. O. Finley. (n.d.). Spatial design for knot selection in knot-based dimension reduction models. Page Spatio-temporal Design: Advances in Efficient Data Acquisition. Wiley.

Gelfand, A. E., and D. K. Dey. 1994. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society: Series B 56:510-514.

Gelfand, A. E., A. M. Schmidt, S. Banerjee, and C. F. Sirmans. 2004. Nonstationary Multivariate Process Modeling Through Spatially Varying Coregionalization. Test 13:263-312.

Gelman, A. E. 2003. A Bayesian formulation of exploratory data analysis an goodness-of-fit testing. International Statistical Review 71:369-382.

Gelman, A. E. 2007. Discussion of "Bayesian checking of the second levels of hierarchical models," by M. J. Bayarri and M. E. Castellanos. Statistical Science 22:349-352.

Gelman, A. E., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2003. Bayesian Data Analysis, 2nd edition. . CRC Press, London.

Gelman, A. E., Y. Goegebeur, F. Tuerlinckx, and I. Van Mechelen. 2000. Diagnostic checks for discrete data regression models using posterior predictive simulations. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 49:247-268.

Gelman, A. E., X.-L. Meng, and H. Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica 6:733-807.

George, E. I., and R. E. McCulloch. 1993. Variable Selection Via Gibbs Sampling. Journal of the American Statistical Association 88:881-889.

Gibbs, H. K., S. Brown, J. O. Niles, and J. A. Foley. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environmental Research Letters 2.

Goetz, S. J., and R. Dubayah. 2011. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. Carbon Management 2.

Gonzalez, O. J., and D. R. Zak. 1994. Geostatistical analysis of soil properties in a secondary tropical dry forest, St. Lucia, West Indies. Plant and Soil 163.

Goodale, C. L., M. J. Apps, R. A. Birdsey, C. B. Field, L. S. Heath, R. A. Houghton, J. C. Jenkins, G. H. Kohlmaier, S. Liu, G.-jan Nabuurs, S. Nilsson, A. Z. Shvidenko, S. E. Applications, and N. Jun. 2002. Forest Carbon Sinks in the Northern Hemisphere. Ecological applications 12:891-899.

Goodale, C. L., M. J. Apps, R. A. Birdsey, C. B. Field, L. S. Heath, R. A. Houghton, J. C. Jenkins, G. H. Kohlmaier, S. Liu, G.-jan Nabuurs, S. Nilsson, A. Z. Shvidenko, S. E. Applications, and N. Jun. 2012. Forest Carbon Sinks in the Northern Hemisphere. America.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. . Oxford University Press, New York.

Green, E. J., and B. Bullock. 2013. Assessing sampling and model error in standing green weight estimation for a Loblolly Pine plantation. Forest Science in press.

Green, E. J., D. W. MacFarlane, and H. T. Valentine. 2000. Bayesian Synthesis for Quantifying Uncertainty in Predictions from Process Models. Tree Physiology 20:415-419.

Green, E. J., D. W. MacFarlane, H. T. Valentine, and W. E. Strawderman. 1999. Assessing Uncertainty in a Stand Growth Model by Bayesian Synthesis. Forest Science 45.

Green, E. J., and H. T. Valentine. 1998. Bayesian analysis of the linear model with heterogeneous variance. Forest Science 42:456-464.

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711-732.

Grimm, R., T. Behrens, M. Marker, and H. Elsenbeer. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island- Digital soil mapping using random forests analysis. Geoderma 146:102-113.

Gruneberg, E., D. Ziche, and N. Wellbrock. 2014. Organic carbon stocks and sequestration rates of forest soils in Germany. Global Change Biology 20:2644-2662.

Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152:195-207.

Guan, Y., and M. Stephens. 2011. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. The Annals of Applied Statistics 5:1780-1815.

Guhaniyogi, R., A. O. Finley, S. Banerjee, and T. Kauranne. 2013. Modeling complex spatial dependencies: low-rank spatially-varying cross-covariances with application to soil nutrient data. Journal of Agricultural, Biological, and Environmental Statistics 18:274-298.

Hasse, J. E., and R. G. Lathrop. 2010. Changing Landscapes in the Garden State: Urban Growth and Open Space Loss in NJ 1986 thru 2007. . New Brunswick, NJ, USA.

Heath, L. S., J. E. Smith, and R. A. Birdsey. 2003. Carbon Trends in U . S . Forestlands : A Context for the Role of Soils in Forest Carbon Sequestration:35-45.

Hengl, T., G. Heuvelink, and A. Stein. 2003. Comparison of kriging with external drift and. Soil Science.

Hengl, T., G. Heuvelink, and A. Stein. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120:75-93.

Herold, M., and M. Skutsch. 2011. Monitoring, reporting and verification for national REDD+ programmes: two proposals. Environmental Research Letters 6.

Hijmans, R., S. Cameron, J. Parra, P. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Journal of Climatology 25:1965-1978.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky. 1999. Bayesian model averaging: a tutorial. Statistical Science 14:382-401.

Hoover, C. 2008. Field Measurements for Forest Carbon Monitoring: A Landscape-Scale Approach. . Springer.

Houghton, R. a. 2003. Why are estimates of the terrestrial carbon balance so different? Global Change Biology 9:500-509.

Huntingford, C., J. a. Lowe, B. B. B. Booth, C. D. Jones, G. R. Harris, L. K. Gohar, and P. Meir. 2009. Contributions of carbon cycle uncertainty to future climate projection spread. Tellus B 61:355-360.

Isaaks, E. H., and M. Srivastava. 1989. An Introduction to Applied Geostatistics. . Oxford University Press, Oxford.

Jeffreys, H. 1961. Theory of Probability. . Oxford University Press, Oxford.

Jenkins, J. C., D. C. Chojnacky, L. S. Heath, and R. A. Birdsey. 2003. National-Scale Biomass Estimators for United States Tree Species. Forest Science 49.

Jenny, H. 1941. Factors of Soil Formation, A System of Quantitative Pedology. . McGraw-Hill, New York.

Jenny, H., A. Salem, and J. Wallis. 1968. Interplay of soil organic matter and soil fertility with state factors and soil properties. Study week on organic matter and soil fertility. Pontif. Acad. Sci. Ser. Varia 32:1-33.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. Trends in ecology & evolution 19:101-8.

Johnson, K., R. Birdsey, A. O. Finley, A. Swantaran, R. Dubayah, C. Wayson, and R. Riemann. 2014. Integrating forest inventory and analysis data into a LIDAR-based carbon monitoring system. Carbon Balance and Management 9.

Jones, R. J. A., R. Hiederer, E. Rusco, P. J. Loveland, and L. Montanarella. 2005. Estimating soil organic carbon in the soils of Europe for policy support. European Journal of Soil Science 56:655-671.

Kadane, J. B., and N. A. Lazar. 2004. Methods and criteria for model selection. Journal of the American Statistical Association 99:279:290.

Kass, R. E., and A. E. Raftery. 1995. Bayes Factors. Journal of the American Statistical Association 90:773-795.

Kaye, J. P., A. Majumdar, C. Gries, A. Buyantuyev, N. B. Grimm, D. Hope, G. D. Jenerette, W. X. Zhu, and L. Baker. 2008. Hierarchical Bayesian scaling of soil properties across urban, agricultural, and desert ecosystems. Ecological applications 18:132-45.

Keenan, T. F., M. S. Carbone, M. Reichstein, and A. D. Richardson. 2011. The model-data fusion pitfall: assuming certainty in an uncertain world. Oecologia 167:587-597.

Keenan, T. F., E. Davidson, A. M. Moffat, W. Munger, and A. D. Richardson. 2012. Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling. Global Change Biology 18:2555-2569.

Keith, H., B. Mackey, S. Berry, D. Lindenmayer, and P. Gibbons. 2009. Estimating carbon carrying capacity in natural forest ecosystems across heterogeneous landscapes: addressing sources of error. Global Change Biology:no-no.

Key, J. T., L. R. Pericchi, and A. F. M. Smith. 1999. Bayesian model choice: What and why? (with discussion). Pages 343-370 *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors. Bayesian Statistics. . Oxford University Press, Oxford.

Kohl, M., A. J. Lister, C. T. Scott, T. Baldauf, and D. Plugge. 2011. Implications of sampling design and sample size for national carbon accounting systems. Carbon Balance and Management 6.

Konen, M. E., P. M. Jacobs, C. L. Burras, B. J. Talaga, and J. A. Mason. 2002. Equations for predicting soil organic carbon using loss-on-ignition for North Central U.S. soils. Soil Science Society of America Journal 66:1878-1881.

Kuha, J. 2004. AIC and BIC: comparisons of assumptions and performance. Sociological Methods Research 91:1343-1370.

Kuo, L., and B. Mallick. 1998. Variable selection for regression models. Sankhya: The Indian Journal of Statistics 60:65-81.

Lal, R. 2004. Soil carbon sequestration to mitigate climate change. Geoderma 123:1-22.

Lal, R. 2005. Forest soils and carbon sequestration. Forest Ecology and Management 220:242-258.

Lal, R. 2008. Carbon sequestration. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 363:815-30.

Lark, R. M. 2000. Regression analysis with spatially autocorrelated error: simulation studies and application to mapping of soil organic matter. Journal of Geographic Information Science 14:247-264.

Latimer, A. M., S. Banerjee, H. Sang, E. S. Mosher, and J. A. Silander. 2009. Hierarchical models facilitate spatial analysis of large datasets: a case study on invasive plant species in the northeastern United States. Ecology Letters 12:144-154.

LeBauer, D., D. Wang, K. Richter, C. Davidson, and M. C. Dietze. 2013. Faciliating feedbacks between field measurements and ecosystem models. Ecological Monographs in press.

Lee, S., G. Wolberg, and S. Y. Shin. 1997. Scattered data interpolation with multilevel B-splines. IEEE Transcations on visualization and computer graphics 3.

Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger. 2008. Mixtures of g Priors for Bayesian Variable Selection. Journal of the American Statistical Association 103:410-423.

Link, W. A., and R. J. Barker. 2006. Model weights and the foundations of multimodel inference. Ecology 87:2626-2635.

Liski, J., D. Perruchoud, and T. Karjalainen. 2002. Increasing carbon stocks in the forest soils of western Europe. Forest Ecology and Management 169:159-175.

Liski, J., and C. J. Westman. 1997. Carbon storage in forest soil of Finland. Biogeochemistry 36:261-274.

Lister, A. J., P. P. Mou, R. H. Jones, and R. J. Mitchell. 2000. Spatial patterns of soil and vegetation in a 40-year-old slash pine (Pinus elliottii) forest in the Coastal Plain of South Carolina, U.S.A. Canadian Journal of Forest Research 30:145-155.

Liu, D., Z. Wang, B. Zhang, K. Song, X. Li, J. Li, F. Li, and H. Duan. 2006. Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China. Agriculture, Ecosystems & Environment 113:73-81.

Lu, D. 2006. The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing 27:1297-1328.

Lunn, D. J., J. C. Whitaker, and N. Best. 2006. A Bayesian toolkit for genetic association studies. Genetic Epimediology 30:231-247.

Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. Clark, and D. Schimel. 2011. Ecological forecasting and data assimilation in a data-rich era. Ecological applications 21:1429-1442.

MacFarlane, D. W., E. J. Green, and H. T. Valentine. 2000. Incorporating Uncertainty into the Parameters of a Forest Process Model. Ecological Modeling 134:27-40.

Malone, B. P., A. McBratney, B. Minasny, and G. Laslett. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 160:614-626.

Maniatis, D., and D. Mollicone. 2010. Options for sampling and stratification for national forest inventories to implement REDD+ under the UNFCC. Carbon Balance and Management 5:9.

Marchetti, A., R. Piccini, R. Francaviglia, S. Santucci, and I. Chiuchiarelli. 2010. Estimating soil organic matter content by regression kriging. Progress in Soil Science 2. Pages 241-254 *in* J. L. Boettinger, D. Howell, A. Moore, A. Hartmemink, and S. Kienast-Brown, editors. Digital Soil Mapping. Briding Research, Environmental Application, and Operation. . Springer, Heidelberg.

McBratney, A., M. Medonca Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117:3-52.

McGrath, D., and C. Zhang. 2003. Spatial distribution of soil organic carbon concentrations in grassland of Ireland. Applied Geochemistry 18:1629-1639.

McKenzie, N. J., and P. J. Ryan. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89:67-94.

McQuarrie, A. D. R., and C. L. Tsai. 1998. Regression and Time Series Model Selection. . World Scientific Publications, Singapore.

Meng, X.-L. 1994. Posterior predictive p-values. The Annals of Statistics 22:1142-1160.

Miehle, P., S. J. Livesley, C. Li, P. M. Feikema, M. a. Adams, and S. K. Arndt. 2006. Quantifying uncertainty from large-scale model predictions of forest carbon dynamics. Global Change Biology 12:1421-1434.

Minasny, B., A. McBratney, A. Malone, and L. Wheeler. 2013. Digital soil mapping of carbon. Advanced Agronomy 118:1-47.

Minasny, B., A. McBratney, M. Medonca Santos, I. Odeh, and B. Guyon. 2006. Prediction and digital mapping of soil carbon storage in Lower Namoi Valley. Australian Journal of Soil Research 44:233-244.

Mishra, U., R. Lal, D. Liu, and M. Van Meirvenne. 2010. Predicting the Spatial Variation of the Soil Organic Carbon Pool at a Regional Scale. Soil Science Society of America Journal 74:906.

Mishra, U., and W. Riley. 2012. Alaskan soil carbon stocks: spatial variability and dependence on environmental factors. Biogeosciences 9:3637-3645.

Moore, I. D., R. B. Grayson, and A. R. Ladson. 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrological Processes 5:3-30.

Mueller, T. G., and F. J. Pierce. 2001. Soil Carbon Maps : Enhancing Spatial Estimates with Simple Terrain Attributes at Multiple Scales. Soil Conservation:258-267.

Murtaugh, P. a. 2009. Performance of several variable-selection methods applied to real ecological data. Ecology letters 12:1061-8.

Ntzoufras, I. 2009. Bayesian Modeling Using WinBUGS. . Wiley.

Ntzoufras, I., J. J. Forster, and P. Dellaportas. 2000. Stochastic search variable selection for log-linear models. Journal of Statisical Computation and Simulation 68:23-37.

Odeh, I., A. McBratney, and D. Chittleborough. 1994. Further results on prediction of soil properties from terrain attributes: heterotropic cokriging and regression-kriging. Geoderma 63:197-214.

Oleson, J. J., D. Hope, C. Gries, and J. Kaye. 2006. Estimating soil properties in heterogeneous land-use patches: a Bayesian approach. Environmetrics 17:517-525.

O'Hagan, A. 1995. Fractional Bayes factors for model comparison. Journal of the Royal Statistical Society: Series B 57:99-138.

O'Hara, R. B., and M. J. Sillanpää. 2009. A review of Bayesian variable selection methods: what, how and which. Bayesian Analysis 4:85-117.

Pan, Y., R. A. Birdsey, J. Fang, R. A. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell, P. Ciais, R. B. Jackson, S. W. Pacala, D. A. McGuire, S. Piao, A. Rautiainen, S. Sitch, and D. Hayes. 2011. A large and persistent carbon sink in the world's forests. Science 333:988-993.

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

Plummer, M. 2008. Penalized loss functions for Bayesian model comparison. Biostatistics 9:523-539.

Poggio, L., and A. Gimona. 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation- An example from Scotland. Geoderma 232-234:284-299.

Poole, D., and A. E. Raftery. 2000. Inference for Deterministic Simulation Models: The Bayesian Melding Approach. Journal of the American Statistical Association 95:1244-1255.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53:793-808.

Post, W. M., and K. C. Kwon. 2000. Soil carbon sequestration and land-use change: processes and potential. Global Change Biology 6:317-327.

Potter, C. 1999. Terrestrial biomass and the effects of deforestation on the global carbon cycle. Bioscience 49:769-778.

Powell, S., W. Cohen, S. Healy, R. Kennedy, G. Moisen, K. Pierce, and J. Ohmann. 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. Remote Sensing of Environment 114:1053-1068.

Purves, D., and S. Pacala. 2008. Predictive models of forest dynamics. Science 320:1452-1453.

Qu, M., W. Li, C. Zhang, and S. Wang. 2012. Effect of land use types on the spatial prediction of soil nitrogen. GIScience & Remote Sensing 49:397-411.

Le Quere, C., M. R. Raupach, J. G. Canadell, and G. Marland. 2009. Trends in the sources and sinks of carbon dioxide. Nature Geosciences 2:831-836.

Radtke, P. J., T. E. Burk, and P. V. Bolstad. 2002. Bayesian Melding of a Forest Ecosystem Model with Correlated Inputs. Forest Science 48:701-711.

Raftery, A. E., G. H. Givens, and J. E. Zeh. 1995. Inference from a Deterministic Population Dynamics Model for Bowhead Whales. Journal of the American Statistical Association1 90:402-416.

Raftery, A. E., D. Madigan, and J. A. Hoeting. 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association 92:179-191.

Rahman, S., L. C. Munn, R. Zhang, and G. F. Vance. 1996. Rocky Mountain forest soils: Evaluating spatial variability using conventional statistics and geostatistics. Canadian Journal of Soil Science 76:501-507.

Ribeiro, P. J., and P. J. Diggle. 2001. geoR: a package for geostatistical analysis. R-NEWS 1:15-18.

Robertson, G. P., J. R. Crum, and B. G. Ellis. (n.d.). The spatial variability of soil resources following long-term disturbance. 1993 96:451-456.

Rubin, D. B. 1984. Bayesianly justifiable and relevant frequency calculation for the applied statistician. Annals of Statistics 12:1151-1172.

Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian inferences for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B 71:319-392.

Ruffo, M. L., G. A. Bollero, R. G. Hoeft, and D. G. Bullock. 2005. Spatial variability of the Illinois soil nitrogen test. Agronomy Journal 97:1485-1492.

Saatchi, S., N. Harris, S. Brown, M. Lefsky, E. Mitchard, W. Salas, B. Zutta, W. Buermann, S. L. Lewis, S. Hagen, S. Petrova, L. White, M. Silman, and A. Morel. 2011. Benchmark map of forest carbon stocks in tropical regions, across three continents. Proceedings of the National Academy of Sciences 108.

Salazar, E., B. Sanso, A. O. Finley, D. Hammerling, I. Steinsland, X. Wang, and P. Delamater. 2011. Comparing and blending regional climate model predictions for the American Southwest. Journal of Agricultural, Biological, and Environmental Statistics 16:586-605.

Schwalm, C. R., C. a. Williams, K. Schaefer, R. Anderson, M. A. Arain, I. Baker, A. Barr, T. A. Black, G. Chen, J. M. Chen, P. Ciais, K. J. Davis, A. Desai, M. Dietze, D. Dragoni, M. L. Fischer, L. B. Flanagan, R. Grant, L. Gu, D. Hollinger, R. C. Izaurralde, C. Kucharik, P. Lafleur, B. E. Law, L. Li, Z. Li, S. Liu, E. Lokupitiya, Y. Luo, S. Ma, H. Margolis, R. Matamala, H. McCaughey, R. K. Monson, W. C. Oechel, C. Peng, B. Poulter, D. T. Price, D. M. Riciutto, W. Riley, A. K. Sahoo, M. Sprintsin, J. Sun, H. Tian, C. Tonitto, H. Verbeeck, and S. B. Verma. 2010. A model-data intercomparison of CO 2 exchange across North America: Results from the North American Carbon Program site synthesis. Journal of Geophysical Research 115:G00H05.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461-464.

Shao, J. 1997. An asymptotic theory for linear model selection (with Discussion). Statistica Sinica 7:221-242.

Shvidenko, A., D. Schepaschenko, I. McCallum, and S. Nilsson. 2010. Can the uncertainty of full carbon accounting of forest ecosystems be made acceptable to policymakers? Climatic Change 103:137-157.

Simbahan, G. C., A. Dobermann, P. Goovaerts, J. Ping, and M. L. Haddix. 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma 132:471-489.

Soil Survey Staff, Natural Resources Conservation Service, U. S. D. of A. (n.d.). Soil Survey Geographic (SSURGO) Database for New Jersey. http://soildatamart.nrcs.usda.gov.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Var Der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64:583-639.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2014. The deviance information criterion 12 years on. Journal of the Royal Statistical Society: Series B 76:485-493.

Stein, M. 1999. Spatial Interpolation, Some Theory for Kriging. . Spring-Verlag, New York.

Steltzer, H., and J. M. Walker. 2006. Modeling the Effect of Photosynthetic Vegetation Properties on the NDVI-LAI Relationship. Ecology 87:2765-2772.

Sutherland, R. A., C. van Kessel, R. E. Farrell, and D. J. Pennock. 1993. Landscape-scale variations in plant and soil Nitrogen-15 abundance. Soil Science Society of America Journal 57:169-178.

De Sy, V., M. Herold, F. Achard, G. P. Asner, A. Held, J. Kellendorfer, and J. Verbesselt. 2012. Synergies of multiple remote sensing data sources for REDD+ monitoring. Current Opinion in Environmental Sustainability 4:696-706.

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. Pages 61-78 *in* R. Forman, editor. Pine Barrens: Ecosystem and Landscape. . Rutgers University Press, New Brunswick, NJ, USA.

Thogmartin, W. E., J. R. Sauer, M. G. Knutson, S. E. Applications, and N. Dec. 2004. A Hierarchical Spatial Model of Avian Abundance with Application to Cerulean Warblers A HIERARCHICAL SPATIAL MODEL OF AVIAN ABUNDANCE WITH. America 14:1766-1779.

Toth, G., A. Jones, and L. Montanarella. 2013a. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. Environmental Monitoring and Assessment 185:7409-7425.

Toth, G., A. Jones, and L. Montanarella. 2013b. LUCAS Topsoil Survey. Methodology, data, and results. . JRC Technical Reports. Publications Office of the European Union, Luxemborg.

Vasques, G. M., S. Grunwald, N. B. Comerford, and J. O. Sickman. 2010a. Regional modelling of soil carbon at multiple depths within a subtropical watershed. Geoderma 156:326-336.

Vasques, G. M., S. Grunwald, and D. B. Myers. 2012. Associations between soil carbon and ecological landscape variables at escalating spatial scales in Florida, USA. Landscape Ecology 27:355-367.

Vasques, G. M., S. Grunwald, J. O. Sickman, and N. B. Comerford. 2010b. Upscaling of Dynamic Soil Organic Carbon Pools in a North-Central Florida Watershed. Soil Science Society of America Journal 74:870.

Vehtari, A., and J. Ojanen. 2012. A survey of Bayesian predictive methods for model assessment, selection and comparison. Statistics Surveys 6:142-228.

De Vos, B., B. Vandecasteele, J. Deckers, and B. Muys. 2005. Capability of loss-on-ignition as a predictor of total organic carbon in non-calcaerous forest soils. Communications in Soil Science and Plant Analysis 36:2899-2921.

Wackernagel, H. 1998. Multivariate Geostatistics: An Introduction With Applications, 2nd Ed. Springer, Berlin.

Wang, H., C. S. Hall, J. D. Cornell, and M. P. H. Hall. 2002. Spatial dependence and the relationship of soil organic carbon and soil moisture in the Luquillo Experimental Forest, Puerto Rico. Landscape Ecology 17:671-684.

Wang, K., C. Zhang, and W. Li. 2013. Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging. Applied Geography 42:73-85.

Wang, L., P. P. Mou, J. Huang, and J. Wang. 2007. Spatial heterogeneity of soil nitrogen in a subtropical forest in China. Plant and Soil 295:137-150.

Wang, Y., and Q. Liu. 2006. Comparison of Akaike information criterion (AIC) and Bayesian informatin criterion (BIC) in selection of stock-recruitment relationships. Fisheries Research 77:220-225.

Ward, E. J. 2008. A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecological Modelling 211:1-10.

Watt, M. S., and D. J. Palmer. 2012. Use of regression kriging to develop a Carbon:Nitrogen ratio surface for New Zealand. Geoderma 183-184:49-57.

Wellbrock, N., and A. Bolte. 2008. Carbon sequestration potential of soils in commercial forests in Germany - contribution of National Forest Monitoring to the advancement of knowledge. Agriculture and Forestry Research 58:247-252.

Wiesmeier, M., J. Prietzel, and F. Barthold. 2013. Storage and drivers of organic carbon in forest soils of southeast Germany (Bavaria)- Implications for carbon sequestration. Forest Ecology and Management 295:162-172.

Wikle, C. 2010. Low-Rank Representations for Spatial Processes. Pages 108-117 *in* A. E. Gelfand, P. J. Diggle, P. Guttorp, and M. Fuentes, editors. Handbook of Spatial Statistics. . CRC Press.

Williams, M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius. 2005. An improved analysis of forest carbon dynamics using data assimilation. Global Change Biology 11:89-105.

Wintle, B. A., M. A. McCarthy, C. T. Volinsky, and R. P. Kavanagh. 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology 17:1579-1590.

Worsham, L., D. Markewitz, and N. Nibbelink. 2010. Incorporating spatial dependence into estimates of soil carbon contents under different land covers. Soil Science Society of America Journal 74:635-646.

Yang, Y. 2005. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. Biometrika 92:937-950.

Zellner, A. 1986. On Assessing Prior Distributions and Bayesian Regression Analysis with g prior Distributions. Pages 389-399 *in* A. Zellner, editor. Bayesian Inference and Decision techniques: Essays in Honor of Bruno de Finetti. . Edward Elgar Publishing Limited.

Zellner, A., and A. Siow. 1984. Basic Issues in Econometrics. . University of Chicago Press, Chicago, USA.

Zhang, C., Y. Tang, X. Xu, and G. Kiely. 2011. Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in Ireland. Applied Geochemistry 26:1239-1248.

Zhang, W., K. Wang, H. Chen, X. He, and J. Zhang. 2012. Ancillary information improves kriging on soil organic carbon data for a typical karst peak cluster depression landscape. Journal of the Science of Food and Agriculture 92:1094-1102.