Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2002 - 14th Annual Conference Proceedings

SPATIAL MODELING OF BIOMASS IN NEBRASKA WINDBREAKS

Qingjiang Hou

Linda J. Young

James R. Brandle

Michele M. Schoeneberger

See next page for additional authors

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)

This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Hou, Qingjiang; Young, Linda J.; Brandle, James R.; and Schoeneberger, Michele M. (2002). "SPATIAL MODELING OF BIOMASS IN NEBRASKA WINDBREAKS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1210>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

Author Information

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

SPATIAL MODELING OF BIOMASS IN NEBRASKA WINDBREAKS

Qingjiang Hou¹, Linda J. Young¹, James R. Brandle², Michele M. Schoeneberger³

Department of Biometry¹ and School of Natural Resource Sciences², University of Nebraska;
National Agroforestry Center³, USDA Forest Service, Rocky Mountain Research Station

Field windbreaks have the potential of sequestering large amounts of carbon. Predicting how much carbon would be sequestered in a newly planted windbreak after ten or more years is of interest. The amount of carbon in a tree depends on its biomass. In a pilot study of Nebraska windbreaks, a Markov random field was used to predict the biomass of green ash in windbreaks as a function of soil and climate conditions. The spatial dependence parameter was significantly different from zero, indicating the presence of small scale variation. In addition to age, the 30-year average summer precipitation and the windbreak growth condition code were included in the final model. Future directions for improving the model are discussed.

Introduction

Field windbreaks are planted throughout the Great Plains region of the United States to protect crops, reduce wind erosion, manage snow, and provide crucial wildlife habitat (Brandle et al. 2000). In addition, windbreaks accumulate large amounts of carbon and store it as wood, serving as a living carbon bank. These windbreaks play an important role in mitigating the greenhouse effect. Estimating the amount of biomass and carbon stored in current windbreaks and predicting how much would be stored in future windbreaks provides a basis for decision making on future land use and arrangement. As a first step toward developing a spatial model to predict windbreak biomass and carbon sequestration in the Great Plains region, a pilot study using existing windbreak data and readily available soil and climate information from the web was conducted on green ash in Nebraska.

A Markov random field model (Cressie, 1993) was fitted to the windbreak, soil, and climate data. Maximum likelihood methods were used. The method can be applied more broadly to green ash and other species from the entire Great Plains region.

MATERIALS AND METHOD

Data Sources and Description

A major goal of this study is to develop a spatial model for predicting windbreak biomass and carbon sequestration for future field windbreaks at a given location within Nebraska. Previous surveys of field windbreaks and sets of soil and climate data, all readily available on the web, are the foundation for this model. These data sets were collected at different times, spatial scales, and coordinate systems. Combining, synthesizing, and extracting useful information from these multi-scale, and spatially misaligned, data sources remains one of the most challenging and fascinating areas in statistical modeling and prediction (Gotway and Young, 2002).

Nebraska windbreak data were obtained from a nationwide survey of field windbreaks conducted during the 1950-1960 period through the USDA Soil Conservation Service. From the website <http://plants.usda.gov/>, a Windbreak Standard Report based on these data can be retrieved for each state of the US (USDA Natural Resource Conservation Service, 2002). This report includes windbreak age, species composition, health condition, and site average height and diameter at breast height (DBH). Site coordinates (Township, Range, Sections), soil types, and annual precipitation from the nearest weather station are also included (Figure 1). To estimate within site variation, individual tree height and DBH were entered in a supplemental data set from the original survey cards. Major tree species in the Nebraska data set include green ash (*Fraxinus pennsylvanica* Marsh), cotton wood (*Populus sp.*), elms (*Ulmus pumila*), Eastern red cedar (*Juniperus virginiana*), and Ponderosa pine (*Pinus ponderosa*).

Green ash was selected for this pilot study because 1) it is one of the dominant windbreak species throughout the Great Plains region, and 2) a model relating above ground volume to growth parameters (height, DBH) has recently been developed (Zhou, 1999). Green ash is native to a large region of central North America. It is sensitive to site and climate variation and subject to large within stand differentiation due to internal competition for light and nutrients. These factors lead to large within and among site variation making estimation of the spatial correlation structure challenging. For these reasons, green ash serves as a valuable example for spatially modeling windbreak biomass for major windbreak species. A list of important parameters included in the windbreak data is given in Table 1.

Nebraska soil data were obtained from the U.S. Soil Conservation Service State Soil Geographic (STATSGO) data base (<http://www.nrc.state.ne.us/databank/>), which was designed for regional, river-basin resource planning and management. STATSGO core data are displayed as polygons (map units) of varying sizes and shapes. Each map unit is a mosaic of up to 21 soil components, but there is no visible distinction as to the circumference of these components on the map. Each map unit has a set of attribute tables containing 60 soil properties, including physical, chemical, biological, taxonomic, and geographical characteristics, pertaining to each type of soil within that unit. These attribute tables are connected to map units through a set of identifier variables. In this study, all attribute tables were imported as SAS datasets and merged together to provide information on available variables at a desired level. A list of some of the climate variables is given in Table 2; some available soil variables are shown in Table 3.

Climate data were downloaded from the web site <http://www.hprcc.unl.edu/> maintained by the High Plains Regional Climate Center (2002). This data set consists of data from 125 weather stations throughout Nebraska with long-term climate records (Figure 1). Data set variables include monthly precipitation, temperature, heating, cooling, and growing degree days. Also available from this web site are monthly and annual means of Palmer Drought Severity Index and the Standard Precipitation Drought Index at each of the 125 weather stations.

Palmer Drought Severity Index (PDI) is widely used as an indicator of regional drought conditions. It measures the accumulated effect of monthly rainfall deficit/surplus relative to the monthly 'climatologically appropriate' rainfall, defined as rainfall needed to maintain adequate

soil water content for normal plant growth in a region (Qi and Wilson, 2000). PDI is a scaled value with a mean of zero. Negative values represent insufficient moisture, and positive values indicate at least adequate moisture. The number of months during the growing season (March through August) over a 30-year period for which PDI indicated drought were calculated at each weather station. Drought was defined at two levels: a PDI less than -2 and a PDI less than -3 . These measurements of drought were used as independent variables for the spatial prediction of volume at each windbreak point. Selecting -2 and -3 as critical values is based on the assumption that woody plants are relatively tolerant to moderate drought conditions due to their deep rooting system and that tree growth is more likely to suffer with increases in both intensity and frequency of drought. Table 2 provides a list of some other important climate variables.

Clearly, these data sources have different spatial support. Although both the windbreak and weather station locations are points (Figure 1), they do not occur at the same points and thus have different supports. The STATSGO data are for polygons, representing a third support. Geographic Information System (ARCMAP) and geostatistical methods were used to synthesize these three spatially misaligned data sources. Despite efforts made to align the data from these three sources, the errors associated with the changes of support were not considered here. The process of putting the three types of data together is displayed in Table 4.

The STATSGO soil map was used as the primary base map for data synthesizing (SCS, 1991). In the first step, the windbreak data set, with coordinates of Township, Range, and Section, was converted into the geographical coordinate system in terms of latitude and longitude. Then, the windbreak data points were overlaid onto STATSGO soil map and a unique map unit was identified for each sampling point with ARCMAP's spatial join function. If the map unit contained the specific type of soil as in the windbreak data, all attributes for that type of soil from STATSGO attribute tables were given to the windbreak data point(s). However, only about half of the windbreak points fall into STATSGO map units containing the same type of soil, possibly due to its relatively lower resolution (Richard et al., 1995). If none of the types of soils in a map unit matched with the windbreak data that fell within it, the mean values from STATSGO attribute tables for that particular type of soil were assigned to the windbreak data point. Consequently, each windbreak sampling point attained a complete set of soil specific attributes from the STATSGO attribute tables. However, the quality of the attributes differs depending on whether or not the windbreak soil type matched one of the soils in the associated STATSGO map unit. This disparity is not considered further here.

The weather data, with coordinates in latitude and longitude, were also overlaid onto the STATSGO soil map. Although the 125 weather stations were roughly evenly distributed over the entire state, their spatial locations did not correspond to the windbreak sampling points. The long-term climate condition at windbreak sampling points was estimated using kriging with inverse distance weighting through ARCMAP's geostatistical analyst. Long-term climate, as opposed to short-term weather conditions, is thought to be more relevant to windbreak growth because the majority of trees sampled in the windbreak data were more than 15 years old. The 30-year average (1961~1990) for all weather attributes (precipitation; mean, maximum, minimum temperatures; cooling, heating degree days; and Palmer Drought Severity Index) were

used as the dependent variables for spatial prediction at the windbreak points. The predicted values for each climate attribute were merged with the windbreak-STATSGO data for spatial modeling.

Spatial Model

Windbreak growth, particularly, height, DBH, above ground biomass, and carbon storage, are spatial variables. Here above ground biomass is modeled as a function of site and climate conditions, as well as windbreak age, species composition and their arrangement, using a Markov random field. 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

$$\{\mathbf{Z}(s_i): i = 1, 2, \dots, n\} \quad (1)$$

where $\{s_i: i = 1, 2, \dots, n\}$ are spatial locations of windbreaks. By the Markov property, the conditional distribution of a windbreak attribute at a specific location $\mathbf{Z}(s_i)$ given all other sampling points $\{\mathbf{Z}(s_j): j \neq i\}$ will only depend on a subset of

$$\{\mathbf{Z}(s_j): j \in N_i\}, \{N_i: i = 1, 2, \dots, n\} \quad (2)$$

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

$$Pr\{\mathbf{Z}(s_1), \mathbf{Z}(s_2), \dots, \mathbf{Z}(s_n)\} \quad (3)$$

is determined by the conditional probability distributions,

$$Pr(z_i | \{z_j: j \in N_i\}), \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m. \quad (4)$$

A Markov random field is the joint probability distribution in (3) as determined by (4) (Cressie and Chan, 1989). Suppose the density function of the conditional distribution in (4) is of exponential family form, that is, the conditional density of (4) can be denoted by

$$Pr(z_i | \{z_j: j \in N_i\}) = \exp\{A_i((z_j: j \in N_i))B_i(z_i) + C_i(z_i) + D_i((z_j: j \in N_i))\}. \quad (5)$$

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

$$Pr\{z_i | (z_j: j \in N_i)\} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-(z_i - \mu_i)^2 / (2\sigma^2)\}, \quad i = 1, 2, \dots, n \quad (6)$$

$$\text{where } \mu_i = E\{z_i | (z_j : j \in N_i)\} = \sigma_i^2 \left\{ \alpha_i + \sum_{j \in N_i} c_{ij} z_j \right\}. \quad (7)$$

is of a linear form. Here $\{\alpha_i : i=1, \dots, n\}$ are large-scale variation parameters. The $\{c_{ij} : j \in N_i\}$ are small-scale variation parameters that model spatial dependence. Note that when $c_{ij} = 0$ in (6) the joint independence model results. Further, $c_{ij} = c_{ji}$. Let $c_{ij} = 0$ and $c_{ij} = 0$, if $j \notin N_i$.

Besag (1974) showed that for the auto-Gaussian case, the expression in (6) and (7) are equivalent to

$$E\{z_i | (z_j : j \in N_i)\} = \mu_i + \sum_{j=1}^n c_{ij} (z_j - \mu_j)$$

and

$$\text{Var } \{z_i | (z_j : j \in N_i)\} = \sigma_i^2. \quad (8)$$

Further, the conditional distribution of $Z(s_i)$ is Gaussian; $i = 1, \dots, n$.

Thus, the spatial process $\mathbf{Z} = (Z(s_1), Z(s_2), \dots, Z(s_n))'$ is normally distributed provided that $\mathbf{M}'(\mathbf{I} - \mathbf{C})$ is symmetric, positive definite, and invertible; that is,

$$\mathbf{Z} \sim N\{\boldsymbol{\mu}, (\mathbf{I} - \mathbf{C})^{-1} \mathbf{M}\} \quad (9)$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ capture the large-scale variation; $\mathbf{M} = \text{Diag}(\sigma_1^2, \dots, \sigma_n^2)$; $\mathbf{I}$ is an $n \times n$ identity matrix; $\mathbf{C} = \{c_{ij}\}$. Here $c_{ij} = 0$ if $j \notin N_i$, and $c_{ij} = \phi f(d_{ij})$, if $j \in N_i$, d_{ij} is a function of distance between S_i and S_j . Notice that $\mathbf{C}$ will capture the small-scale spatial dependence.

Consequently, the negative log-likelihood for a data set from a distribution in (9) is

$$L(\boldsymbol{\mu}, \mathbf{M}, \mathbf{C}) = (n/2) \log(2\pi) + (1/2) \log |\mathbf{M}(\mathbf{I} - \mathbf{C})^{-1}| + (1/2) (\mathbf{Z} - \boldsymbol{\mu})' \mathbf{M}^{-1} (\mathbf{I} - \mathbf{C}) (\mathbf{Z} - \boldsymbol{\mu}) \quad (10)$$

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:

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \quad (11)$$

In (11), 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 (Alemdag, 1991; Sander, 1971).

Modeling small-scale variation

The spatial dependence matrix 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 (11). 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 ($j \in N_i$)

$$\text{if } d_{ij} = \text{abs}\{(x_i - x_j)^2 + (y_i - y_j)^2\}^{1/2} \leq R \quad (12)$$

where the x 's and y 's are coordinates of sample locations; and R stands for the estimated range of spatial correlation.

As defined in (9), C represents spatial dependence with $c_{ij} = 0$ if $j \notin N_i$, and $c_{ij} = \phi f(d_{ij})$, if $j \in N_i$. Now, for $j \in N_i$ define

$$f(d_{ij}) = C(k) d_{ij}^{-k} \quad (13)$$

where $C(k) = \text{minimum } (d_{ij}; j \in N_i, i = 1, 2, \dots, n)^k$, and d_{ij} is as defined in (12). 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 β and σ^2 as

$$\hat{\beta}(\phi) = (X'(I - \phi H)X)^{-1} X'(I - \phi H)Y \quad (14)$$

$$\hat{\sigma}^2(\phi) = Y'(I - \phi H) \{I - X(X'(I - \phi H)^{-1}X)^{-1}X'(I - \phi H)\}Y/n \quad (15)$$

where $H = f(d_{ij}) = C(k) d_{ij}^{-k}$, and all other terms are as defined earlier.

Substituting (14) and (15) back into (10), the MLE for ϕ can be obtained as

$$L(\phi) = (n/2)\log(2\pi) + (n/2) - (1/2) \sum_{i=1}^n \log(n_i) + (n/2)\log(\hat{\sigma}^2(\phi)) - (1/2) \sum_{i=1}^n \log(1 - \phi\delta_i) \quad (16)$$

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

The maximum likelihood ratio test statistic can be used to test $H_0: \phi = 0$ vs. $H_1: \phi \neq 0$:

$$G = 2 \{L(\hat{\phi}) - L(\phi = 0)\} \sim \chi^2_1 \quad (17)$$

The maximum likelihood ratio test statistic can also be used to test the composite hypotheses $H_0: k'\beta = 0$ vs. $H_1: k'\beta \neq 0$ since

$$G = 2 \{(n - p - q) / n\} \{L_p - L_{p+q}\} \sim \chi^2_q \quad (18)$$

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

Spatial prediction

The variance of the MLE of β can be estimated by

$$\widehat{Var}(\hat{\beta}) = (X'(I - \hat{\phi}H)^{-1}X)^{-1}\hat{\sigma}^2. \quad (19)$$

The predicted value can be obtained through

$$\mathbf{Y}^p = X\hat{\beta} + (\hat{\phi}H)(Y - X\hat{\beta}). \quad (20)$$

The variance of the predicted value is estimated using

$$\widehat{Var}(\mathbf{Y}^p) = \mathbf{F}\hat{\Sigma}\mathbf{F}' \quad (21)$$

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

In (20), $\mathbf{Y}^p$ is a vector of predicted mean values of above ground 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; β 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 $\mathbf{Y}$ is the vector of observed values for the response variable from the previous data set. Obviously, a confidence interval for the prediction depends on all estimated parameters in the prediction equation, especially the large-scale parameters β and spatial dependence parameter ϕ . Accuracy of these parameter estimates will increase as more data become available.

Results

Maximum likelihood based fitting of the spatial model for the site mean data

Before attempting to fit the model, the green ash windbreak data were checked for possible outliers. To fully capture the inherently large variation within and among windbreaks, no observations were excluded unless the tree was dead, physically damaged, or indicated as having been replanted or a sprout.

The above ground volumes for site means were used as the dependent parameter. Conversion from height and DBH to volume was based on equations proposed by Zhou (1999). A close look at the windbreak data indicates that sample trees from the same sites are not normally distributed in terms of height, DBH, and above ground volume, possibly caused by severe within-site competition. In most cases the data are skewed to the left, suggesting some trees might be under suppression from neighboring trees. A logarithmic transformation for the above ground volume was found to be effective in overcoming the non-normal nature of the data in that it helped stabilize the variance across the spectrum of diameters at breast height.

The Markov random field model (9) is fitted to the site mean data with the log-transformed mean volume at each site as the response variable. The linear predictor in (11) was modeled as a linear function of windbreak characteristics (age, spacing, species composition, growth condition), climate variation (30-year average precipitation, temperature, Palmer Drought Severity Index, cooling degree days, heating degree days), and site conditions (windbreak suitability group, soil texture, soil reaction, organic content, soil depth, permeability):

$$\mu = W\varphi + Y\gamma + Z\kappa \quad (22)$$

In (22), μ represents the mean of the natural log of above ground volume in cubic centimeters of the sampled windbreak; φ , γ , κ are row vectors of coefficients and collectively they make up the row vector β in (11); and W , Y , Z represent the windbreak, climate, and soil parameters, respectively, collectively they form the design matrix X .

Based on the SAS type III sum of squares from the GLM procedure, the following linear model was considered for capturing the large-scale variation with the Markov random field model:

$$\text{Model 1:} \quad \text{Log}(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \varepsilon_i$$

where $y_i = \log(\text{volume})$ at site i ; $x_1 = \log(\text{age})$; x_2 = cumulative number of months with Palmer Drought Severity Index < -2 ; x_3 = 30-year average summer precipitation; x_4 = Windbreak within-row spacing; x_5 = windbreak growth condition code; x_6 = windbreak neighboring row species.

Small-scale variation was modeled as a function of separating distance between sample sites as described in (13). The maximum range of spatial correlation R in (12) was selected as 15

miles. Three different values for the scale parameter ($k = 0, 1, 2$) in (13) $f(d_{ij}) = C(k) d_{ij}^{-k}$ were tested, where $C(k) = \text{minimum}(d_{ij} : j \in N_i, i = 1, 2, \dots, n)^k$. Viewing the maximum of the likelihood as a function of the spatial dependence parameter (ϕ) suggests that ($k = 1$) provides a better fit than either ($k = 0$) or ($k = 2$) because the minimum of negative log likelihood for ($k = 1$) is the smallest (See Figure 3). This indicates that the spatial dependence among windbreaks can be best modeled as an inverse linear function of distance between sites. For $k = 1$, the log likelihood changes with the spatial dependence parameter and reaches a maximum at $\phi = 0.65$. However, the maximum likelihood estimates for β_2 , β_4 , and β_6 in model 1 are not significantly different from zero ($p = 0.27, 0.44$, and 0.20 , respectively). Parameter estimates and their statistics for model 1 are given below.

Effect	Estimate	STDERR	t-Value	Pr > t-Value
Intercept	-7.269	1.530	-4.751	<0.0001
Log (age)	2.113	0.451	4.682	<0.0001
PDI2	-0.006	0.011	-0.614	0.270
Summer Rain	0.404	0.060	6.737	<0.0001
Tree Spacing	0.005	0.032	0.148	0.441
Condition Code	-0.420	0.101	-4.141	<0.0001
Neighbor Species	-0.103	0.121	-0.849	0.198

Since maximum likelihood estimates for the effects of cumulative Palmer Drought Severity Index, tree spacing, and neighbor row species are not significantly different from zero, they are removed from the linear model for large-scale variation, leading to the reduced model:

$$\text{Model 2: } \log(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_3 x_{3i} + \beta_5 x_{5i} + \varepsilon_i$$

where x_1, x_3, x_5 as defined in Model 1.

Again, the scale parameter ($k = 1$) provides the best fit with the negative log likelihood attaining a minimum of 136.37 at $\phi = 0.575$ (Figure 3). The maximum likelihood ratio test statistic can be used to test for $H_0: \phi = 0$ vs. $H_a: \phi \neq 0$. From (17), the test statistic is $G = 3.91$. Under the null hypothesis (H_0), G has an approximate χ^2_1 distribution. Since $3.91 > \chi^2_{1, 0.05} = 3.84$, the spatial dependence is significant at the 5% level.

Based on $\phi = 0.575$, β is estimated as in (14). Tests for $\beta_i = 0$ ($i = 0, 1, 2, 3$) are all significant at 0.0001 level.

Effect	Estimate	STDERR	t-Value	Pr > t-Value
Intercept	-6.903	1.624	-4.260	<0.0001
Log (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

According to Whittle (1954, p 441), the confidence interval for the spatial dependence parameter ϕ can be approximated by,

$$\Pr\{L(\phi) \leq L(\hat{\phi}) + (n/(n-q-2)) * \chi^2_{1,(\alpha)} / 2\} \approx 1 - \alpha \quad (24)$$

where $\hat{\phi}$ is the MLE for model 2 and q is the number of large-scale parameters fitted in the model.

From (24), the 95% confidence interval for ϕ becomes (0.15, 0.8), which suggests the spatial dependence parameter is significantly different from zero in Model 2. The spatial dependence parameter for the site-mean model is significantly different from zero ($p < 0.05$), suggesting the site-mean windbreak volumes are positively correlated over space. Notice that, in Figure 3, the log likelihood function converges to the same value of 139.56 for all three scale parameters ($k = 0, 1, 2$) when the spatial dependence parameter is set to zero. This value corresponds to the covariance structure of independence; that is, a linear model without spatial dependence parameter. As indicated by Figure 3, the spatial independence model is not the best choice because the log likelihood reaches the maximum at $\phi = 0.575$ with scale parameter being one.

The residuals from Model 2 (Figure 2) do not suggest any trend or non-constant variance. A plot of the predicted mean volume for each location from Model 2 versus the observed mean volume at that location is shown in Figure 4. The extra parameters in Model 1 as compared to Model 2 do not significantly improve the fit of the model as evidenced by comparing the maximum log likelihood values of the two models: $G = 135.65 - 134.82 = 0.83$. Under H_0 , the G statistic in (18) follows the χ^2_{p-q} distribution with p and q being the number of parameters in Model 1 and Model 2, respectively. Since $G = 0.83 < \chi^2_{(3, 0.05)} = 5.99$, the more parsimonious model (Model 2) is appropriate.

Using equation (20) with spatial dependence parameter at 0.575 the predicted site mean volumes at age 36 for all sample sites based on model 2 is illustrated in Figure 5. The predicted site mean volume is larger in the southeast than in the northwest. Windbreaks grown in the eastern half of the state produce more above ground tree volume than those from the western half. This is reasonable since eastern Nebraska is generally more productive with more favorable climatic conditions, compared to the western panhandle region. The uncertainty of the prediction can be quantified through the prediction errors obtained from (21).

Discussion

The readily available soil and climate information from the web makes it possible to model windbreak growth with quantitative and qualitative soil and climate variables in place of site indices. This approach overcomes the limitations of earlier approaches used in forestry in the following aspects. First, it models windbreak growth as a function of site and climate condition, thus capturing the large-scale trend as a smooth curve over space rather than an abrupt change

from location to location. Second, it enables the prediction of windbreak growth at any site over an entire region regardless of previous tree growth information at that particular location. Third, small-scale variation can be modeled with certain spatial correlation structure and then used to improve prediction.

As has been noted throughout this work, some simplifying assumptions have been made. Different parameterizations of the covariance matrix lead to different results (Figure 3). Other covariance structures should be considered. The change of support needs to be more explicitly addressed. The varying quality of the soil data at the windbreak sites should be accounted for in the modeling process. These will be addressed in future research.

Acknowledgments

This research was supported in part by funds provided by the National Agroforestry Center, part of the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture.

References

- Alemdag, I.S. 1991. National site-index and height-growth curves for white spruce growing in natural stands in Canada. *Canadian Journal of Forest Research*. **21**: 1466-1474.
- Besag, J.E. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Ser. B*. **35**: 192-236.
- Brandle, J.R., Laurie, H. and Bruce, W. 2000. Windbreak practices. *North American Agroforestry: An Integrated Science and Practices*. American Society of Agronomy.
- Cressie, N. 1993. *Statistics for Spatial Data*. New York: John Wiley.
- Cressie, N and Chan, N.H. 1989. Spatial modeling of regional variable. *Journal of the American Statistical Association*. **84**: 393-401.
- Gotway, C.A. and Young, L.J. 2002. Combining Incompatible spatial data. *Journal of the American Statistical Association* **97**: 632-648.
- Qi, H. and Willson, G.D. 2000. Effects of temperature anomalies on the Palmer Drought Severity Index in the central United States. *International Journal of Climatology* **20**: 1899-1911.
- Richard, G.L. Jr., John, D.A. and John, A.B. 1995. Spatial variability of digital soil maps and its impact on regional ecosystem modeling. *Ecological Modeling* **82**: 1-10.
- Sander, D.H. 1971. Soil properties and Siberian elm tree growth in Nebraska windbreaks. *Soil Science* **112**: 357-363.
- SCS, 1991. *State soil geographic data base (STATSGO) data users guide*. U.S. Soil Conservation Service Miscellaneous Publication 1492.
- STATSGO (State Soil Geographic data base). Nebraska Department of Natural Resources. <http://www.nrc.state.ne.us/>.
- High Plains Regional Climate Center 2002. US High Plains Monthly Coop Climate Data. University of Nebraska-Lincoln. <http://www.hprcc.unl.edu/>.

USDA Natural Resource Conservation 2002. Service.Windbreak Site Standard Plot Reports.

<http://plants.usda.gov/>.

Whittle, P. 1954. On stationary processes in the plane. *Biometrika* **41**: 434-449.

Zhou, X.H. 1999. *On the three dimensional aerodynamic structure of shelterbelts*. Ph.D. dissertation. University of Nebraska-Lincoln.

Figure 1. Spatial distributions of green ash windbreak sites (dot), weather stations (flag), and STATSGO map units (polygon) that contain windbreak sites in Nebraska.

Figure 2

Figure 3

Figure 4

Figure 5. Smoothed map for the predicted site mean volume of green ash at age 36. Circles represent sample locations. The diameter of each circle is proportional to site mean volume.

Table 1 List of 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	Above ground volume (stem, branches, leaves)

Table 2 List of climate variables from STATSGO data.

Code	Description
PDI2	Total number of months in which the Palmer Drought Severity Index was below -2 over a 30 year period
Arain	30-year average annual precipitation
Avget1	30-year average January temperature
Srain	30-year average summer precipitation
Avgt7	30-year average July temperature
Cddall	30-year average annual cooling degree day
Hddall	30-year average annual heating degree day
Meantall	30-year average annual mean temperature

Table 3 List of 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

Table 4 Data source coordinates and synthesizing process

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