**Quantifying and Mapping Spatial Variability
in Simulated Forest Plots**

By

Gavin R. Corral

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Forestry

Harold E. Burkhart, Chair

Pierre Legendre

Inyoung Kim

Philip J. Radtke

Thomas R. Fox

December 7, 2015

Blacksburg, VA

Keywords: Forestry, multivariate, canonical analysis, cluster analysis, redundancy
analysis, simulations, PTAEDA 4.1

Quantifying and Mapping Spatial Variability
in Simulated Forest Plots

Gavin R. Corral

ABSTRACT

Spatial analysis is of primary importance in forestry. Many factors that affect tree development have spatial components and can be sampled across geographic space. Some examples of spatially structured factors that affect tree growth include soil composition, water availability, and growing space. Our goals for this dissertation were to test the efficacy of spatial analysis tools in a forestry setting and make recommendations for their use. Reliable spatial analysis tools will lead to more effective statistical testing and can lead to useful mapping of spatial patterns.

The data for this project is from simulated even aged loblolly pine stands (*Pinus taeda L.*). These simulated stands are grown at regular spacing and we impose a range of parameters on the stands to simulate many possible scenarios. In chapter 3 of this dissertation we perform a sensitivity analysis to determine if our methods are suitable for further research and applications. In chapter 4 we perform our analysis on more realistic data generated by a spatially-explicit stand simulator, PTAEDA 4.1.

In chapter 3 we performed a statistical simulation of plantation stands without effects of competition and mortality. We used redundancy analysis (RDA) to quantify spatial variability, partial redundancy analysis (pRDA) to test for spatial dependence, and spatially constrained cluster analysis to map soil productivity. Our results indicated that RDA and pRDA are reliable methods and future evaluation is appropriate. The results from the spatially constrained cluster analysis were less clear. The success or failure of the clustering algorithm could not be disentangled from the success or failure of the selection criterion used to predict the number of clusters. Further investigations should address this concern.

In chapter 4 we used PTAEDA 4.1, a loblolly stand simulator, to simulate a range of site conditions and produce data that we could use for analysis. The results showed that RDA and pRDA were reliable methods and ready for the field. Spatially constrained cluster analysis performed poorly when more realistic data was used and because of this further use was uncertain . It was clear from the results that levels of variation and spatial pattern complexity of microsites influenced the success rate of the methods. Both RDA and pRDA were less successful with higher levels of variation in the data and with increased spatial pattern complexity.

In chapter 5 we related the coefficient of variation from our simulations in (chapters 3 and 4) to two sets of real plot data, including a clonal set and open pollinated set. We then implemented a spatial analysis of the real plot data. Our spatial analysis results of the two comparable data sets were unaffected by genetic variability indicating that the primary source of variability across plots appears to be soil, not genetic related.

Acknowledgements

I am deeply thankful for the mentoring, guidance, and unwavering support of my advisor Dr. Harold Burkhart. His style, mentorship, and technical proficiency are exemplary and will be a standard I will carry with me throughout my professional career. This dissertation would not be possible without his thoroughness and exceptional knowledge in the field of forestry. I would also like to thank my committee members Drs. Inyoung Kim, Pierre Legendre, Thomas Fox, and Philip Radtke for their support throughout my PhD program. Their input has made my research more productive and thorough than I could have anticipated.

My success has been made possible through the continuous and unwavering love and support of my family and friends. I would like to thank Joe C. for his friendship and willingness to help. I am thankful to my best friend and honorary brother, Brandon Herman, for his encouragement, humor, and unbroken loyalty. My continued success would not have been possible without the love and encouragement of my mom and dad. I am so thankful to have parents that will love and support me no matter where I find myself in life. I am especially thankful to my brother and sister – thank you for giving me the courage when I needed it most. Lastly, I thank my beautiful girlfriend, Ana Ortega, for always believing in me. Ana, you are a bright, shinning light in my life.

Table of Contents

List of Figures

Preface

This Dissertation is organized under what is referred to as manuscript format. The potential individual manuscripts make up chapters 3, 4, and 5. Dr. Burkhart and I plan to submit these manuscripts for publication to Forest Science, Canadian Journal of Forest Research, and/or Forest Ecology and Management journals. The manuscripts will be revised and condensed before submission.

Chapter 1

Introduction and Objectives

1.1 Introduction

A multitude of factors influence forest growth dynamics. The ability to disentangle the singular effects of these factors from each other is often difficult and requires careful investigation. Among the more prominent theoretically, but less studied factors that influence growth dynamics, are spatial effects. Broadly, tree attributes are spatially dependent if those attributes are correlated to spatial coordinates. Correlation between tree structure and spatial location results in spatially recognizable patterns (SRP's). SRP's occur because tree size becomes dependent on its location in space. Stands with positive spatial correlation occur when trees of similar size tend to be near each other. Conversely, negative spatial correlation occurs when trees of dissimilar size tend to be near each other.

There is a complex interaction of variables that can cause SRP's, specifically interactions below ground which affect soil productivity. Oliver & Larson (1996) note that variations in soil productivity cause variation in height-age growth trajectories. The varying shapes of growth curves caused by soil characteristics lead to SRP's and are subject to spatial analysis. There are distinguishable patterns of soil productivity within stands (Weber 1983), which through spatial analysis, could be mapped and the variability quantified.

When evaluating site productivity for even-aged forests the site index concept – that is average height of the trees that represent the dominant part of the canopy indexed

at a specified index age is commonly used (Burkhart & Tomé 2012, p. 131). The Site curves for different site qualities are sometimes defined as being proportional at all ages, resulting in what are commonly termed "anamorphic" site index curves. This assumption of a common shape for all sites is not biologically reasonable as height growth accumulates earlier on the better quality sites. When the shape of the height growth pattern is allowed to vary by site quality the resulting site index curves are commonly called "polymorphic" in the forestry literature (Burkhart & Tomé 2012, p. 138).

Spatial analysis can be a powerful tool used by practitioners and investigators. Understanding spatial structures in data may allow for more accurate sampling through stratification and more accurate statistical testing. Accurate estimation of microsites (discontinuities in productivity within a site) has applications in precision silviculture and forest health. More accurate statistical testing is possible with spatial analysis because variation caused by spatial components is accounted for and not confounded with random variation. Research conducted in this study used multivariate statistical methods to quantify, test, and map spatial variability.

Redundancy analysis (RDA) is a multivariate statistical method used to explain, among other things, the variance of a response matrix. Redundancy analysis quantifies the variability explained by a set of variables. With RDA, one can perform variation partitioning of the response variables into meaningful components such as pure spatial, environmental, or chemical. Using a method called partial redundancy analysis, one could test the significance of the individual components.

Partial redundancy analysis (pRDA) is a method used in multivariate analysis to test the significance of a set of variables. The pRDA method tests the response variable **Y**

by the explanatory variables in **X** while in the presence of additional explanatory variables called covariables (Legendre & Legendre 2012). In practical terms, this method can test if tree structures and spatial location are linearly related while controlling for the effects of the other.

Cluster analysis is the search for groups in data (Everitt et al. 2011). Trees growing in the same microsites may develop similarly. This is because trees in the same microsite will have similar resources. Burkhart and Tennent (1977) note that families of polymorphic site index curves, which exhibit different shapes for varying site quality, are generally preferred in forest modeling. This concept can be extended to microsites where site quality refers to each microsite ranging microsites can have different curves within a single stand.

Cluster analysis is a method used to discover these groups and map them. Furthermore, external information about the study site can be used in the clustering algorithm to facilitate solutions. Spatial constraints, for example, allow clusters to form spatially contiguous groups on a landscape. Spatial constraints are useful for situations were spatial proximity of neighbors is important or even necessary for interpretation.

There is a paucity of information regarding methods for capturing spatial variability in forest at a plot level. Furthermore, many approaches fail to disentangle confounding factors of variation, namely genetic, environmental, spatial, temporal, and competition. Simulations offer a controlled way to test the efficacy of spatial methodologies in forest settings and are a reasonable first step to real world applications.

1.2 Overall Research Objectives and Specific Research Questions

The goal of this dissertation is to test the efficacy of spatial analysis methods and make recommendations on their use. These methods have potential for many applications in forest science and forest management, but the efficacy for quantifying, testing, and mapping spatial variability is not well evaluated and documented in forestry. Thus this research progressed through three stages of increasing complexity. As an initial stage, spatial patterns were specified and their varying levels of means and relative variability for normal distributions without negative spatial correlation (due to intraspecific competition) or gaps (due to mortality). These highly controlled simulations provided a baseline of the effectiveness of methods of spatial dependence testing, quantifying spatial variability, and identifying number of groups in the data and to which groups individuals belong. From this point a distance-dependent stand simulator was employed to simulate tree diameters and heights with intraspecific competition and tree mortality included. These simulations were done for the same microsite spatial patterns as those specified for subpopulations with specified properties, thus enabling an evaluation of loss of discriminatory power with more realistic data. As a final step, the spatial methods chosen for evaluation were applied to actual field data.

We hypothesize that soil productivity alters tree characteristics at fine scales and that spatial analysis to characterize and map areas of alternating productivity will aid in many future investigations by helping to better understand stand uniformity. This dissertation examines the efficacy of multivariate spatial analysis and mapping through the following specific analytical steps which imply a number of underlying hypothesis, namely:

1)   Test spatial dependency with partial redundancy analysis, quantify

       spatial variability with redundancy analysis, and map microsites with

       cluster analysis.

   a. Will the outcome of spatial dependence tests be affected by microsite

       configuration, variability of diameter distributions, or relative size of

       the average tree among microsites?

   b. Does estimated spatial variability vary with microsite configuration,

       variability of diameter distribution, or coefficient of variation of

       average tree?

   c. Can cluster analysis reliably estimate the correct number of groups in

       simulated stands and do estimates vary with microsite configuration,

       variability of diameter distributions, and relative size of the average

       tree among microsites?

       This part of the study will be carried out with simulated populations of

       specified properties (i.e. without confounding effects of competition

       and mortality). (chapter 3)

2)   Test spatial dependency with partial redundancy analysis, quantify

       spatial variability with redundancy analysis, and map microsites with

       cluster analysis using a spatially-explicit simulator for planted stands of

       loblolly pine (*Pinus taeda* L.).

a. Will outcome of spatial dependence test be affected by microsite configuration, stand age, and site index variability?

b. Will estimated spatial variability vary by microsite configuration, stand age, and site index variability?

c. Will cluster analysis estimate the correct number of groups in each stand and do the results vary with microsite configuration, site index variability, and stand age?

This analysis will be carried out using a spatially –explicit loblolly pine stand simulator PTAEDA 4.1. (chapter 4)

3)   Conduct a spatial analysis of comparable sets of plot data from clonal and open-pollinated sources using partial redundancy analysis to test for spatial dependence, quantify spatial variability with redundancy analysis, and map microsites with cluster analysis. The relevant research questions are:

a. Without the confounding of genetic and microsite variability, will spatial patterns in clonal plots be more identifiable?

b. Will field plots contain spatial dependence?

c. Will there be detectable spatial variability in the height and diameter data?

d. Will cluster analysis adequately map spatial heterogeneity in tree growth?   (chapter 5).

1.3 Literature Cited

Burkhart, H. E., & Tennent, R. B. (1977). Site index equations for radiata pine in New Zealand. *New Zealand Journal of Forestry Science*, *7*(3), 408-416.

Burkhart, H. E., & Tomé, M. (2012). *Modeling forest trees and stands*. Springer Science & Business Media.

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). Cluster analysis. Wiley Series in Probability and Statistics.

Legendre, P. (2008). Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *Journal of Plant Ecology*, *1*(1), 3-8.

Oliver, C. D., & Larson, B. C. (1990). *Forest stand dynamics*. McGraw-Hill, Inc.

Weber, C. D. (1983). *Height growth patterns in a juvenile Douglas-fir stand, effects of planting site, microtopography and lammas occurrence* (Doctoral dissertation, University of Washington).

Chapter 2

Literature Review

Forestry data are commonly analyzed using ordinary least squares (Zhang et al. 2005). However, forestry data are often collected across space, which qualify them as spatial data sets. Like other environmental data, forestry data may include space-structured environmental phenomena that classical statistics cannot adequately study (Legendre & Fortin 1989). For this reason, methods of spatial analysis should be applied to account for the special nature of spatial data.

Modern modeling of tree growth remains overly simplified (Fox et al. 2001) and many modelers have come to accept the large amounts of unexplained variability as occupational hazards (Burkhart & Gregoire 1994). However, recent research has demonstrated that incorporating unexplained variability (i.e. spatial and/or temporal variability) into individual tree growth models is theoretically correct and produces more realistic predictions of actual tree growth (Fox et al. 2001). Stochastic structures, such as spatial dependencies, have largely been ignored in forest growth and yield modelling (Fox et al. 2001), however, Burkhart and Gregoire (1994) note that stochastic structures are prevalent in many applications of forest growth and yield models. In fact, Fox et al. (2001) note that a clear reluctance exists on the part of practitioners of forest growth modeling to characterize stochastic structures. The field of spatial modelling has undergone rapid development in recent years (Liu & Burkhart 1994a), but forest scientists have largely failed to incorporate these new developments, despite their applicability to growth and yield modeling (Fox et al. 2001).

From a sampling perspective, trees are treated as individual sampling units and are assumed to be independent of each other. This assumption is made to satisfy the requirements needed to implement Ordinary Least Squares (OLS) which is the predominate means of developing individual tree growth models (Fox et al. 2001). Researchers have begun to realize that treating individual trees within a single stand as independent is erroneous (Ford and Diggle 1981, Penttinen et al. 1992, Mateu et al. 1998). The presence of spatial dependence in the data violates the OLS assumption of independence and results in biased estimates of parameters on standard errors (Legendre & Fortin 1989).

Complications arise in defining a spatial stochastic structure for individual tree models because of the confounding effects of microsites, genetics, and competition (Matérn 1960, Liu & Burkhart 1994a, Liu & Burkhart 1994b). Furthermore, Magnussen (1990) notes that complex patterns of spatial dependence change with stand development. The complicated nature of spatial variability has created a focus on characterizing and modeling spatial dependence rather than the explicit modeling of spatial dependence in growth models (Fox et al. 2001). Procedures have been developed to diagnose spatial autocorrelation (Liu & Burkhart 1994b) such that further analysis can be done to characterize and model spatial dependence.

An important component of spatial regression modeling is accounting for spatial autocorrelation in the model residuals (Anselin & Griffith 1988). Spatial autocorrelation in the model residuals is a violation of the OLS independence assumption which directly affects resulting statistics, will bias the estimation of parameter standard errors, and will consequently cause misleading significance tests (Openshaw & Taylor 1979, Krämer

1980, Zhang et al. 2005, Meng et al. 2009). To identify spatial autocorrelation researchers (i.e. Reed & Burkhart 1985, Liu & Burkhart 1994a, Zhang et al. 2005, Meng et al. 2009) have used spatial indices to test the hypothesis of no spatial autocorrelation.

The most commonly used statistical techniques to detect spatial dependence such as autocorrelation are Moran's I, Geary's C, and Getis' G (Dale & Fortin 2002, Paez & Scott 2005) and are present in forestry literature (e.g., Reed & Burkhart 1985, Köhl & Gertner 1997, Meng et al. 2009) (Zhang et al. 2009). These indices have become popular in research for describing the degree of spatial autocorrelation between locations (Anselin 1995, Paez and Scott 2005) and to test the null hypothesis of no spatial autocorrelation (Fortin et al. 1989). Other researchers use the indices as a measure of model effectiveness in dealing with the presence of spatial autocorrelation (e.g. Bullock & Burkhart 2005, Zhang et al. 2005).

There are a number of different modeling techniques that can be used to account for the spatial autocorrelations among data observations (Zhang et al. 2005). Some examples of common modeling techniques that are used as spatial models are linear mixed models, generalized additive models, geographically weighted least squares, spatial lag models, spatial error model, and Durbin process models (Zhang et al. 2005, Meng et al. 2009). These models can be useful in dealing with autocorrelation, but they are not well suited for testing of spatial dependence and estimating spatial variability. Redundancy analysis is well suited for these tasks. Redundancy analysis is also useful when dealing with fine spatial scales. Borcard et al. (2004) found that through RDA one can model spatial structures at all spatial scales that can be perceived by the data set.

RDA can then be used to investigate environmental phenomena that happen at fine scales such as microsite variations.

Discontinuities in soil productivity often happen at fine scales. The patches formed by these discontinuities are called microsites in this report. These microsites cause trees to grow along different height-age trajectories (Oliver & Larson 1996). Polymorphism occurs when one microsite may favor rapid early growth and another may favor rapid late growth (Oliver & Larson 1996). Unlike anamorphic differences where curves develop proportionally to one another, polymorphic patterns imply that structural differences may develop at different points in time and not necessarily be exhibited evenly through time. Weber (1983) recognized that the height-age curves for sites with varying microsite productivity could be polymorphic (i.e tree height-age trajectories are among microsites are not proportional to each other). Carmean & Lenthall (1988) note that polymorphic tree height growth patterns have been observed for forest trees growing in areas of contrasting productivity. Burkhart & Tennent (1977) commented that families of site index curves which exhibit different shapes for varying site quality are commonly termed polymorphic. Oliver & Larson (1996) explain that trees on good microsites are more likely to become dominant if they grow next to trees on poor microsites. This phenomenon gives rise to spatially recognizable structures (patches), which may be spatially analyzed (Borcard et al. 1992, Legendre 1993). The methods of canonical ordination are well suited to analyze and account for spatial effects in environmental data.

Methods of canonical ordination belong to one of two general groups: symmetric and asymmetric. The asymmetric group consists of methods that identify two types of data sets, the response (Y) data set and the explanatory (X) data set. Types of asymmetric

analysis include RDA, canonical correspondence analysis (CCA), and linear discriminant analysis (LDA) (Legendre & Legendre 2012). Symmetric analysis, conversely, does not distinguish between independent and dependent variables. Therefore, the two data sets are considered Y_1 and Y_2. Examples of these methods include canonical correlation analysis (CCA), co-inertia analysis (CoIA), and procrustes analysis (PrA) (Legendre & Legendre 2012).

Ordination is also a common technique that is used widely by ecologists (Borcard et al. 1992, Legendre & Legendre 2012). At its most basic, ordination is simply the arrangement of units into some order (Goodall 1954). Bray & Curtis (1957) observed that a rise in the use and application of ordination techniques began in the 1950's. Work such as Motyka et al. (1950), Curtis & McIntosh (1951), Brown & Curtis (1952), Vries (1952), Webb (1954) and Poore (1955) began applying and developing quantitative techniques for plant community classification.

Through the years ordination has been used in forestry as a useful tool in studying the variance of some response, generally vegetation composition, across a landscape (see citation Greig-Smith et al. 1967, Peet 1981, Lahti & Väisänen 1987, Martel et al. 2007, Grimaldi et al. 2014). Carleton et al. (1985) used canonical ordination techniques to explain variance components of forest understory composition. Kolb and Diekmann (2004) used RDA to assess the relative importance of habitat quality, habitat configuration, and habitat continuity for the distribution of forest species. The authors then followed the approach by Borcard et al. (1992) to test effects and partition the variation.

Borcard et al. (1992) proposed a new method, called variation partitioning, based on ordination techniques. This method utilized preexisting methods of asymmetric canonical analysis to partition the variation of a response data matrix into independent components. Borcard et al. (1992) used RDA to partition the variation of Y among two explanatory data sets (X and W) and partial RDA (pRDA) to test the significance of the unique fraction of variation explained by the variables in X while controlling for covariables in W.

Redundancy analysis is the investigation of explained variance (Gittins 1985, Legendre & Legendre 2012). Rao (1964) was the first to describe RDA and it was later rediscovered and presented by van den Wollenberg (1977). Since then, RDA has been applied widely in many different aspects of ecology (Peres-Neto et al. 2006). The search for causes of spatially recognizable structures is of great importance to environmental scientists and RDA is an important tool for this purpose (Peres-Neto et al. 2006). Redundancy analysis provides the means for conducting direct explanatory analysis in which the association among growth measures can be studied according to their shared relationship with environmental and spatial variables (Peres-Neto et al. 2006). Peres-Neto et al. (2006) noted that over 1500 studies have been published applying canonical correspondence analysis (an extension of RDA) and RDA in modeling species-environment relationships.

Borcard et al. (1992) introduced variation partitioning using two sets of explanatory variables. One set of variables was called environmental. The environmental variables were described as measured values that do not describe the spatial positions explicitly. The other set of explanatory variables were spatial and described the spatial

positions of the sampling units (Borcard et al. 1992). Borcard et al. (1992) described

applications of RDA that could be used to partition the variation of observed responses

into components of variation. The partitioning included pure environmental, pure spatial,

spatially-structured environmental, and undetermined components of the variation.

Spatial relationships can be represented in a number of different ways.

Polynomial functions of geographic coordinates can be used as regressors to generate

trend surfaces (Gittins 1968) and have been used in multivariate analysis to model spatial

relationships (Dray et al. 2006). More recent developments include principal coordinates

of neighbor matrices (PCNM) which has provided a new and more powerful way of

studying spatial variation (Borcard & Legendre 2002, Pere-Neto et al. 2006). Dray et al.

(2006) developed the framework of Moran's Eigenvector Maps (MEM's) which is a

generalization of the PCNM approach (Blanchet et al. 2008). MEM's are relatively new

but are becoming quite common even in variation partitioning (Peres-Neto & Legendre

2010).

Hypothesis testing is possible with RDA, because RDA is a linear model that

partitions the variation of **Y** into explained and residual components (Pere-Neto et al.

2006, Borcard et al. 2011, Legendre et al. 2011). In order to test the effects of m

explanatory variables while controlling for the linear effects of q covariables, partial

redundancy analysis (pRDA) can be used (Legendre 2008). In pRDA the linear effects of

the explanatory variables on the response variable are adjusted for the covariables

(Legendre et al. 2011). Permutation F-tests are used in RDA because the response data

are standardized and the error distribution is often non normal (Legendre & Legendre

2012, Borcard et al. 2011). These methods can bring insight into whether or not spatial

variables are related to growth variables of trees. In other words, using pRDA we can test for the effects of *m* environmental variables while controlling for the linear effects of *q* spatial variables (Desdevises et al. 2003).

Cluster analysis is the search for groups in data (Everitt et al. 2011). There are several examples in the literature of cluster analysis with forest data, but a notable lack of literature pertaining to regularly spaced, managed forest. In the 1980's, a pulse of forestry literature utilizing cluster analysis emerged. Lorimer (1985) described how cluster analysis could be used to improve sampling designs to better understand disturbance history in the forest. Guevara et al. (1986) used cluster analysis to examine the spatial analysis of forest succession and habitat patterns for bird species. Applications of cluster analysis remained persistent through the years with focus on forest patch diversity, canopy patterns, or disturbance patterns (see Fraver 1994, Oliveira-Filho & Fontes 2000, Plotkin et al. 2002, Steane et al. 2006). These applications are appropriate as spatial heterogeneity of populations and communities play a central role in many ecological theories, in particular theories of succession, adaption, maintenance of species diversity, community stability, competition, parasitism, epidemics, and natural catastrophes (Legendre & Fortin 1989). There are a great deal of algorithms and methods of clustering so, for brevity, we describe our general approach to clustering by introducing some terms:

Sequential- This algorithm works by applying a recurrent sequence of operations to the objects (trees).

Agglomerative- Agglomerative procedures begin with all objects being considered separate from one and another. This method successively groups the objects into larger and larger clusters until a single all encompassing cluster is obtained.

Hierarchical- Hierarchical methods allocate members of inferior ranking clusters to larger, higher ranking clusters.

Non-Probabilistic- do not use parametric or non-parametric methods for estimating density functions in multivariate space. There are no probabilities linked with the association matrices.

The Lance & William general model (Borcard et al. 2011) for clustering encompasses many agglomerative methods and is easy translatable into computer programs (Legendre & Legendre 2012). The Lance & Williams algorithm can be modified for constrained cluster analysis that looks for clusters made of contiguous objects (Legendre & Legendre 2012). The modified method is based on a similarity (opposed to dissimilarity) matrix whose elements are values that describe how related two objects are. The Lance & Williams algorithm is known as a combinatorial agglomerative method (Legendre & Legendre 2012). This clustering procedure is used widely for classification of landscape and cover types (see Drewa et al. 2002, Urban et al. 2002, Snelder et al. 2004, Perrin et al. 2006, Divíšek et al. 2014). Through the clustering process and analysis one can develop maps of spatially correlated data (Legendre et al. 2009, Legendre & Legendre 2012).

Currently, there are several criteria to determine the appropriate number of clusters in cluster analysis. Among these are the adjusted R^2, AICc, and cross validation residual error (CVRE). The most informative criterion for our needs will be the Calinski-

Harabasz statistic (CH). In one study, Corral (2014) examined the probability of success of these selection criterion and found that the CH statistic had the highest rate of success in detecting the correct number of clusters in simulated plantation stands. Cluster analysis and methods of RDA can be of great practical use to forestry researchers to better understand stand uniformity and productivity.

Greater stand uniformity can lead to greater resource-use efficiency and enhanced productivity (Nilsson & Allen 2003, Binkley et al. 2010, Aspinawall et al. 2011). In one study, Stape et al. (2010) found that uniform stands were 13% more productive than stands with higher levels of heterogeneity. Allen et al. (2005), note that successful intensive silviculture requires an understanding of what resources limit production on temporal and spatial scales. Redundancy analysis is an all inclusive spatial analysis tool which includes the ability to test spatial dependence while controlling for the linear effects of other variables in the model (pRDA), estimating the total amount of pure spatial variability through variation partitioning (RDA), and has been shown by Bougeard et al. (2011) to have better predictive ability than some competing models. Furthermore, since RDA is a multivariate extension of partial linear regression, it is very easy to transition from univariate to multivariate response data (Desdevises et al. 2003). Cluster analysis offers a way to map spatial heterogeneity in stands thereby locating potential areas of varying resources availability.

2.4 Literature Cited

Allen, H. L., Fox, T. R., & Campbell, R. G. (2005). What is ahead for intensive pine plantation silviculture in the south? *Southern Journal of Applied Forestry*, *29*(2), 62-69.

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, *27*(2), 93-115.

Anselin, L., & Griffith, D. A. (1988). Do spatial effects really matter in regression analysis?. *Papers in Regional Science*, *65*(1), 11-34.

Aspinwall, M. J., King, J. S., McKeand, S. E., & Bullock, B. P. (2011). Genetic effects on stand-level uniformity and above-and belowground dry mass production in juvenile loblolly pine. *Forest Ecology and Management*, *262*(4), 609-619.

Binkley, D., Stape, J. L., Bauerle, W. L., & Ryan, M. G. (2010). Explaining growth of individual trees: light interception and efficiency of light use by Eucalyptus at four sites in Brazil. *Forest Ecology and Management*, *259*(9), 1704-1713.

Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology*, *89*(9), 2623-2632.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*(1), 51-68.

Borcard, D., Legendre, P., Avois-Jacquet, C., & Tuomisto, H. (2004). Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, *85*(7), 1826-1832.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, *73*(3), 1045-1055.

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical ecology with R*. Springer.

Bougeard, S., Qannari, E. M., Lupo, C., & Hanafi, M. (2011). From multiblock partial least squares to multiblock redundancy analysis. A continuum approach. *Informatica*, *22*(1), 11-26.

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, *27*(4), 325-349.

Brown, R. T., & Curtis, J. T. (1952). The upland conifer-hardwood forests of northern Wisconsin. *Ecological Monographs*, 217-234.

Bullock, B. P., & Burkhart, H. E. (2005). An evaluation of spatial dependency in juvenile loblolly pine stands using stem diameter. *Forest Science*, *51*(2), 102-108.

Burkhart, H. E., & Gregoire, T. G. (1994). Forest biometrics. *Handbook of Statistics*, *12*, 377-407.

Burkhart, H. E., & Tennent, R. B. (1977). Site index equations for radiata pine in New Zealand. *New Zealand Journal Foresry Science*, *7*(3), 408-416.

Carleton, T. J., Jones, R. K., & Pierpoint, G. (1985). The prediction of understory vegetation by environmental factors for the purpose of site classification in forestry: an example from northern Ontario using residual ordination analysis. *Canadian Journal of Forest Research*, *15*(6), 1099-1108.

Carmean, W. H., & Lenthall, D. J. (1989). Height-growth and site-index curves for jack pine in north central Ontario. *Canadian Journal of Forest Research*, *19*(2), 215-224.

Corral, G. R. (2014). *Investigating Selection Criteria of Constrained Cluster Analysis: Applications in Forestry* (Master Thesis, Virginia Polytechnic Institute and State University).

Curtis, J. T., & McIntosh, R. P. (1951). An upland forest continuum in the prairie-forest border region of Wisconsin. *Ecology*, *32*(3), 476-496.

Dale, M. R., & Fortin, M. J. (2002). Spatial autocorrelation and statistical tests in ecology. *Ecoscience*, 162-167.

Desdevises, Y., Legendre, P., Azouzi, L., & Morand, S. (2003). Quantifying phylogenetically structured environmental variation. *Evolution*, *57*(11), 2647-2652.

Divíšek, J., Chytrý, M., Grulich, V., & Poláková, L. (2014). Landscape classification of the Czech Republic based on the distribution of natural habitats. *Preslia* 86, 209-231.

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3), 483-493.

Drewa, P. B., Platt, W. J., & Moser, E. B. (2002). Community structure along elevation gradients in headwater regions of longleaf pine savannas. *Plant Ecology*, *160*(1), 61-78.

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. Wiley Series in Probability and Statistics.

Ford, E. D., & Diggle, P. J. (1981). Competition for light in a plant monoculture modelled as a spatial stochastic process. *Annals of Botany*, *48*(4), 481-500.

Fortin, M. J., Drapeau, P., & Legendre, P. (1989). Spatial autocorrelation and sampling design in plant ecology. *Vegetatio*, *83*(1-2), 209-222.

Fox, J. C., Ades, P. K., & Bi, H. (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, *154*(1), 261-276.

Fraver, S. (1994). Vegetation responses along edge to interior gradients in the mixed hardwood forests of the Roanoke River Basin, North Carolina. *Conservation Biology*, *8*(3), 822-832.

Gittins, R. (1968). Trend-surface analysis of ecological data. *The Journal of Ecology*, 56(3), 845-869.

Gittins, R. (1985). Canonical correlations and canonical variates. In *Canonical Analysis* (pp. 13-36). Springer Berlin Heidelberg.

Goodall, D. W. (1954). Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Australian Journal of Botany*, *2*(3), 304-324.

Greig-Smith, P., Austin, M. P., & Whitmore, T. C. (1967). The Application of Quantitative Methods to Vegetation Survey: I. Association-Analysis and Principal Component Ordination of Rain Forest. *The Journal of Ecology*, 483-503.

Grimaldi M, Oszwald J, Doleˊdec S, Hurtado MP, Miranda IS, de Sartre XA, de Assis WS, Castan͂eda E, Desjardins T, Dubs F, Guevara E, Gond V, Lima TTS, Marichal R, Michelotti F, Mitja D, Noronha NC, Oliveira MND, Ramirez B, Rodriguez G, Sarrazin M, da Silva Jr ML. (2014). Ecosystem services of regulation and support in Amazonian pioneer fronts: searching for landscape drivers. *Landscape Ecology*, *29*(2), 311-328.

Guevara, S., Purata, S. E., & Van der Maarel, E. (1986). The role of remnant forest trees in tropical secondary succession. *Vegetatio*, *66*(2), 77-84.

Köhl, M., & Gertner, G. (1997). Geostatistics in evaluating forest damage surveys: considerations on methods for describing spatial distributions. *Forest Ecology and Management*, *95*(2), 131-140.

Kolb, A., & Diekmann, M. (2004). Effects of environment, habitat configuration and forest continuity on the distribution of forest plant species. *Journal of Vegetation Science*, *15*(2), 199-208.

Krämer, W. (1980). Finite sample efficiency of ordinary least squares in the linear regression model with autocorrelated errors. *Journal of the American Statistical Association*, *75*(372), 1005-1009.

Lahti, T., & Väisänen, R. A. (1987). Ecological gradients of boreal forests in South Finland: an ordination test of Cajander's forest site type theory. *Vegetatio*, *68*(3), 145-156.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm?. *Ecology*, *74*(6), 1659-1673.

Legendre, P., & Fortin, M. J. (1989). Spatial pattern and ecological analysis. *Vegetatio*, *80*(2), 107-138.

Legendre, P. (2008). Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *Journal of Plant Ecology*, *1*(1), 3-8.

Legendre, P., & Legendre, L. F. (2012). *Numerical Ecology* (Vol. 20). Elsevier.

Legendre, P., Mi, X., Ren, H., Ma, K., Yu, M., Sun, I. F., & He, F. (2009). Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology*, *90*(3), 663-674.

Legendre, P., Oksanen, J., & ter Braak, C. J. (2011). Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, *2*(3), 269-277.

Liu, J., & Burkhart, H. E. (1994a). Modelling inter-and intra-specific competition in loblolly pine (Pinus taeda L.) plantations on cutover, site-prepared lands. *Annals of Botany*, *73*(4), 429-435.

Liu, J., & Burkhart, H. E. (1994b). Spatial characteristics of diameter and total height in juvenile loblolly pine (Pinus taeda L.) plantations. *Forest Science*, *40*(4), 774-786.

Lorimer, C. G. (1985). Methodological considerations in the analysis of forest disturbance history. *Canadian Journal of Forest Research*, *15*(1), 200-213.

Magnussen, S. (1990). Application and comparison of spatial models in analyzing tree-genetics field trials. *Canadian Journal of Forest Research*, *20*(5), 536-546.

Meng, Q., Cieszewski, C. J., Strub, M. R., & Borders, B. E. (2009). Spatial regression modeling of tree height-diameter relationships. *Canadian Journal of Forest Research*, *39*(12), 2283-2293.

Martel, N., Rodriguez, M. A., & Berube, P. (2007). Multi-scale analysis of responses of stream macrobenthos to forestry activities and environmental context. *Freshwater Biology*, *52*(1), 85-97.

Matérn, B. (1960). Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Survey and Other Sampling Investigations. Esselte.

Mateu, J., Uso, J. L., & Montes, F. (1998). The spatial pattern of a forest ecosystem. *Ecological Modelling*, *108*(1), 163-174.

Motyka, J., Dobrzanski, B., & Zawadzki, S. (1950). Preliminary studies on meadows in the south-east of Lublin province. *Ann. Univ. Mariae Curie-Sklodowska 5E.*, 367-447.

Nilsson, U., & Allen, H. L. (2003). Short-and long-term effects of site preparation, fertilization and vegetation control on growth and stand development of planted loblolly pine. *Forest Ecology and Management*, *175*(1), 367-377.

Oliveira-Filho, A. T., & Fontes, M. A. L. (2000). Patterns of Floristic Differentiation among Atlantic Forests in Southeastern Brazil and the Influence of Climate1. *Biotropica*, *32*(4b), 793-810.

Oliver, C. D., & Larson, B. (1996). Forest Stand Dynamics (update edition) Wiley. *New York*.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Sciences*, *21*, 127-144.

Páez, A., & Scott, D. M. (2005). Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, *61*(1), 53-67.

Peet, R. K. (1981). Forest vegetation of the Colorado front range. *Vegetatio*, *45*(1), 3-75.

Penttinen, A., Stoyan, D., & Henttonen, H. M. (1992). Marked point processes in forest statistics. *Forest Science*, *38*(4), 806-824.

Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, *87*(10), 2614-2625.

Peres-Neto, P. R., & Legendre, P. (2010). Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, *19*(2), 174-184.

Perrin, P. M., Martin, J. R., Barron, S. J., & Roche, J. R. (2006, January). A cluster analysis approach to classifying Irish native woodlands. In *Biology & Environment: Proceedings of the Royal Irish Academy* (Vol. 106, No. 3, pp. 261-275). The Royal Irish Academy.

Plotkin, J. B., Chave, J., & Ashton, P. S. (2002). Cluster analysis of spatial patterns in Malaysian tree species. *The American Naturalist*, *160*(5), 629-644.

Poore, M. E. D. (1955). The use of phytosociological methods in ecological investigations: III. Practical application. *The Journal of Ecology*, 606-651.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, 329-358.

Reed, D. D., & Burkhart, H. E. (1985). Spatial autocorrelation of individual tree characteristics in loblolly pine stands. *Forest Science*, *31*(3), 575-587.

Snelder, T. H., Cattanéo, F., Suren, A. M., & Biggs, B. J. (2004). Is the River Environment Classification an improved landscape-scale classification of rivers?. *Journal of the North American Benthological Society*, *23*(3), 580-598.

Stape, J. L., Binkley, D., Ryan, M. G., Fonseca, S., Loos, R. A., Takahashi, E. N., ... & Azevedo, M. R. (2010). The Brazil Eucalyptus Potential Productivity Project: Influence of water, nutrients and stand uniformity on wood production. *Forest Ecology and Management*, *259*(9), 1684-1694.

Steane, D. A., Conod, N., Jones, R. C., Vaillancourt, R. E., & Potts, B. M. (2006). A comparative analysis of population structure of a forest tree, Eucalyptus globulus (Myrtaceae), using microsatellite markers and quantitative traits. *Tree Genetics & Genomes*, *2*(1), 30-38.

Urban, D., Goslee, S., Pierce, K., & Lookingbill, T. (2002). Extending community ecology to landscapes. *Ecoscience*, *9*(2), 200-212.

Wollenberg, A. L. van den (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, *42*(2), 207-219.

Vries, D. D. (1952). Objective combinations of species. *Acta Botanica Neerlandica*, *1*(4), 497-499.

Webb, D. A. (1954). Is the classification of plant communities either possible or desirable? *Bot. Tidsskr*, *51*, 362-370.

Weber, C. D. (1983). *Height growth patterns in a juvenile Douglas-fir stand, effects of planting site, microtopography and lammas occurrence* (Doctoral dissertation, University of Washington).

Zhang, L., Gove, J. H., & Heath, L. S. (2005). Spatial residual analysis of six modeling techniques. *Ecological Modelling*, *186*(2), 154-177.

Zhang, L., Ma, Z., & Guo, L. (2009). An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models. *Forest Science*, *55*(6), 533-548.

Chapter 3

Testing the Efficacy of Statistical Methods to Detect, Quantify, and Map Spatial
Heterogeneity in Simulated Forest Plots Under Ideal Conditions

Abstract

Trees are often treated as individual sampling units and are assumed to be independent of

each other. This assumption is made to meet the ordinary least squares assumption (Fox

et al. 2001). However, researchers have begun to realize that treating individual trees

within a single plot as independent is erroneous (Penttenen et al. 1992, Mateu et al.

1998). Our goal was to reproduce the induced spatial dependence of plot data and

measure the efficacy of spatial analysis techniques to test, quantify and map spatial

heterogeneity. We aim to better understand forest plot heterogeneity with the information

gathered in this study. We designed a range of conditions for plots that included

manipulation of variation, microsite productivity, and spatial pattern complexity.

Simulated populations without confounding effects of competition and mortality were

analyzed using redundancy analysis (RDA) to quantify spatial variability, partial

redundancy analysis (pRDA) to test for spatial dependence, and spatially constrained

cluster analysis was used to map microsites. We found that increased variation in growth

and decreased average tree size differences among microsites resulted in decreased

efficacy of variation partitioning by RDA and decreased efficacy of spatial dependence

testing by pRDA. However, spatial pattern complexity of microsites within simulated

plots resulted in mixed effects in variation partitioning by RDA and tended towards

increased efficacy in spatial dependence testing by pRDA. The results indicated that

RDA and pRDA are reliable methods of analysis and are suitable for more realistic

applications. Spatially constrained cluster analysis produced uncertain results and was heavily affected by changes in variation levels but should also be considered for investigation on more realistic data.

Keywords: Forestry, Simulation, Redundancy analysis, cluster analysis, spatial analysis.

3.1 Introduction

A lack of information exists on dependable methods to detect, quantify, and map spatial heterogeneity in small-scale forest plots (Marignani et al. 2007, Hou et al. 2015). An increased effort to improve plot uniformity has led to questions as to why spatial patterns of tree growth emerge. Spatially dependent growth of trees results from many factors; in particular, genotypic variation and microsite conditions are thought to greatly influence variation of tree structure within a plot (Schweingruber 1987, Downes et al. 2002). Variation in height-age trajectories amongst trees growing together in the same plot but of different genetic stocks is noted in the literature (Buford & Burkhart 1987, Magnussen & Kremer 1993, Tang et al. 2001). Carmean and Lenthall (1988) use the term "polymorphism" to describe that variation in height-age trajectories which have been observed for forest trees growing in areas of contrasting productivity. Likewise, Oliver and Larson (1996) note that polymorphism is likely to occur due to variations in microsite productivity. Both conceptually and mathematically, if non-clonal plots are planted, there will always exist, to some extent, a confounding of genetic and microsite effects on tree growth (Buford & Burkhart 1987). A logical first step in understanding the underlying causes of spatially dependent tree growth is to understand how spatial patterns

emerge in plots and to quantify those effects. In order to investigate spatial effects on tree growth we developed a method to simulate forest plots, then tested the efficacy of statistical methods to detect, quantify and map spatial heterogeneity among the plots.

Simulations based on model forests are used by scientists to gain greater insights into a host of forestry issues (Tokolo & Shrestha 1999, Tommpo 2006, Hou et al. 2015). A large number of publications utilize simulations to better understand treatment outcomes, economic feasibility, and carbon sequestration with respect to climate change (Betts et al. 1997, Galbraith et al. 2010, Haynes et al. 1994, Huntingford et al. 2013, Prentice 1993, Scheller & Mladenoff 2005, Solomon 1986, Webster et al. 2003). Furthermore, simulations have been used by environmental scientist to better understand, compare, and verify statistical methods to be applied in the field (i.e Borcard & Legendre 2002, Legendre et al. 2011, Peres-Neto et al. 2006, Peres-Neto & Legendre 2010, Hou et al. 2015). Simulations offer a way of understanding future scenarios so that uncertainty can be minimized and management can adapt. One can use simulations to control environmental and spatial factors such that statistical methods can be applied in field-like conditions with the benefit of knowing actual plot conditions.

Induced spatial dependency can be produced by functional dependence of the response variables on explanatory variables that are themselves spatially correlated (Borcard & Legendre 2002). In our study the response variables were the height and diameter of each tree and the explanatory variable is the site index (a measure of site productivity). We simulated plots with imbedded patterns of microsite productivity and, therefore, there is a functional dependence of height and diameter on the site index values that are themselves spatially correlated. In our study, it is assumed that **Y** (height and

diameter) has acquired the spatial structure of **X** (site index). If all important spatially-structured explanatory variables are included in the analysis, equation (1) correctly accounts for the spatial structure induced in **Y.**

$$y_i = f(X_j) + \varepsilon_j \tag{1}$$

In equation (1), y_i is the value of the dependent variable y at site j and ε_j is the error term whose value is independent from site to site (Dray et al. 2012, Dale & Fortin 2014). If induced spatial dependency is occurring, then it should be properly accounted for in the model in order to perform correct standard statistical tests of the relationship between **Y** and **X**. The first step to accounting for spatial variability in a model is to test for its presence.

Univariate methods of testing for spatial dependency are well known and documented. Geary's c (Geary 1954) and Moran's I (Moran 1950) coefficient are among the most popular methods for detecting presence of spatial autocorrelation. Coefficients like Geary's c and Moran's I examine the spatial correlation among the error components of the observed data (Legendre & Legendre 2012). Geary's c coefficient varies from 0 to some unspecified number larger than 1. Positive spatial correlation produces c values from 0 to 1 and negative spatial correlation produces C values greater than 1. Moran's I formula is related to the Pearson correlation coefficient in that the numerator is a covariance, comparing the values at all pairs of points in turn, while its denominator is the maximum likelihood estimator of the variance. Unlike Moran's I, Pearson's correlation coefficient contains the same number of terms in the sums of the numerator and denominator (Legendre & Legendre 2012). Moran's I coefficients can also be used spatial modeling to develop spatial variables (Griffith & Peres-Neto 2006).

Testing of spatial variables requires that the spatial structure be expressed by meaningful spatial variables. The most straightforward method for modeling spatial relationships is polynomial regression where the geographic coordinates are used to generate trend surfaces (Legendre 1990, Borcard & Legendre 2002). However, trend surface is only satisfactory when the sampling area is approximately homogenous, the sampling design is closely regular, the number of spatial locations is "reasonable" (Norcliffe 1969), and the spatial structure to be modeled is somewhat simple (Dray et al. 2006). A more modern method of creating meaningful spatial variables is Moran's eigenvector maps (MEM's) (Dray et al. 2006). Moran eigenvector maps rely on eigenvector decomposition of truncated geographic distance matrix (Griffith & Peres-Neto 2006).

Redundancy analysis (RDA) (Rao 1964) is a method that extends multiple linear regression, to multivariate linear regression involving multiple response variables and a common matrix on predictors (Peres-Neto & Legendre 2010). Variation partitioning for RDA with multivariate response and two sets of predictor matrices (X for environmental factors and W for spatial) are straightforward and applications for multiple matrices are possible (Peres-Neto & Legendre 2010). Methods such as RDA utilize the spatial variables created with MEM's to estimate the contribution of spatial variation in the response data and filter out the effects of spatial correlation when testing the importance of ecological factors (Peres-Neto & Legendre 2010). RDA is well known to statisticians and ecologist alike.

Simple RDA's of response Y_{nxp} (n objects with p variables) by an explanatory matrix X_{nxm} (n objects and m variables) and W_{nxr} (n objects and r variables) is done to

partition the variation of the response data into components (Borcard et al. 1992). The general steps of simple RDA are to preform a multivariate regression on the centered values of Y by X and obtain the matrix of fitted values $\hat{Y}$. The second step of RDA is a principal components analysis (PCA) of $\hat{Y}$ (Laliberté et al. 2009). Akin to the fitted values in multiple regression, the canonical axes produced by the PCA are linear combinations of the explanatory variables in X. The canonical axes being linear combinations of the explanatory matrix is a fundamental property of RDA (ter Braak & Prentice 1988). The Bimultivariate Redundancy Statistic (RDA R-square) is used to quantify the variation of each explanatory matrix (Peres-Neto et al 2006). Simple RDA and partial redundancy analysis (pRDA) are often used concurrently to partition the variance of response data into components and to test the significance of those components in the presence of other covariables, respectively (Legendre et al 2011.).

Partial RDA can test an individual explanatory matrix while controlling for the linear effects of a second matrix containing covariables (Legendre et al 2011). In pRDA we are able to isolate and test for pure effects (ter Braak & Smilauer 2002). The most compelling function of pRDA is that one can test for pure effects of spatial, environmental, chemical, treatment, and other components while controlling for the linear effects of other covariables (Legendre et al 2011).

Testing significance in pRDA is done by permutation F-tests (Legendre 2008). We permuted the residuals of the reduced models. In this approach, the matrix of fitted values and the matrix of residuals are computed from Y, on a matrix of covariables (W). The rows of the matrix of residuals are permuted, which produces a new matrix of residuals. Then the matrix of permuted Y is obtained by fitting values to the matrix of

permuted residuals (Anderson & Legendre 1999). F – statistics are then collected and this process is repeated to determine extreme values. Figure 3.1 illustrates how RDA quantifies variability from multiple explanatory matrices. After RDA and pRDA are performed and a significant spatial relationship is formed, mapping techniques can be used to capture the spatial variability of each plot.
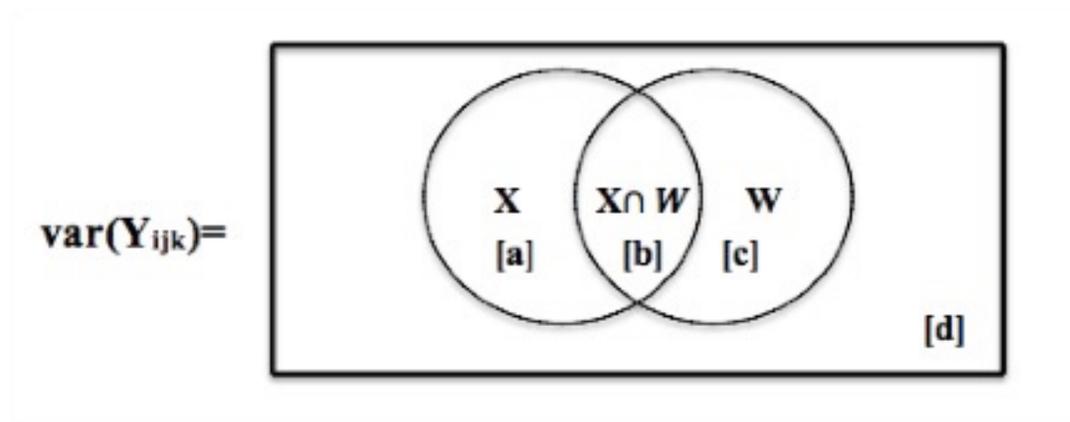
Figure 3.1 Illustrates the variation partitioning by RDA. Here, **X** and **W** are two explanatory matrices and [a], [b], [c], and [d] are fractions of variance. If matrix **X** is a matrix of environmental variables and **W** is a matrix of spatial variables then the fraction [a] is pure environmental variability, [b] is confounded variability of environmental and spatial, [c] is pure spatial, and [d] is the residual component. Fractions [a] and [c] are considered testable fractions and significance tests can be carried out by pRDA.

Cluster analysis is the search for groups in data (Everitt et al. 2011). Clustering is a way to classify objects into groups based on some similarity or dissimilarity (Murtagh & Legendre 2011). The classification of objects is usually done based on some partitioning criterion. A partition is the division of objects of a set into subsets, such that each object belongs only to one subset. Exceptions exist in cluster analysis and theory because cluster analysis is not a set procedure with formal statistical testing but rather a broad topic with many interpretations and related statistics. One such exception is the classification of objects into one and only one subset. Fuzzy clustering allows objects to

be members to more than one subset at a time. This paper focuses on a method of hard clustering where each object belongs to only one group at a time. We used a method of agglomerative, spatially-constrained cluster analysis described by Legendre and Legendre (2012).

A first step to clustering is deciding on a distance measure to use for the clustering method. Depending on the type of algorithm and data used, practitioners choose between similarity or dissimilarity measures. Similarity and dissimilarity measures measure how quantitatively akin or contrasting each pair of objects is to each other (Murtagh & Legendre 2011). In the clustering algorithm agglomerative clustering was applied where each object is grouped into larger more inclusive groups based on which objects are least dissimilar.

Agglomerative styles of clustering describe a category of algorithms, which assign n objects to be clustered to k groups. After each iteration of the algorithm, objects are assigned to larger, more inclusive groups. The number of groups (k) after each iteration approaches min(k)=2 groups. This iterative process is what defines sequential clustering algorithms. Constraints can be imbedded within the algorithm to account for successive temporal relationship and spatial constraints.

There are a number of ecological reasons why practitioners may want to group objects together if and only if they occur in succession to each other (in time series) or if objects in each group are neighbors to each other in space. We used spatial constraints such that the resulting map groups together trees growing under similar conditions. This makes the map more interpretable for mapping spatial heterogeneity of tree growth resulting from microsites.

The objectives of this chapter are to measure the efficacy of RDA. pRDA and spatially constrained cluster analysis to quantify spatial variability, detect spatial variability, and map microsites. To achieve these objectives, we will use site index as a single explanatory variable and use MEM's to construct spatial variables.

3.2 Methods

3.2.1 Study material and simulated data

This study simulates 5 hypothetical plots, each with a unique microsite design (Figure 3.2). The simulation produces tree height and diameter measurements similar to loblolly pine (*Pinus taeda L.*). Each plot consisted of 25 rows and 25 columns of simulated trees. Trees are "planted" on 10-foot square spacing on ~1.43 acre plots. Each plot was assigned 1 of 5 microsite patterns. Microsite arrangements are distinct formations of high or low areas of productivity within each plot.
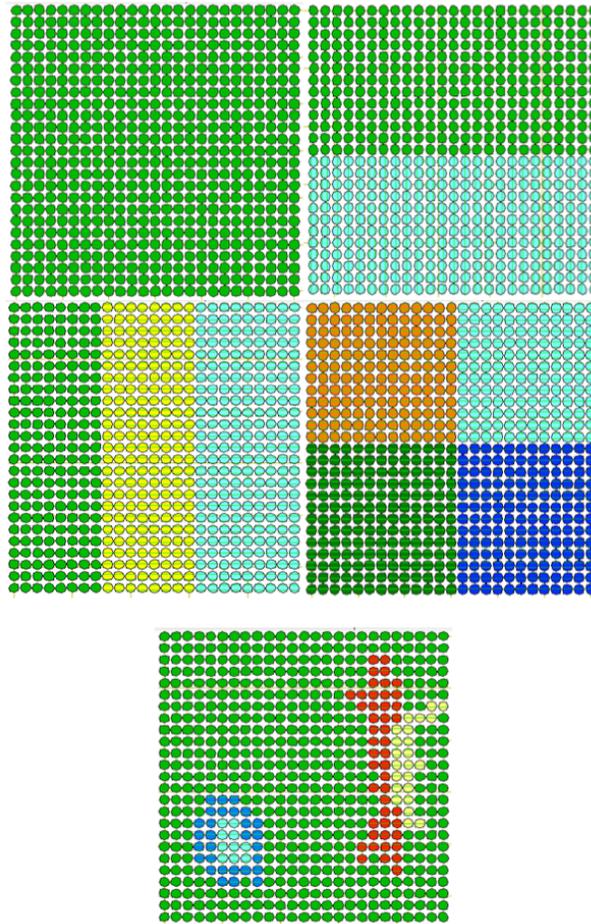
Figure 3.2 Illustrates the 5 spatial patterns used in this study Microsites are distinguished by color. For left to right: Control, biplot, triplot, quadplot, and freeplot.

Plots were developed to represent a range of possible microsite patterns and most importantly, to induce spatial dependence of tree growth in a increasingly complex fashion. By creating areas of high and low productivity, the size of each tree is related to its location in space. There are 1 to 5 microsites per plot represented by the control (uniform productivity), biplot (2 microsites), triplot (3 microsites), quad plot (4 microsites), and the free plot (5 microsites). All 5 site patterns are "grown" without the confounding effects of competition and mortality based spatial dependence.

The first step to developing plot conditions is to assign microsite productivity

(*Refer to Figure 3.2*)

Control- The control plot has a single site index value because of microsite homogeneity. The SI value is 70 feet average dominant height at base age 25.

Biplot- There are two microsites on the biplot, 70 feet average dominant height at base age 25 (green) and 80 feet average dominant height at base age 25 (blue).

Triplot- There are three microsites of the triplot have a SI value of 60 feet average dominant height at base age 25 (yellow), 70 average dominant height feet at base age 25 (green), or 80 feet average dominant height at base age 25 (blue).

Quadplot- The quadplot has four microsites with SI value of 50 feet average dominant height at base age 25 (orange), 60 feet average dominant height at base age 25 (yellow), 70 feet average dominant height at base age 25 (green), or 80 feet average dominant height at base age 25 (blue).

Free plot- The free plot has five microsites. The large green area of the free plot is SI 70 feet average dominant height at base age 25. The small light blue area is SI 60 feet average dominant height at base age 25, the dark blue area circumscribing the blue area is SI 65 average dominant height feet at base age 25. The red area is SI 75 feet average dominant height at base age 25and the yellow area is SI 80 feet average dominant height at base age 25.

All calculations and simulations were done using R software. Since SI values are averages, we can draw SI values for each individual tree, microsite, and simulate from an assumed distribution. Assuming normality we state: $SI_{ijk} \sim N(\mu_j, \sigma_j^2)$ where the *i*th simulation ranges from i=1…S, the *j*th microsite ranges from j=1…k, and the *k*th tree ranges from 1=1…K. For estimating SI values we hold this condition true for all

microsites: $CV_j = 0.10 = (\sigma_j \div \mu_j) * (100\%)$ where μ_j and CV_j are fixed and known

for each microsite. We can then use SI as our environmental variable $X_{ijk}=[SI_{ijk}]$. For the

trees to reflect the spatial dependence (size of tree dependent on its location in space)

caused by SI fluctuations within plots we imposed different mean diameter at breast

height (DBH) values for each microsite.

For each microsite in a plot we draw from a normal distribution of a

specified mean DBH and coefficient of variation (CV). For K microsites we have $\mu_1 <$

$\mu_2 < \cdots < \mu_K$, where μ_K is the mean DBH for the Kth microsite. For each simulation the

CV was equal among all microsites. That is, $DBH_{ijk} \sim N(\mu_j, \sigma^2)$, where σ^2 is chosen to

achieve a prespecified value of $CV = (\sigma \div \mu) * (100\%)$. Table 3.1 displays all 15

possible combinations examined for each spatial pattern. It should be noted, however,

that the control plot only has 5 possible combinations (5 levels of CV 5%-25% and 0

levels of size difference) because there are no differences in mean DBH specified. As an

example of treatment application, we can look at the biplot. In Table 3.1 the cell where

CV=10% and difference in mean DBH=1 in. intersect, represents 1 of the 15 scenarios to

be used in simulations for the biplot. In this example, the green portion of the biplot (see

Figure 3.2) will contain trees with DBH values drawn from a normal distribution, N(5,

0.25), and the blue portion will contain trees with DBH values drawn from the

distribution N(6, 0.36). These two distributions reflect the difference in mean DBH=1 in

between the two microsites and variance calculations based on a 10% CV.

Table 3.1 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%)

From DBH values we can then estimate total tree height. Heights are calculated based on established diameter-height relationships. Equation (2) from Sabatia and Burkhart (2013) was used to model heights.

$$H_{ijk} = \frac{\beta_0 e^{\beta_1 DBH_{ijk}^{-1}} + \theta_{ijk}}{3.28} \tag{2}$$

The parameter estimates from Sabatia and Burkhart (2013) are from a loblolly pine study of similar planting density. Where H_{ijk} is the total height (feet) and DBH_{ijk} is the diameter at breast height (inches) of the kth tree in the jth microsite of the ith simulation. The index values range from, i=1…S, j=1…k, and k=1…K. For equation (2), $\boldsymbol{\beta}_0$ is the upper asymptote parameter, $\boldsymbol{\beta}_1$ is the rate parameter, and $\boldsymbol{\theta}_{ijk}$ is the random error due to the kth tree [$\boldsymbol{\theta}_{ijk} \sim N(0, \sigma_\theta^2)$] (Sabatia & Burkhart 2013). The estimated equation from Sabatia and Burkhart (2013) is then:

$$H_{ijk} = \frac{20.382 e^{-7.309 DBH_{ijk}^{-1}}}{3.28} + \boldsymbol{\theta}_{ijk} , \ \boldsymbol{\theta}_{ijk} \sim N(0, 0.5186) \tag{3}$$

We now have our response matrix $Y_{ijk}=[DBH_{ijk}, H_{ijk}]$.

The last data we compute are the spatial variables. We simulated trees on a 10ft by 10ft planting grid and collected spatial information based on an (X,Y) coordinate system. The tree in the top left position of the plot (origin position) has spatial coordinates (0,0), the tree to the immediate right of the origin has coordinates (10,0), the tree immediately beneath the origin and to the right one position has coordinates (10,10). A pairwise matrix of distances (D_{ij}) constructed. D_{is} contains the distances between each tree. A threshold value is chosen to truncate matrix D_{ij}. We used R software package PCNM() (Legendre et al. 2012) for this analysis with dbMEM option which use the longest edge of the minimum spanning tree (Legendre et al. 2012), which in our case is 10ft. The following rule is then used to truncate D_{ij} :

$$D_{trunc(ij)} \begin{cases} \boldsymbol{D_{trunc\,(ij)} = D_{ij},} & \boldsymbol{if \quad D_{ij} \leq threshold} \\ \boldsymbol{D_{trunc\,(ij)} = 4 \times thresh} & \boldsymbol{if \qquad\qquad i = j} \\ \boldsymbol{D_{trunc\,(ij)} = 4 \times threshold} & \boldsymbol{if \quad D_{ij} \geq Threshold} \end{cases}$$

Next, a principal coordinate analysis (PCoA) is performed on $D_{trunc(ij)}$. The resulting principal coordinates are spatial eigenfunctions that model multiscale distance relationships among trees within each plot (Dray et al. 2006). We retain only the eigenvectors with positive eigenvalues (spatial variables) because they model positive spatial autocorrelation. Lastly, a forward selection of the positive spatial variables is done to select significant spatial variables which the compose the matrix W.

It should be noted that each iteration of the simulation process examines a "new" plot. For example, the biplot was simulated thousands of times throughout this study.

Each iteration of the biplot holds only microsite pattern and spatial location of trees constant, but the trees themselves take on unique values for every simulation.

3.2.2 Redundancy Analysis

Redundancy analysis was done using R software's "VEGAN" package (Oksanen et al. 2013). Redundancy analysis is synonymous with explained variance (Legendre & Legendre 2012). The first step to RDA is to standardize the response variables. The response variables mentioned hereafter have been standardized. This symmetric form of analysis utilizes a response matrix Y_{625x2} (hereafter Y) with explanatory vector X_{625x1} (hereafter X) and covariables W_{625x2} (hereafter W). In RDA, the ordination axes are obtained by a PCA of $\widehat{Y}$, which is computed by fitting the Y variables to X by multivariate linear regression. An important characteristic of the RDA process is that the ordination of Y produces ordination axes that are linear combinations of X. Multiple linear regression on all variables in X is done for each variable in Y and $\boldsymbol{\beta} = [\boldsymbol{X^t X}]^{-1} \boldsymbol{X^t Y}$ is calculated (Peres-Neto et al. 2006).

The canonical R^2, called the bimultivariate redundancy statistic by Miller & Farr (1971), quantifies the strength of the linear relationship between variables Y and X, where $\widehat{\mathbf{Y}}$ are the multivariate estimated values of the response:

$$R^2_{Y|X} = \frac{ss(\widehat{Y})}{ss(Y)} \tag{4}$$

In equation (4), $SS(\hat{Y})$ is the total sum of squares of $\hat{Y}$ and $SS(Y)$ is the total sum of squares of Y. Using the well known adjusted R^2 (rR^2) from Ezekiel (1930), we use the equation 5 (Peres-Neto et al. 2006):

$$rR^2 = 1 - \left(1 - R_{Y|X}^2\right)\frac{n-1}{(n-m-1)} \tag{5}$$

Referring to Figure 3.1, we can begin to calculate the fractional explained variance components (a,b, and c) as well as the unexplained (d) by following steps outlined by Legendre (2008).

1. Perform the canonical analysis of Y with respect to X. Use equations 4 and 5 to calculate the $R_{Y|X}^2$ and rR^2. The rR^2 corresponds to the circle on the left side of Figure 3.1. It contains fractions [a] and [b].

2. Perform the canonical analysis of Y with respect to W. Use equations 4 and 5 to calculate the $R_{Y|W}^2$ and rR^2. The rR^2 corresponds to the circle on the right side of Figure 3.1. It contains fractions [b] and [c].

3. Perform the canonical analysis of Y with respect to X and W. Use equations 4 and 5 to calculate the $R_{Y|X,W}^2$ and rR^2. The rR^2 corresponds to both circles in Figure 3.1. It contains fractions [a], [b], and [c].

4. From these results we can compute [b]=[a+b]+[b+c]-[a+b+c].

5. Next, compute [a]=[a+b]-[b]

6. Next, compute [c]=[b+c]-[b]

7. Next, compute [d], which represents the residual component of variance, [d]=1-[a+b+c].

Variation partitioning of Y results from three simple RDA's each with a different independent matrix. For example, we used X , W, and X|W to estimate variation from environmental, spatial, and environmental and spatial variables, respectively.

We followed this procedure for each iteration of our simulation. Referring to Table 3.1, there are 5 scenarios for the control plot (only changes in CV can be used since there are no microsites) and 15 combinations of CV and difference in mean DBH for the biplot, triplot, quadplot, and free plots. For each of these scenarios 1000 simulations were run and the rR^2 values were collected. After the 1000 simulation runs an estimate of pure spatial variation is obtained for given spatial patterns and scenarios.

3.2.3 Partial Redundancy Analysis

Partial redundancy analysis was done using R software's "VEGAN" package (Oksanen et al. 2013). Partial redundancy analysis is the partial canonical redundancy analysis of response matrix Y on matrix X while controlling for the linear effect of matrix W of covariables. Matrices X and W are interchangeable, depending on which partial variance is to be tested. The same F-statistic used in partial linear regression (PLR) can be used in pRDA. For our specific objectives, we tracked the significance testing of Y~W|X. This is the hypothesis of spatial dependence and tests the significance of fraction [c], representing pure spatial variation. A two-stage approach to testing the significance of the pure spatial component was taken. First, the software calculated the R^2 statistic of partial regression:

$$R^2_{Y \sim W|X} = \frac{SS(Y \sim W|X)}{SS(Y \sim W)} \qquad (6)$$

Following this calculation, the F-statistic was used to test the overall significance of the

partial regression relationship is:

$$F = \frac{R^2_{Y \sim W|X}}{m} \cdot \frac{1 - R^2_{Y \sim W|X}}{n - 1 - m - q} \qquad (7)$$

Significance of the F-statistic may be tested with the F-distribution if the assumption of

normality of the residuals holds and the data are standardized, however this is rarely the

case in many ecological studies and for this reason permutation tests are preferred and are

used in this study. Specifically, we used a permutation of the residuals of the reduced

model (R software default options VEGAN'S anova.cca) (Oksanen et al 2013).

The null hypothesis for the partial F-test of X|W states that the response is not

spatially dependent. For each of the spatial patterns all applicable scenarios in Table 3.1

were simulated 1000 times, and for each iteration of the simulation a permutation F-test

was performed (999 permutations per test). Each iteration of the simulation was

considered a Bernoulli trial and assigned a 1 if the null hypothesis was rejected and 0

otherwise. After simulations, the efficacy of pRDA in detecting spatial dependency was

estimated by calculating the probability of success (p) as the average of all trials. Under

these conditions, the central limit theorem provides the distribution of p:

$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$. From this property we can calculate confidence intervals of the

probability of detecting spatial dependency for each spatial pattern and scenario.

3.2.4 Cluster Analysis

The clustering algorithms used are part of R software package "const.clust" (Legendre 2011). 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, and the second is to pick a connection method and build a contiguity matrix.

Distance metrics are used to measure the association between two objects (trees). The smaller a distance value or closer it is to zero the more related objects are structurally. In our data, two trees that are identical would have a distance value of 0. The most common metric measure is the Euclidean distance (Legendre & Legendre 2012). We used Euclidean distances (a measure if dissimilarity) (eq. 8) among objects using non-geographic information to create our dissimilarity matrix (D).

$$D(\boldsymbol{y_r}, \boldsymbol{y_{r+1}}) = \sqrt{\sum_{c=1}^{p}(\boldsymbol{y_{rc}} - \boldsymbol{y_{(r+1)c}})^2} \qquad (8)$$

For equation (8), r=row of matrix Y, c=column of matrix Y, and p is the number of variables in matrix Y. For one of our simulations, r=1…625, and p=2. This step is typical in many clustering algorithms, but in the next steps we impose spatial constraints on the dissimilarity matrix, which is information typically not incorporated into clustering analysis.

Prior to performing spatially constrained clustering it is important to state which trees are neighbors in space. In order for trees to be grouped together they have to be neighbors in space. The only admissible clusters in a spatially constrained analysis are those that obey the contiguity scheme. We equate clusters to microsites by constraining clusters so they are spatially defined in the same way as a microsite. A cluster is then a contiguous patch or group of trees that are structurally unlike surrounding trees. Microsites create spatially recognizable patterns in tree attributes due to contrasts in soil productivity. This is the link from clusters to microsite and because of this we expect cluster locations to parallel microsite locations.

The Delaunay triangulation uses spatial coordinates to identify neighbors and to define contiguity. To determine neighbors, we produce a list of connection edges to create a contiguity matrix containing 1's for connected and 0's elsewhere (based on spatial coordinates of plot map). The contiguity matrix is how we spatially constrain the cluster analysis. The 1's and 0's are used to define neighbors and create connections among the trees. The Delaunay triangulation method states that for any triplet of non-collinear points A, B, and C, the three edges connecting these points are included if and only if the circle passing through these points (Figure 3.3) includes no other point (Legendre & Legendre 2012). This criterion is a robust method for defining contiguity. This connection scheme works well with regular grids and is adaptable to various patterns of planting grids and will transfer well to real plots that are slightly irregular.
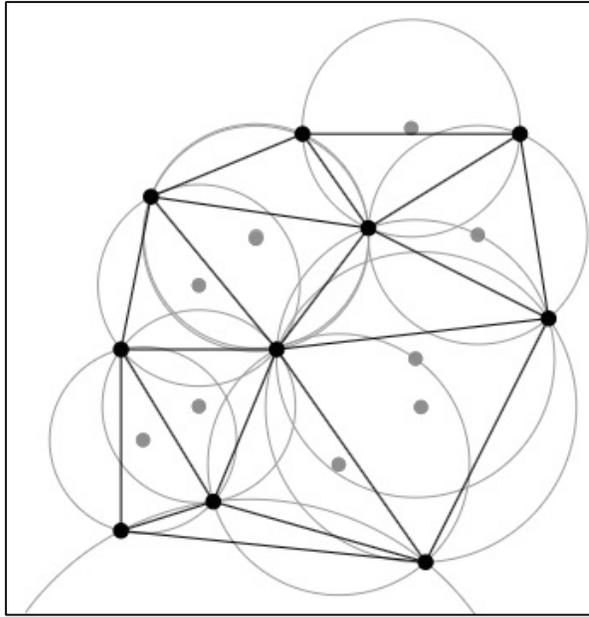
Figure 3.3 Illustrates the circles used in Delaunay triangulation. Black dots represent the objects (trees in our data), red dots represent the center of each circle used to circumscribe three points, and the thick black lines connect "neighbors" (Wikipedia, November 3rd, 2014).

The spatial constraint allows only connected trees to be clustered together. This prevents a scattering of cluster assignments on the map. Instead, clusters form distinct clumps. The cluster analysis results can be mapped with the spatial coordinates of the trees. The resulting map shows cluster assignment of each tree. When compared to our map of known microsites, we expect clusters to form over microsites and for trees within a microsite to be assigned the same cluster number.

The Hadamard product between the dissimilarity matrix and contiguity matrix creates a constrained dissimilarity matrix where distance values exist only where neighbors were previously defined by the contiguity matrix. Our data file consists of growth information on 625 trees. A 625x625 dissimilarity matrix is created from equations (6). The more akin any two trees are in structure the closer to 0 is their dissimilarity value. We then create a 625x625 contiguity matrix of 1's and 0's where 1's

mark neighbors as defined by Delaunay triangulation and 0's elsewhere. The Hadamard product for our data is the dissimilarity in growth among neighboring trees.

We used the Calinski-Harabasz (CH) statistics to estimate the number of groups in the data. Corral (2014) found that the CH statistic performed well when estimating the number of groups in a forestry setting. This process was iterated 100 times for each combination of spatial pattern and parameter values. After the iteration process was completed, we examined if the CH statistic correctly identified the number of groups in each the spatial pattern (2 for biplot, 3 for triplot, 4 for quadplot, and 5 for free plot). If the number of groups was correctly identified a 1 was saved, otherwise a 0 was saved. Once the simulations were finished we estimated the probability of correctly identifying the number of microsites in each pattern under all 15 scenarios. After this was complete we ran misclassification simulations.

The cluster map should reflect the microsite map if each tree is assigned correctly from the clustering algorithm. This is a simple process where we count 1 if tree i is incorrectly grouped into a cluster and 0 otherwise. Probability of misclassification for any tree can then be estimated at the end of simulations. We ran 100 misclassification simulations for each scenario that has a probability of successfully detecting the correct number of groups greater than 80%, which was chosen as the threshold for reliability.

3.3 Results

3.3.1 Redundancy Analysis

We used RDA to perform a variation partitioning to estimate the pure spatial variability. Table 3.2 contains the estimated rR^2 value associated with the pure spatial component. As discussed earlier, there are 15 scenarios per spatial pattern. The scenarios

are a combination of differences in mean DBH and CV. Table 3.1 contains all possible

combinations of CV and difference in mean DBH. Notably, the rR^2 values for the control

plot are 12-13%. The largest rR^2 values are in the quadplot which is one of the more

complex patterns. More variation tends to be captured in scenarios with larger differences

in DBH and lower values of CV. No obvious patterns emerge in the amount of variation

captured across the 5 different microsite patterns.

Table 3.2 Estimated rR^2 for each spatial pattern and all scenarios.

Coefficient of Variation	Control	Biplot	Triplot	Quadplot	Free	
5%	13%	6%	46%	24%	24%	0.5 Inch
10%	13%	10%	29%	19%	20%	
15%	13%	11%	22%	16%	16%	
20%	13%	12%	18%	15%	15%	
25%	12%	13%	16%	14%	13%	
Coefficeint of Variaiton	Control	Biplot	Triplot	Quadplot	Free	
5%	13%	3%	62%	76%	22%	
10%	13%	7%	52%	67%	20%	1 Inch
15%	13%	9%	41%	55%	18%	
20%	13%	10%	33%	44%	16%	
25%	12%	12%	27%	36%	15%	
Coefficeint of Variaiton	Control	Biplot	Triplot	Quadplot	Free	
5%	13%	0%	68%	76%	20%	2 Inch
10%	13%	3%	63%	72%	19%	
15%	13%	5%	55%	66%	19%	
20%	13%	7%	47%	59%	18%	
25%	12%	8%	40%	50%	17%	

3.3.2 Partial Redundancy Analysis

Partial redundancy analysis was used to detect spatial dependence. The results

reported are the probability of detecting spatial dependence for all combinations of spatial

patterns and scenarios. Table 3.3 contains the results from the pRDA simulations.

Table 3.3 Illustrates the results from pRDA. The values represent the probability of detecting spatial dependence.

Coefficient of Variation	Control	Biplot	Triplot	Quadplot	Free	
5%	4%	100%	100%	99%	100%	0.5 Inch
10%	6%	88%	86%	99%	100%	
15%	4%	77%	51%	90%	100%	
20%	6%	58%	39%	80%	100%	
25%	4%	51%	31%	80%	100%	
Coefficeint of Variaiton	Control	Biplot	Triplot	Quadplot	Free	
5%	4%	100%	100%	100%	100%	
10%	6%	100%	98%	99%	100%	1 Inch
15%	4%	96%	92%	98%	100%	
20%	6%	93%	85%	86%	100%	
25%	4%	86%	56%	77%	100%	
Coefficeint of Variaiton	Control	Biplot	Triplot	Quadplot	Free	
5%	4%	100%	100%	100%	100%	
10%	6%	100%	100%	100%	100%	2 Inch
15%	4%	100%	100%	100%	100%	
20%	6%	99%	97%	98%	100%	
25%	4%	100%	83%	92%	100%	

The highest probabilities for detection are for the free plot and with larger differences in

DBH among microsites. The probability for detecting spatial dependence in the control

plot is consistently at approximately 5%. 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.

3.3.3 Cluster Analysis

Cluster analysis was performed to examine the efficacy in detecting the number of

microsites in our plots and to correctly allocate each tree to the correct microsite.

Spatially constrained cluster analysis was used to estimate the number of microsites and

to create maps. Next, we ran misclassification simulations with the results from the

cluster analysis and compared those results to the actual data to estimate probability of

misclassification.

Table 4 illustrates the probability of detecting the correct number of microsites for all combinations of spatial patterns and CV. The control pattern is excluded because it is homogeneous and the clustering algorithm does not test for k=1 group in the data. Homogeneous plots were 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.

In Table 4 the probabilities for detecting the correct number of microsites are given. A threshold for reliability was set at 80% success. Therefore, in all cells of Table 4 where the probability is greater than or equal to 80% misclassification simulations were run. Table 5 shows the probability of misclassifying any given tree when using constrained cluster analysis. Values are generally low, but range from approximately 0% to 32%.

Table 3.4 Illustrates the probability of correctly identifying the number of microsites in all combinations of spatial patterns and scenarios.

Coefficient of Variation	BIPLOT	TRIPLOT	QUADPLOT	FREE	
5%	90%	52%	28%	0%	
10%	80%	22%	0%	0%	0.5 inch
15%	68%	26%	4%	4%	
20%	58%	28%	4%	2%	
25%	50%	10%	0%	0%	
Coefficient of Variation	BIPLOT	TRIPLOT	QUADPLOT	FREE	
5%	100%	100%	100%	0%	
10%	97%	84%	14%	27%	1 inch
15%	80%	12%	0%	14%	
20%	74%	12%	0%	2%	
25%	70%	10%	0%	0%	
Coefficient of Variation	BIPLOT	TRIPLOT	QUADPLOT	FREE	
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 3.5 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%. (*) indicates a combination where the probability from Table 4 was less than 80%.

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

3.4 Discussion

3.4.1 Redundancy Analysis

The variation partitioning and quantification of spatial variability performed as expected. Referring to Table 3.2, we see there is no distinct pattern in the amount of spatial variation captured. This confirms that spatial pattern complexity and perhaps the shape of microsites can influence the efficacy of RDA to capture spatial variability. Even if there are no obvious patterns in how rR^2 changes with spatial pattern complexity, one can still see that large differences among spatial pattern exist in rR^2. It is possible that the spatial processes, in the biplots for example, show fortuitous correlations with the

environmental variable (Bell et al. 2006). The rR^2 of the control plots ranges from 12-13%. This observed spatial variation is simulated by chance, but it happens to reflect realistic measures of variability in a random or aggregated tree growth. Lepš and Kindlmann (1987) noted in their study that it is incorrect to deduce the independence of individuals within a population from an observed random pattern. The results from this portion of the simulation study were very promising. The values from Table 3.2 are rR^2 values and we expected them to be rather low (<10%), but found values as high as 76%. The largest portions of variation explained by the spatial component were found in the scenarios with the greatest difference in mean DBH and in more complex spatial patterns.

The difference in mean DBH among microsites was examined at three levels (0.5in, 1in, 2in). As the difference in mean DBH increases among microsites, the rR^2 values increase as well. This trend is intuitive. By implementing larger differences in mean DBH we are essentially creating stronger and more obvious spatial dependencies. Redundancy analysis is detecting the stronger relationship between size and spatial location. Coefficient of variation is very influential in how much spatial variability is captured.

In Table 3.2, as CV increase the rR^2 decreases. This is an expected effect. The amount of variation used to draw DBH values for each microsite blurs the line, so to speak, between microsites. Graphing the DBH distributions of each microsite simultaneously shows increased overlap of the distributions with increased values ofCV. Conceptually, as overlap among the distributions increases there is an increase towards uniformity and thus spatial dependencies become less obvious.

3.4.2 Partial Redundancy Analysis

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. These results are promising with many scenarios showing greater than 80% success rate and some with 100% success rate. These results are promising because we observed higher success rates in the most complex spatial pattern which contained smaller more frequent microsites. This may be the most realistic spatial pattern. Palmer (1980), who quantified spatial patterns of plant environment relationships in hardwood plots. He found that most of the spatial dependence measured in his variables were at small scales (within 10 meter subplots). Our results indicate that spatial patterns with smaller, more frequent microsites more often had significant spatial variables. The trends observed are surprising. An increase in spatial pattern complexity increases the probability of successful detection. More influential than pattern complexity is difference in mean DBH.

As difference in mean DBH increases the probability of detecting spatial dependency increases. The increase in mean difference in DBH creates more obvious spatial dependencies and less overlap among distributions of microsites. The largest effect seems to result from increases in CV.

As CV increases, the probability of detecting spatial dependency decreases. The increased amount of variation in each microsite actually increases the uniformity of the plot and assuages the effects of spatial dependencies. Therefore, it becomes increasingly more difficult to detect spatial dependency with higher amounts of variation.

The probability of detecting spatial dependence of the control plot for all

scenarios was approximately 5%. This is a reassuring statistic because it validates the

statistical methods used. An α=0.05 for the permutation F-tests was specified before

running the simulations. Therefore, about 5% of the time we would expect the test to

detect spatial dependence when there is no spatial dependence.

3.4.3 Cluster Analysis

Table 4 shows a sharp decrease from all three factors: mean difference in DBH,

spatial complexity, and CV. There are many things that can affect the outcome of cluster

analysis and because cluster analysis is not a formal method of statistical testing it is

difficult to identify exactly why the trends develop so rapidly. In particular, the efficacy

of the spatially constrained cluster algorithm and the CH statistic are confounded in the

results. It is impossible, in this analysis, to determine whether the success or failure of the

clustering simulations are attributed to the algorithm or to the CH statistic. We do see,

however, that the parameters manipulated in our experiment greatly influence the success

rate of our clustering algorithm to correctly detect the number of microsites in the data.

The success of the clustering algorithm depends on how distinct each microsite is

and how similar each tree is to others within a given microsite. For example, when

simulating the biplot, we are looking at tree sizes drawn from two distributions. The

greater the difference in means and the smaller the CV, the more similar trees are to each

other within a microsite and the more dissimilar they are to trees from other microsites.

As 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 to be created and

the clustering algorithm may identify these overlapped trees as unique groups (given they are also neighbors) and disturb the accuracy of estimates.

Misclassification simulations were done for scenarios where the probability of successful detection was greater than 80%. For scenarios where probability of successful detection was greater than 80%, the misclassification of trees was fairly low ranging from about 0%-32%. 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 this method's ability to detect microsites accurately.

3.5 Conclusions and Management Recommendations

The spatial analyses we examined have many applications to forestry. Many types of multivariate regression analysis done in forestry can potentially benefit from spatial analysis. Redundancy analysis is a method that allows for partitioning of spatial components to better understand what is affecting the response, but more specifically it accounts for variation otherwise confounded in the linear model that may alter the statistical outcome and produce inaccurate results. Mapping of spatial variability has applications in precision silviculture where managers may benefit from a richer understanding of contributing factors of plot development. A better understanding of spatial patterns can lead to targeted applications of resources such as fertilizers and herbicides. Most importantly, we need to continue investigations into how these methods work on different types of data. We propose future investigations to include use of plot simulation models that incorporate distance dependent growth relationships and

stochastic mortality. Lastly, applications to real plot data are needed to examine how incorporating spatial analysis influences results and conclusions.

Successful intensive silviculture requires an understanding of what resources are spatially limited (Allen et al. 2005). This, in part has to do with understanding plot uniformity. Stape et al. (2010) found that uniformity in plot structure substantially influences wood growth and that uniform plots were up to 13% more productive than non-uniform plots. Redundancy analysis is well suited for studying tree growth variations that can be caused my resource limitations. Specifically, RDA can be used to better understand and measure plot uniformity. Greater plot uniformity can lead to greater resource-use efficiency and enhanced productivity (Nilsson & Allen 2003). Cluster analysis is a possible tool for mapping the variability studied with redundancy analysis and may be valuable for determining areas of low productivity within a plot.

3.6 Literature Cited

Allen, H. L., Fox, T. R., & Campbell, R. G. (2005). What is ahead for intensive pine plantation silviculture in the south?. *Southern Journal of Applied Forestry*, *29*(2), 62-69.

Anderson, M. J., & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, *62*(3), 271-303.

Bell, G., Lechowicz, M. J., & Waterway, M. J. (2006). The comparative evidence relating to functional and neutral interpretations of biological communities. *Ecology*, *87*(6), 1378-1386.

Betts, R. A., Cox, P. M., Lee, S. E., & Woodward, F. I. (1997). Contrasting physiological and structural vegetation feedbacks in climate change simulations. *Nature*, *387*(6635), 796-799.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*(1), 51-68.

Buford, M. A., & Burkhart, H. E. (1987). Genetic improvement effects on growth and yield of loblolly pine plantations. *Forest Science*, *33*(3), 707-724.

Carmean, W. H., & Lenthall, D. J. (1989). Height-growth and site-index curves for jack pine in north central Ontario. *Canadian Journal of Forest Research*, *19*(2), 215-224.

Corral, G. R. (2014). *Investigating Selection Criteria of Constrained Cluster Analysis: Applications in Forestry* (Masters Thesis, Virginia Polytechnic Institute and State University).

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3), 483-493.

Dray, S., Pélissier, R., Couteron, P., Fortin, M. J., Legendre, P., Peres-Neto, P. R., ... & Wagner, H. H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, *82*(3), 257-275.

Dale, M. R., & Fortin, M. J. (2014). *Spatial analysis: a guide for ecologists*. Cambridge University Press.

Downes, G. M., Wimmer, R., & Evans, R. (2002). Understanding wood formation: gains to commercial forestry through tree-ring research. *Dendrochronologia*, *20*(1), 37-51.

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3), 483-493.

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. Wiley Series in Probability and Statistics.

Ezekiel, M. (1930). Methods of correlation analysis. 427 pp., illus. *New York and London*. Chicago

Fox, J. C., Ades, P. K., & Bi, H. (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, *154*(1), 261-276.

Galbraith, D., Levy, P. E., Sitch, S., Huntingford, C., Cox, P., Williams, M., & Meir, P. (2010). Multiple mechanisms of Amazonian forest biomass losses in three dynamic global vegetation models under climate change. *New Phytologist*, *187*(3), 647-665.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5(3), 115-146.

Griffith, D. A., & Peres-Neto, P. R. (2006). Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, *87*(10), 2603-2613.

Haynes, R. W., Alig, R. J., & Moore, E. (1994). Alternative simulations of forestry scenarios involving carbon sequestration options: Investigation of impacts on regional and national timber markets. Forest Service general technical report (No. PB-95-128856/XAB; FSGTR-PNW--335). Forest Service, Portland, OR (United States). Pacific Northwest Research Station.

Hou, Z., Xu, Q., Hartikainen, S., Antilla, P., Packalen, T., Maltamo, M., & Tokola, T. (2015). Impact of plot size and spatial pattern of forest attributes on sampling efficacy. *Forest Science*, 61(5), 847-860.

Huntingford, C., Zelazowski, P., Galbraith, D., Mercado, L. M., Sitch, S., Fisher, R., & Cox, P. M. (2013). Simulated resilience of tropical rainforests to CO2-induced climate change. *Nature Geoscience*, *6*(4), 268-273.

Laliberté, E., Paquette, A., Legendre, P., & Bouchard, A. (2009). Assessing the scale-specific importance of niches and other spatial processes on beta diversity: a case study from a temperate forest. *Oecologia*, *159*(2), 377-388.

Legendre, P. (1990). Quantitative methods and biogeographic analysis. In *Evolutionary Biogeography of the Marine Algae of the North Atlantic* (pp. 9-34). Springer Berlin Heidelberg.

Legendre, P. (2008). Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *Journal of Plant Ecology*, *1*(1), 3-8.

Legendre, P., Oksanen, J., & ter Braak, C. J. (2011). Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, *2*(3), 269-277.

Legendre, P. (2011). const. clust: space-and time-constrained clustering package. R package version 1.2. *URL: http://adn. biol. umontreal. ca/~ numericalecology/Rcode*.

Legendre, P., Borcard, D., Blanchet, F. G., & Dray, S. (2012). PCNM: MEM spatial eigenfunction and principal coordinate analyses. R package version 2.1-2/r106.

Legendre, P., & Legendre, L. F. (2012). *Numerical ecology* (Vol. 24). Elsevier.

Lepš, J., & Kindlmann, P. (1987). Models of the development of spatial pattern of an even-aged plant population over time. *Ecological Modelling*, *39*(1), 45-57.

Magnussen, S., & Kremer, A. (1993). Selection for an optimum tree growth curve. *Silvae Genetica*, *42(6)*, 322-322.

Marignani, M., Del Vico, E., & Maccherini, S. (2007). Spatial scale and sampling size affect the concordance between remotely sensed information and plant community

discrimination in restoration monitoring. *Biodiversity and Conservation*, *16*(13), 3851-3861.

Mateu, J., Uso, J. L., & Montes, F. (1998). The spatial pattern of a forest ecosystem. *Ecological Modelling*, *108*(1), 163-174.

Miller, J. K., & Farr, S. D. (1971). Bimultivariate redundancy: a comprehensive measure of interbattery relationship. *Multivariate Behavioral Research*, *6*(3), 313-324.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17-23.

Murtagh, F., & Legendre, P. (2011). Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. *Statistics*, *1050*, 11.

Nilsson, U., & Allen, H. L. (2003). Short-and long-term effects of site preparation, fertilization and vegetation control on growth and stand development of planted loblolly pine. *Forest Ecology and Management*, *175*(1), 367-377.

Norcliffe, G. B. (1969). On the use and limitations of trend surface models. *The Canadian Geographer/Le Géographe Canadien*, *13*(4), 338-348.

Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G., Solymos, P., Stevens, M., and Wagner, H. (2013)."Package 'vegan'." *R Packag ver* 254: 20-8.

Palmer, M. W. (1990). Spatial scale and patterns of species-environment relationships in hardwood forest of the North Carolina Piedmont. *Coenoses*, *5*(2), 79-87.

Penttinen, A., Stoyan, D., & Henttonen, H. M. (1992). Marked point processes in forest statistics. *Forest Science*, *38*(4), 806-824

Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, *87*(10), 2614-2625.

Peres-Neto, P. R., & Legendre, P. (2010). Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, *19*(2), 174-184.

Prentice, I. C., Sykes, M. T., & Cramer, W. (1993). A simulation model for the transient effects of climate change on forest landscapes. *Ecological Modelling*, *65*(1), 51-70.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4), 329-358.

Sabatia, C. O., & Burkhart, H. E. (2013). Height and diameter relationships and distributions in loblolly pine stands of enhanced genetic material. *Forest Science*, *59*(3), 278-289.

Scheller, R. M., & Mladenoff, D. J. (2005). A spatially interactive simulation of climate change, harvesting, wind, and tree species migration and projected changes to forest composition and biomass in northern Wisconsin, USA. *Global Change Biology*, *11*(2), 307-321.

Schweingruber, F.H. (1987). Tree Rings: Basics and Applications of Dendrochronology. Kluwer Academic Publishers, Dordrecht.

Solomon, A. M. (1986). Transient response of forests to CO2-induced climate change: simulation modeling experiments in eastern North America. *Oecologia*, *68*(4), 567-579.

Stape, J. L., Binkley, D., Ryan, M. G., Fonseca, S., Loos, R. A., Takahashi, E., Azevedo, M. R. (2010). The Brazil Eucalyptus Potential Productivity Project: Influence of water, nutrients and stand uniformity on wood production. *Forest Ecology and Management*, *259*(9), 1684-1694.

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), 140-145.

ter Braak, C. J., & Prentice, I. C. (1988). A theory of gradient analysis. *Advances in Ecological Research*, *18*, 271-317.

ter Braak, C. J., & Smilauer, P. (2002). *CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5)*. www. canoco. com.

Tokola, T., & Shrestha, S. M. (1999). Comparison of cluster-sampling techniques for forest inventory in southern Nepal. *Forest Ecology and Management*, *116*(1), 219-231.

Tomppo, E. (2006). The Finnish national forest inventory. In *Forest Inventory* (pp. 179-194). Springer Netherlands.

Webster, M., Forest, C., Reilly, J., Babiker, M., Kicklighter, D., Mayer, M., Prinn, R., Sarofim, M., Sokolov, A., Stone, P. & Wang, C. (2003). Uncertainty analysis of climate change and policy response. *Climatic Change*, *61*(3), 295-320.

Chapter 4

Quantifying and Mapping Spatial Heterogeneity in Simulated Plot Data

Abstract

Spatial analysis is appropriate for data that is collected across geographic space and may be influenced by spatially structured variables or influenced by neighboring spatial structures (Legendre 1993). A spatially explicit stand simulator PTAEDA 4.1was used to investigate the efficacy of spatial analysis methods on simulated forest plot data. We simulated regularly spaced, plantation stands of loblolly pine (*Pinus taeda L.*) at stand age 10, 20 and 30 years old. Spatial dependence was induced via site index tools in the PTAEDA 4.1 software package. We utilized three methods of spatial analysis to quantify, test, and map the spatial variability in the simulated stands. The methods include redundancy analysis, partial redundancy analysis, and spatially constrained cluster analysis. The results indicated that redundancy analysis and partial redundancy analysis were useful tools and appropriate for forest data. Spatially constrained cluster analysis was found unreliable, and the results indicated that it is not useful in accurately mapping microsites. Age, levels of variation, and microsite pattern complexity were all found to affect the outcome of the analysis. At both young and old ages, the probability of detecting spatial relationships was lower than for stands 20 years old. As levels of variation in the data increased there was a decrease in the effectiveness of the methods. Lastly, as spatial pattern complexity increased there was a decrease in the effectiveness of the methods to test and quantify spatial variability. These results provide evidence that RDA and pRDA have potential to serve as useful tools for quantifying and modeling spatial heterogeneity in forests.

Keywords: Forestry, loblolly pine, simulations, redundancy, cluster, spatial analysis.

Introduction 4.1

Little research has been done on the phenomenon of spatially dependent growth patterns in small-scale forest plantation plots. Zhang et al. (2005) point out that most data analysis carried out in forestry is done by ordinary least squares (OLS) and Fox et al. (2001) note that OLS is the predominate means of developing individual tree growth models. The use of OLS is carried out with the assumption that individual trees within plots are independent of each other. However, researchers have begun to recognize that treating individual trees within a given plot as independent is inaccurate (Penttinen et al. 1992, Mateu et al. 1998). Thus, modern modeling of tree growth is overly simplified (Fox et al. 2001) and modelers have become accustom to accepting large amounts of unexplained variance as an occupational hazard (Burkhart & Gregoire 1994). Classical methods of OLS, for example, may fail to detect spatial-structures in the data (Legendre & Fortin 1989).

Understanding the spatial variability of individual tree growth is challenging due to the confounding effects of contrasting productivity within a stand and the individual fitness of trees (Liu & Burkhart1994a, Liu & Burkhart 1994b). When evaluating site productivity, the site index (SI) concept – that is average height of the trees that represent the dominant part of the canopy at a specific index age- is commonly used (Burkhart & Tomé 2012, p. 131). In forest plots, areas of contrasting productivity tend to produce separate SI curves (Carmean & Lenthall 1989). Burkhart & Tennent (1977) note that groups of SI curves which exhibit different shapes for varying site quality are commonly termed "polymorphic". Thus, areas of contrasting productivity within a stand, termed

"microsites" in this report, are expected to demonstrate polymorphism (Buford &

Burkhart 1987, Magnussen & Kremer 1993, Tang et al. 2001). Due to the size of the trees

being dependent on their location within a stand, tree size is spatially dependent.

Generally speaking, spatially structured environmental data may arise from two

independent processes (Legendre 1993, Fortin & Dale 2005): induced spatial dependence

and spatial autocorrelation (Dray et al. 2006). Induced spatial dependency can be

produced by functional dependence of the response variables on explanatory variables

that are themselves spatially correlated (Legendre et al. 2009). If all important spatially-

structured explanatory variables are included in the analysis, equation (1) correctly

accounts for the spatial structure induced in Y.

$$y_i = f(X_j) + \varepsilon_j \tag{1}$$

For formula (1), y_i is the value of the dependent variable $\mathbf{y}$ at site j and $\boldsymbol{\varepsilon_j}$ is the error term

whose value is independent from site to site (Dray et al. 2012, Dale & Fortin 2014,

Legendre 1993, Legendre & Legendre 2012). Spatial autocorrelation produces

relationships between values observed at neighboring points in space and therefor the

observed variables lack independence (Legendre et al. 2009). Spatial autocorrelation

(equation 2) may occur in the response data $\boldsymbol{y_j}$ observed at site j on the geographic

surface and is assumed to be the overall mean of the process $\boldsymbol{\mu_y}$ in the study area plus a

weighted sum of the centered values $(\boldsymbol{y_j} - \boldsymbol{\mu_y})$ at surrounding sites i, plus the error term

ε_j.

$$y_j = \mu_y + \sum w_i(y_i - \mu_y) + \varepsilon_j \tag{2}$$

The y_i 's are the values of y of tree i located within the zone of spatial influence of the process generating spatial autocorrelation (Legendre 1993).

Tests for presence of spatial dependence are well documented. Geary's c (Geary 1954) and Moran's I (Moran 1950) coefficient are among the most popular methods for detecting presence of spatial dependence (Dale & Fortin 2002, Paez & Scott 2005). Coefficients like Geary's c and Moran's I examines the spatial correlation among the error components of the observed data (Anselin 1995). Geary's c coefficient varies from 0 to some unspecified number larger than 1. Positive spatial correlation produces c values from 0 to 1 and negative spatial correlation produces c values greater than 1. (Legendre & Legendre 2012). Interestingly, Moran's I coefficients can be used to develop spatial eigenfunctions (hereafter spatial variables) through an analysis known as Moran eigenvector maps (MEM's) (Borcard & Legendre 2002, Griffith & Peres-Neto 2006).

Testing for a spatial relationship in the data requires that the spatial structure be expressed by meaningful variables. Trend surface analysis, which uses polynomial transformations of geographic coordinates, is the most straight forward method to model spatial relationships (Borcard & Legendre 2002). However, Norcliffe (1969) cautions that trend surface analysis is only satisfactory when the sampling area is approximately homogenous with a regular sampling design and the number of spatial locations is "reasonable". Dray et al. (2006) also points out that trend surface analysis is appropriate only if the spatial structure to be modeled is somewhat simple and a more modern method for creating meaningful spatial variables is with MEM's. These spatial variables can be used for variation partitioning (Borcard et al. 1992) and significance testing (Legendre et al. 2011). Stand simulators can be used to produce spatial variables.

Forest growth simulators are flexible tools for predicting tree growth and are commonly applied in the United States (Pretzsch et al. 2002). One particular type of forest growth model simulates individual tree growth and interactions while considering tree position (Munro 1974). This class of forest growth models depends on a system of equations which dictate growth behavior of individual trees based on competition indices which are used to predict heights and diameters (Wensel & Koehler 1985). PTAEDA 4.1 is an example of an individual tree growth, spatially-explicit, stand simulator for planted stands of loblolly pine. The model structure was first published by Daniels & Burkhart (1975). An updated version based on a region-wide data set was released by Burkhart et al. (1987); subsequently enhancements in software and data have been made, resulting in the current version designated, PTAEDA 4.1.

PTAEDA 4.1 is in many respects a typical individual tree, distance-dependent model (Burkhart & Tomé 2012). The PTAEDA 4.1 stand simulator incorporates distances between competing trees in competition index calculations for each subject or target tree. Competition is measured as a ratio of the division of the DBH of a competitor tree to the DBH of the subject tree, divided by the distance of the subject tree to the competitor tree. The PTAEDA 4.1 simulator consists of two main subsystems. The first subsystem deals with the creation of the precompetitive stand. Following the onset of competition. the second subsystem deals with the growth dynamics of the stand. A number of management subroutines are available for simulation including hardwood competition control, fertilizer, and thinning (Burkhart 2008). Other specifications can be assigned prior to implementing simulations including spacing, variation in growth around mean height and DBH, and delineation of microsites based on SI. After simulation of the

precompetitive stand, competition is evaluated and growth is calculated individually on an annual basis. Generally speaking, growth is assigned based on theoretical growth potentials with adjustments to growth based on competitive status and vigor, site productivity, and a random component is assigned to account for unknown effects of genetic and microsite variation. A general description of the PTAEDA model structure is contained in Burkhart and Tomé (2012), pg. 3113-317; additional information on model structure is available in Burkhart et al. (1987); Burkhart et al. (2014) contains details on PTAEDA 4.1 used in this study.

The data generated in this chapter is limited by variables simulated by PTAEDA 4.1. For example, PTAEDA 4.1 summarizes effects of environmental variables by a single variable SI. Thus we will use SI as a single explanatory variable; SI is taken as a proxy variable for a number of environmental factors related to tree growth. PTAEDA 4.1 also provides X and Y coordinates of individual trees. We use MEM's to construct spatial variables so that complex spatial patterns could be detected. Raw coordinates and polynomial functions do not model complex spatial patterns well.

PTAEDA 4.1 can be a valuable tool for understanding processes which lead to spatial correlations of tree attributes. For example, a better understanding of spatial processes and spatial analysis techniques is essential for quantifying how silvicultural practices affect stand uniformity. Stape et al. (2010) note a need for more research on understanding how silvicultural practices lead to higher uniformity in stand structure Binkley et al. (2002) noted that increased yield was more associated with higher uniformity of tree sizes than with genetics. Other scientist (i.e. Nilsson & Allen 2003, Aspinwall et al. 2011, Binkley et al 2010, Ryan et al. 2010) have noted the importance of

stand uniformity, noting that it may lead to increase resource use efficiency and enhanced productivity. Spatial analysis techniques are well suited for studying variations in tree growth across a stand.

Unfortunately, in field data it is unclear how multiple forms of spatial processes influence tree attributes or how the relative strength of the processes determine positive, negative, or no spatial correlation in the data. Stand simulators are a useful tool for studying spatial processes that may otherwise be too difficult or impossible to control and measure. Simple statistical simulations which only consider a limited number of variables and outputs can be run on most software packages, but for more complex simulations, stand simulator software is appropriate.

In chapter 3, we utilized statistical simulations with known distributions to simulate stands with varying microsite conditions. We simulated reasonably simple stand conditions where tree attributes were correlated to microsites, creating a spatial process known as induced spatial dependence. The purpose was to measure the efficacy of a set of statistical methods to detect spatial dependence, quantify spatial variability, and map microsite without the more complicated and confounding effects of a second type of spatial process known as spatial autocorrelation.

The goal of this investigation was to assess the efficacy of spatial analysis techniques to detect spatial dependence, quantify spatial variability, and map microsites from data simulated with PTAEDA 4.1. Specifically, our objectives were to: 1) evaluate the efficacy of pRDA to detect spatial dependence in simulated stands which range in microsite configuration, stand age, and coefficient of variation of SI, 2) evaluate the efficacy to RDA to estimate spatial variability in simulated stands which range in

microsite configuration, stand age, and coefficient of variation of SI, and 3) evaluate the efficacy of spatially constrained cluster analysis to estimate the number of microsites in simulated stands which range in microsite configuration, stand age, and coefficient of variation in SI.

4.2 Methods

4.2.1 Study material and data

The study material was composed of data generated from the PTAEDA 4.1 stand simulator for loblolly pine. We created five spatial patterns, each with a unique microsite configuration (Figure 4.1), and evaluated simulations at 3 age levels (10, 20, 30 years) and at varying levels of SI coefficient of variation (CV) (5%, 10%, 15%, 20%, 25%). Each stand consisted of 25 rows and 25 columns, a total of 625 trees, all on an 8ft by 8ft spacing, on approximately 0.92 acres. We created spatial patterns by using the site index (SI) tool in PTAEDA 4.1 to select groups of trees and assign unique SI values to those areas. Five microsite patterns ranging from one microsite to five microsites were created so that we could represent a range of spatial pattern complexity.

The first step was to determine how to vary the SI values within each stand. The simplest plot was assigned one SI value, while the most complex has five different site index values.
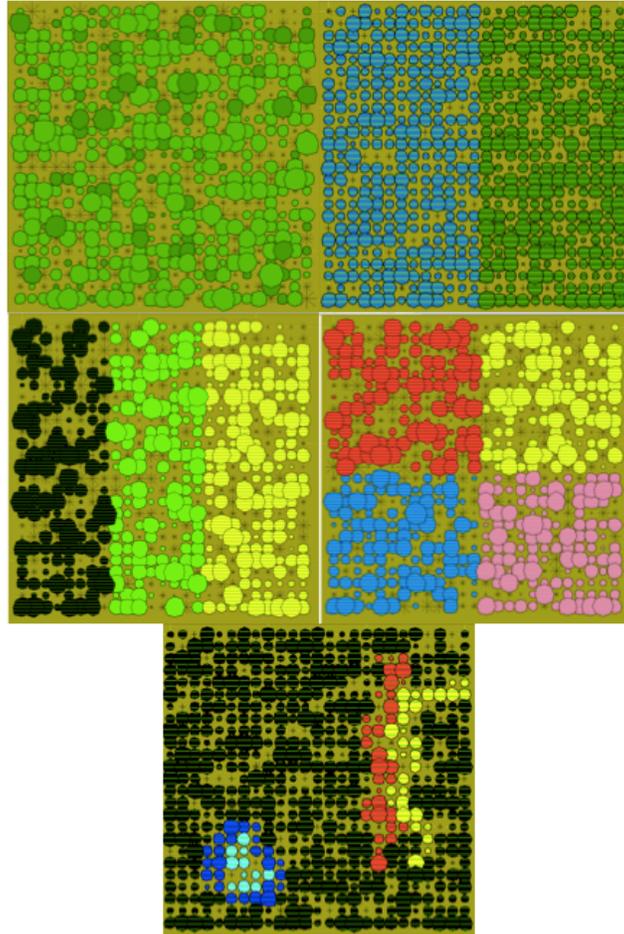
Figure 4.1 Illustrates the 5 spatial patterns used in this study. Microsites are distinguished by color. For left to right: Control, biplot, triplot, quadplot, and freeplot.

Site index is the average height of dominant and codominant trees at a selected base age (Burkhart & Tomé 2012, p.132). Higher site index values are associated with larger trees than areas with lower site index values, given that the trees are of similar age. Site index values and plot names were assigned as follows:

 (*Refer to Figure 4.1*)

Control- The control plot has a single site index value because of microsite homogeneity. The SI value is 70 feet average dominant height at base age 25.

Biplot- There are two microsites on the biplot, 70 feet average dominant height at base age 25 (green) and 80 feet average dominant height at base age 25 (blue).

Triplot- There are three microsites of the triplot have a SI value of 60 feet average dominant height at base age 25 (yellow), 70 average dominant height feet at base age 25 (green), or 80 feet average dominant height at base age 25 (blue).

Quadplot- The quadplot has four microsites with SI value of 50 feet average dominant height at base age 25 (orange), 60 feet average dominant height at base age 25 (yellow), 70 feet average dominant height at base age 25 (green), or 80 feet average dominant height at base age 25 (blue).

Free plot- The free plot has five microsites. The large green area of the free plot is SI 70 feet average dominant height at base age 25. The small light blue area is SI 60 feet average dominant height at base age 25, the dark blue area circumscribing the blue area is SI 65 average dominant height feet at base age 25. The red area is SI 75 feet average dominant height at base age 25and the yellow area is SI 80 feet average dominant height at base age 25.

We ran simulations for all combinations of five spatial patterns, at three age levels (10, 20, 30 years), and five CV levels (5, 10, 15, 20, 25%). After each iteration was run, the resulting data were stored in Excel files and used later for analysis with R statistical software.

4.2.2 Redundancy Analysis

Redundancy analysis RDA (Rao 1964, Wollenberg 1977) can be used to partition the variance of the response into into meaningful categories (i.e spatial, environmental) (Borcard et al. 1992, Legendre 1993). Redundancy analysis can be used as a method of

variation partitioning to decompose ecological variation into four components (pure

spatial, pure environmental, explained both by space and environment, and unexplained

(Figure 4.2) (Borcard & Legendre 2002). Significance testing in RDA examines the

overall relationship between **Y** and **X**, where **X** is represented by components [a] and [b],

or [b] and [c] in Figure 4.2.
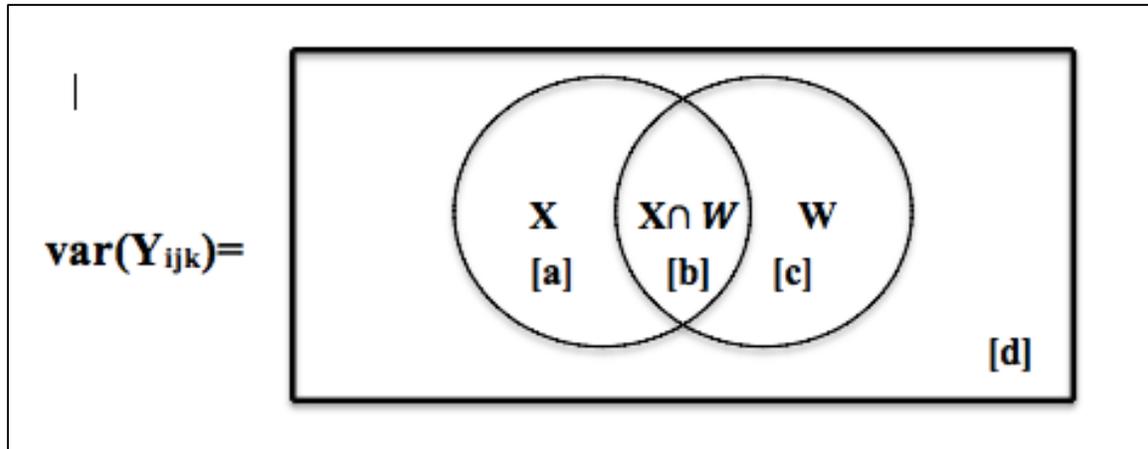
Figure 4.1 Illustrates the partitioning of variation by RDA. Here, X and W are two
explanatory matrices and [a], [b], [c], and [d] are fractions of variance. If matrix X is an
matrix of environmental variables and W is a matrix of spatial variables then the fraction
[a] is pure environmental variability, [b] is confounded variability of environmental and
spatial, [c] is pure spatial, and [d] is the residual component. Fractions [a] and [c] are
considered testable fractions and significance tests can be carried out by pRDA.

In this study, redundancy analysis was done using R software's "VEGAN"

package (Oksanen et al. 2013). The RDA method is used to partition the total percentage

of variation (R^2) explained by the model (Peres-Neto et al. 2006). As a prerequisite to

analysis we standardized all the response variables by subtracting each observation by its

mean and dividing the difference by its standard deviation. The standardization of

response variables applies to all analyses done in this report. This symmetric form of

analysis utilizes a response matrix Y_{nx2} (hereafter Y) with explanatory vector X_{nx1}

(hereafter X) and covariables W_{nx2} (hereafter W). The covariables we used are spatial

variables.

To create W, we started with the XY coordinates of each tree and created a matrix

C_{nx2} where *n* is the number of live trees at time *t*. The software (package PCNM in R by

Legendre et al. 2012) used a minimum spanning tree to connect all live trees in the stand.

The longest edge is considered our threshold value. It then truncates C_{nx2} with the

following criteria:

$$C_{trunc(ij)} \begin{cases} \boldsymbol{C_{trunc\ (ij)} = C_{ij},} & \boldsymbol{if\ \ \ C_{ij} \leq threshold} \\ \boldsymbol{C_{trunc\ (ij)} = 4 \times thresh} & \boldsymbol{if} & \boldsymbol{i = j} \\ \boldsymbol{C_{trunc\ (ij)} = 4 \times threshold} & \boldsymbol{if\ \ C_{ij} \geq Threshold} \end{cases}$$

Computer software packages typically replace the value with *4*threshold.* The diagonal

values of $\mathbf{C_{trunc}}$, which were zero, were also replaced by *4*threshold*. Next, a principal

coordinate analysis (PCoA) was performed on the truncated matrix. The result is n-1

eigenvalues and n-1 eigenvectors. Some eigenvalues are positive and some are negative.

For our simulation study, it was only important to keep the positive eigenvalues and their

associated vectors, because we are only interested in positive spatial correlation among

the trees. Lastly, we took the set of eigenvectors with positive eigenvalues and performed

a forward selection on Y. The eigenvectors found significant were then placed in matrix

W, called the spatial matrix (also called spatial eigenfunctions). Following the creation of

the spatial matrix, variation of Y was partitioned between X (site index values) and W

(spatial variables).

The canonical R^2, called the bimultivariate redundancy statistic by Miller & Farr (1971), quantifies the strength of the linear relationship between variables Y and X:

$$R^2_{Y|X} = \frac{SS(\hat{Y})}{SS(Y)} \tag{3}$$

In equation (3), $SS(\hat{Y})$ is the total sum of squares of $\hat{Y}$ (predicted values) and $SS(Y)$ is the total sum of squares of Y. Using the well known adjusted R^2 (rR^2) from Ezekiel (1930), given in equation 4, we can estimate the variability explained by X or W. In equation 4, Y has n objects and p variables and X has n objects and m variables:

$$rR^2 = 1 - \left(1 - R^2_{Y|X}\right)\frac{n-1}{(n-m-1)} \tag{4}$$

Referring to Figure 4.2, we can begin to calculate the components of explained variance (a,b, and c) as well as the unexplained (d) by following steps outlined by Legendre (2008).

1. Perform the canonical analysis of Y with respect to X. Use equations 3 and 4 to calculate the $R^2_{Y|X}$ and rR^2. The rR^2 corresponds to the circle on the left side of Figure 4.2. It contains fractions [a] and [b].

2. Perform the canonical analysis of Y with respect to W. Use equations 4 and 5 to calculate the $R^2_{Y|W}$ and rR^2. The rR^2 corresponds to the circle on the right side of Figure 4.2. It contains fractions [b] and [c].

3. Perform the canonical analysis of Y with respect to X and W. Use equations 4 and 5 to calculate the $R^2_{Y|X,W}$ and rR^2. The rR^2 corresponds to both circles in Figure 4.2. It contains fractions [a], [b], and [c].

4. From these results compute [b]=[a+b]+[b+c]-[a+b+c].

5. Next, compute [a]=[a+b]-[b]

6. Next, compute [c]=[b+c]-[b]

7. Next, compute [d], which represents the residual component of variance, [d]=1-[a+b+c].

To test individual components, [a] or [c], a method called partial redundancy analysis is used (pRDA) (Legendre et al. 2011).

4.2.3 Partial Redundancy Analysis

Partial RDA considers three sets of variables: Y a matrix of response variables, X a matrix of explanatory variables, and W a matrix of covariables (Legendre et al. 2011). Partial RDA tests the significance of explanatory variables in X while controlling for the linear effect of the covariables in W (Legendre 2008, Legendre et al. 2011). In pRDA we are able to isolate and test for pure effects. In pRDA permutation F-tests are used because the data in the response matrix is standardized and because error terms are often non normal. By testing the significance of individual components ([a] and [c] from figure 4.2), we can more fully understand the underlining causes of variation in the response data. For example, if a significant pure spatial relationship is detected, one could conclude that there are significant patterns present in the data (Legendre 2008).

Partial redundancy analysis tests the significance of a matrix of variables in the presence of covariables. In our analysis we tested for a pure spatial effect by testing the

significance of W while controlling for the linear effects of X. To achieve this goal, we

used the $R^2_{Y\sim W|X}$ statistic (equation 5) and a partial F-statistic (equation 6):

$$R^2_{Y\sim W|X} = \frac{SS(Y\sim W|X)}{SS(Y\sim W)}$$ (5)

The F-statistic used to test the overall significance of the partial regression relationship is:

$$F = \frac{R^2_{Y\sim W|X}}{m} \cdot \frac{1-R^2_{Y\sim W|X}}{n-1-m-q}$$ (6)

For equation 6, n is the number of live trees at time t, m is the number of variables in X,

and q is the number of variables in W. Significance of the F-statistic may be tested with

the F-distribution if the assumption of normality of the residuals holds; however, this is

rarely the case in many ecological studies and for this reason permutation tests are

preferred and used in this study (Legendre & Legendre 2012, p. 25). Specifically, we use

a permutation of the residuals of the reduced model (R software, method="reduced" in

VEGAN'S anova.cca()).

4.2.4 Cluster Analysis

Cluster analysis is the search for groups in data (Everitt et al. 2011). Clustering is

a way to classify objects into groups based on some similarity or dissimilarity (Legendre

1989). The classification of objects is usually done based on some partitioning criterion.

A partition is the division of objects of a set into subsets, such that each object belongs

only to one subset. Exceptions exist, of course, throughout cluster analysis and theory

because cluster analysis is not a set procedure with formal statistical testing but rather a broad topic with many interpretations and related statistics. One such exception is the classification of objects into one and only one subset. Fuzzy clustering allows objects to be members to more than one subset at a time. This paper focuses on a method of hard clustering where each object belongs to only one group at a time. We will use a method of agglomerative, spatially-constrained cluster analysis described by Legendre & Legendre (2012, p. 840) to map the spatial structures in our stand.

A first step to clustering is deciding on a distance measure to use for the clustering method. Depending on the type of algorithm and data used, practitioners choose between similarity or dissimilarity measures. Similarity and dissimilarity measures indicate how quantitatively similar or dissimilar each pair of objects are to each other. In the clustering algorithm used, each object is grouped into larger more inclusive groups based on which objects are least dissimilar. This approach is called agglomerative clustering.

Agglomerative styles of clustering describe a category of algorithms which, as an initial step, assign n objects to be clustered to n groups. After each iteration of the algorithm, objects are assigned to larger more inclusive groups. The number of groups (k) after each iteration approaches min(k)=2 groups. This iterative process is what defines sequential clustering algorithms. Constraints can be imbedded within the algorithm to account for successive temporal relationships and for objects which may be neighbors in space.

There are a number of ecological reasons why practitioners may want to group objects together if and only if they occur in succession to each other or if objects in each

group must be neighbors to each other in space (Legendre 1989). We used spatial constraints such that the resulting map groups together trees growing under similar conditions. This makes the map more interpretable for mapping spatial heterogeneity of tree growth resulting from microsites.

The clustering algorithms used are part of R software package "const.clust". Our method of clustering is an agglomerative approach with spatial constraints. Agglomerative methods begin with examining the data as n-1 objects and with each iteration of the algorithm, the data are put into larger more inclusive groups such that the first iteration shows n-2 groups, the second iteration n-3 groups, and this continues up to 2 groups. The spatial constraints restrict non-neighboring trees from entering larger groups. In order for a tree to enter a group it must be a neighbor to it in space. Trees that are most similar are grouped together (or trees that are least dissimilar are grouped together). The similarity or dissimilarity is based on distance metrics. These metrics quantify likeness of objects. Dissimilarity and spatial constraints require two important steps for this type of clustering. We must determine a distance metric to develop a dissimilarity matrix, and determine which connectivity method to use and build the contiguity matrix.

For our type of data, which is continuous and contains no zeros, a Euclidean distance metric (equation 7) is appropriate. This type of distance measures the association between two objects (trees in our study) and uses only non-geographic information. Euclidean distance values have a range [0,1], where values ≈ **0** indicate high likeness whereas values ≈ **1** indicate low likeness (high dissimilarity). We used Euclidean

distances among objects using non-geographic information to create our dissimilarity

matrix (D).

$$D(y_r, y_{r+1}) = \sqrt{\sum_{c=1}^{p}(y_{rc} - y_{(r+1)c})^2} \qquad (7)$$

For equation (7), r=row of matrix Y, c=column of matrix Y, and p is the number

of variables in matrix Y. For our simulations, r=1…n, where n is the number of live trees

at time t, and p=2. This step is typical in many clustering algorithms, but in the next steps

we impose spatial constraints on the dissimilarity matrix, which is information typically

not incorporated into clustering analysis.

Spatial constraints are important in this analysis because they offer a more

interpretable map. We hope to map microsites by proxy. In other words, by mapping

groups of trees that are growing similarly we hope to identify microsites, which are

assumed to be causing groups of trees to grow differently than the surrounding trees.

Therefor, by identifying these groups of trees we are identifying the location of the

microsites. The spatial constraints we impose are defined by a connection scheme that

identifies which trees are considered neighbors to each other. We used a method called

Delaunay triangulation.

The Delaunay triangulation uses spatial coordinates of each tree to identify

neighbors. Matrix C (the contiguity matrix) of zeros is created that is *nxn* in size, where n

is the number of live trees at time t. Matrix C was used to identify all neighbors in the

stand and so columns 1..n include all trees and rows 1..n also include all trees. The

properties of matrix C allow us to put a 1 in all elements were trees$_i$ and tree$_j$ are

76

neighbors and a zero otherwise, an obvious result is that all diagonal elements (where

tree$_i$=tree$_j$) will be a 1. Edges are created between two objects that are considered

neighbors. The Delaunay triangulation method states that for any triplet of non-collinear

points A, B, and C the three edges connecting these points are included if and only if the

circle passing through these points (Figure 4.3) includes no other point (Legendre &

Legendre 1998). This criterion is a robust method for defining contiguity. This

connection scheme works well with regular grids and is adaptable to various patterns of

planting grids and will transfer well to field plots that are slightly irregular.
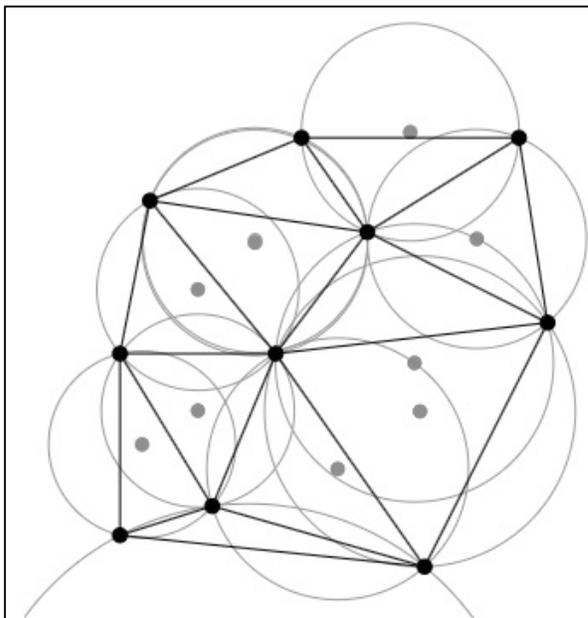
Figure 4.2 Illustrates the circles used in Delaunay triangulation. Black dots represent the objects (trees in our data), red dots represent the center of each circle used to circumscribe three points, and the thick black lines connect "neighbors" (Wikipedia, November 3rd, 2014).

In constrained clustering, the Hadamard product of the dissimilarity matrix and

contiguity matrix produce a dissimilarity matrix for constrained clustering. An *nxn*

dissimilarity matrix is created from equations (7). The more akin any two trees are in

structure the closer to 0 is their dissimilarity value. We then create an *nxn* contiguity

matrix of 1's and 0's where 1's mark neighbors as defined by Delaunay triangulation and 0's elsewhere. The Hadamard product for our data is the dissimilarity in growth among neighboring trees.

The Calinski-Harabasz (CH) statistics are used to estimate the number of groups in the data. Corral (2014) found that the CH statistic performed well when estimating the number of groups for tree plantation data. We examined if the CH statistic correctly identified the number of groups in each the spatial pattern (2 for biplot, 3 for triplot, 4 for quadplot, and 5 for free plot).

4.3 Results

4.3.1 Redundancy Analysis

We used RDA to partition the variation of the response data and estimate the proportion of variability explained by spatial variables. Table 4.1 lists the rR^2 values for each simulation completed in PTAEDA 4.1. The rR^2 values range from 11% (control age=20, CV=5%) to 0% in several cases. Generally, there is a trend with age and CV. As age increases there is a decrease in percent of variation explained and a decrease in variation explained as CV increases. Relative to the other age groups, the simulations at stand age of 10 years tend to have the higher rR^2 values. The average changes in rR^2 values from stand ages 10-20 years, 20-30 years, and 10-30 years are lowest in the control plot, with an average change of 4.8% and highest in biplot with an average change of 8.4%

Table 4.1 Estimated rR^2 for each spatial pattern and all scenarios.

AGE	CV	Control	Biplot	Triplot	Quad	Free
10	*5*	10%	5%	11%	8%	7%
10	*10*	3%	5%	6%	8%	6%
10	*15*	3%	3%	3%	3%	3%
10	*20*	1%	1%	2%	1%	0%
10	*25*	1%	7%	1%	0%	3%
20	*5*	11%	5%	10%	5%	2%
20	*10*	4%	2%	7%	1%	2%
20	*15*	0%	2%	4%	1%	3%
20	*20*	2%	1%	0%	1%	5%
20	*25*	2%	2%	0%	1%	0%
30	*5*	5%	2%	5%	7%	3%
30	*10*	3%	9%	1%	2%	4%
30	*15*	0%	0%	8%	0%	2%
30	*20*	1%	0%	2%	0%	2%
30	*25*	2%	0%	0%	1%	0%

4.3.2 Partial Redundancy Analysis

Partial redundancy analysis was performed to detect spatial dependence. The significance of spatial variables indicates presence or absence of spatial dependence. Table 4.2 contains the p-values from the permutation F-tests, which test for a significant spatial component in the data. Values less than (0.05) indicate the presence of a spatial component. Failure to reject the null hypothesis seems to be more related to an increase in stand age rather than to CV or complexity of spatial pattern. Even the control stand, where we did not impose spatial dependence, tested as significant in most cases.

Table 4.2 Pvalue from pRDA test of significant spatial component for each spatial pattern and all scenarios. (*) indicates pvalues greater than the specified alpha level.

AGE	CV	Control	Biplot	Triplot	Quad	Free
10	5	0.01	0.01	0.01	0.01	0.01
10	10	0.01	0.01	0.01	0.01	0.01
10	15	0.01	0.01	0.01	0.01	0.01
10	20	0.03	0.01	0.01	0.01	0.01
10	25	0.01	0.01	0.01	0.03	0.01
20	5	0.01	0.01	0.01	0.01	0.01
20	10	0.01	0.01	0.01	0.02	0.01
20	15	0.23*	0.01	0.01	0.11*	0.01
20	20	0.01	0.01	0.03	0.01	0.01
20	25	0.01	0.01	0.41*	0.03	0.24*
30	5	0.01	0.02	0.01	0.01	0.01
30	10	0.01	0.01	0.11*	0.01	0.01
30	15	0.14*	0.20*	0.01	0.20*	0.01
30	20	0.07*	0.14*	0.01	0.16*	0.01
30	25	0.01	0.13*	0.24*	0.04	0.08*

4.3.3 Cluster Analysis

Cluster analysis was performed for all scenarios. Table 4.3 lists the outcomes of the cluster analysis. An (*) is used to highlight each successful trial where the cluster analysis estimated the correct number of groups. The numbers listed in the cells are the estimated number of groups by the constrained cluster analysis. Generally, there is a low success rate throughout the trials with an exception for the biplot at stand age 20 years. All other spatial patterns were detected correctly only one or two times throughout all the scenarios. A trend seen in Table 4.3 is a tendency toward underestimation of the number of groups for the triplot, quadplot, and free plot.

Table 4.3 Illustrates the number of estimated microsites in all combinations stand age, microsite configuration, and coefficient of variation (CV). (*) indicate that cluster analysis correctly identified each plot.

AGE	CV	Biplot	Triplot	Quad	Free
10	5	2*	2	2	2
10	10	4	2	5	3
10	15	3	2	3	3
10	20	3	4	3	3
10	25	2*	4	2	3
20	5	2*	2	2	9
20	10	2*	2	4*	2
20	15	2*	2	2	2
20	20	2*	5	3	10
20	25	2*	6	2	5*
30	5	2*	2	2	2
30	10	3	2	9	2
30	15	3	3*	3	2
30	20	7	3*	10	4
30	25	5	4	2	4

4.4 Discussion

4.4.1 Redundancy Analysis

The lack of apparent pattern or trend in the results may indicate that the spatial structure is correctly accounted for with the site index variable. Thus, partitioning out the variability associated with the SI variable leaves only spatial variability associated with other processes such as growth response to mortality, but not the induced spatial dependence of microsites. This indicates that PTEADA 4.1 may not be a suitable simulator to test efficacy of RDA.

The amount of variability captured remained low throughout the simulations, but it is important to keep in mind that these are rR^2 (fractions of the total explained variation) values and therefore they are expected to be much lower than regular R^2 values. For example, Gray & Harding (2011) reported significant spatial structuring with

rR^2 values of less than 4%. Similarly, Heikkinen et al. (2004) reported significant spatial structures in data sets with rR^2 (of pure spatial component) ranging from 2.5-3.4%. Obvious trends do not appear in the results with the exception of decreasing rR^2 values as CV increases and slight decreases as age increases. We expected CV to affect the rR^2 values because increasing the random noise in the system makes the spatial variability less obvious and more difficult to disentangle from the residual variation. The RDA attempts to sift out the pure spatial component of the total variation, but by increasing the randomness we end up hiding the spatial variability.

As age increases a number of different factors contribute to the variability in the system. Two factors in particular are competition and mortality. Competition introduces a form of fine scale spatial dependence called negative spatial correlation. Negative spatial correlation occurs because competition among trees creates "winners" and "losers". As trees compete for resources some trees outcompete their neighbors and grow larger and move into more dominant positions. The negative spatial correlation results from differences in DBH of neighboring trees. For example, large trees will tend to have smaller neighbors and smaller trees will tend to have larger neighbors. The occurrence of negative spatial correlation becomes more prevalent as the stand matures and has more time for the competitive effects to set in.

PTAEDA 4.1 is designed with a random mortality function that eliminates trees throughout the life of the stand. The elimination of trees creates gaps of growing space that the software is designed to acknowledge by increasing growth capacity of adjacent trees. The increased growth of trees around gaps may create positive spatial correlation. This can occur if the removal of the smaller tree increases the similarity among the group

of trees that surround the newly created gap. Also, the growth of trees surrounding the gap can increase with newly created growing space.

The effects of competition and mortality are important because the defined groups within each simulation plot will exhibit positive spatial correlation due to microsite effects. Competition creates negative spatial correlation which can cancel out positive spatial correlation. The RDA results show that the amount of spatial variability decreases over time. The higher levels of spatial variability at the young ages may represent the microsite effects which decrease due to competition over time. The control plot also shows this trend. However, the average change across stand age is smallest (4.8%) in the control plot. This is similar to results described by Liu & Burkhart (1994b) who found positive spatial correlation in simulated height. Liu & Burkhart (1994b) used an earlier version of the simulator (PTAEDA 2). The authors noted that microsite effects were largely ignored in their simulations and could be the causing the positive spatial correlation. The spatial variability measured in the control may be due to pockets of increased growth due to gap formation. The trees surrounding these pockets will respond via growth functions as a result of decreased competition. It may be possible that competition based mortality is creating spatial structures in the simulated data, including the control. All spatial patterns would experience a response to mortality, but since the control does not have spatial correlation from microsite effects the average change in rR^2 overtime is smaller than for the others spatial patterns.

4.4.2 Partial Redundancy Analysis

Partial redundancy analysis tests for the significance of a pure spatial component in the response data. P-values less than α=0.05 indicate the presence of spatial

dependence. Table 4.2 consists mostly of p-values below the alpha threshold, which is expected based on how spatial variables were constructed. During the construction of the spatial matrix we did a forward selection (selection criteria α=0.05) of the variables on the response. The variables selected from this process were considered significant before the pRDA.

An interesting point to mention is that we failed to reject the null hypothesis in the control plot a similar amount of times as we did the other spatial patterns. This is interesting because the control has no induced spatial dependence. Initially, we reasoned that that during the forward selection process we were retaining variables in the control plots as a result of an inflated type 1 error noted by other scientist (Blanchet et al. 2008, Legendre et al. 2011). However, it is unclear if the significant spatial variables are due to type 1 errors or if the spatial variables are detecting patterns of growth due to mortality. Furthermore, since it is possible that SI correctly accounts for all the spatial structure occurring in Y, the significant variables detected in pRDA are modeling something other than microsites. This would also explain why the success of detection

Spatial dependence of tree attributes is significant in most cases and no apparent trends across stand age exist. Bullock & Burkhart (2005) examined examined 16 plots of loblolly pine trees and used simultaneous autoregressive models to evaluate the amount of influence stems in loblolly pine plantation have on each other and also reported no significant spatial trends with stand age.

4.4.3 Cluster Analysis

Results from the cluster analysis are not promising. Other researchers using the same same algorithm (Trakhtenbrot & Kadamon 2005) have found that cluster analysis

was promising and worked well for their data, indicating it's the complex spatial nature of tree attributes that does not work well with clustering. Only in the least complex spatial pattern (the biplot) and at a stand age of 20 years was the cluster analysis able to consistently detect the correct number of groups in the data. The "Goldie Locks" zone in this case occurs when the stand is not too young or too old and only occurs in the biplot. There are two possible reasons why the Goldie Locks zone occurs. The first is that it may take greater than 10 years for differences among groups to be large enough that they can be categorized correctly. The other possible reason, indicated in the 30 year simulations, is that confounding effects of competition and mortality create too much noise in the system for cluster analysis to work. Other researchers have reported cluster analysis useful in the context of forestry data. Morsdorf et al. (2004) concluded it was feasible to segment single trees in LIDAR raw data using cluster analysis and accurately match the raw data to the field data. Furthermore, Heliölä et al. (2001) used cluster analysis to show strong relationships among beetle species in edge habitats and forest interior habitats.

Throughout the age of the stand trees are competing with each other and taking over growing space that may open randomly through mortality. Spatially constrained cluster analysis works by grouping pairs of neighboring trees together that have the highest similarity value. Competition creates dissimilarity among pairs of neighboring trees. The presence of negative autocorrelation caused by competition may decrease the distinctiveness among microsites by creating "winners" and "losers" within microsites (Magnussen 1994). By year 30 the effects of competition and mortality are too large for the process of spatially constrained cluster analysis to successfully detect which trees are growing similarly due to microsite processes (Ford 1975). Conversely, Thomson (1986)

found in his investigation on spatial pattern, microsite effects, and competitive stress that competitive effects had little effect on the spatial pattern of tree size.

Lastly, the identity of the source of success and failure, i.e. clustering algorithm or selection critierion used (CH statistic), remains unclear. The CH statistics itself is a ratio of variances and can be sensitive to changes in noise among observations.

4.5 Conclusions and Management Recommendations

Redundancy analysis is a powerful tool for understanding the variance components of a data set. Breakthroughs in designing and interpreting spatial variables have made RDA more useful to scientists and managers than before. The ability to create categories for variables and analyze them accordingly is a major benefit. However, our study is not sufficient in making management recommendations for the use of RDA to plantation forestry. PTAEDA 4.1 has limited capabilities in terms of producing environmental variables and using SI alone was found to sufficiently account for the spatial structure in Y and thus PTAEDA 4.1 was not suitable for developing data to study variation partitioning.

We found that partial redundancy analysis succeeded in testing of the pure spatial component in the response data, but probably preformed poorly for what we intended it to do. In particular, pRDA detected spatial patterns but did so equally amoung all spatial patterns and the homogonues control plot. This indicates to us that the positive spatial correlation pRDA detected was from sources present in all plots including the control. An example being growth response due to mortality.

The spatially constrained cluster analysis was unreliable. It appears the clustering algorithm or the CH statistic are too sensitive to levels of variation common in forests. Many algorithms exist for cluster analysis and the results imply that refinement or alteration of the existing methods may lead to the development of clustering approaches well suited for our needs. Further testing of selection criteria may also improve the success of the our approach to mapping. Another method that can be used to map microsites based on tree size include quartile or quintile plots. This approach assigns trees to groups according to their relative size. This is not a spatially constrained method, but based on some preliminary trials it may still be useful in the field.

Lastly, the results failed to show that RDA and pRDA are reliable methods due to the inability to our inability to simulate relevant environmental variables. Our analysis indicates that these methods should be preformed on real data where SI is not the only determinate of microsite. The method of spatially constrained cluster analysis applied here was not fully satisfactory but further work in refining the algorithm or another selection criterion may make this approach a useful tool.

4.6 Literature Cited

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, *27*(2), 93-115.

Aspinwall, M. J., King, J. S., McKeand, S. E., & Bullock, B. P. (2011). Genetic effects on stand-level uniformity and above-and belowground dry mass production in juvenile loblolly pine. *Forest Ecology and Management*, *262*(4), 609-619.

Binkley, D., Stape, J. L., Ryan, M. G., Barnard, H. R., & Fownes, J. (2002). Age-related decline in forest ecosystem growth: an individual-tree, stand-structure hypothesis. *Ecosystems*, *5*(1), 58-67.

Binkley, D., Stape, J. L., Bauerle, W. L., & Ryan, M. G. (2010). Explaining growth of individual trees: light interception and efficiency of light use by Eucalyptus at four sites in Brazil. *Forest Ecology and Management*, *259*(9), 1704-1713.

Blanchet, F. G., Legendre, P., & Borcard, D. (2008b). Forward selection of explanatory variables. *Ecology*, *89*(9), 2623-2632.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, *73*(3), 1045-1055.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*(1), 51-68.

Buford, M. A., & Burkhart, H. E. (1987). Genetic improvement effects on growth and yield of loblolly pine plantations. *Forest Science*, *33*(3), 707-724.

Burkhart, H. E., & Tennent, R. B. (1977). Site index equations for radiata pine in New Zealand. *New Zealand Journal of Forest Science*, *7*(3), 408-416.

Burkhart, H. E., Farrar, K. D., Amateis, R. L., & Daniels, R. F. (1987). Simulation of individual tree growth and stand development in loblolly pine plantations on cutover, site-prepared areas. *Virginia Polytechnic Institute and State University, Blacksburg. Pub. FWS-1-87*.

Burkhart, H. E., & Gregoire, T. G. (1994). Forest biometrics. *Environmental Statistics Handbook of Statistics*, *12*, 377-407. Elsevier,

Burkhart, H. E. (2008). Modelling growth and yield for intensively managed forests. *Journal of Forest Science*, *24*(3), 119-126.

Burkhart, H. E., & Tomé, M. (2012). *Modeling forest trees and stands*. Springer Science & Business Media, Berlin.

Burkhart, H. E., Amateis, R. L., Westfall, J. A., & Daniels, R. F. (2014). PTAEDA 4.1: Simulation of individual tree growth, stand development and economic evaluation in loblolly pine plantations. Blacksburg, Virginia, 24061.

Bullock, B. P., & Burkhart, H. E. (2005). An evaluation of spatial dependency in juvenile loblolly pine stands using stem diameter. *Forest Science*, *51*(2), 102-108.

Carmean, W. H., & Lenthall, D. J. (1989). Height-growth and site-index curves for jack pine in north central Ontario. *Canadian Journal of Forest Research*, *19*(2), 215-224.

Corral, G. R. (2014). *Investigating Selection Criteria of Constrained Cluster Analysis: Applications in Forestry* (Masters Thesis, Virginia Polytechnic Institute and State University).

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3), 483-493.

Dray, S., Pélissier, R., Couteron, P., Fortin, M. J., Legendre, P., Peres-Neto, P. R., Wagner, H. H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, *82*(3), 257-275.

Dale, M. R., & Fortin, M. J. (2002). Spatial autocorrelation and statistical tests in ecology. *Ecoscience*, 162-167.

Dale, M. R., & Fortin, M. J. (2014). *Spatial analysis: a guide for ecologists*. Cambridge University Press.

Daniels, R. F., & Burkhart, H. E. (1975). *Simulation of individual tree growth and stand development in managed loblolly pine plantations. Virginia Polytechnic Institute and State University, Blacksburg. Pub. FWS-5-75.*

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. Wiley Series in Probability and Statistics.

Ezekiel, M. (1930). Methods of correlation analysis. 427 pp. *New York and London*.

Ford, E. D. (1975). Competition and stand structure in some even-aged plant monocultures. *The Journal of Ecology*, 63(1),311-333.

Fortin, M-J, & Dale, M.B. (2005). Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge.

Fox, J. C., Ades, P. K., & Bi, H. (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, *154*(1), 261-276.

Gray, D. P., & Harding, J. S. (2011). Multi-scaled environmental determinism of benthic invertebrate communities in braided rivers: evidence for complex hierarchical controls on local communities. *Fundamental and Applied Limnology/Archiv für Hydrobiologie*, *179*(1), 3-15.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 115-146.

Griffith, D. A., & Peres-Neto, P. R. (2006). Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, *87*(10), 2603-2613.

Heikkinen, R. K., Luoto, M., Virkkala, R., & Rainio, K. (2004). Effects of habitat cover, landscape structure and spatial variables on the abundance of birds in an agricultural–forest mosaic. *Journal of Applied Ecology*, *41*(5), 824-835.

Heliölä, J., Koivula, M., & Niemelä, J. (2001). Distribution of carabid beetles (Coleoptera, Carabidae) across a boreal forest–clearcut ecotone. *Conservation Biology*, *15*(2), 370-377.

Kantavichai, R., Gallagher, T. V., & Teeter, L. D. (2014). Assessing the economic feasibility of short rotation loblolly biomass plantations. *Forest Policy and Economics*, 38, 126-131.

Legendre, P., & Fortin, M. J. (1989). Spatial pattern and ecological analysis. *Vegetatio*, *80*(2), 107-138.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm?. *Ecology*, *74*(6), 1659-1673.

Legendre, P. (2008). Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *Journal of Plant Ecology*, *1*(1), 3-8.

Legendre, P., Mi, X., Ren, H., Ma, K., Yu, M., Sun, I. F., & He, F. (2009). Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology*, *90*(3), 663-674.

Legendre, P. (2011). const. clust: space-and time-constrained clustering package. R package version 1.2. *URL: http://adn. biol. umontreal. ca/~ numericalecology/Rcode*.

Legendre, P., Oksanen, J., & ter Braak, C. J. (2011). Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, *2*(3), 269-277.

Legendre, P., Borcard, D., Blanchet, F. G., & Dray, S. (2012). PCNM: MEM spatial eigenfunction and principal coordinate analyses. R package version 2.1-2/r106.

Legendre, P., & Legendre, L. F. (2012). *Numerical ecology* (Vol. 24). Elsevier.

Liu, J., & Burkhart, H. E. (1994a). Spatial characteristics of diameter and total height in juvenile loblolly pine (Pinus taeda L.) plantations. *Forest Science*, *40*(4), 774-786.

Liu, J., & Burkhart, H. E. (1994b). Modelling inter-and intra-specific competition in loblolly pine (Pinus taeda L.) plantations on cutover, site-prepared lands. *Annals of Botany*, *73*(4), 429-435.

Magnussen, S., & Kremer, A. (1993). Selection for an optimum tree growth curve. *Silvae Genetica*, *42*, 322-322.

Magnussen, S. (1994). A method to adjust simultaneously for spatial microsite and competition effects. *Canadian Journal of Forest Research*, *24*(5), 985-995.

Mateu, J., Uso, J. L., & Montes, F. (1998). The spatial pattern of a forest ecosystem. *Ecological Modelling*, *108*(1), 163-174.

Miller, J. K., & Farr, S. D. (1971). Bimultivariate redundancy: a comprehensive measure of interbattery relationship. *Multivariate Behavioral Research*, *6*(3), 313-324.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 17-23.

Morsdorf, F., Meier, E., Kötz, B., Itten, K. I., Dobbertin, M., & Allgöwer, B. (2004). LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, *92*(3), 353-362.

Munro, D. D. (1974). Forest growth models—a prognosis. In *Growth Models for Tree and \Stand Simulation* (Vol. 30, pp. 7-21). Research Note 30. Department of Forest Yield Research, Royal College of Forestry, Stockholm.

Nilsson, U., & Allen, H. L. (2003). Short-and long-term effects of site preparation, fertilization and vegetation control on growth and stand development of planted loblolly pine. *Forest Ecology and Management*, *175*(1), 367-377.

Norcliffe, G. B. (1969). On the use and limitations of trend surface models. *The Canadian Geographer/Le Géographe Canadien*, *13*(4), 338-348.

Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G., Solymos, P., Stevens, M., and Wagner, H. (2013)."Package 'vegan'." *R Packag ver* 254: 20-8.

Páez, A., & Scott, D. M. (2005). Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, *61*(1), 53-67.

Penttinen, A., Stoyan, D., & Henttonen, H. M. (1992). Marked point processes in forest statistics. *Forest Science*, *38*(4), 806-824.

Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, *87*(10), 2614-2625.

Pretzsch, H., Biber, P., & Ďurský, J. (2002). The single tree-based stand simulator SILVA: construction, application and evaluation. *Forest Ecology and Management*, *162*(1), 3-21.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, 329-358.

Ryan, M. G., Stape, J. L., Binkley, D., Fonseca, S., Loos, R. A., Takahashi, E. N., & Silva, G. G. (2010). Factors controlling Eucalyptus productivity: how water availability and stand structure alter production and carbon allocation. *Forest Ecology and Management*, *259*(9), 1695-1703.

Stape, J. L., Binkley, D., Ryan, M. G., Fonseca, S., Loos, R. A., Takahashi, E. N., & Azevedo, M. R. (2010). The Brazil Eucalyptus Potential Productivity Project: Influence of water, nutrients and stand uniformity on wood production. *Forest Ecology and Management*, *259*(9), 1684-1694.

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), 140-145.

Thomson, A. J. (1986). Trend surface analysis of spatial patterns of tree size, microsite effects, and competitive stress. *Canadian Journal of Forest Research*, *16*(2), 279-282.

Trakhtenbrot, A., & Kadmon, R. (2005). Environmental cluster analysis as a tool for selecting complementary networks of conservation sites. *Ecological Applications*, *15*(1), 335-345.

Wensel, L. C., & Koehler, J. R. (1985). A Tree growth projection system for Northern California coniferous forests. Northern California Forest Yield Cooperative, Department of Forestry and Resource Management, University of California, Berkeley, CA.

Van Den Wollenberg, A. L. (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, *42*(2), 207-219.

Zhang, L., Gove, J. H., & Heath, L. S. (2005). Spatial residual analysis of six modeling techniques. *Ecological Modelling*, *186*(2), 154-177.

Chapter 5

Spatial Analysis of Field Data of Managed Loblolly Pine

Abstract

This study extends results from simulated data sets to real plot data. In chapter 3,
simulated data was from statistical simulations where whole plots and areas of
contrasting productivity within plots, called microsites in this report, have "tree" values
drawn from distributions with specified means and variances. The statistical simulations
consisted of subpopulations of specified means and variances without spatial
autocorrelation from competition and mortality. In chapter 4, simulated data is from the
PTAEDA 4.1 spatially-explicit individual tree stand simulator for loblolly pine (*Pinus
taeda L.*). The real plot data analyzed comes from open-pollinated, as well as clonal
plantings of intensively managed loblolly pine. We compared coefficient of variation
(CV) for height and diameter at breast height among simulated (chapters 3 and 4) and
real data sets. We found that CV of DBH differs most among the data sets. The PTAEDA
4.1 stand simulator had the largest CV values while statistical simulations generally
resulted in CV values that were lower than what was found in the field plots. The CVs for
heights were very similar across ages for all data sets. Second, we implemented a full
spatial analysis on the real data sets. A full spatial analysis consists of testing for spatial
dependence, modeling spatial variables, and mapping spatial microsites within a stand.
The full spatial analysis did not yield distinctive results between data sets from clonal and
open-pollinated stands. The results of the spatial analysis indicate that stand variability is
is largely a function of site, not genetic variability.

Keywords: Forestry, Simulation, Redundancy analysis, cluster analysis, spatial analysis.

5.1 Introduction

Spatial analysis is relevant in environmental research because measurements taken across geographic space can be influenced by spatial components (ter Braak & Prentice 1988, Leduc et al. 1992, Peres-Neto et al. 2006). Spatial structures arise from two independent processes in environmental data (Legendre 1993, Dale & Fortin 2005). Environmental factors that are spatially structured can indirectly create what's called spatial dependence (Dray et al. 2006). A number of environmental factors are recognized as being spatially structured including soil composition (Rayment & Jarvis 1999), soil nutrients (Gross et al. 1995), and weather events (Mock 1996). Spatial structures can also be created directly by biotic processes such as growth, mortality and competition (Dray et al. 2006).

Microsite (areas of contrasting productivity) effects are the main cause of positive spatial correlation (i.e. similar sized trees are clumped) in forestry (Zhang et al. 2009). Negative spatial correlation (i.e. large trees tend to neighbor smaller trees) results from intertree competition among neighboring trees (Magnussen 1994, Fox et al. 2001). Planted stands at establishment will have a spatial correlation of near zero. Overtime, spatial correlation will increase as neighboring trees respond correspondingly to their shared microsites (Magnussen 1994). Eventually the spatial correlation will decrease as intertree competition begins and trees become either "winners" of "losers" at a local scale (Ford 1975, Magnussen 1994). If competition interacts without interference the spatial correlation may approach zero as the distribution of tree sizes becomes more uniform (Ford 1975, Ghent & Franson 1986). Since forest attributes are often spatially correlated (Hou et al. 2015) it is beneficial to better understand these spatial aspects of forestry data.

Carmean & Lenthall (1989) note that groups of trees growing on separate

microsites may develop different height-age curves, resulting in what is termed

"polymorphic" curves (Burkhart & Tomé 2012, p. 138). When evaluating site

productivity, the site index (SI) concept- the average height of the dominant and

codominant part of the canopy indexed at a specific age, is commonly used (Burkhart &

Tomé 2012, p. 131). Thus, in forest plots which contain microsites, we would expect

there to be polymorphic SI curves. A consequence of polymorphism is that tree size is

dependent on its location in space, or more specifically, average tree size is dependent on

which microsite it is located in, and microsites themselves are spatially structured.

This report relates results and insights learned from a three stage approach to

understanding spatial variability in forest plots. The three-stage examined in this

dissertation include analysis from simulations of forest plots with specified means and

variances which do not include spatial autocorrelation due to competition and mortality

(chapter 3), simulations with PTAEDA 4.1 a spatially-explicit, individual tree stand

simulator which includes autocorrelation due to competition and mortality (chapter 4),

and analysis of actual field plot data. The coefficient of variation from simulations results

in chapters 3 and 4 were compared with data from plantings with open-pollinated and

clonal stock.

The simulations with groups of specified statistical properties were designed to

investigate the efficacy of spatial analysis techniques in a structured, simple forestry

setting without spatial autocorrelation caused by competition or mortality. For the

purposes of this chapter we use a control plot to compare against the PTAEDA 4.1 data

and the real plot data. The control plot of our simulated stand is designed to be

homogeneous (productivity across the site is approximately equal), but we specify mean and variance parameters from which the diameter at breast height (DBH) are drawn and used to calculate heights.

The PTAEDA 4.1 spatially explicit stand simulator was developed to model growth in loblolly pine plantations. The modeling units are individual trees, which is a characteristic of all individual tree growth simulators. Annual growth of the trees is a function of tree size, the site quality, and competition from neighbors. Furthermore, growth increments can be adjusted by stochastic elements representing genetic and microsite variability (Burkhart 2008). For our purposes, we simulated tree growth for plots of uniform productivity and across a range of 20 years and quantified CV at each age. Additional variation for site index (SI) was not specified and the random variation option for diameter and height increment was not selected.

Stand data from two sets of plot measurements were used in this chapter. The first set consists of open-pollinated loblolly pine trees and is part of a region-wide study (hereafter OP stands). These data were collected from permanent plots planted across the natural range of loblolly pine (Burkhart & Amateis 2012, Amateis et al. 2006). The second sets of data come from plantings of clonal stock of loblolly pine (hereafter CL stands). Both CL and OP plots come from plantings in the same general part of the Coastal Plain region so site and climatic conditions are similar (Figure 5.1). The OP and CL data are from stands selected by common criteria including established on a site that received intensive site preparation, free from heavy damage, trees per acre ranging from 200-700 stems per acre, and unthinned, unpruned and free from interplanting. Plot establishment guidelines for OP and CL required that plots should be located to avoid

roads, windrows, large fail spots, and free of competing arborescent vegetation.

Guidelines for plot establishment also specified rectangular plots with corners established

halfway between rows to avoid biasing the plot area. Similar data was collected at OP

and CL plots including stem maps with X and Y coordinates of each tree. Given that

these plots share similar growing conditions, stand selection criteria, measurement

procedures and protocols, these data sets were considered comparable.
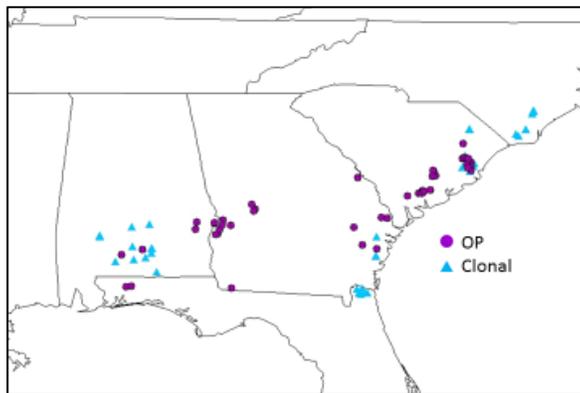

Figure 5.1 Plot locations of OP and CL stands with field plot instillations.

The spatial analysis executed consists of three steps. In the first step, we used

Moran eigenvector maps (MEM) to construct spatial eigenfunctions (also called spatial

variables) and test for spatial effects while controlling for the linear effect of our

environmental variable. The first step consists primarily of a method called partial

redundancy analysis (pRDA). In the second step we perform a variation partitioning

through 3 simple redundancy analyses (RDA) of our variables. In the last step, we map

the spatial patterns with spatially constrained cluster analysis to identify subgroups in the

stand (microsites). Finally, we validate some of our results with univariate tests and

descriptive statistics.

Our investigation applied spatial analysis techniques to field data and compared those results with findings from previous chapters. We compared coefficient of variation of data from statistical simulations, PTAEDA 4.1, OP, and CL, then implemented a spatial analysis on two comparable sets of plot measurements (OP, CL). Our hypothesis is: Without confounding of microsite and genetic variability spatial patterns will be more identifiable.

5.2 Methods

The data for this study comes from four sources; statistical simulations (called ideal stands) from R software, PTAEDA 4.1 simulator data, OP stand data, and CL stand data. The purpose of using the four data sources is to demonstrate the relationship between simulated data and real data. The simulated stands were constructed with no specified spatial dependence for consistency and efficiency.

Stand data comes from two comparable sets of plot measurement: (1) OP plots with microsite and genetic variability confounded in the data, and (2) CL plots without confounding of genetic variability. Both sets of plot measurements are from similar Coastal Plan soils and climatic zones (Figure 5.1). OP and Cl plots had the same stand selection standards for installing plot in a given plantation, measurement procedures, and protocols.

The ideal stands are statistical simulations and were simulated based on average size and not age, per se. Ideal stands were simulated with 625 trees (25 rows and 25 columns) on an 8ft by 8ft spacing. Means and variances were specified for tree attribute distributions. We specified a mean DBH of 5 inches for the ideal stand. Next, we calculated DBH variance for five scenarios of variability based on specified levels of CV

(5%, 10%, 15%, 20%, and 25%). The resulting levels of DBH variance were σ^2=(

0.0625, 0.25, 0.5625, 1, 1.5625). For one scenario at a time, we randomly sampled 625

DBH values with mean 5 inches and at a given variance. Then for each DBH we

calculated height from equations by Sabatia and Burkhart (2013). Next, site index values

were drawn randomly from a single SI distribution. The ideal was assigned only one

distribution for SI of 70ft to reflect uniform productivity. Site index values for each tree

were then drawn from a normal distribution, where $SI_i \sim N(70, 7^2)$. Site index values

were then assigned to each tree. Higher values of SI were assigned to larger trees and

smaller values of SI were assigned to smaller trees. Finally, each of the 625 trees were

assigned an X and Y spatial coordinate on a map of the stand (Figure 5.2).
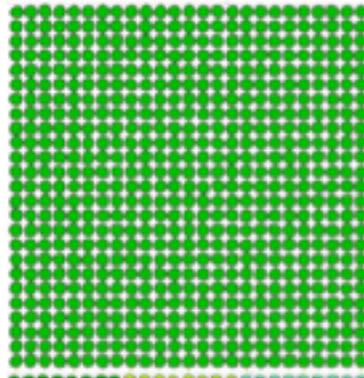

Figure 5.2. Illustrates the ideal stand. All trees are present and the stand and spatial
dependence is not induced..

PTAEDA 4.1 simulations were conducted for ages 8-23 years and data were

summarized for each year. In the PTAEDA 4.1 simulations the SI of the whole stand was

specified as 70 and grew trees were on an 8ft by 8ft spacing (Figure 5.3). For our

purposes, we did not specify CV for SI or opt for adding variation to increment growth.

The purpose of this was to establish a baseline for CV that represents the inherent error in

the modeling equations for growth. For each of the sixteen years we examined the CV for TH and DBH.
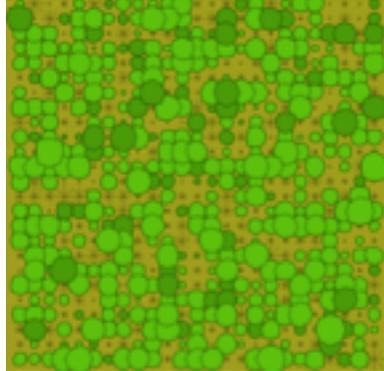

Figure 5.3. PTAEDA 4.1 simulated plot with competition and mortality effects.

For the actual field plot data (CL and OP) we calculated the total CV for TH and DBH at each age. For the CL plots the age range was 2 years to 9 years since plot establishment and the OP data range was 3 years to 23 years since plot establishment. After calculations of CV were complete we did a full spatial analysis with data from CL and OP using *R* software packages *Vegan* (Oksanen et al. 2013)*, Const.Clust* (Legendre 2011)*,* and *PCNM* (Legendre et al. 2010) for three CL stands and three IMP stands.

Our analysis was designed with multivariate response Y=(DBH, HT). For all analyses done in this chapter response variables have been standardized. We used Moran eigenvector maps (MEM's), as described by Borcard & Legendre (2002) and Dray et al. (2006), to create spatial eigenfunctions, (also called spatial variables. Spatial variables result from a principal coordinate analysis (PCoA) of the truncated pairwise distance matrix. Dray et al (2006) provide a clear description of statistical details for MEM's. The matrix of spatial variables was then $W_{nxp}=(\vec{v_1} ... \vec{v_p})$ Site index was set as an environmental variable X=(SI). Site index is a measure of site productivity and is related

to a number of environmental variables that contribute to tree growth. For this reason SI is used as a proxy variable for environmental factors which contribute to tree growth.

The analysis included variation partitioning using RDA to estimate the canonical R^2 (rRDA), testing spatial variables while controlling for the linear effects of SI with pRDA, and mapping of microsites using spatially constrained cluster analysis. Mapping results were compared with quartile plots using *R* software *geoR* package, which is univariate and mapped TH. Partial RDA results was compared with Moran's I test for spatial dependence, which is a univariate method and is also part of *R* software package *geoR*.

5.3 Results

The results for comparison of CV values are best illustrated in two plots. Figure 5.4 illustrates how CV changes over time for DBH of all stands of the two real plot data sources (OP, CL), PTAEDA 4.1 data, and ideal stands data (fixed at 10 years) with varying levels of CV specified for simulations. The PTAEDA 4.1 simulator has the largest CV over time and the CL plots have the smallest CV over time. However, at stand age 6 there is a sharp increase in CV for the CL plots, which by age 8 becomes approximately equal to CV of the OP plots. The ideal stand has comparable CV as CL and OP plots when specified CV (20% and 25%). Similarly, Figure 5.5 illustrates how CV for TH changes over time for the two data sources, PTAEDA 4.1 and the ideal stand across time.
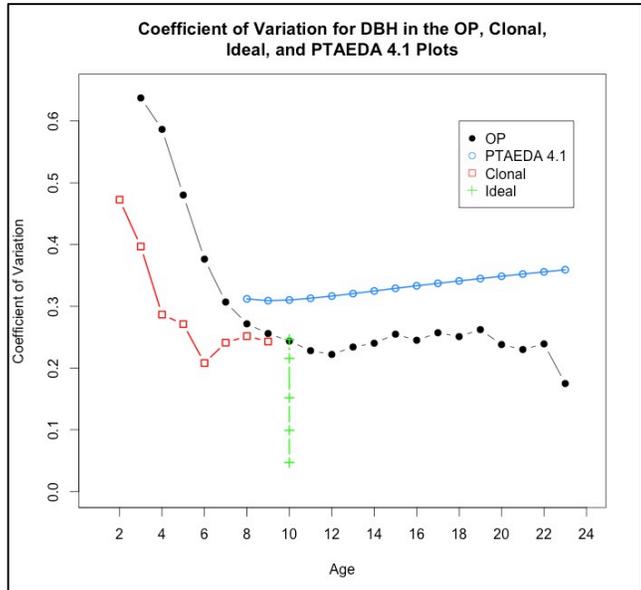
Figure 5.4. Illustrates the relationship between DBH CV of each source of data. The statistical simulation of the ideal stand (age 10 at CV 5%, 10%, 15%, 20% and 25%), PTAEDA 4.1 stand simulator (age 8-23), and OP (age 3-23) and Clonal (age 2-9) data sets.
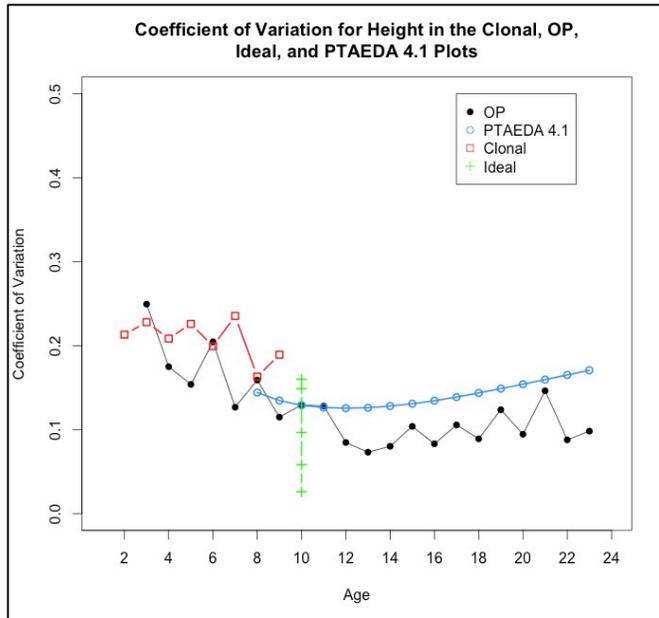

Figure 5.5. Illustrates changes in CV of height across time for OP (age 3-23), Clonal (age 2-9) and PTAEDA 4.1(age 8-23) data sets, and ideal stand (age 10 at CV 5%, 10%, 15%, 20% and 25%).

In Figure 5.5 CV values for OP, CL, and PTAEDA 4.1 are very similar and the ideal

stand reaches similar values to the other data sources at its three highest levels (15%,

20%, 25%).

Next, we implemented a full spatial analysis. Table 5.1 displays the results from

the spatial analysis. We used pRDA to test spatial variables while controlling for the

effects of SI. We found no significant relationship with the spatial variables even though

Moran's I detected spatial dependence in four of the six plots. Also, of note is the

indication of positive spatial correlation on all six plots. We also list the results from

RDA, and cluster analysis in Table 5.1.

Table 5.1. This table lists the relevant statistics from the spatial analysis. Canonical R^2
(pure spatial variability) from the RDA, pvalue from the pRDA (spatial dependence test),
estimated number of groups from cluster analysis, Moran's I pvalue of spatial dependence
(* indicates significance), and estimates of positive or negative autocorrelation.

Plot	Age	R Squared	pRDA pvalue	Groups	Moran's Pvalue	Spatial Correlation
CL 8	5	0	0.09	>10	0.94	+
CL 17	7	0.17	0.26	2	0.004*	+
CL 9	9	0.12	0.96	2	0.311	+
OP 3101	5	0.02	0.26	9	0*	+
OP 3106	7	0.09	0.16	3	0.017*	+
OP 7106	9	0.56	0.89	2	0*	+

Figures 5.6-5.11 are maps illustrating the spatial heterogeneity of growth patterns. Each

figure contains a quartile plot and a map from the cluster analysis. The quartile plots were

constructed by grouping the height data into 4 height levels. The heights are put in

ascending order based on height. Level 1 has the shortest trees, level 2 has the trees larger

than level 1 but shorter than level 3, level three contains trees taller than level 2 but

shorter than level 4, and level for has the tallest trees. Each level consists of 25% of the

data. The maps generally seem to line up with large groupings indicating patches of

higher or lower site productivity.
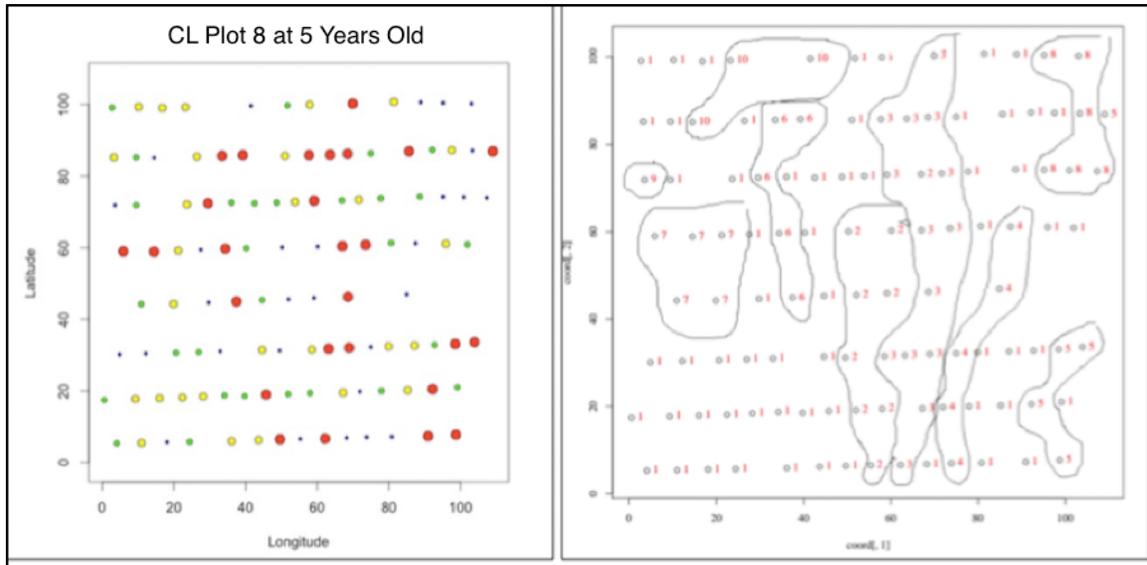
Figure 5.6. Clonal plot 8 at age 5. Left side: Quartile plot shows trees color coded into 4 groups (red=largest, yellow=second largest, green= third largest, black=smallest). Spatially constrained cluster map is displayed on right (number of microsites >10)
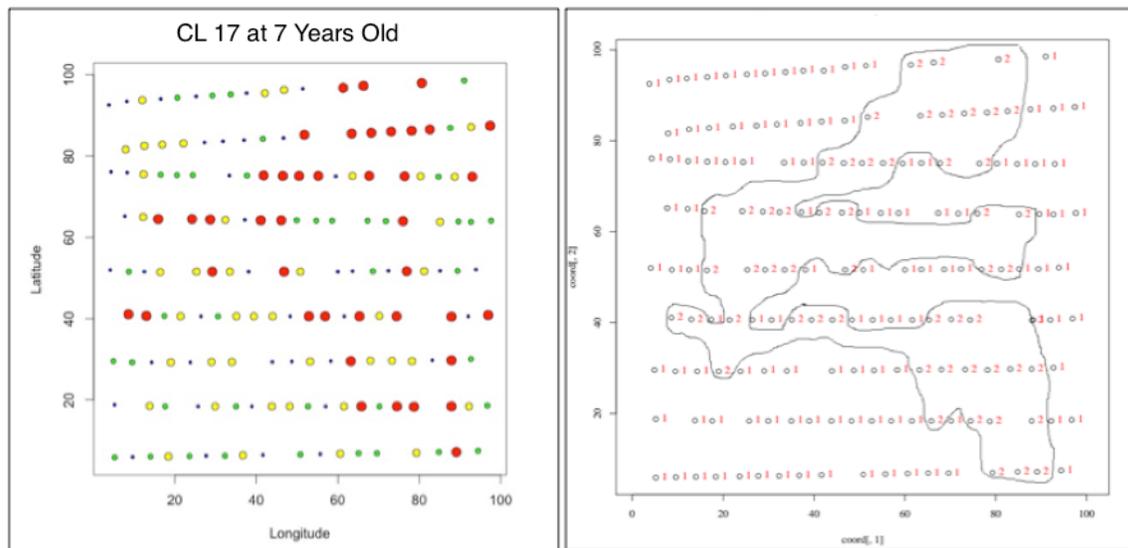
Figure 5.7. Clonal plot 17 at age 7. Left side: Quartile plot shows trees color coded into 4 groups (red=largest, yellow=second largest, green= third largest, black=smallest). Spatially constrained cluster map is displayed on right (number of microsites 2).
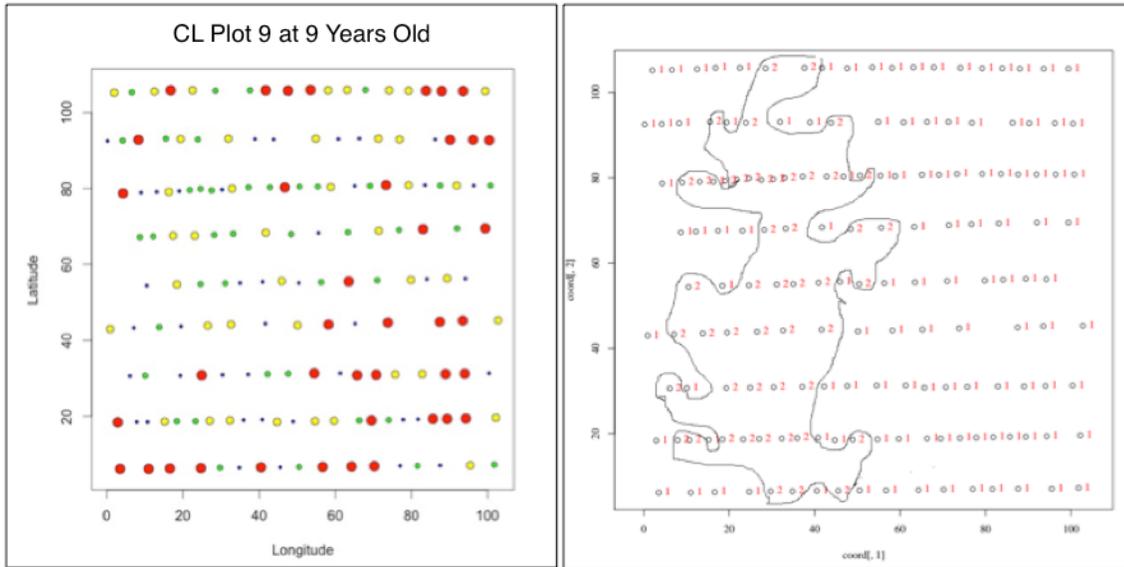
Figure 5.8. Clonal plot 9 at age 9. Left side: Quartile plot shows trees color coded into 4 groups (red=largest, yellow=second largest, green= third largest, black=smallest). Spatially constrained cluster map is displayed on right (number of microsites 2).
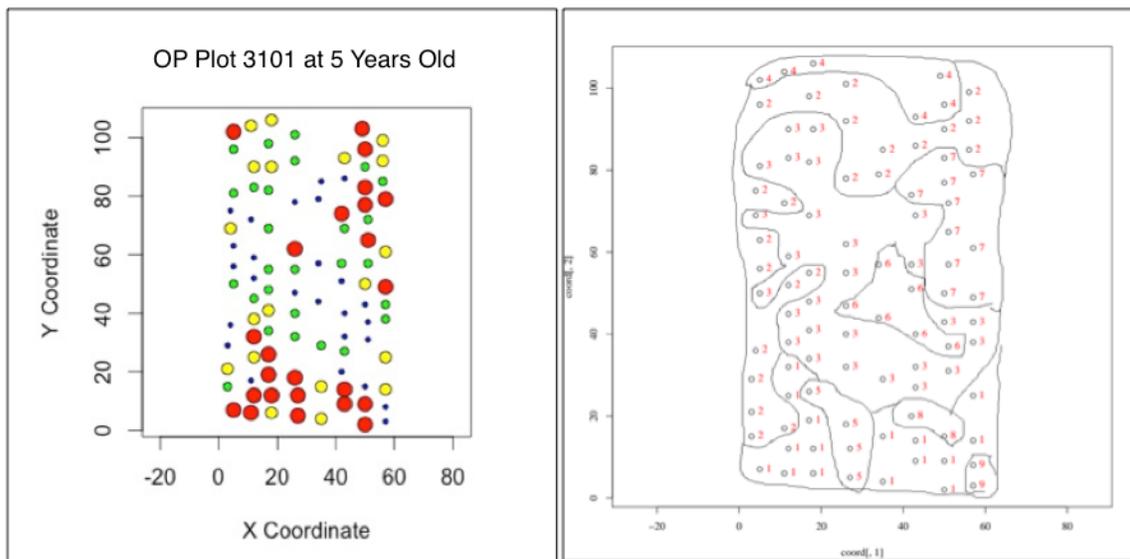
Figure 5.9. OP plot 3101 at age 5. Left side: Quartile plot shows trees color coded into 4 groups (red=largest, yellow=second largest, green= third largest, black=smallest). Spatially constrained cluster map is displayed on right (number of microsites 9).
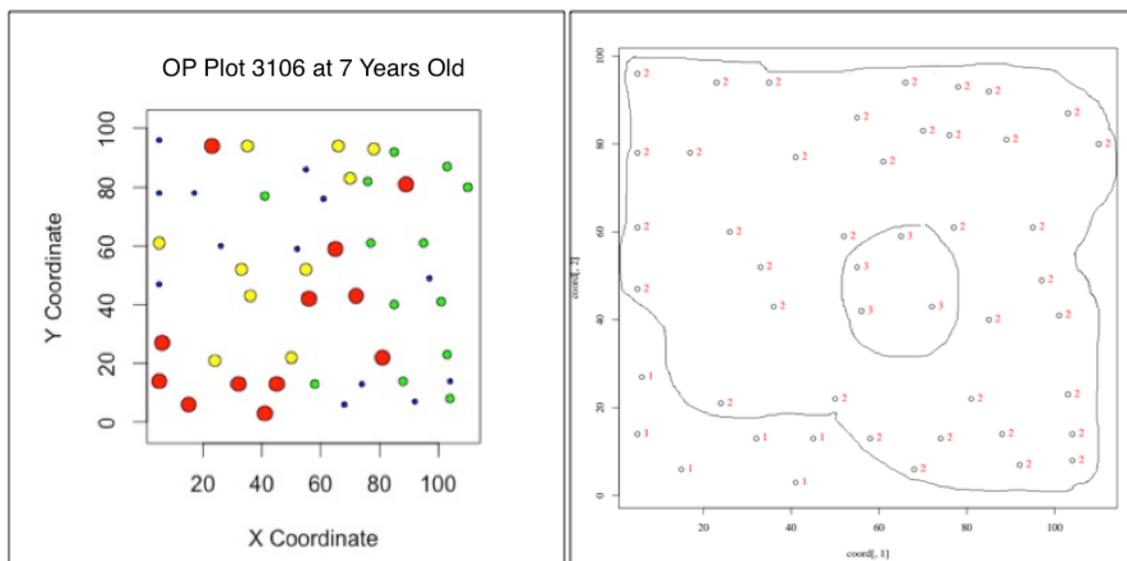
Figure 5.10. OP plot 3106 at age 7. Left side: Quartile plot shows trees color coded into 4 groups (red=largest, yellow=second largest, green= third largest, black=smallest). Spatially constrained cluster map is displayed on right (number of microsites 3).
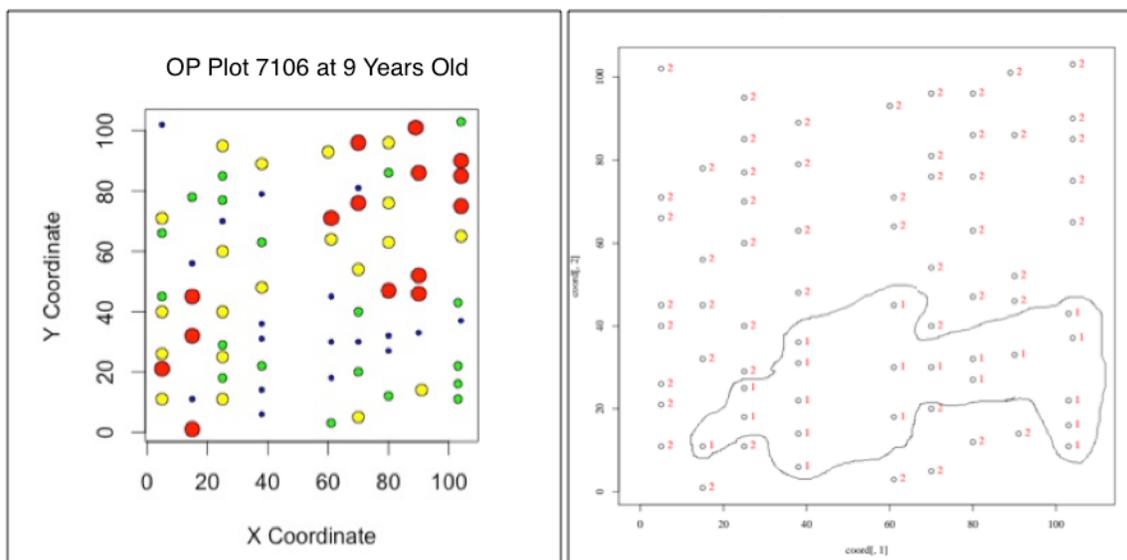
Figure 5.11. OP plot 7106 at age 9. Left side: Quartile plot shows trees color coded into 4 groups (red=largest, yellow=second largest, green= third largest, black=smallest). Spatially constrained cluster map is displayed on right (number of microsites 2).

5.4 Discussion

Coefficient of variation for DBH behaved as expected across age. The OP stands were generally higher in CV than the Clonal stands which is expected. The clonal stands are expected to have higher uniformity in growth than the nonclonal stands because they capture additive and non-additive genetic variance (Baltinus et al. 2007). The PTAEDA 4.1 simulations yielded the highest values of CV across stand age. A number of causes could be involved, but it likely results from DBH variability incorporating error from multiple sources into the system (i.e. the sampling DBH distribution, competition, and random mortality). The sensitivity analysis had a fairly low level of CV, but the values reflect the CV levels incorporated into the sensitivity analysis (5% ,10%, 15%, 20%, and 25%) with the top two points (20% and 25%) representing the actual data.

The analysis of height data yielded interesting results. First, we see both the PTAEDA 4.1 data and the sensitivity analysis represent the actual data well. A surprising result was higher CV values for CL stands versus OP stands. Our results here are similar to those of Aspinwall et al. (2011) who reported varieties were less uniform in tree height than open pollinated and control pollinated families. Increased uniformity in varietal stands is often only supported by theoretical reasons. Empirical evidence for increased uniformity is inconsistent. Sabatia & Burkhart (2013) reported that more intense genetic selection decreased variance of height distributions at a given DBH.

The full spatial analysis of the six stands yielded similar results for OP and CL plots. In Figures 5.4 and 5.5, CV values are highest at young ages and decrease over time. The large amounts of variation in the CL and OP plots make it difficult to detect significant relationships. Magnussen (1994) found that spatial autocorrelation was near

zero due to competition effects not setting in and microsite effects not developing, but Figures 5.6-5.11 do not appear to have a uniform distribution of tree sizes, rather identifiable clump of big or small trees appear in each. Magnussen (1994) points out that spatial dependence due to microsites shows up before competition effects. According to our analysis all plots have indications of positive spatial correlation, which would support Magnussen's (1994) claim because positive spatial correlation is associated with microsite effects. Due to the young age of these stands, it is reasonable to assume that the spatial dependence detected in some of the plots is likely due to microsites. The rR^2 values increase up to 55%, similar to values reported by Cottenie (2005), but the pRDA test fail to detect significance. The Morans I coefficient for height, however, detected significant spatial dependence in four of the six plots. The high levels of CV are likely the reason pRDA tests did not detect significant relationships, Chapter 3 of this dissertation showed that successful detection of microsites declines with increased levels of CV.

Constrained cluster analysis yielded some interesting results. Each map, in at least one instance, reflects the same patterns seen in the plots for height quartiles plots. This validates some results seen in the cluster analysis. That is, cluster analysis is grouping trees of similar size together. With that said, constrained cluster analysis is not a statistical test, but rather a method for finding similar objects in a data set and grouping them together. These maps came out better than expected, in part because we tested these methods in simulations that had, in some instances, more variability than the actual data, or because of the change in temporal scale and plot size from the simulated data (Chapters 3 & 4) to field data. The smaller, younger filed plots are less intricate. In field

plots, after establishment, tree growth is dependent on the productivity of the microsite it is growing on (Magnussen 1993). As competition sets in and trees become "winners" or "losers" and negative autocorrelation sets in - obscuring positive autocorrelation caused by microsites and moving the stand towards negative spatial correlation (Perry 1985). Eventually, competition based mortality occurs removing the "loser" trees from the stand which creates a uniformity and the stand becomes more uniform (Ford 1975). This uniformity, or random spatial structure, causes researchers to incorrectly deduce that tree attributes are spatially independent (Lepš & Kindlmann 1987). Larger plots which develop over up to 30 years (chapter 4) have forces of positive and negative correlation occurring concurrently over time and tree attributes dependent on microsite may only be detectable in small windows of time. Furthermore, the larger plot sizes have competition and mortality occurring within microsites which also makes microsite detection difficult. These maps, particularly the maps with fewer groups, represent structural patterns well and may be useful for future applications.

It is evident from the results of this chapter that removing genetic variability in plots will not necessarily make the spatial patterns more identifiable. Therefore, the hypothesis of this investigation is rejected. Evidence suggests that the primary source of variability in plots appears to be from soils or other site factors, not genetics. Figures 5.6-5.11, display plots with identifiable clumps of large trees and small trees in a nonrandom pattern. This indicates microsite effects are present were groups of trees in the same microsite grow similarly. A positive spatial correlation is indicative of this phenomena and Table 5.1 provides evidence of positive spatial correlation in all plots. Despite being able to visually identify patterns on the maps, which are likely areas of contrasting

productivity, the techniques we used for spatial analysis were still unable to perform reliably. Based on results form early chapters we contribute the inability of the methods used to perform reliably to large amounts of variability within microsites.

5.5 Conclusions and Management Recommendations

Simulations offer a controlled way to test the efficacy of statistical methods, study possible management scenarios/outcomes, and study stand dynamics. Our approach for simulations was found to have a reasonably close CV compared to actual plot data. These methods of simulation appear viable and should be useful to managers for a variety of purposes.

The results from the full spatial analysis are a reasonable start, but more replications across longer time periods are needed to fully understand the usefulness of the spatial approach. The methods we used were not reliable for the stands we examined and the probable cause is the young age (5, 7, and 9) and high amounts of variation in the data. We found that microsite variability was likely the primary cause for large variation in data and that data from clonal plantings did not sufficiently reduce the variability.

Spatially constrained cluster analysis provided useful maps which grouped trees of similar size together. The spatial pattern provided by the constrained clustering offers some insight into microsite conditions of the stand. Managers may consider using these maps in young, small plots to locate clumps of of smaller trees. This may be a useful way to target areas of low productivity within the plot. This information could be useful for management goals concerned with increasing uniformity.

In conclusion, it is necessary to continue research on spatial processes in plots over longer periods of time and under multiple conditions. In particular, more research on how the relative strengths of different spatial processes like those created by microsites and by competition influence tree growth and stand structure. This chapter highlighted that variability in plot data is not sufficiently reduced by removing the confounding of genetic effects for methods to reliably detect spatial patterns.

5.6 Literature Cited

Amateis, R.L., Prisley, S.P., Burkhart, H.E., and Liu, J. (2006). The effect of physiographic region and geographic locale on predicting the dominant height and basal area of loblolly pine plantations. *Southern Journal of Applied Forestry*. 30(3):147–153.

Aspinwall, M.J., King, J.S., McKeand, S.E., Bullock, B.P., (2011). Genetic effects on stand-level uniformity and above- and belowground dry mass production in juvenile loblolly pine. *Forest Ecology and Management*. 262(4), 609-619.

Baltunis, B.S., Huber, D.A., White, T.L., Goldfarb, B., Stelzer, H.E., (2007). Genetic analysis of early field growth of loblolly pine clones and seedlings from the same full-sib families. *Canadian Journal of Forest Research*. 37, 195-205.

Burkhart, H. E. (2008). Modelling growth and yield for intensively managed forests. *Journal of Forest Science*, *24*(3), 119-126.

Burkhart, H. E., & Amateis, R. L. (2012). Plot Installations for Modeling Growth and Yield of Loblolly Pine Plantations. In: *Forest Observational Studies, Proceedings of an International Workshop (Zhao Xiu Hai, Zhang Chun Yu, and Klaus V. Gadow, eds.), Beijing Forestry University, China, p.17-34.*

Burkhart, H. E., Amateis, R. L., Westfall, J. A., & Daniels, R. F. (2014). PTAEDA 4.1: Simulation of individual tree growth, stand development and economic evaluation in loblolly pine plantations. Blacksburg, Virginia, 24061.

Burkhart, H. E., & Tomé, M. (2012). *Modeling forest trees and stands*. Springer Science & Business Media.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*(1), 51-68.

Carmean, W. H., & Lenthall, D. J. (1989). Height-growth and site-index curves for jack pine in north central Ontario. *Canadian Journal of Forest Research*, *19*(2), 215-224.

Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, *8*(11), 1175-1182.

Dale, M. R., & Fortin, M. J. (2014). *Spatial analysis: a guide for ecologists*. Cambridge University Press.

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3), 483-493.

Ford, E. D. (1975). Competition and stand structure in some even-aged plant monocultures. *The Journal of Ecology*, 63(1), 311-333.

Fox, J. C., Ades, P. K., & Bi, H. (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, *154*(1), 261-276.

Ghent, A. W., & Franson, S. E. (1986). Changes in mortality and size-class spatial distribution patterns in pre-closure and post-closure conifer plantations. *Forest Science*, *32*(3), 559-575.

Gross, K. L., Pregitzer, K. S., & Burton, A. J. (1995). Spatial variation in nitrogen availability in three successional plant communities. *Journal of Ecology*, 83(3), 357-367.

Hou, Z., Xu, Q., Hartikainen, S., Antilla, P., Packalen, T., Maltamo, M., & Tokola, T. (2015). Impact of plot size and spatial pattern of forest attributes on sampling efficacy. *Forest Science*, 61(5), 847-860

Leduc, A., Drapeau, P., Bergeron, Y., & Legendre, P. (1992). Study of spatial components of forest cover using partial Mantel tests and path analysis. *Journal of Vegetation Science*, 3(1), 69-78.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm?. *Ecology*, *74*(6), 1659-1673.

Legendre, P., Borcard, D., Blanchet, G., & Dray, S. (2010). PCNM: PCNM spatial eigenfunction and principal coordinate analyses. R package version 2.1/r82.

Legendre, P. (2011). const. clust: space-and time-constrained clustering package. R package version 1.2. *URL: http://adn. biol. umontreal. ca/~ numericalecology/Rcode*.

Lepš, J., & Kindlmann, P. (1987). Models of the development of spatial pattern of an even-aged plant population over time. *Ecological Modelling*, *39*(1), 45-57.

Magnussen, S. (1994). A method to adjust simultaneously for spatial microsite and competition effects. *Canadian Journal of Forest Research*, *24*(5), 985-995.

Mock, C. J. (1996). Climatic controls and spatial variations of precipitation in the western United States. *Journal of Climate*, *9*(5), 1111-1125.

Perry, D. A. (1985). The competition process in forest stands. *Attributes of trees as crop plants*, 481-506.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G.L., Solymos, M.H., Stevens, H. & Wagner, H. (2013). Package 'vegan'. *R Packag ver*, *254*, 20-8.

Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, 87(10), 2614-2625.

Rayment, M. B., & Jarvis, P. G. (2000). Temporal and spatial variation of soil CO_2 efflux in a Canadian boreal forest. *Soil Biology and Biochemistry*, *32*(1), 35-45.

Sabatia, C. O., & Burkhart, H. E. (2013). Height and diameter relationships and distributions in loblolly pine stands of enhanced genetic material. *Forest Science*, *59*(3), 278-289.

Sabatia, C. O., & Burkhart, H. E. (2014). Predicting site index of plantation loblolly pine from biophysical variables. *Forest Ecology and Management*, *326*, 142-156.

ter Braak, C.J., & Prentice, I.C. (1988). A theory of gradient analysis. *Advances in Ecological Research*, 18, 271-317.

Zhang, L., Ma, Z., & Guo, L. (2009). An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models. *Forest Science*, *55*(6), 533-548.

Chapter 6

Summary, Implications, and Recommendations

6.1 Summary and Implications

Our approach to studying spatial variability included three analytical

stages of investigation. First, we studied simulated populations which were reasonably

simple and included only induced spatial dependence caused by microsite. Induced

spatial dependence occurs when the response data takes on the spatial structure of

environmental variables that are, themselves, spatially structured. Second, we

investigated simulated stands with PTAEDA 4.1 which included both induced spatial

dependence and spatial autocorrelation caused by competition and mortality. Spatial

autocorrelation occurs when observations in geographic space are dependent on each

other. In the third stage of investigation we examined actual plot data which,

theoretically, included both induced spatial dependence and spatial autocorrelation. Our

approach brought to light a few important points. Spatial dependence of tree attributes is

a balancing act between two processes: induced spatial dependence of microsites and

spatial autocorrelation due to competition and mortality. Moreover, it is evident that the

relative strengths of these processes determines efficacy of the spatial analysis methods

utilized.

6.1.1 Summary

The statistical simulations (chapter 3) investigated how variation partitioning

(Borcard et al. 1992) via redundancy analysis (RDA) (Rao 1964) responded to different

levels of coefficient of variation (CV), range of differences in average tree diameter

within microsites, and spatial pattern complexity. The response to varying levels of the parameters was unexpected. In particular, the variation explained by the pure spatial component, or canonical adjusted R^2 (rR^2), was lowest in the biplots and increased as spatial pattern complexity increased, while dropping some in the free plot.

Decreasing the difference in the size parameter among microsites decreased the amount of detectable spatial variation, with the exception of the free plot where detection levels remained fairly constant as average diameter differences increased. When average size of the trees among microsites is more similar, the overlap of distributions among microsites increased and the variation caused by spatial location becomes indistinguishable from random noise. The consequences are that the magnitude of growth potential and site productivity play an important role in determining spatial variation (see also Magnussen 1990, Magnussen 1994, Dutkoski 2002).

Increases in the CV of tree diameter distributions decreased the amount of spatial variability RDA can detect. This occurred in large part because the statistical distributions of tree size flatten out. In other words, the normal distributions become less peaked and narrow. If we were to graph the distributions of each microsite, we would see an increased amount of overlap of the distributions. The point being, that if we increase the CV we loose ability to detect differences in individual tree characteristics. The consequences here are rather general and build on known concepts. In order to better detect spatial patterns of stands we need have relatively large differences in means and relatively small CV values.

Partial RDA (pRDA) (ter Braak 1988) was used to test for pure spatial effects in the data. The probability of successfully detecting spatial dependence decreased with

increased levels of CV, with the exception of the free plot which remained 100% for all simulations. Spatial pattern complexity had no clear trend for detection except, higher complexity did not seem to decrease success rates. Furthermore, we observed a decrease in probability of detection as mean differences among microsites decreased. The reason these factors influence the probability of success of pRDA are the same as RDA. Any factor that increases the random variation of the system or dilutes and flattens group distributions is going to influence the outcome of pRDA testing.

Spatially constrained cluster analysis appeared to be very sensitive to changes in CV, mean difference of parameters among microsites, and spatial pattern complexity. Constrained cluster analysis groups trees together based on their similarity. The method works best when groups are well defined and have low variability. Small changes appear to create sufficient ambiguity such that separate groups have sufficient overlap to enter the same cluster. The sensitivity of cluster analysis to variation and spatial complexity were concerning, but with low misclassification rates it was decided more analysis would be necessary. Given the results of the sensitivity analysis we deemed clustering to have potential for accurate mapping and decided to move forward with more testing.

Simulations with PTAEDA 4.1 were inconclusive. In these simulations we investigated RDA, pRDA, and constrained cluster analysis for five spatial patterns. For each of the five spatial patterns we applied our methods to a combination of CV (5 levels) and stand age (3 levels). Spatial pattern complexity and CV had little effect on the results (except for clustering), but stand age was influential.

Redundancy analysis and pRDA results were not valid due to the use of the SI as a single explanatory variable. The SI variable accounted for the spatial structure in the Y

data rendering the spatial variables extraneous. The rR^2 values measured were not large, but that was expected. Stand age appeared to affect the outcome of pRDA as most p-values that were larger than the specified alpha level occurred in the simulated 30-year-old stand. The 30-year-old stand was influenced over the years by competition and mortality. More prominent and cumulative effects of competition and mortality overshadowed the spatial dependencies, which resulted from variations of productivity.

Spatially constrained cluster analysis was not a reliable method to map microsites in the data. There were a small number of successful simulations and those occurred only in the 20-year old stands of the biplot. The clustering algorithm appeared to be very sensitive to increases in spatial pattern complexity and age. The results indicate that the 10 year old stand had not displayed enough growth differences among microsites to be efficiently grouped and stands at 30 years old exhibit the cumulative effects of years of competition and mortality which overshadowed the growth patterns of spatial dependence.

Comparison of simulated data to field data yielded some interesting results. We found that CV for both the statistical simulations and PTAEDA 4.1 were similar to field data, particularly with CV of height. The CV for DBH tended to be underestimated by the statistical simulations and overestimated with PTAEDA 4.1. The higher variation levels present in the simulations lead to some new insights into clustering analysis. Clustering analysis was found to be very sensitive to changes in CV in earlier chapters of this dissertation. When the field data yielded CV's that were somewhat lower than expected it became a good opportunity to use cluster analysis. The maps created with clustering appeared to be accurate, for the most part, in terms of grouping trees of similar structure

together. The cluster maps can be compared to quartile maps for validation. The full spatial analysis was unreliable and unaffected by data source. This result may stem from a number of reasons, but based on the results of the earlier chapters it was most likely due to the higher levels of CV at the young ages (less than 10 years).

6.1.2 Implications

We found strong effects of competition on spatial structure of plots. In particularly, the ability of competition to produce negative spatial correlation that overcomes, or hides the induced spatial dependence caused by microsites. Further complicating one's ability to detect sources of spatial variation is mortality. Specifically, the resulting stand uniformity created by the dying of "loser" trees. Our results also indicate that cluster analysis may be most effective on smaller scale plots or only for plots that are within a specific "age window". This unexpected result may be due to the balancing between induced spatial dependence and spatial autocorrelation effects caused by mortality.

Multiple authors have noted the concurrent effects of multiple sources of spatial autocorrelation due to microsites, competition, and mortality (see Ford 1975, Magnussen 1994, Fu et al. 1999, Harper et al. 2011). More specifically, Fox et al. (2007) notes that the magnitude and sign of spatial dependence in a stand depend on the relative strength of two processes: induced spatial dependence caused by microsites and spatial autocorrelation due to competition. The transition from plots with induced spatial dependence to plots with induced spatial dependence plus spatial autocorrelation due to competition and mortality (chapter 3-4) brought forth a decrease in the amount of spatial variability estimated by RDA. It is important to acknowledge that this is not a perfect side

by side comparison where all was held equal except the presence or absence of spatial autocorrelation due to competition and mortality. However, our data is in line with the theory that microsite effects become less recognizable with the onset of competition effects and competition based mortality. In chapter 5, rR^2 values ranged from 0-55%. These values are indicative of stands that may contain one or both types of spatial dependence. The stands are young enough (5-9 years after plot establishment) that microsite effects may be distinguishable and, competition and mortality effects have not greatly affected tree attributes. Unfortunately, it is also apparent that while microsite effects may be present, small differences in average size among microsites and high CV values make testing of the pure spatial component difficult as demonstrated in chapters 3 and 4.

Testing of individual components, namely the pure spatial components, via pRDA is a little more difficult to interpret. In chapter 3, pRDA was often successful in detecting spatial dependence and spatial pattern complexity did not seem to influence the results. Also, the probability of detecting spatial dependence in control (homogenous) plots stayed around 5% which is expected. However, the introduction of spatial autocorrelation due to competition and mortality in chapter 4, according to the statistical evidence, that all plots – even those control plots without induced spatial dependence showed evidence of a spatial component the data. It is unclear, however, if the pRDA results in chapter 4 are due to inflated type 1 error rates as reported by Blanchet et al. (2008), or if the spatial eigenfunctions (also called spatial variables) are modeling the positive spatial autocorrelation caused by competition based mortality (Reed & Burkhart 1985). The simulator in chapter 4, PTAEDA 4.1, calculates the probability of any tree remaining

alive in given year as a function of its competitive stress and individual vigor. The probability is then used in Bernoulli trials to stochastically determine annual mortality. Thus after the establishment of the juvenile stand in PTAEDA 4.1, smaller trees die more frequently as a result of having a larger competition index (more competitive stress). The point being, positive spatial autocorrelation due to competition based mortality may be more pronounced and occur more quickly (immediately after generated juvenile stand age 8) in the simulated stand because mortality is a function of competitive stress in PTAEDA 4.1. Whereas real plot data shows positive spatial correlation only in the early stages and late stages of stand development, but often exhibits negative spatial autocorrelation after the onset of competition and before competition based mortality (Reed & Burkhart 1985). The field data of chapter 5 did not have a significant spatial component according to pRDA tests, but the data (4 out of 6 plots) were significant for spatial dependence according to Moran's I. Based on the results from chapters 3 and 4, as well as Figures 5.4 and 5.5, the disagreement between the Moran's I result and pRDA can and large part be explained by high levels of CV at young ages and possibly small differences in average size of trees in different microsites.

Cluster analysis had moderate success in detecting and mapping microsites in stands with only induced spatial dependence (chapter 3). The inclusion of spatial autocorrelation in PTAEDA 4.1 (chapter 4) simulations brought a sharp decline in the success of detecting the correct number of groups. However, many of the cluster maps in chapter 5 closely resembled the quartile plots and appeared to make useful mappings. The smaller size of the field plots and younger age of trees may work well with cluster

analyses despite higher levels of CV. Less severe competition at this stage allows cluster analysis to find large groups of similar size trees.

Understanding spatial variability of tree attributes in forest plots is essential for future management of planation stands. Before model building, testing the data for spatial dependence should be the standard. It has become apparent that treating trees as individual samples which are independent from each other is erroneous (Fox et al. 2001). Researchers simply ignore spatial dependence in order to meet the assumptions of ordinary least squares (OLS). Redundancy analysis is a multi-tool. Unlike other methods, RDA provides a straight forward method for variation partitioning, partial redundancy analysis can test the significance of spatial variables while controlling the linear effect of other covariables, RDA is multivariate linear model and can be used for prediction equations, and RDA is a multivariate extension of partial linear regression (PLR) so if univariate methods are preferred it is an easy transition and can be performed with the same software. Cluster analysis may also be useful in a more limited capacity, but more research is needed. The results of these investigations provide useful insights into understanding how relative strengths of two processes of spatial dependence (induced spatial dependence and spatial autocorrelation) determine spatial variability of a plot, or conversely spatial uniformity.

Spatial uniformity of tree attributes, also called stand uniformity, can lead to greater resource use efficiency and enhanced productivity (Nilsson & Allen 2003). In fact, Allen et al. (2005) noted that successful intensive silviculture requires an understanding of what resources limit production on temporal and spatial scales. These methods can be useful tools in better understanding and measuring changes in spatial

variability that may result from intensive management. While all methods of studying

spatial variability and, conversely, stand uniformity have limitations so too does RDA,

pRDA, and cluster analysis more research is needed.

In conclusion, this dissertation explored capabilities of modern techniques of

spatial analysis to to detect, quantify and map spatial variability in forest plots. Our three

stage analytical approach progressed through increasingly more complex scenarios.

Through this progression we were able to identify limitations of the spatial analysis

techniques utilized. Furthermore, we studied how multiple spatial processes affects

spatial patterns and spatial analysis results. Ultimately, we applied spatial analysis

techniques to comparable data sets from clonal and open-pollenated stands. We found

that the primary source of variability in forest plots, and within identifiable groups,

appears to be soil, and not genetic. Plantings with clones generally exhibited somewhat

lower levels of CV, but these small decreases were not sufficient to make detection of

patterns more reliable than what was found in plots of similar ages planted with open-

pollenated seedlings.

6.3 Recommendations for Future Research

Our study indicates that redundancy analysis and partial redundancy analysis have

promising applications in forestry. The methods of RDA and pRDA, which are

commonly used in ecology to examine community composition, apply well to studying

plot variability of tree characteristics. Our study also looked at the efficacy of using a

spatially constrained cluster analysis to map microsites based on tree characteristics;

however, we found that this was inefficient and considerations for future work should

look at applying mixtures of three factors. We suggest that future research should focus

on simulating amounts of induced spatial dependence via microsites (high, medium, low), controlling spatial autocorrelation due to competition via density (high, medium, low), and examining a range of ages after establishment (precrown closure-post competition based mortality). A study that examines the interaction effects of these factor levels would have an impact in forestry. This type of research would not only develop a greater understanding of pRDA, RDA, and cluster analysis, but could quantify how relative strengths of two processes of spatial dependence (induced spatial dependence and spatial autocorrelation) determine whole plot spatial autocorrelation (positive or negative). Furthermore, the results would help identify when transitions from positive to negative (and vice versa) spatial autocorrelation occurs and how factors relate to when the transitions occur.

RDA was applied to examine how tree attributes are explained by an environmental variable (SI) and spatial variables (spatial eigenfunctions). Redundancy analysis is not limited to these categories or even to two explanatory matrices. For future research we recommend a thorough examination of how genetic, spatial, and competition effects explain variability of tree attributes. Tree attribute variability explained by genetic and microsite factors is typically confounded. Adding an explanatory matrix for genetics and competition would be interesting. Researchers would be able to quantify the pure effects of genetics, microsites, and competition. Depending on the experimental design, the explanatory matrix for genetics could consist of dummy variables, or if more detailed genetic information is available then a principal coordinate analysis of a genetic distance matrix can be computed. The first few principal coordinates, or those selected by forward selection could be used to represent genetic variability in variation partitioning (Pierre

Legendre, pers. comm. August 11, 2015). A matrix of competition variables could be constructed with spatial eigenfunctions and selected to model the negative spatial autocorrelation effects of competintion. Analysis would be done in the same way that ours was done; variation could be partitioned by matrix type. This would give valuable insight into how genotypes perform and if there is a significant "pure" genotype component.

Another interesting research project would be to use RDA to examine space-time interactions (STI). A STI is produced by the analysis of the interaction of space and time. An STI indicates if there is a spatial structural change when time is controlled. Significant STI's are a marker of climate change (Legendre & Legendre 2012). The concept of STI has been around for some time, but a more modern approach is to use eigen decompositions for the spatial and temporal variables. Variable selection through eigen decomposition would create better variables for space and time than would otherwise exist. An investigation of this type could certainly be examined in planted even-age stands of trees, but with the flexibility of this method it could be applied to forest community composition data. Furthermore, the data could be collected at forest sites within transitional zones that may be more sensitive to climate change.

Cluster analysis was initially thought to be unreliable for our needs, but when the field data was analyzed we found that the variation levels were not so high that mapping could not be done. The cluster maps with real data were validated with quartile plots and appear, for the most part, to reflect the spatial structure of the stand. More research in selection criteria for the number of groups in this type of agglomerative cluster is needed. Also, we believe that studying the effects of extent (plot size) on the success rate of

cluster analysis would be beneficial and would help to identify when cluster analysis is

most useful.

6.4 Literature Cited

Allen, H. L., Fox, T. R., & Campbell, R. G. (2005). What is ahead for intensive pine plantation silviculture in the south?. *Southern Journal of Applied Forestry*, *29*(2), 62-69.

Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology*, *89*(9), 2623-2632.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, *73*(3), 1045-1055.

Dutkowski, G. W., Silva, J. C. E., Gilmour, A. R., & Lopez, G. A. (2002). Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research*, *32*(12), 2201-2214.

Fox, J. C., Ades, P. K., & Bi, H. (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, *154*(1), 261-276.

Fox, J. C., Bi, H., & Ades, P. K. (2007). Spatial dependence and individual-tree growth models: I. Characterizing spatial dependence. *Forest Ecology and Management*, *245*(1), 10-19.

Fu, Y. B., Yanchuk, A. D., & Namkoong, G. (1999). Spatial patterns of tree height variations in a series of Douglas-fir progeny trials: implications for genetic testing. *Canadian journal of forest research*, *29*(6), 714-723.

Harper, K. A., Danby, R. K., De Fields, D. L., Lewis, K. P., Trant, A. J., Starzomski, B. M., Savidge, R. & Hermanutz, L. (2011). Tree spatial pattern within the forest-tundra ecotone: a comparison of sites across Canada This article is a contribution to the series Tree recruitment, growth, and distribution at the circumpolar forest-tundra transition. *Canadian Journal of Forest Research*, *41*(3), 479-489.

Legendre, P., & Legendre, L. F. (2012). *Numerical ecology* (Vol. 24). Elsevier.

Magnussen, S. (1994). A method to adjust simultaneously for spatial microsite and competition effects. *Canadian journal of forest research*, *24*(5), 985-995.

Nilsson, U., & Allen, H. L. (2003). Short-and long-term effects of site preparation, fertilization and vegetation control on growth and stand development of planted loblolly pine. *Forest Ecology and Management*, *175*(1), 367-377.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, 329-358.

Reed, D. D., & Burkhart, H. E. (1985). Spatial autocorrelation of individual tree characteristics in loblolly pine stands. *Forest Science*, *31*(3), 575-587.

ter Braak, T. C. (1988). Partial canonical correspondence analysis. In *Classification and related methods of data analysis: proceedings of the first conference of the International Federation of Classification Societies (IFCS), Technical University of Aachen, FRG, 29 June-1 July 1987*. North-Holland.