

QUANTIFYING AND MAPPING SPATIAL VARIABILITY IN SIMULATED FOREST PLOTS

Gavin R. Corral and Harold E. Burkhart¹

Abstract—We used computer simulations to test the efficacy of multivariate statistical methods to detect, quantify, and map spatial variability of forest stands. Simulated stands were developed of regularly-spaced plantations of loblolly pine (*Pinus taeda* L.). We assumed no effects of competition or mortality, but random variability was added to individual tree characteristics. The purpose of simulating stands without these complex interactions was to provide a controlled situation to measure the efficacy of our methods. We examined redundancy analysis, partial redundancy analysis, and spatially constrained cluster analysis for detecting spatial patterns and found that redundancy analysis and partial redundancy analysis were reliable methods to quantify and test spatial dependence, respectively. Spatially constrained cluster analysis had moderate success in mapping variability, but its application to more complex situations may be limited.

INTRODUCTION

Information is lacking on reliable methods to detect, quantify, and map spatial heterogeneity in small-scale forest plots. This study focused on individual plots of even-aged, regularly-spaced plantation stands. Increased efforts to improve stand uniformity have led to questions as to why spatial patterns of tree growth emerge. Spatially recognizable growth patterns of trees may be caused by many factors, in particular the effects of genotype and microsite variation are thought to greatly influence tree characteristics within a plot. Tree growth within a stands of different genetic stocks is an important topic (Buford and Burkhart 1987, Magnusson and Kremer 1993, Tang and others. 2001). Moreover, Oliver and Larson (1996) noted that polymorphism is likely to occur due to variations in microsite productivity. Both conceptually and mathematically, there exist, to some extent, a confounding of genetic and microsite effects on tree growth. A logical first step in understanding the underlying causes of structural dissimilarities in tree growth is to better understand how spatial patterns emerge and to quantify those effects. In order to investigate spatial effects on tree growth we first simulated then tested the efficacy of statistical methods to detect, quantify and map spatial heterogeneity.

METHODS

The study material was made up of five simulated plots of planted loblolly pine (*Pinus taeda* L.) (fig. 1). Plots consisted of 25 rows and 25 columns of even-aged loblolly pine. Each plot was assigned 1 of 5 microsite patterns. Microsite patterns are distinct formations of

high or low areas of productivity within each stand. The patterns of microsites were chosen to represent a range of possible site conditions and, most importantly, to induce spatial dependence of tree structure. By creating areas of high and low productivity, we make the characteristics (DBH and height) of trees dependent on their location in space. Microsite patterns consisted of 1 to 5 microsites per plot: a control (uniform productivity), biplot (2 microsites, triplot (3 microsites), quad plot (4 microsites), and the free plot (5 microsites). All five site patterns were established without the complexities of competition and mortality for an initial assessment of sensitivity analysis of statistical methods for assessing patterns before progressing to increasingly more complex situations.

Site index (SI) and diameter at breast height (DBH) values were drawn from normal distributions. Site index and DBH distributions were unique to microsites within plots. Microsites with higher SI values have higher DBH values. Height and diameter relationship equations from loblolly pines were used to allocate heights. Lastly we used grid coordinates to assign spatial variables to each tree. Our model with covariates was then:

$$Y_{625 \times 2} = X_{625 \times 1} + W_{625 \times 2} \rightarrow [\text{DBH, Height}] = [\text{SI}] + [X_{\text{coordinate}}, Y_{\text{coordinate}}]$$

Redundancy analysis which is synonymous with explained variance (Legendre & Legendre 2012) was implemented using R software's "VEGAN" package. The first step to RDA is to center all variables. The variables mentioned hereafter are considered centered.

¹PhD Candidate, Virginia Tech, Department of Forest Resources and Environmental Conservation; University Distinguished Professor of Forestry, Virginia Tech, Department of Forest Resources and Environmental Conservation.

Citation for proceedings: Schweitzer, Callie J.; Clatterbuck, Wayne K.; Oswalt, Christopher M., eds. 2016. Proceedings of the 18th biennial southern silvicultural research conference. e-Gen. Tech. Rep. SRS-212. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 614 p.

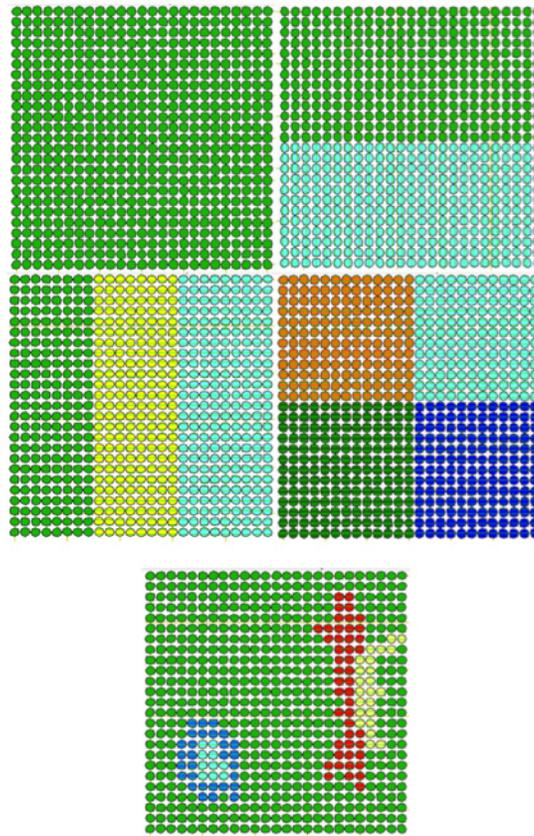


Figure 1—Illustrates the 5 spatial patterns used in this study. Microsites are distinguished by color. From left to right: Control, biplot, triplet, quadplot, and freeplot.

This symmetric form of analysis utilizes a response matrix $\mathbf{Y}_{625 \times 2}$ (hereafter \mathbf{Y}) with explanatory vector $\mathbf{X}_{625 \times 1}$ (hereafter \mathbf{X}) and covariables $\mathbf{W}_{625 \times 2}$ (hereafter \mathbf{W}). Three different simple RDA's were performed to estimate the pure spatial, pure environmental, spatially structured environmental, and residual variability. The pure spatial variability could be isolated and tested for significance by partial redundancy analysis (pRDA).

Partial redundancy analysis was done using R software's "VEGAN" package, is the partial canonical redundancy analysis of our response matrix \mathbf{Y} on matrix \mathbf{X} while controlling for the linear effect of matrix \mathbf{W} of covariables. For our specific objectives, we tracked the significance testing of $\mathbf{Y} \sim \mathbf{W} | \mathbf{X}$. This is the hypothesis of spatial dependence and tests the significance of a pure spatial component, representing a hypothesis of spatial dependence.

The clustering algorithms we used come from the R software package "const.clust". This clustering algorithm is an agglomerative approach with a constraint of spatial contiguity. The spatial contiguity constraint allows only trees that are neighbors to each other in space to be grouped together. There are

two important steps that must be performed in order to execute this algorithm. The first is to determine a dissimilarity metric and build a dissimilarity matrix; the second is to pick a connection method and build a contiguity matrix. We used Euclidean distance to build our dissimilarity matrix and Delaunay triangulation to define neighbors.

For each of our spatial patterns we imposed a range of scenarios (table 1). Each scenario specifies a difference in mean diameter at breast height (DBH) values among microsites and the value of coefficient of variation used to draw diameter values from normal distributions for individual trees.

RESULTS

Using RDA we performed a variation partitioning and estimate of the pure spatial variability. Table 2 contains the estimated partial R^2 value associated with the pure spatial component. As discussed earlier, there are 15 scenarios per spatial pattern. The scenarios are a combination of difference in mean DBH and a specified coefficient of variation (CV). Table 1 contains all possible combinations of CV and difference in mean DBH. Notably, the partial R^2 values for the control plot

Table 1—Illustrates 15 possible scenarios for each spatial pattern with k>1 microsites. Each scenario will be simulated for each spatial pattern. The control plot will be simulated with all 5 levels of CV

	Coefficient of Variation				
	5%	10%	15%	20%	25%
Difference in mean DBH (inches)					
0.5	(0.5,5%)	(0.5,10%)	(0.5,15%)	(0.5,20%)	(1,25%)
1	(1,5%)	(1,10%)	(1,15%)	(1,20%)	(1,25%)
2	(2,5%)	(2,10%)	(2,15%)	(2,20%)	(2,25%)

Table 2—Estimated partial R2 for each spatial pattern and all scenarios

Coefficient of Variation	BIPLOT	TRILOT	QUADPLOT	FREE	Control	
5%	29%	12%	5%	1%	0%	0.5 inch
10%	11%	7%	3%	1%	0%	
15%	5%	4%	2%	1%	0%	
20%	3%	2%	1%	0%	0%	
25%	2%	1%	1%	0%	0%	
Coefficient of Variation	BIPLOT	TRILOT	QUADPLOT	FREE	Control	
5%	47%	18%	6%	1%	0%	1 inch
10%	27%	14%	5%	1%	0%	
15%	16%	10%	4%	1%	0%	
20%	10%	8%	3%	1%	0%	
25%	6%	5%	2%	1%	0%	
Coefficient of Variation	BIPLOT	TRILOT	QUADPLOT	FREE	Control	
5%	56%	20%	7%	1%	0%	2inch
10%	45%	19%	7%	1%	0%	
15%	33%	16%	6%	1%	0%	
20%	24%	13%	5%	1%	0%	
25%	17%	11%	0%	0%	0%	

of 0 percent reflected the lack of spatial dependence. The largest R² values are in the biplot where spatial dependency was more pronounced in the data.

Partial redundancy analysis was used to detect spatial dependence. The results we report are the probability of detecting spatial dependence for all combinations of spatial patterns and scenarios. Table 3 contains the results from our pRDA simulations. The highest probabilities for detection are for the biplot and with larger differences in DBH among microsites. The probability for detecting spatial dependence in the control plot is consistently at about 5 percent. This reflects the type 1 error rate of $\alpha=0.05$, which is the probability of rejecting the null hypothesis when the null hypothesis is true.

Cluster analysis was performed to examine the efficacy in detecting the number of microsites in simulated plots and to correctly allocate each tree to the correct microsite. Table 4 illustrates the probability of detecting the correct number of microsites for all combinations of spatial pattern and scenario. The control spatial pattern is excluded because it is homogeneous and

the clustering algorithm does not test for k=1 groups in the data. Homogeneous stands are tested with spatial dependency tests such as the pRDA method described earlier. Much like the other methods, cluster analysis worked best in the biplot and where DBH differences were greatest.

Table 4 displays the probabilities for detecting the correct number of microsites. We used a threshold for reliability is 80 percent success. Therefore, in all cells of table 4 where the probability is greater than or equal to 80 percent we ran misclassification simulations. Table 5 shows the probability of misclassifying any given tree when using constrained cluster analysis. Values are generally low, but range from approximately 0 percent to 32 percent.

DISCUSSION

The variation partitioning and quantification of spatial variability performed as expected. Referring to table 2, we see that there is a general decrease in the amount of spatial variation captured as complexity of spatial pattern increases. This confirms that spatial pattern complexity and perhaps the shape of microsites

Table 3—Illustrates the results from pRDA. The values represent the probability of detecting spatial dependence

Coefficient of Variation	<i>BI PLOT</i>	<i>TRI PLOT</i>	<i>QUAD PLOT</i>	<i>FREE</i>	<i>Control</i>	
5%	100%	100%	100%	94%	5%	0.5 inch
10%	100%	100%	100%	64%	5%	
15%	100%	99%	91%	36%	7%	
20%	98%	92%	67%	24%	6%	
25%	88%	72%	49%	19%	5%	
Coefficient of Variation	<i>BI PLOT</i>	<i>TRI PLOT</i>	<i>QUAD PLOT</i>	<i>FREE</i>	<i>Control</i>	
5%	100%	100%	100%	99%	6%	1 inch
10%	100%	100%	100%	97%	4%	
15%	100%	100%	100%	89%	5%	
20%	100%	100%	100%	78%	6%	
25%	100%	100%	98%	64%	4%	
Coefficient of Variation	<i>BI PLOT</i>	<i>TRI PLOT</i>	<i>QUAD PLOT</i>	<i>FREE</i>	<i>Control</i>	
5%	100%	100%	100%	99%	6%	2inch
10%	100%	100%	100%	99%	4%	
15%	100%	100%	100%	95%	5%	
20%	100%	100%	100%	93%	6%	
25%	100%	100%	100%	85%	4%	

Table 4—Illustrates the probability of correctly identifying the number of microsites in all combinations of spatial patterns and scenarios

Coefficient of Variation	<i>BI PLOT</i>	<i>TRI PLOT</i>	<i>QUAD PLOT</i>	<i>FREE</i>	
5%	90%	52%	28%	0%	0.5 inch
10%	80%	22%	0%	0%	
15%	68%	26%	4%	4%	
20%	58%	28%	4%	2%	
25%	50%	10%	0%	0%	
Coefficient of Variation	<i>BI PLOT</i>	<i>TRI PLOT</i>	<i>QUAD PLOT</i>	<i>FREE</i>	
5%	100%	100%	100%	0%	1 inch
10%	97%	84%	14%	27%	
15%	80%	12%	0%	14%	
20%	74%	12%	0%	2%	
25%	70%	10%	0%	0%	
Coefficient of Variation	<i>BI PLOT</i>	<i>TRI PLOT</i>	<i>QUAD PLOT</i>	<i>FREE</i>	
5%	100%	100%	100%	0%	2 inch
10%	100%	100%	100%	0%	
15%	96%	34%	62%	10%	
20%	96%	4%	6%	6%	
25%	74%	0%	0%	0%	

Table 5—Illustrates the probability of misclassifying trees for all combinations of spatial patterns and scenarios where the probability of correctly identifying the correct number of microsites (Table 4) was above 80 percent. (*) indicates a combination where the probability from table 4 was less than 80 percent

Coefficient of Variation	BI PLOT	TRI PLOT	QUAD PLOT	
5%	0.068	*	*	0.5 inch
10%	0.316	*	*	
15%	*	*	*	
20%	*	*	*	
25%	*	*	*	
Coefficient of Variation	BI PLOT	TRI PLOT	QUAD PLOT	
5%	0.004	0.012	0.009	1 inch
10%	0.080	0.276	*	
15%	0.159	*	*	
20%	*	*	*	
25%	*	*	*	
Coefficient of Variation	BI PLOT	TRI PLOT	QUAD PLOT	
5%	0.000	0.000	0.000	2 inch
10%	0.007	0.013	0.036	
15%	0.084	*	*	
20%	0.136	*	*	
25%	*	*	*	

can influence the efficacy of RDA to capture spatial variability. The results from this portion of the experiment were very promising. The values from table 2 are partial R² values range to a high of 56 percent. The largest portions of variation explained by the spatial component were found in the scenarios with the greatest difference in mean DBH.

In pRDA we measured the probability of successfully detecting spatial dependency with permutation F-tests which looked at the significance of the “pure spatial” component. Generally, the results are promising with many scenarios > 80 percent success. Not surprising, an increase in spatial pattern complexity decreases the probability of successful detection; however, the decreases tend to be small. More influential than pattern complexity is difference in mean DBH.

The success of our clustering algorithm depends on how distinct each microsite is and how similar each tree is to others within a single microsite. For example, when we simulate the biplot we are looking at trees from two different distributions. The greater the difference in means and the smaller the CV, the more similar trees are within a microsite and the more dissimilar they are to trees from other microsites. As the mean difference in DBH decreases and/or CV increases the more overlap there is in the distributions of the microsites. This overlap can create false groups and the clustering algorithm may identify these overlapped trees as unique

groups (given they are also neighbors) and affect the accuracy of estimates.

Misclassification simulations were done for scenarios where the probability of successful detection was greater than 80 percent. For scenarios where probability of successful detection was greater than 80 percent, the misclassifications of trees ranged from about 0 percent-32 percent. This indicates that when cluster analysis works with high reliability the probability of misclassification is low. The cluster analysis was moderately successful. Sharp decreases in probability of success leave some lack of confidence in the ability of these methods to detect microsites patterns accurately.

LITERATURE CITED

- Buford, M.A.; Burkhart, H.E. 1987. Genetic improvement effects on growth and yield of loblolly pine plantations. *Forest Science*. 33(3): 707-724.
- Legendre, P.; Legendre, L.F. 2012. *Numerical ecology*. Vol. 20. The Netherlands: Elsevier. 989 p.
- Magnussen, S.; Kremer, A. 1993. Selection for an optimum tree growth curve. *Silvae Genetica*. 42: 322-322.
- Oliver, C.D.; Larson, B.C. 1996. *Forest stand dynamics*. Updated ed. New York: Wiley and Sons. 520 p.
- Tang, S.; Meng, F.R.; Bourque, C.P.A. 2001. Analyzing parameters of growth and yield models for Chinese fir provenances with a linear mixed model approach. *Silvae Genetica*. 50(3-4): 140-14.