A Spatial Model Approach for Assessing Windbreak Growth and Carbon Stocks

Qingjiang Hou, Linda J. Young, James R. Brandle,* and Michele M. Schoeneberger

Agroforestry, the deliberate integration of trees into agricultural operations, sequesters carbon (C) while providing valuable services on agricultural lands. However, methods to quantify present and projected C stocks in these open-grown woody systems are limited. As an initial step to address C accounting in agroforestry systems, a spatial Markov random field model for predicting the natural logarithm (log) of the mean aboveground volume of green ash (*Fraxinus pennsylvanica* Marsh.) within a shelterbelt, referred to as the log of aboveground volume, was developed using data from an earlier study and web-available soil and climate information. Windbreak characteristics, site, and climate variables were used to model the large-scale trend of the log of aboveground volume. The residuals from this initial model were correlated among sites up to 24 km from a point of interest. Therefore, a spatial dependence parameter was used to incorporate information from sites within 24 km into the prediction of the log of the aboveground volume. Age is an important windbreak characteristic in the model. Thus, the log of aboveground volume can be predicted for a given windbreak age and for values of other explanatory variables associated with a site of interest. Such predictions can be exponentiated to obtain predictions of aboveground volume for windbreaks without repeated inventory. With the capability of quantifying uncertainty, the model has the potential for large regional planning efforts and C stock assessments for many deciduous tree species used in windbreaks and riparian buffers once it is calibrated.

J. Environ. Qual. 40:842–852 (2011)
doi:10.2134/jeq2010.0098
Posted online 16 Feb. 2011.
Received 8 Mar. 2010.
*Corresponding author (jbrandle@unl.edu).
© ASA, CSSA, SSSA
5585 Guilford Rd., Madison, WI 53711 USA

Agroforestry, the deliberate integration of trees into crop and livestock operations, sequesters substantial amounts of carbon (C) on agricultural lands while providing the production and conservation services for which it was designed (Korn et al., 2003; Nair et al., 2009; Schoeneberger, 2009; Verchot et al., 2007). The Global Research Alliance on Agricultural Greenhouse Gases, established at the 2009 climate change meetings in Copenhagen (http://www.globalresearchalliance.org/home.aspx), explicitly includes agroforestry as a viable C sequestering option for agricultural operations. Of the five main agroforestry practices used in the United States (windbreaks, riparian buffers, alley cropping, silvopasture, and forest farming), windbreaks are especially appealing as a C sequestering option on private lands. Windbreaks, also referred to as shelterbelts, are linear plantings consisting of trees and shrubs. They are used throughout the United States to protect and improve crop yields, reduce wind erosion, manage snow, reduce energy consumption by homesteads and other buildings, and protect livestock. In so doing, they provide additional wildlife habitat in areas dominated by agriculture as well as other benefits afforded by the altered microclimate and landscape structure created by the plantings (Brandle et al., 2009). Although a small portion (about 2 to 5%) of an agricultural field is dedicated to the windbreak, this small amount of land is able to sequester greater amounts of C per unit land area than many of the other agricultural options, thereby contributing significantly to overall greenhouse gas mitigation within a farming operation (Schoeneberger, 2009; USEPA, 2006). Furthermore, the very purpose for windbreak plantings—the use of perennials and the additional services they provide to the landowner—adds a level of permanence not necessarily present in other practices.

Being able to estimate current and future amounts of biomass and C sequestered in agroforestry plantings, such as windbreaks, provides a basis for directing conservation programs and policy development as well as future land management decisions by landowners. Initial estimates made for windbreaks in the north-central United States (USDA NAC, 2001) and for riparian buffers, woody plantings in the unfarmed corners of center pivot fields, and living snow fences in Nebraska (Nebraska Department of Natural Resources, 2001) indicate that agroforestry has tremendous potential as a C sequestering option for these areas. However, more reliable means for generating these estimates are

Q. Hou, Univ. of Kansas Medical Center, Dep. of Biostatistics, Mail-Stop 1026, 3901 Rainbow Blvd., Kansas City, KS 66160; L.J. Young, Dep. of Statistics, Univ. of Florida, 408 McCarty Hall C, P.O. Box 110339, Gainesville, FL 32611-0339; J.R. Brandle, School of Natural Resources, Univ. of Nebraska, Lincoln, NE 68583-0814; M.M. Schoeneberger, USDA FS/NRCS National Agroforestry Center, Southern Research Station, East Campus-UNL, Lincoln, NE 68583-0822. Assigned to Associate Editor Scott Chang.

Abbreviations: DBH, diameter at breast height; FIA, USDA Forest Inventory and Analysis; PDSI, Palmer Drought Severity Index; STATSGO, State Soil Geographic database.

needed as cost-share and C credit programs are argued and formulated and as modeling efforts are developed for C accounting in agroecosystems.

Like a forest stand, windbreak growth is a function of site quality and climate conditions as well as species genetics, composition, arrangement, and age. Site quality and climate conditions change from location to location. Even within a single windbreak, site quality may vary dramatically depending on the microtopographical conditions. Species characteristics, as influenced by the seed source (genetic potential) and environmental conditions, are also quite variable, reflecting the variety of plant materials generally used in conservation plantings and the wide range of settings into which they are planted (Cunningham, 1988). Considerable effort has been made to develop methods to estimate the site quality of individual forest stands so their future growth and development can be estimated (Alemdag, 1991). Two principle approaches have been used for this purpose. The ecological approach classifies site quality based on plant communities because plant associations reflect both climatic and topographic factors that contribute to a stand's growth potential. Such classifications are largely descriptive, difficult to quantify, and predominantly used for broad classification and comparisons across large geographical scales. The other approach uses site index as a comprehensive indicator of site quality and is the most commonly used method for estimating site quality in North American forests (Carmean et al., 1989). Site index refers to the predicted height of the dominant and codominant trees of a stand at an index age (usually 50 yr for hardwood species). This approach treats site condition as a single variable with additive effects and models height growth as a nonlinear function of age. Theoretically, each site should have a unique site index, and abrupt, rather than smooth, changes should be expected from site to site. The nonlinear functional relationship between height and age provides the basis to forecast future growth through extrapolation.

Despite great effort, especially in commercial forest production systems (Alemdag, 1991), it is impractical to develop a site index for each stand. In many cases, the estimation of site index remains impossible due to the lack of a woody plantation for measurement. Approximations and assumptions must therefore be made to make use of index curves for other sites with similar site conditions. Because site index is a function of age and the average height of the dominant and codominant trees in a stand, at least one measurement of age and height must be made to estimate site index. Consequently, site indices for projected plantings on reclaimed, agricultural, or pasture sites are rarely available, and predictions for future growth based on site index at potential windbreak sites are essentially impossible.

The objective of this work is to develop a spatial model of the natural logarithm of the mean aboveground volume of trees within a windbreak that provides spatial prediction at any point within Nebraska. To do this, we propose using a spatial Markov random field model that uses web-available data. This approach differs from the traditional forestry practice in that the discrete qualitative site classification is replaced with a continuous linear predictor based on a series of quantitative and qualitative soil and climate variables (Lundergren and Dolid, 1970). With a Markov random field model, the spatial variation in the natural logarithm of the mean aboveground volume of trees within a windbreak is attributed to two sources: (i) large-scale variation or trend across the region and (ii) small-scale variation due to correlation among nearby sites. The linear predictor can be used to capture the trend of natural logarithm of the windbreak mean aboveground volume (the response variable) over space, with soil, climate, and windbreak-related parameters serving as the predictor variables. The spatial dependence parameter quantifies the correlation among sites as a function of their distance from each other, thereby capturing the small-scale correlation among neighboring locations. The Markov random field model approach is independent of site index and allows the user to combine data over a geographical area of interest, providing spatial prediction at existing and at new locations.

Materials and Methods

Data Sources and Description

The primary focus of this study was on developing a model to predict the log of aboveground volume of green ash (*Fraxinus pennsylvanica* Marsh.) windbreaks as a first step for getting estimates of aboveground volume and then ultimately for use in estimating potential woody biomass and C in future plantings. The woody component in these afforestation-like practices represents the dominant component of C sequestered, with the aboveground portion generally representing the majority of new C sequestered in these systems (Nui and Duicker, 2006) as well as being used to estimate roots in forestry projects (Brown, 2002). Such a model could begin to provide estimates of agroforestry's current and future contributions for reporting and management planning purposes. The model could be applied at any site within the research area for which soil and climate data are available. The data used in this study were obtained from three different sources and are described below.

Windbreak Data

Green ash windbreak data were obtained from the Windbreak Site Standard Plot Reports (USDA NRCS, 2002a), which have windbreak characteristics (age, species composition, health condition, and site average height and diameter at breast height [DBH]) for each site. In addition, site coordinates (Township, Range, and Section), soil types within the windbreak, and annual precipitation from the nearest weather station are listed. To estimate within-site variation, individual tree height and DBH were entered into a supplemental database. Exploratory data analyses were conducted. Although potential outliers were identified, no observations were excluded unless the record indicated the tree was dead, physically damaged, replanted, or a sprout so that the full variation within and among windbreaks could be captured. Unlike a forest inventory of the USDA Forest Service Inventory and Analysis Program (FIA), the Windbreak Site Standard Plot Reports were not designed to measure and monitor the total aboveground volume per unit area. No undergrowth and shrubs were measured, and the size of each sample plot was not necessarily uniform (Table 1). Each windbreak was sampled at only one point in time.

Of the major tree species listed in the Nebraska standard report (e.g., green ash, cottonwood [*Populus sp.*], elm [*Ulmus pumila* L.], eastern red cedar [*Juniperus virginiana* L.], and ponderosa pine [*Pinus ponderosa* Dougl. Ex Laws.]), only green ash and eastern red cedar were dominant and present at most sites. We selected green ash for this study because it was widely used in many agroforestry and conservation plantings throughout the Great Plains region. (Since our study, the spread of the emerald ash borer [*Agrilus planipennis* Fairmaire] into the Midwest and toward the Great Plains has begun to threaten the survival of this species, and it is no longer recommended as a key conservation tree.) Additionally, green ash was selected because we have a model relating the log of aboveground volume to height and DBH developed from field windbreak measurements (Zhou et al., 2002). This model reflects the more open-grown tree form attained in windbreaks, as opposed to the generally available models developed from forest stands (Smith et al., 2004). Of the 235 windbreaks surveyed in Nebraska, 96 contained green ash as a major component, and these windbreaks were selected for this study (Table 1). Green ash is native to a large region of north-central United States, is sensitive to site and climate variation, and is subject to large within-stand differentiation due to internal competition for light and nutrients (Kennedy, 1990). These factors lead to large within- and between-site variation, making estimation of the spatial dependence structure especially challenging. Table 2 lists the variables in the windbreak data used in our model. Among these variables was the windbreak growth condition code (Cocode), which used G, F, P, and D for good, fair, poor and deteriorating growing status of the windbreak as a result of maintenance and management. For spatial modeling, we recoded G, F, P, or D as 1, 2, 3, and 4, respectively, considering the ordinal nature of the original survey classifications. Based on the assumption that the distance between consecutive condition codes is the same, this variable was entered as a continuous variable in the model.

Soil Data

Nebraska soil data were downloaded from the USDA Natural Resources Conservation Service State Soil Geographic (STATSGO) database maintained by the Nebraska Department of Natural Resources (State Soil Geographic [STATSGO] database for Nebraska). Designed for regional, river-basin resource planning and management, STATSGO core data are available for map units, which are polygons of various shapes and sizes. Depending on the region, these map units average from 7000 to 60,000 acres in size; the minimum size is 1544 acres (USDA SCS, 1991).

Each STATSGO map unit may have up to 21 soil types. However, the location of the specific individual soil types within each map unit is not given. Instead, the proportion of the map unit's area covered by each soil type is provided. In addition, each map unit has a set of attribute tables containing 60 soil properties that include physical, chemical, biological, taxonomic, and geographical characteristics of each soil type within that unit. These attribute tables are connected to map units through a set of identifier variables. All attribute tables were merged to form one SAS data set (SAS Institute, 1990). A list of the soil variables included in our model is given in Table 3.

Climate Data

Climate data for the windbreak sites were obtained from the High Plains Regional Climate Center website (High Plains Regional Climate Center, 2002). This site has short-term weather records and long-time climate measurements from 125 weather stations throughout Nebraska. Climate variables include monthly precipitation; temperature; and heating, cooling, and growing-degree-days. Also available from this data source are monthly and annual means of the Palmer Drought Severity Index (PDSI) for each weather station. Widely used as an indicator of regional drought conditions, the PDSI provides an estimate of the accumulated effect of monthly rainfall deficit or surplus relative to the monthly climatologically "appropriate" rainfall, defined as precipitation

Table 1. Background information on windbreak sites used in this study.

County	Year established	No. of trees	No. of sites
Antelope	1967	12	1
Blaine	1963	12	1
Box Butte	1965	12	1
Chase	1967	12	1
Custer	1959	258	16
Dundy	1967	12	1
Franklin	1965	60	5
Gauge	1980	14	1
Hitchcock	1964	60	5
Holt	1961	242	12
Johnson	1983	16	1
Lancaster	1965	24	1
Lancaster	1989	45	2
Madison	1960	209	15
Morrill	1965	7	2
Seward	1984	21	1
Sheridan	1967	30	4
Stanton	1960	125	12
Thurston	1963	14	1
Webster	1964	24	4
Wheeler	1964	30	4
Wheeler	1965	41	5
Total (16)		1280	96

Table 2. Windbreak variables from the windbreak survey data.

Code	Description
Age	windbreak age at time of survey
DBH	tree diameter at breast height
Ht	tree height at time of survey
Wthnrow	windbreak within-row spacing (tree spacing)
Btwnrow	windbreak between-row spacing (row spacing)
Neighbor	neighbor row species
Post	row position (interior or side)
Cocode	windbreak growth condition code (1, 2, 3, or 4)
Volume	log of aboveground volume (stem, branches, leaves)

Table 3. Soil properties included in the tree model and site mean model.

Code	Description
Group	windbreak suitability group
Om1	first-layer organic matter content
Shrin1	second-layer soil shrinkage
Clay2	second-layer clay content
Wei	soil wind erodibility index
PH1	first-layer soil reaction
Perm1	first-layer soil permeability
Text2	second-layer soil texture
Sdep	soil depth of the first and second layers
Liq	soil available water content
Cec2	second-layer cation exchange capacity
Kfactor	soil erodibility factor

needed to maintain adequate soil water content for normal plant growth in a particular region (Qi and Willson, 2000). The PDSI is a scaled value with a mean of zero. Negative values represent insufficient moisture, and positive values indicate at least adequate moisture. To better reflect climate impact on the multi–decade-long growth of trees in windbreaks, we calculated the number of months during each growing season (March–August) over a 30-yr period (1961–1990) for which the PDSI indicated drought based on records from each weather station. In this study, we have defined drought for trees as a PDSI of less than −2. Selection of −2 as a critical value is based on the assumption that woody plants are relatively tolerant to moderate drought conditions due to their deep root system and that tree growth is more likely to suffer with increases in both intensity and frequency of drought periods. This measurement of drought was used as an independent variable for the spatial prediction of volume at each windbreak point. Table 4 provides a list of the climate variables included in the model.

Combining Windbreak with Soil and Climate Data

The windbreak, soil, and climate data used in this study are spatially misaligned (Gotway and Young, 2002); that is, they have been collected on different observational units. The windbreak information was observed at the windbreak sites, which are points on a map. The soil data are recorded on polygons. The climate data are recorded at weather stations, which are also points on a map but different from those of the windbreaks. The first challenge was to combine all of the data at the windbreak site level (see Fig. 1).

To combine the soil and windbreak data, the coordinates for the windbreak sites expressed in terms of Township, Range, and Section (USDA NRCS, 2002a) were converted into latitude and longitude. Then, the windbreak data points were overlaid onto the STATSGO soil map, and a unique map unit was identified for each sampling point using ArcMap's spatial join function (ESRI, 2001; State Soil Geographic [STATSGO] data base for Nebraska). If the map unit contained the specific type of soil identified in the windbreak data set, all attributes for that soil type from STATSGO attribute tables were assigned to the windbreak data point (Fig. 2).

An exact match occurred for only about half the windbreak points, possibly because of differences in sampling scales or changes in terminology in soil taxonomy. If none of the soil types in a map unit matched those identified in the windbreak data set, the weighted mean values from STATSGO attribute tables for all soil types in the corresponding map unit were assigned to the corresponding windbreak data point, where the weight for a soil type was the proportion of that soil type in the map unit.

Table 4. Climate variables included in the weather data used in developing the tree model.

Code	Description
PDSI2	total number of months in which the Palmer Drought Severity Index was below −2 over a 30-yr period
Arain	30-yr average annual precipitation
Avget1	30-yr average Jan. temperature
Srain	30-yr average summer precipitation
Avgt7	30-yr average July temperature
Cddall	30-yr average annual cooling degree day
Hddall	30-yr average annual heating degree day
Meantall	30-yr average annual mean temperature

Consequently, for each windbreak sampling point, we obtained a complete set of soil specific attributes. However, the quality of the attributes differed depending on whether the windbreak soil type matched one of the soils in the associated STATSGO map unit. This disparity in quality was not considered further.

Because windbreaks grow for years, long-term climate, as opposed to short-term weather conditions, was thought to be more relevant to overall biomass accumulation. We calculated 30-yr averages (1961 to ~1990) for precipitation; mean, maximum, and minimum temperatures; cooling and heating degree days; and PDSI for each weather station. Although the 125 weather stations were roughly uniformly distributed over the state (Fig. 1), their spatial coordinates in latitude and longitude did not match those of the windbreak sampling points. We obtained the spatial predictions for the long-term climate attributes at each windbreak sampling point using inverse distance weighting (Cressie, 1993), which assigns weight to nearby stations according to their proximity to the target point (i.e., the closer the station, the greater its weighted value). Finally, the predicted values for all climate variables were merged with the windbreak-STATSGO data for spatial modeling.

Spatial Markov Random Field Model

Historically, multiple regression models have been developed for predicting quantities of interest (Searle, 1971; Neter et al., 1996; Draper and Harry, 1998). Here, the natural logarithm of the aboveground volume is to be predicted from potential explanatory variables (from soil, windbreak, and climatic data). A regression model can account for large-scale trends over a region. However, it is difficult, if not impossible, to identify all explanatory variables that are influential in predicting the log of aboveground volume. Because they tend to vary over space and are omitted from the modeling process, spatial dependencies among the errors are often present. Furthermore, because sites close together tend to be more alike than sites further apart, using nearby sites to inform predictions can make them more precise. Thus, in the modeling process, we account for large-scale trends and borrow information from nearby sites to capture small-scale trends. This leads to a correlated, and not an independent, error structure.

We hypothesized that a measure of windbreak tree growth (the natural logarithm of aboveground volume), $Z(\mathbf{s}_i)$, for a site located at $\mathbf{s}_i$ is spatially correlated with the same windbreak tree growth measure at all sites within a certain range of the one of interest. (Although the aboveground volume is of interest, its distribution is highly skewed. Thus, the natural logarithm, referred to here as the log, of aboveground tree volume is modeled so that the assumption of normality is more nearly met. The aboveground tree volume can then be obtained by exponentiation.) This spatial dependence can be quantified by fitting a spatial Markov random field model (Cressie and Lele, 1992). For a site i, N_i represents a neighborhood with a set of sampled windbreaks within some distance from site i, but not including site i. Thus,

$$[Z(\mathbf{s}_i) : i \notin \mathrm{N}_i, \{\mathrm{N}_i : i = 1, 2, \ldots, n]$$

where n is the number of sites for which predictions are required. In the simple linear regression of a response y on an explanatory variable x, it is assumed that there is a population of y's and each x and that population has a mean, which is linearly related to

x, and a variance, which is often assumed to be the same for all x's. A similar assumption is made in spatial random processes. Here a conceptual population of possible log-transformed aboveground volumes is assumed for each windbreak. A particular realization of these volumes has been observed at each site, and that realization reflects any inherent spatial dependence among sites. Let μ_i and σ_i^2 be the mean and variance of the log of aboveground volume at site i. Suppose that the mean varies over space so that $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)$ captures this large-scale variation. Further, define $\mathbf{M} = \mathrm{Diag}\,(\sigma_1^2,\ldots,\sigma_n^2)$ to be a diagonal matrix of the unknown site variances and $\mathbf{C} = \{c_{ij}\}$ to be a matrix that captures the small-scale dependence among sites. Here, the small-scale dependence is assumed to be 0 if a site is not in the neighborhood ($c_{ij} = 0$ if $j \in \mathrm{N}_i$) and to decrease with distance between the windbreaks when a site is in the neighborhood ($c_{ij} = \varphi f(d_{ij})$, if $j \in \mathrm{N}_i$ where d_{ij} is a function of distance between $\mathbf{s}_i$ and $\mathbf{s}_j$). Let $\mathbf{I}$ be an $n \times n$ identity matrix and $\mathbf{C} = \{c_{ij}\}$ be a spatial dependence matrix. Assuming the auto-Gaussian model (an autoregressive model with Gaussian (normally) distributed responses) and under certain conditions (Cressie, 1993; Griffith, 2003), the log of the aboveground volume at sites $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$, namely the spatial random process $\mathbf{Z} = [Z(s_1), Z(s_2),\ldots,Z(s_n)]'$ is normally distributed (see Appendix for details). That is,

$$\mathbf{Z} \sim \mathrm{N}[\boldsymbol{\mu}(\mathbf{I} - \mathbf{C})^{-1}\,\mathbf{M}] \qquad [1]$$

For the combined windbreak data, the large-scale variation is assumed to be a linear function of the explanatory site and climate variables; that is,

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \qquad [2]$$

where the design matrix $\mathbf{X}$ includes all potential quantitative and qualitative predictor variables ranging from windbreak attributes, soil measurements, and long-term climate data, and $\boldsymbol{\beta}$ is a vector of coefficients (Searle, 1971). The qualitative variables enter the model through indicator variables (Neter et al., 1996; Draper and Harry, 1998). We assume that all effects on the log of aboveground volume by predictor variables are additive because soil, climate, and windbreak characteristics tend to have linear effects on site, as indicated in the various site index curves found in the forestry literature (Alemdag, 1991; Sander, 1971).

The spatial dependence matrix $\mathbf{C}$ in Eq. [1] can be modeled as a function of distance between windbreaks. In this study, a semivariogram was used to quantify the spatial dependence. The empirical semivariogram is a function of distance and estimates half the variation among pairs of points a specified distance (lag) apart. Because the empirical semivariogram is comprised of a series of estimates (one estimate for each lag), it generally does not appear to be smooth, just as data do not all lie on the line in a linear regression. To obtain a smooth function of the relationship between the semivariogram and distance, so that the semivariogram can be estimated for distances for which no pairs were observed, various semivariogram models, such as spherical, exponential, and power semivariogram models, are used (Cressie, 1993). For windbreaks in Nebraska, the range (R^*) of spatial dependence must first be estimated. We fitted a semivariogram model to the residual from the regression model in Eq. [2] and then used the estimated range of correlation from the semivariogram model to identify the neighborhoods of the Markov random field. That is, a spatial dependence is assumed to exist only within a distance R^* of a given point, and all windbreaks within R^* of the point are in its neighborhood. Observations from windbreaks further than R^* apart are uncorrelated. The value of R^* is unknown and must be estimated (see Appendix for details). Then windbreak j is in the neighborhood of point i ($j \in \mathrm{N}_i$) if

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq R$$

where the x's and y's are coordinates of sample locations, and R represents the estimated range of spatial dependence.

We have a neighborhood for each site, and we now turn to quantifying the spatial dependence within a neighborhood. To do this, define $f(d_{ij}) = C(k)d_{ij}^{-k}$ for $j \in \mathrm{N}_i$. Then:

$$c_{ij} = \varphi C(k)d_{ij}^{-k} \qquad [3]$$

where c_{ij} is the ijth element of the spatial dependence matrix $\mathbf{C}$, φ is the spatial dependence parameter, $C(k)$ = minimum d_{ij}^{k} : $j \in \mathrm{N}_i$, $i = (1, 2,\ldots,n)$, and k is a scale parameter that controls how rapidly the spatial dependence changes with distance of separation. Cressie and Chan (1989) showed that the spatial

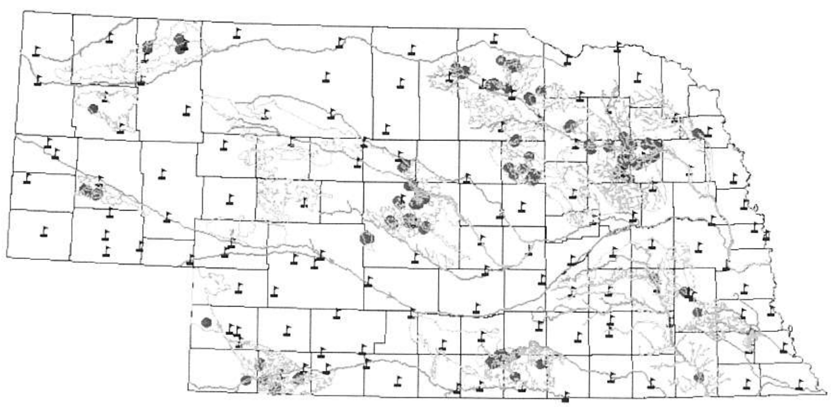

Fig. 1. Spatial distributions of green ash windbreak sampling points (dots), weather stations (flags), and State Soil Geographic database map units (polygons).

Source	Coordinate system	Attribute
Windbreak	Township-Range System	Points
STATSGO	UTM, NAD83	Polygons
Climate	Latitude & Longitude	Points

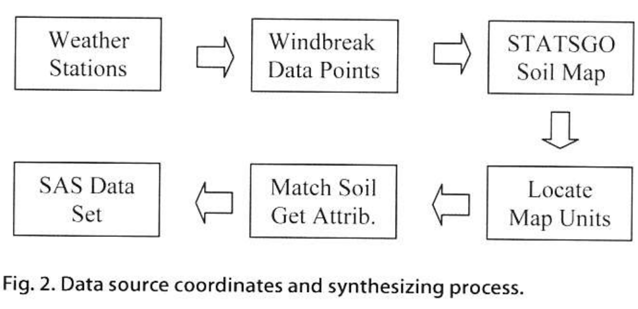

Fig. 2. Data source coordinates and synthesizing process.

dependence parameter φ can be estimated and tested using maximum likelihood methods (see Appendix for details).

The predicted log of aboveground volume at a site includes a term for large-scale trend and a term that accounts for small-scale variation:

$$\mathbf{Y}^p = \mathbf{X}\boldsymbol{\beta} + (\varphi\mathbf{H})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad [4]$$

where $\mathbf{H} = f(d_{ij}) = C(k)d_{ij}^{-k}$. The first term, $\mathbf{X}\boldsymbol{\beta}$, in Eq. [4] accounts for large-scale variation that comes from windbreak characteristics, such as age, species composition, and the arrangement of the trees in the windbreak. The second term, $(\varphi\mathbf{H})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$, in Eq. [4] accounts for small-scale variation that occurs because the response (the log of aboveground volume) at sites closer together tend to be more alike than that at sites further apart, leading to spatial dependence among responses from neighboring sites. The variance of the predicted log of the volume is estimated using Eq. [5] (see Appendix for details).

$$\mathrm{Var}(\mathbf{Y}^p) = \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}' \quad [5]$$

where $\mathbf{F} = (\mathbf{I} - \varphi\mathbf{H})(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma} = (\mathbf{I} - \varphi\mathbf{H})^{-1}\sigma^2$.

The Markov random field model is especially useful because its predictive power at a given point increases as more data within the spatial neighborhood become available.

Results and Discussion

Because trees within a windbreak varied considerably in size, the distribution of the aboveground volume of individual trees was highly skewed. Thus, the natural logarithm of individual tree volume was taken as the response so that the assumption of normality would be more nearly met. Two separate Markov random field models were fitted to the spatially aligned climate, soil, and windbreak data. The first used the log of aboveground volume of individual trees as the response variable (individual tree model), while the other took the average of the log of individual tree volume first and then modeled the mean log tree volume on each site (site mean model). The individual tree variability in the log of aboveground volume on the log scale among and within sites was further explored through analysis of variance. The estimated within-site variance and between-site variance was 0.36 and 0.26, respectively, suggesting within windbreak variation was even larger than between windbreaks. All of the explanatory variables, except neighbor row species and row position indicator, considered in this study were observed at the site level and thus were constant for any given windbreak. As a consequence, no explanatory variable was capable of explaining any of the tree-to-tree variability within a site. The large residual variance, possibly due to severe internal competition, indicated that the individual tree model obscured any spatial dependence that might exist. As a result, the site mean model was found to be more useful. Thus, the log of the mean aboveground volume of all green ash trees within a windbreak, which we refer to as the log of the aboveground volume, is the focus of our modeling and is discussed in the rest of the paper.

Large-Scale Variation

Because of the heavily skewed distribution, we fitted the Markov random field model with site mean volume of green ash on the log scale as the response variable using soil, climate, and windbreak characteristics as explanatory variables. The model for large-scale trend in the log of aboveground volume for green ash windbreaks across Nebraska was estimated with multivariate regression. Backward elimination (using $p \le 0.05$) was used to identify the most important explanatory variables (SAS Institute, 1990). The resulting linear model was as follows:

$$\ln(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon \text{ for } i = 1, 2, \ldots\ldots n$$

where i is an indicator of the windbreak site, y_i is the log of aboveground tree (more specifically, the natural logarithm of the mean aboveground tree volume of all green ash trees within the windbreak), x_{1i} is the natural logarithm of windbreak age, x_{2i} is the 30-yr average summer precipitation at the windbreak location, x_{3i} is the windbreak growth condition code (1, 2, 3, or 4 for good, fair, poor, or deteriorating, respectively), and ε_i represents random error, which is assumed to be independently and identically distributed $N(0, \sigma^2)$.

Small-Scale Variation

The maximum range of spatial dependence R was 24 km as estimated by semivariogram analysis with the SAS KRIGE2D procedure (SAS Institute, 2008; Drignei, 2009). Three different values for the scale parameter ($k = 0, 1, 2$) in $f(d_{ij}) = C(k) d_{ij}^{-k}$ were tested, where $C(k) = \text{minimum } (d_{ij}^{k} : j \in N_i, i = 1, 2, \ldots n)$. Maximum likelihood methods were used to choose among the three values of k. The maximum likelihood estimate is that value of the parameter that maximizes the likelihood function, which is the probability of the observed data as a function of the parameter(s). The logarithm of the likelihood function is often mathematically more tractable, and, because maximizing a function is equivalent to maximizing the log of that function, usually the log likelihood function is maximized or, equivalently, the negative of the log-likelihood function is minimized, to find the maximum likelihood estimates. Because the minimum of the negative log-likelihood function (Chap, 2003) for $k = 1$ is smaller than $k = 0$ or $k = 2$ (Fig. 3a), the spatial dependence function $f(d_{ij})$ based on the inverse function of distance between sites ($k = 1$) provided a better fit than a power function ($k = 2$) or a uniform function ($k = 0$) of distance. For $k = 1$, the log-likelihood changed with the spatial dependence parameter and reached a maximum at the estimated value of 0.575, which implied a positive correlation among neighboring windbreaks. Biologically, this result was appealing because soil and climatic conditions tend to be similar for sites within close proximity and thus positively correlated. As a consequence, windbreak growth at sites within close proximity tends to be similar. Here "close" was estimated to be within 24 km. By the maximum likelihood ratio test, the spatial dependence parameter was found to be statistically significant ($\varphi = 0$ vs. $\varphi \neq 0$) at the 5% level. Based on an estimated φ of 0.575, $\boldsymbol{\beta}$ was estimated as shown in the Appendix. Tests for H_0: $\beta_i = 0$ ($i = 0, 1, 2, 3$) were all significant at the 0.01% level (Table 5).

Prediction of Windbreak Growth

The residuals from the model (Fig. 3b) showed no trend and were approximately normally distributed. A graph of the predicted mean volume for each location from the model versus the observed mean volume at that location, both on the logarithmic scale, is shown in Fig. 3c.

Using the estimated spatial dependence parameter, the model predicted mean site volumes (obtained by exponentiating the predicted values from the model, which are on the log scale) in cubic centimeters at age 36 for all sample sites. A smoothed map of predictions within the target region is illustrated in Fig. 4. The predicted mean site volume in Nebraska increased from the southeast to the northwest. This is reasonable because eastern Nebraska, with its more favorable climatic conditions, is generally more productive compared with the western panhandle region. The uncertainty of the prediction was quantified through the prediction errors (see Appendix for details). These too had to be back transformed from the log to the original scale using the delta method (Rice, 1994), and we refer to those on the original scale as the prediction errors of aboveground tree volume. The prediction error of aboveground tree volume was smallest near the sample sites (Fig. 5) because more information was available for modeling at these sites. The highest prediction errors of aboveground tree volume occurred at points farthest from the sample sites. The model's predictive accuracy would improve and prediction errors would be reduced if additional data become available.

Existing C stock models for forests are site and type specific, mostly using periodic inventory data to obtain volume estimates on a unit area basis, which are then converted to biomass with a set of nonlinear volume-to-biomass equations (Brown, 2002; Smith et al., 2004; Von Mirbach, 2000). Prediction for future C stocks is based on results from forest simulation models that project inventory, growth, and harvest on timberland changes from consecutive inventories. Smith et al. (2003) provided inventory-based calculations for major forest types in the United States. However, for the six forest types in the northern prairie states region that includes Nebraska, none included green ash or eastern red cedar, which are two of the major windbreak species, as the dominant species. Because trees in agroforestry systems are not explicitly inventoried within the FIA program (Perry et al., 2009), recent attempts at estimating current or predicting future sequestration of C by these systems have mainly used modifications to models developed for major forest types.

Without periodic forest inventories, site index curves can be a useful alternative to predicting tree growth, and thus biomass and C stock for a single stand, or multiple stands with uniform or similar site and climate conditions. Its use for inferences over large geographical regions, however, is limited because many site curves are lacking and site and climatic conditions are variable. This is particularly true for use in predicting growth for estimating C stocks in current and potential windbreak locations throughout entire states and regions as well as for other high C-sequestering agroforestry plantings, such as riparian forest buffers.

Conclusions

The readily available soil and climate information from online sources makes it possible to model windbreak growth with quantitative and qualitative soil and climate variables in place of site

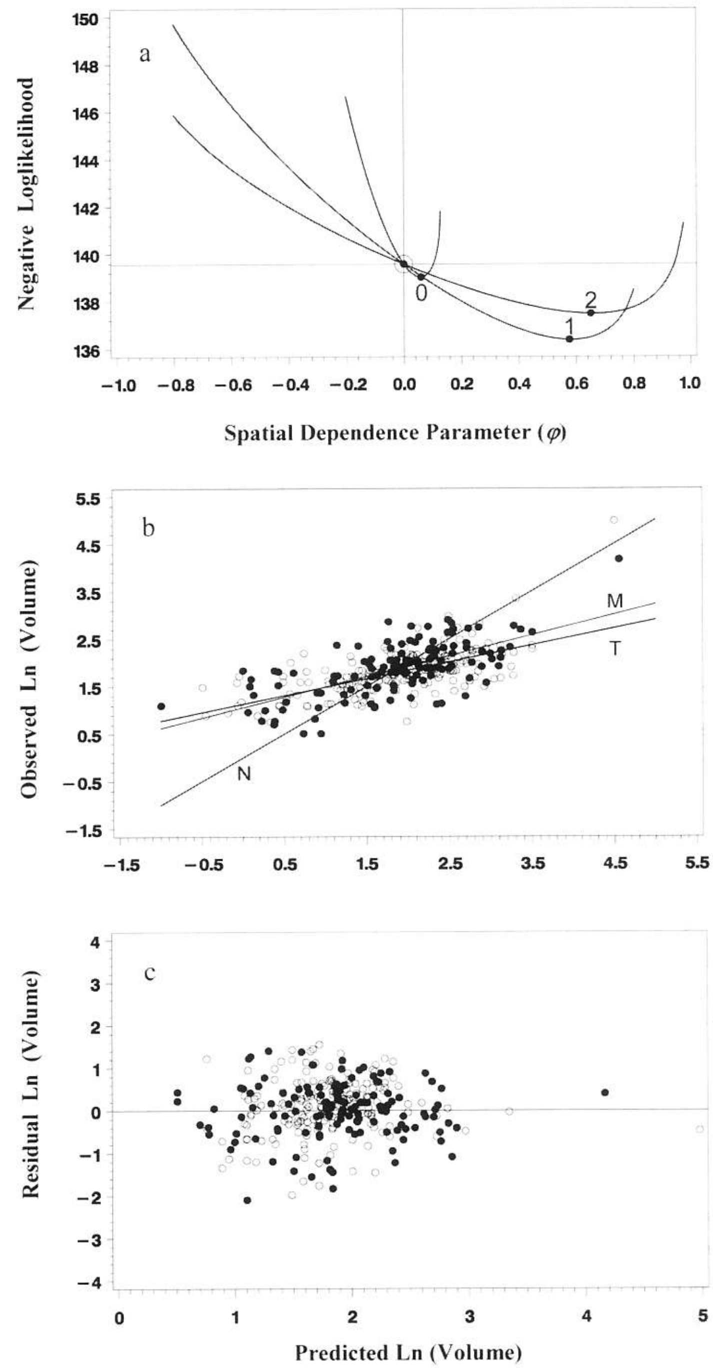

Fig. 3. (a) Negative log-likelihood as a function of spatial dependence parameter (φ) and scale parameter (k) for the site mean model. The log-likelihood reaches maximum with $k = 1$ and $\varphi = 0.575$. (b) Residual distributions for the site mean model with spatial dependence parameter $\varphi = 0.575$ (dots) and for the individual tree model with independence model with $\varphi = 0$ (circles). (c) Predicted vs. observed site mean volume on natural log scale for the site mean model (dot) and individual tree model (circle). The linear regression line for the site mean model (M) is closer to the 1:1 reference line (N) than the individual tree model (T).

indices. This approach overcomes the limitations of the stand method by capturing the large-scale trend as a smooth curve over space rather than an abrupt change from location to location. Further, it enables the prediction of windbreak growth at any location over an entire region regardless of previous tree growth information at that particular location. On the large-scale trend, we found that the log of windbreak age ($p < 0.001$), the 30-yr average summer precipitation ($p < 0.001$), and growth condition

Table 5. Parameter estimates and test for statistics for site mean model 2.

Effect	Parameter estimate	SE	t value	Pr > t value
Intercept	−6.903	1.624	−4.260	<0.0001
Ln (age)	1.933	0.458	4.227	<0.0001
Summer rain	0.404	0.062	6.559	<0.0001
Condition code†	−0.441	0.092	−4.799	<0.0001

† Growth condition code is inversely correlated with the natural log of aboveground volume. One degree lower in condition code, from good to fair, for example, will lead to a drop in the log of aboveground volume of 1.554 cm^3 on average (if every other predictor variable remains the same).

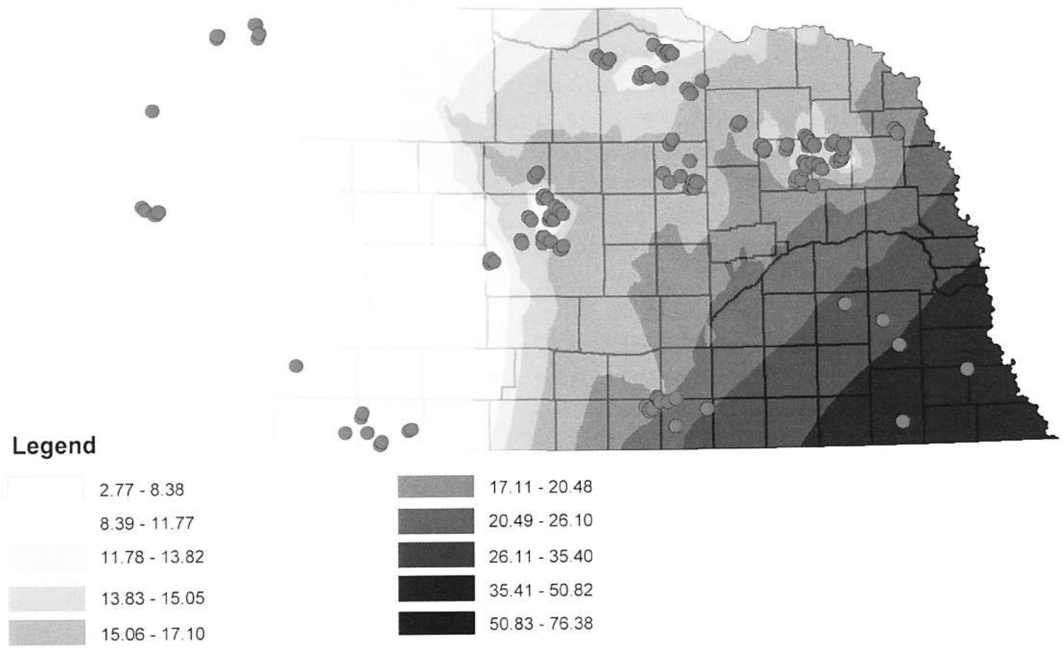

Fig. 4. Smooth map for the predicted log of the aboveground volume (cm^3) per tree at sample points (circles) with the site mean model (spatial dependence parameter $\varphi = 0.575$; scale parameter $k = 1$; tree age, 36 yr). At this fixed age, the average volume for a green ash field windbreak increases from the northwest to southeast in the state of Nebraska.

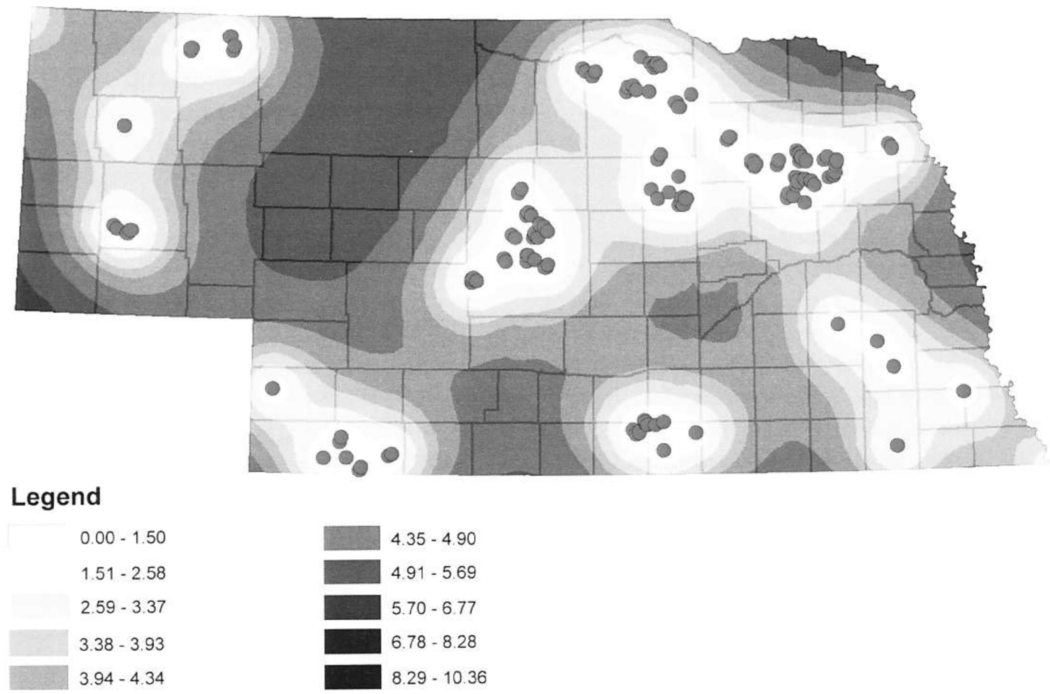

Fig. 5. Smooth map for the standard error for predicted log of aboveground volume (cm^3) per tree at sample points (circles) with the site mean model (spatial dependence parameter $\varphi = 0.575$; scale parameter $k = 1$; tree age, 36 yr). The error is lower at or around densely sampled points and becomes larger when moving away from sample points and densely sampled areas.

($p < 0.001$) were the most important predictors for the log of aboveground volume, indicating that climate and management practices play major roles in determining windbreak C yield within a given time period. For small-scale variation, the model estimated that the log of windbreak aboveground volumes were positively correlated ($\varphi = 0.575$; $p < 0.05$) for sites within 24 km of each other. By capturing this spatial dependence, the spatial model approach makes it possible to improve predictions of C stocks for present and future windbreaks by incorporating information from neighboring sites. Because the maximum likelihood method is used for parameter estimates and statistical inferences and all the input information is readily available, this model provides an alternative that can be further improved with updated data sources for green ash windbreaks for larger regional assessments, as well as for all major windbreak tree species, with limited adjustment.

Limitations

There are a number of limitations with the spatial Markov random field model. First, we modeled the natural logarithm of the site mean tree volume rather than biomass per unit area like other forest C stock models. Unlike the FIA program, repeated measurement of windbreak yield from permanent sampling plots was not available in the agroforestry intensive Midwest region and is unlikely to be in the near future. The limited windbreak standard report data we used in this study may underestimate the total C stocks because the original survey was not designed to measure total biomass per unit area but rather was a survey of major windbreak tree species with no estimates of understory growth. An attempt to convert individual tree data to biomass per unit area may induce uncertainty due to varied species composition, unequal sample size, and lack of understory measurements. Second, quantification of net change over time for the same sites was unrealistic because of the lack of repeated sampling on the same windbreaks. Prediction for future C stocks based on modeling tree growth as a function of time along with site, climate, and available management condition may not fully capture the potentially large variation over space, especially at sites with few sample points in the surrounding area. Third, climate data and about half of the soils data used for the windbreak sites in this study were predicted. Fourth, the quality of the soil attributes differed depending on whether the windbreak soil type matched one of the soils in the associated STATSGO map unit. This disparity in quality was not considered. Finally, because the STATSGO soil data and the long-time records by High Plains Regional Climate Center are quality data sources and the prediction process leads to predictions that tend to be smoother than the true responses, the use of these predicted values in the modeling process should result in unbiased predictions of the response (Gryparis et al., 2009; Lopiano et al., 2010). However, the errors associated with the predicted volumes reported here are biased downward because the extra variation associated with the predicted soil and climate variables was not considered.

With the method developed here, future studies on agroforestry C stocks could benefit by focusing on (i) increasing the extent and intensity of data collection from the target region, (ii) using an updated network for climate measurements, and (iii) using the more detailed Soil Survey Geographic Database. All of these steps would decrease potential error associated with the geographical mismatch from different data sources, thus improving the model fit and leading to better predictions of C stocks.

Appendix

The spatial model used to fit the windbreak data is explained in this Appendix. The notation of Cressie and Chan (1989) and Cressie (1993) is followed closely.

A windbreak growth variable can be regarded as a spatial random process

$$[Z(\mathbf{s}_i) : i = 1, 2, \ldots, n] \qquad \text{[A1]}$$

where $\{\mathbf{s}_i : i = 1, 2, 3, \ldots, n\}$ represents a windbreak spatial locations. By the Markov property, the conditional distribution of a windbreak attribute at a specific location $Z(\mathbf{s}_i)$ given all other sampling points $\{\mathbf{Z}(\mathbf{s}_j) : j \neq i\}$ depends only on a subset of

$$[Z(\mathbf{s}_j) : j \in \mathrm{N}_i; i = 1, 2, \ldots, n] \qquad \text{[A2]}$$

where the N_i's are a set of neighborhood sample locations determined by the distance between point i and j with $j \neq i$. By deriving the Hammersley-Clifford theorem, Besag (1974) showed that the joint distribution

$$\Pr[Z(\mathbf{s}_1), Z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_n)] \qquad \text{[A3]}$$

is determined by the conditional probability distributions

$$\Pr[z_i | (z_j : j \in \mathrm{N}_i)],\ i = 1, 2, \ldots n;\ j = 1, 2, \ldots, m \qquad \text{[A4]}$$

A Markov random field is the joint probability distribution in Eq. [A3] as determined by Eq. [A4] (Cressie and Chan, 1989). Suppose the density function of the conditional distribution in Eq. [A4] is of exponential family form, that is, the conditional density of Eq. [A4] can be denoted by

$$\Pr[z_i \mid (z_j : j \in \mathrm{N}_i)] = \exp[A(z_j)B_i(z_i) + C_i(z_i) + D_i(z_j)] \qquad \text{[A5]}$$

Then a consequence of the Hammerley-Clifford theorem is that, under regularity conditions, the auto-Gaussian model is a spatial model for continuous data and has conditional densities

$$\Pr[z_i \mid (z_j : j \in N_i)] = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(z_i - \mu_i)^2 / (2\sigma^2)] \qquad \text{[A6]}$$
$$i = 1, 2, \ldots, n$$

where

$$\mu_i = \mathrm{E}\{z_i \mid (z_j : j \in \mathrm{N}_i)\} = \sigma^2\{\alpha_i + \sum_{j \in \mathrm{N}_i} c_{ij} z_j\} \qquad \text{[A7]}$$

is of a linear form. Here $\{\alpha_i : i = 1, \ldots, n\}$ are large-scale variation parameters. The $\{c_{ij} : j \in \mathrm{N}_i\}$ are small-scale variation parameters that model spatial dependence. Note that when $c_{ij} = 0$ in Eq. [A6], the joint independence model results. Furthermore, $c_{ij} = c_{ji}$. Let $c_{jj} = 0$ and $c_{ij} = 0$ if $j \notin \mathrm{N}_i$.

Besag (1974) showed that for the auto-Gaussian case, the expression in Eq. [A6] and [A7] are equivalent to $E[z_i \mid (z_j : j \in N_i)] = \mu_i + \sum_{i=1}^{n} c_{ij}(z_j - \mu_j)$ and

$$\mathrm{Var}[z_i \mid (z_j : j \in N_i)] = \sigma_i^2 \qquad \text{[A8]}$$

Further, the conditional distribution of $Z(\mathbf{s}_i)$ is Guassian; $i = 1,\dots, n$.

Thus, the log of the aboveground volume at sites $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$, $\mathbf{Z} = [Z(\mathbf{s}_1), Z(\mathbf{s}_2),\dots,Z(\mathbf{s}_n)]'$ is normally distributed provided that $\mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$ is symmetric, positive definite, and invertible; that is,

$$\mathbf{Z} \sim \mathrm{N}[\boldsymbol{\mu},(\mathbf{I}-\mathbf{C})^{-1}\mathbf{M}] \qquad \text{[A9]}$$

where $\boldsymbol{\mu} = (\mu_1,\dots, \mu_n)'$ captures the large-scale variation; $\mathbf{M} = \mathrm{Diag}\,(\sigma_1^2,\dots, \sigma_n^2)$; $\mathbf{I}$ is an $n\times n$ identity matrix; $\mathbf{C} = \{c_{ij}\}$ with $c_{ij} = 0$ if $j \notin \mathrm{N}_i$; $c_{ij} = \varphi f(d_{ij})$; and if $j \in \mathrm{N}_i$, d_{ij} is a function of distance between $\mathbf{s}_i$ and $\mathbf{s}_j$. Notice that $\mathbf{C}$ captures the small-scale spatial dependence.

Consequently, the negative log-likelihood for a data set from a distribution in Eq. [A9] is

$$\begin{aligned} \mathrm{L}(\boldsymbol{\mu},\mathbf{M},\mathbf{C}) &= (n/2)\ln(2\pi)+(1/2)\ln|(\mathbf{I}-\mathbf{C})^{-1}\mathbf{M})| \\ &+(1/2)(\mathbf{Z}-\boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{I}-\mathbf{C})(\mathbf{Z}-\boldsymbol{\mu}) \end{aligned} \qquad \text{[A10]}$$

which can be minimized with respect to the parameters $\boldsymbol{\mu}$, $\mathbf{M}$, and $\mathbf{C}$.

Modeling Large-Scale Variation

To account for large-scale variation in site and climate conditions, the linear predictor ($\boldsymbol{\mu}$) is used to capture the mean response at a given site with given responses at neighboring windbreaks:

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \qquad \text{[A11]}$$

In Eq. [A11], the design matrix $\mathbf{X}$ includes all significant quantitative and qualitative predictor variables ranging from windbreak attributes, soil measurements, and long-term climate data. We assume that all effects are additive. Site productivity tends to be linear in effects of soil, climate, and windbreak characteristics, as indicated in the various site index curves found in the forestry literature (Carmean et al., 1989; Alemdag, 1991).

Modeling Small-Scale Variation

The spatial dependence matrix $\mathbf{C}$ is modeled as a function of the distance between windbreaks. To do this, the range (R^*) of spatial dependence must first be determined. The correlation structure can be approximated by fitting an appropriate semivariogram model to the residual from the regression in Eq. [A11]. The range parameter R^* from the semivariogram model identifies the neighborhoods of the Markov random field. That is, the spatial dependence is assumed to exist only within a distance of R^* of a given point, and all windbreaks within R^* of the point are in its neighborhood. Windbreaks further than R^* apart are independent. Thus, windbreak j is in the neighborhood of point i ($j \in \mathrm{N}_i$) if

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \le R \qquad \text{[A12]}$$

where x and y are coordinates of sample locations, and R stands for the estimated range of spatial dependence.

As defined in Eq. [A9], $\mathbf{C}$ represents spatial dependence with $c_{ij} = 0$ if $j \notin \mathrm{N}_i$, and $c_{ij} = \varphi f(d_{ij})$ if $j \in \mathrm{N}_i$. For $j \in \mathrm{N}_i$ define

$$f(d_{ij}) = C(k)\, d_{ij}^{-k} \qquad \text{[A13]}$$

where $C(k)$ = minimum ($d_{ij}^k : j \in \mathrm{N}i$, $i = 1, 2,\dots n$), and d_{ij} is as defined in Eq. [A12]. Notice that k is a scale parameter, controlling the speed of changes for the spatial dependence with distance of separation.

Likelihood Based Fitting of the Spatial Model

Cressie and Chan (1989) showed that the maximum likelihood estimator (MLE) for spatial dependence parameter φ can be obtained by first assuming φ as fixed and determining the MLEs of $\boldsymbol{\beta}$ and σ^2 as

$$\boldsymbol{\beta}(\varphi) = (\mathbf{X}`(\mathbf{I}-\mathbf{H})\mathbf{X})^{-1}\mathbf{X}`(\mathbf{I}-\varphi\mathbf{H})\mathbf{Y} \qquad \text{[A14]}$$

$$\sigma^2(\varphi) = \mathbf{Y}(\mathbf{I}-\mathbf{H})[\mathbf{I}-\mathbf{X}(\mathbf{X}`(\mathbf{I}-\mathbf{H})^{-1}\mathbf{X})^{-1}\mathbf{X}(\mathbf{I}-\mathbf{X})]\mathbf{Y}/n \qquad \text{[A15]}$$

where the ijth element of $\mathbf{H}$ is $f(d_{ij}) = C(k)\, d_{ij}^{-k}$, and all other terms are as defined earlier.

Substituting Eq. [A14] and [A15] back into Eq. [A10], the MLE for φ can be obtained as

$$\begin{aligned} L(\varphi) &= (n/2)\ln(2\pi)+(n/2)-(1/2)\sum_{i=1}^{n}\ln(n_i) \\ &+(n/2)\ln(\sigma^2(\varphi))-(1/2)\ln(1-\varphi\delta_i) \end{aligned} \qquad \text{[A16]}$$

where δ_i, $i = 1, 2,\dots, n$ are the eigenvalues of the symmetric $\mathbf{H}$ matrix.

The maximum likelihood ratio test statistic can be used to test H_0: $\varphi = 0$ vs. H_i: $\varphi \neq 0$:

$$G = 2[L(\varphi) - L(\varphi = 0)] \sim \chi_1^2 \qquad \text{[A17]}$$

The maximum likelihood ratio test statistic can also be used to test the composite hypotheses H_0: $k`\boldsymbol{\beta} = 0$ vs. $\mathrm{H}_1$: $k`\boldsymbol{\beta} \neq \mathbf{0}$ because

$$G = 2[(n-p-q)/n][L_p - L_{p+q}] \sim \chi_q^2 \qquad \text{[A18]}$$

where $(p+q)$ and p are numbers of parameters in the full and reduced models, respectively.

Spatial Prediction

The variance of the MLE of $\boldsymbol{\beta}$ can be estimated by

$$\mathrm{Var}(\boldsymbol{\beta}) = (\mathbf{X}`(\mathbf{I}-\mathbf{H})^{-1}\mathbf{X})^{-1}\sigma^2 \qquad \text{[A19]}$$

The predicted value can be obtained through

$$\mathbf{Y}^p = \mathbf{X}\boldsymbol{\beta} + (\varphi\mathbf{H})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \qquad \text{[A20]}$$

The variance of the predicted value is estimated using

$$\mathrm{Var}(\mathbf{Y}^p) = \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}' \qquad \text{[A21]}$$

where $\mathbf{F} = (\mathbf{I}-\varphi\mathbf{H})(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}) + \varphi\mathbf{H}$ and $\boldsymbol{\Sigma} = (\mathbf{I}-\varphi\mathbf{H})^{-1}\sigma^2$.

In Eq. [A20], $\mathbf{Y}^p$ is a vector of predicted mean values of log of aboveground volume at a given age and a given set of locations; $\mathbf{X}$ is a matrix of predictor variables assembled from the climate and STATSGO data sets based on locations, as well as planting arrangements in terms of spacing, species composition, direction, and the expected health condition and survival rates; $\boldsymbol{\beta}$ and φ are parameters estimated from previous data; $\mathbf{H}$

is a matrix whose elements are functions of distance between the target locations and previous sites; and **Y** is the vector of observed values for the response variable from the previous data set. A confidence interval for the prediction depends on all estimated parameters in the prediction equation, especially the large-scale parameters $\boldsymbol{\beta}$ and spatial dependence parameter φ. Accuracy of these parameter estimates will increase as more data become available.

Acknowledgments

We thank the anonymous reviewers for their critical review and helpful comments and suggestions, which greatly improved the quality of this manuscript. We also thank colleagues at the USDA National Agroforestry Center, especially Bruce Wight (NRCS), Richard Straight (FS), Richard Carman (FS), Gary Bentrup (FS), and Gary Kuhn (NRCS), along with Johnny Patterson (USDA NRCS) for their discussions and assistance on procuring and working with the windbreak and soils data sets. Early comments by Ron McRobert (USFS–Northern Research Station) greatly improved the manuscript. The first author would like to acknowledge Matthew S. Mayo, PhD., Professor and Chair, Department of Biostatistics at Kansas University Medical Center for his advice on statistical analysis and for providing time and support for this work. This research was supported by the Research Joint Venture Agreement RMRS-99132 through the USFS Rocky Mountain Research Station, Forest Service and the McIntyre-Stennis Forestry Research Program at the University of Nebraska.

References

Alemdag, I.S. 1991. National site-index and height-growth curves for white spruce growing in natural stands in Canada. Can. J. For. Res. 21:1466–1474.

Besag, J.E. 1974. Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc. Ser. B. Methodological 35:192–236.

Brandle, J.R., H. Laurie, J. Tyndall, and R. Sudmeyer. 2009. Windbreak practices. p. 75–104. *In* H.E. Garrett (ed.) North American agroforestry: An integrated science and practices. 2nd ed. ASA, Madison, WI.

Brown, S. 2002. Measuring, monitoring, and verification of carbon benefits for forest-based projects. Philos. Trans. R. Soc. Lond. A 360:1669–1683.

Carmean, W.H., J.T. Hahn, and R.D. Jacobs. 1989. Site index curves for forest tree species in the Eastern United States. General Technical Report NC-128. USDA, Forest Service, North Central Forest Exp. Station, St. Paul, MN.

Chap, T.L. 2003. Introductory biostatistics. John Wiley & Sons, Hoboken, NJ.

Cressie, N.A.C. 1993. Statistics for spatial data. John Wiley & Sons, Hoboken, NJ.

Cressie, N., and N.H. Chan. 1989. Spatial modeling of regional variable. J. Am. Stat. Assoc. 84:393–401.

Cressie, N., and S. Lele. 1992. New models for Markov random fields. J. Appl. Probab. 29:877–884.

Cunningham, R. 1988. Genetic improvement of trees and shrubs used in windbreaks. Agric. Ecosyst. Environ. 22/23:483–498.

Draper, N.R., and S. Harry. 1998. Applied regression analysis. John Wiley & Sons, Hoboken, NJ.

Drignei, D. 2009. A kriging approach to the analysis of climate model experiments. J. Agric. Biol. Environ. Stat. 14:99–114.

ESRI. 2001. ARCMAP. Environmental Systems Research Inst., Redlands, CA.

Gotway, C.A., and L.J. Young. 2002. Combining incompatible spatial data. J. Am. Stat. Assoc. 97:632–648.

Griffith, D.A. 2003. Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization. Springer, New York.

Gryparis, A., C.J. Paciorek, A. Zeka, J. Schwartz, and B.A. Coull. 2009. Measurement error caused by spatial misalignment in environmental epidemiology. Biostatistics 10:258–274.

High Plains Regional Climate Center. 2002. US high plains monthly coop climate data. University of Nebraska-Lincoln. Available at http://www.hprcc.unl.edu/data/historical/ (verified 3 Feb. 2011).

Kennedy, H.E., Jr. 1990. Green ash (Fraxinus pennsylvanica Marsh.). p. 348–354. *In* Silvics of North America. Agriculture Handbook 654. USDA Forest Service, Washington, DC.

Korn, H., P. Ntayombya, O. Berghall, J. Cotter, R. Lamb, G. Ruark, and I. Thompson. 2003. Climate change mitigation and adaptation options: Links to, and impacts on biodiversity. p. 48–87. *In* Secretariat of the Convention on Biological Diversity. Interlinkages between biological diversity and climate change: Advice on the integration of biodiversity considerations into the implementation of the United Nations Framework Convention on Climate Change and its Kyoto protocol. CBD Tech. Series 10. Montreal, QC.

Lopiano, K.K., L.J. Young, and C.A. Gotway. 2011. A comparison of errors in variables methods for use in regression models with spatially misaligned data. Stat. Methods Med. Res. 20:29–47.

Lundergren, A.L., and W.A. Dolid. 1970. Biological growth functions describe published site index curves for Lake States timber species. Research Paper NC-36. U.S. Forest Service, Washington, DC.

Nair, P.K.R., B.M. Kumor, and V.D. Nair. 2009. Agroforestry as a strategy for carbon sequestration. J. Plant Nutr. Soil Sci. 172:10–23.

Nebraska Department of Natural Resources. 2001. Carbon sequestration, carbon emissions, and Nebraska: Background and potential. A report relating the requirements of LB 957 of the 2000 Session of the Nebraska Unicameral and containing the recommendations of the Carbon Sequestration Advisory Committee. Available at http://www.neo.ne.gov/reports/CarbonSequestration.pdf (verified 3 Feb. 2011).

Neter, J., W. Wasserman, M.H. Kutner, and C. Nachtsheim. 1996. Applied linear statistical models. McGraw-Hill Higher Education, Chicago, IL.

Nui, X., and S.W. Duiker. 2006. Carbon sequestration potential by afforestation of marginal agricultural land in the midwestern U.S. For. Ecol. Manage. 223:415–427.

Perry, C.H., C.W. Woodall, G.S. Liknes, and M.M. Schoeneberger. 2009. Filling the gap: Improving estimates of working tree resources in agricultural landscapes. Agrofor. Syst. 75:91–101.

Qi, H., and G.D. Willson. 2000. Effects of temperature anomalies on the Palmer Drought Severity Index in the central United States. Int. J. Climatol. 20:1899–1911.

Rice, J. 1994. Mathematical statistics and data analysis. 2nd ed. Duxbury Press, Belmont, CA.

Sander, D.H. 1971. Soil properties and Siberian elm tree growth in Nebraska windbreaks. Soil Sci. 112:357–363.

SAS Institute. 1990. The GLM procedure. SAS/STAT user's guide. Vol. 2. Version 6. 4th ed. SAS Inst., Cary, NC.

SAS Institute. 2008. The KRIGE2D procedure. SAS/STAT user's guide http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_krige2d_sect002.htm (verified 3 Feb. 2011).

Searle, S.R. 1971. Linear models. John Wiley & Sons, Hoboken, NJ.

Schoeneberger, M.M. 2009. Agroforestry: Working trees for sequestering carbon on agricultural lands. Agrofor. Syst. 75:27–37.

Smith, J.E., L.S. Heath, and J.C. Jenkins. 2003. Forest volume-to-biomass models and estimates of mass for live and standing dead trees of U.S. forests. Gen. Tech. Rep. NE-298. USDA Forest Service, Northeastern Research Station, Newtown Square, PA.

Smith, J.E., L.S. Heath, and P.B. Woodbury. 2004. How to estimate forest carbon for large areas from inventory data. J. For. 102:5–31.

State Soil Geographic (STATSGO) data base for Nebraska. Available at http://cohyst.dnr.ne.gov/metadata/m004statsgo_94.html (verified 5 Jan. 2011).

USDA NAC. 2001. Working trees for carbon: Windbreaks in the U.S. Available at http://www.unl.edu/nac/brochures/wbcarbon/index.html (verified 3 Feb. 2011).

USDA NRCS (Natural Resource Conservation Service). 2002a. Windbreak site standard plot reports. USDA Natural Resource Conservation Service. Available at http://esis.sc.egov.usda.gov/StandardReports/pgWindbreakSelect.aspx (verified 3 Feb. 2011).

USDA SCS. 1991. State soil geographic data base (STATSGO) data user's guide. Miscellaneous publication 1492. USDA SCS, Washington, DC.

USEPA. 2006. Representative carbon sequestration rates and saturation periods for key agricultural and forestry practices. Available at http://www.epa.gov/sequestration/rates.html (verified 7 Sept. 2010).

Verchot, L.V., M. Van Noordwijk, S. Kanji, T. Tomich, C. Ong, A. Albrecht, J. Mackensen, C. Bantilan, K.V. Anupama, and C. Palm. 2007. Climate change: Linking adaption and mitigation through agroforestry. Mitig. Adapt. Strategies Glob. Change 12:901–918.

Von Mirbach, M. 2000. Canada's model forest program, carbon budget accounting at the forest management unit level: An overview of issues and methods. Canada Forest Service, Ottawa.

Zhou, X.H., J.R. Brandle, E.S. Takle, and C.W. Mize. 2002. Estimation of the three-dimensional aerodynamic structure of a green ash shelterbelt. Agric. For. Meteorol. 111:93–108.