Chijien Lin

GENERATING FOREST STANDS WITH SPATIO-TEMPORAL DEPENDENCIES

PREFACE

I was told that there is a long story behind each doctorate degree. I finally understood why when I started to write my own story in 1993. I had just received my master degrees in quantitative forestry and applied statistics, and I decided to investigate methods to generate hypothetical forests for my Ph.D. research at the Faculty of Forestry, State University of New York. In 1997 I met Professor Juha M. Alho who accepted me as a doctorate student at the Department of Statistics, University of Joensuu. Since 2000 I have received studentships from the Faculty of Social Sciences and the Department of Statistics, University of Joensuu, and the Finnish Graduate School in Stochastics. This allowed me to work on my dissertation full time. It was a difficult challenge for me to reinitiate the research after having been away from academia for nearly six years. However, Professor Alho reminded me of the beauty of statistical modeling, and provided me with all the resources I needed to complete my study, including long hours of tutoring and supervision. I owe a debt of great gratitude to him for his continuous encouragement and support. I am also indebted to Professors Timo Pukkala and Antti Penttinen, who have offered invaluable comments in developing the statistical models. I thank Mr. Tim Green for correcting the English of the manuscript. There are many other people who have helped me during my study: too many to mention here. I thank you all for your help. The story ended with the completion of the dissertation, but this is the start of another story. Finally, I dedicate this dissertation to Mrs. Francis Farnsworth whose strength, elegance and kindness will always be remembered.

Joensuu, November 2003

Chijien Lin

ABSTRACT

In forestry hypothetical tree populations are commonly generated for
studying statistical sampling strategies, forest development and management
planning purposes. The notion of Poisson (random) process has been
dominant in the construction of hypothetical forests since the 1960s.
However, spatio-temporal dependencies that result from growth, mortality
and reproduction processes are not reflected in the generated tree locations
and characteristics. The objective of this study is to develop methods for
simulating hypothetical forests with tree characteristics that are dependent on
the locations, that have arbitrary analytical or empirically estimated
distributions, and that have an isotropic or anisotropic spatial structure.
Statistical criteria for evaluating either homogeneous or inhomogeneous
point pattern are discussed. Transformation effects on correlations of tree
characteristics are examined. Empirical generalized linear mixed models are
developed for generating correlated tree characteristics conditioned on tree
locations. A stand simulator is developed to carry out iterative generation of
hypothetical forests. As shown in examples, spatio-temporal dependencies
can be generated through iterative simulation of growth, mortality and
reproduction processes, as an alternative to reproducing the dependencies via
covariance functions. In a forestry application, the generated hypothetical
forests are used to study the accuracy of location independent thinning
models.

TIIVISTELMÄ

Chijien Lin, Tila-aika -riippuvien metsiköiden generointi.
Yhteiskuntatieteellisiä julkaisuja 64, Joensuun yliopisto 2003, 123 s.

Avainsanat: Hypoteettinen metsä, Tila-aika -riippuvuus, Epähomogeenisuus,
Ekologiset prosessit, Yleistetyt lineaariset sekamallit, Muutoksen vaikutus,
Iteratiivinen simulointi, SPATE.

Metsätieteissä käytetään hypoteettisia puupopulaatioita tilastollisen otannan,
metsän suunnittelun ja metsän kasvun tutkimuksen apuvälineenä.
Hypoteettisten metsien luominen on perustunut 1960-luvulta lähtien
Poisson-prosessiin. Puiden kasvusta, kuolemasta ja uudistumisesta
aiheutuvia tila-aika -riippuvuuksia ei ole yleensä otettu huomioon puiden
sijainneissa eikä ominaisuuksissa. Tämän tutkimuksen tarkoituksena on
kehittää menetelmiä hypoteettisten metsien simulointiin, joissa puiden
ominaisuudet ovat riippuvia niiden sijainneista. Ominaisuudet voivat
noudattaa mielivaltaisia teoreettisia tai empiirisesti estimoituja jakaumia.
Niillä voi olla joko isotrooppinen tai ei-isotrooppinen spatiaalinen rakenne.
Pistekuvio voidaan valita homogeeniseksi tai epähomogeeniseksi tilastollisin
kriteerein. Puiden sijainnista riippuvien ominaisuuksien luontiin käytetään
empiirisiä yleistettyjä lineaarisia sekamalleja. Hypoteettisten metsien
iteratiiviseen simulointiin on kehitetty metsikkösimulaattori. Esimerkeissä
osoitetaan, että tila-aika -riippuvuudet voidaan generoida kasvun, kuolemien
ja uudistumisen avulla iteratiivisesti vaihtoehtona riippuvuuksien
tuottamiselle kovarianssifunktioilla. Metsätieteellisessä sovelluksessa
tutkitaan generoitujen hypoteettisten metsien avulla paikasta
riippumattomien harvennusmallien oikeellisuutta.

中文摘要

林及人，用时空相关性模拟森林林相，社会科学出版系列第64号，芬兰约恩苏大学 2003，123页。

关键字：虚拟森林，时空相关性，非均衡性，生态过程，广义化的混合线性模型，变换影响，巡回模拟，SPATE。

在林学上，为了研究统计取样、森林生长和林业经理，普遍地建造虚拟林木族群。自从1960年以来，波松(随机)程序为主要模拟族群的建造方法。但是生长、死亡和再繁殖程序所产生的时空相关性并未反应模拟林木的空间位置及特性。本论文研究的目的是创建虚拟森林的模拟方法，使林木的特性是依赖其位置并且由理论的或由实验推算出的统计分布及各向同性与非各向同性所产生的空间结构。论文中讨论了统计判断的标准以用来选择均衡的和非均衡的点模式，测验了变换效应对林木特性的依赖性，发展了基于林木的位置来模拟相关的林木特性的经验概括混合线性模式，首创了用来巡回建造虚拟森林的林分模拟器。如同范例所示，由巡回模拟生长、死亡及再繁殖过程来产生的时空相关性可以替代经由协方差函数产生的时空相关性。虚拟森林可用来检定空间独立疏伐模形的精确性。

TABLE OF CONTENT

GLOSSARY OF NOTATIONS

Mathematical Notations

$log(y)$	Natural logarithm of y for $y > 0$
$exp(y)$	Exponential function; $exp(log(y)) = y$
$logit(y)$	Logistic function; $logit(y) = log(y/(1-y))$
$max\ A$	Maximum of set A

Matrix Notations

$\mathbf{Y}$	$n \times 1$ vector; $\mathbf{Y} = (Y_1,\ldots,Y_n)^T$
$\underline{\mathbf{Y}}$	$(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)^T$
$\mathbf{I}$	Identity matrix
$\otimes$	Kronecker (direct) product

Notations for Point Processes

$\mathbb{R}^2$	2-dimensional Euclidean space
$\mathbf{x}$	Location of a tree; $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$
A	A bounded (Borel) set, $A \subset \mathbb{R}^2$
$N(A)$	A random variable that represents the number of points in A
$|A|$	The area (Lebesgue measure) of A
$\mu(\mathbf{x})$	Intensity at $\mathbf{x}$
$\mu(A)$	$\int_A \mu(\mathbf{x})d\mathbf{x}$
$\lambda(A)$	$\mu(A)/|A|$
$\mathcal{X}$	A finite set of points in $\mathbb{R}^2$
$\|\mathbf{x}_i - \mathbf{x}_j\|$	Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$
$\prod_n$	A product over pairs $\mathbf{x}_i \neq \mathbf{x}_j$, $i, j = 1,\ldots,n$
σ^2	Dispersion parameter
$U_r(\mathbf{x})$	A disk of radius r centered at point $\mathbf{x}$
$A(r)$	$\{\mathbf{x} \in A | U_r(\mathbf{x}) \subset A\}$
A_0	Background area; $A_0 = A \setminus \bigcup_{ij} A_{ij}$
A_{ij}	The jth concentric ring of the ith tree
$A_{ij \cap kl}$	Intersection of concentric rings A_{ij} and A_{kl}

Notations for Statistical Distributions

$f_Y(y)$	The probability density function of Y
F_Y	The cumulative distribution function of Y
$F_Y(y)$	$P[Y \le y] = \int_{-\infty}^{y} f_Y(x)dx$
$E[Y]$	$\int_{-\infty}^{+\infty} y f_Y(y) dy = \mu_Y$
$E[\mathbf{Y}]$	$\left(\mu_{Y_1}, ..., \mu_{Y_n}\right)^T = \boldsymbol{\mu_Y}$
$\Phi(x)$	$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}u^2\right) du$
$G(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	A distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$U[0,1]$	Uniform distribution
$Po(\mu(A))$	Poisson distribution with expectation $\mu(A)$
$Var(Y)$	The variance of Y ; $Var(Y) = \sigma_Y^2$
$Cov(Y_i, Y_j)$	$E\left[(Y_i - \mu_Y)(Y_j - \mu_Y)\right] = \sigma_{Y_i Y_j}$
$\boldsymbol{\Sigma_Y}$	$E\left[(\mathbf{Y} - \boldsymbol{\mu_Y})(\mathbf{Y} - \boldsymbol{\mu_Y})^T\right] = Cov(\mathbf{Y})$
τ	Random variable that represents initial heterogeneity
α, ϕ	Parameters (standard deviation) of initial heterogeneity
ξ, ψ	Random variables that represent environmental effects
γ, κ	Parameters (standard deviation) of environmental effects
ε	Regression residuals
π, ω	Probabilities for a Bernoulli random variable that takes values 0 or 1

GLOSSARY OF FORESTRY TERMS

Allele	One of a pair of genes located at the same locus in homologous chromosomes and controlling the same character. One of a pair of characters controlled by such genes and alternative to each other in inheritance (Wright 1976, p. 439).
Basal area	The cross-sectional area of a tree's bole measured at a predefined point above ground, usually breast height (Dunster and Dunster 1996).
Cohort	A group of trees regenerating after a single disturbance (Oliver and Larson 1990, p. 142).
Current annual increment	The growth increment added in the past year, or in the past few years, in which case, it would be a periodic annual increment (Dunster and Dunster 1996).
Heritability	That portion of the total variance due to genetic factors. In a broad sense, that portion of the total variance due to all genetic factors. In a narrow sense, that portion of the total variance due to genes with additive effects and most indicative of the superiority that can be transmitted by seed (Wright 1976, p. 444).
Low thinning	Trees are removed from lower crown classes (Smith 1962, p. 64).
Recruitment	(synonym of ingrowth) The number of trees or the volume of trees that have grown past a predetermined threshold in a set period (Dunster and Dunster 1996).
Regeneration	The renewal of a forest or stand of trees by natural or artificial means, or the stand of young trees under *1.3* meters high that results (Dunster and Dunster 1996).

Sapling	Typically one to two meters tall and two to four centimeters in diameter, with vigorous growth, no loose, dead bark, and few (if any) dead branches (Dunster and Dunster 1996).
Seedling	A young tree growing from seed is a seedling from the time of germination through to the sapling stage (Dunster and Dunster 1996).
Selection thinning	To remove dominant trees in order to stimulate the growth of smaller trees (Smith 1962, p. 77).
Silvicultural practices	The set of field techniques and methods implemented to modify and manage a forest stand over time to meet defined management goals and objectives, which will depend on the silvicultural system [type of cuttings, the stand treatments, and any intermediate cuttings] being used (Dunster and Dunster 1996).
Site	An area's potential for tree growth; site usually incorporates an area's soil and climate conditions (Oliver and Larson 1990, p. 1).
Stand	A spatially continuous group of trees and associated vegetation having similar structures and growing under similar soil and climatic conditions (Oliver and Larson 1990, p. 1).
Stand development	The part of stand dynamics concerned with changes in stand structure over time (Oliver and Larson 1990, p. 1).
Stand dynamics	The study of changes in forest stand structure with time, including stand behavior during and after disturbances (Oliver and Larson 1990, p. 1).

Stand structure	The physical and temporal distribution of trees in a stand. The distributions can be described by species; by vertical or horizontal spatial patterns; by size of trees or tree parts, including the crown volume, leaf area, stem, stem cross section, and others; by tree ages; or by combinations of the above (Oliver and Larson 1990, p. 1).
Thinning	The removal of selected stems from a developing stand in order to salvage potential mortality, and promote silvicultural or other objectives (enhanced growth in the remaining trees, etc.) (Dunster and Dunster 1996).
Tree crown	The upper part of a tree or other woody plant that carries the main system of branches and the foliage (Dunster and Dunster 1996).

1. INTRODUCTION

Computer generation of hypothetical forests for forestry and statistical studies has developed rapidly during the last 40 years. For example, Palley and O'Regan (1961) were only able to generate a small amount of trees in a limited area using early computers, while Mackisack and Wood (1990) were able to generate a 1000 ha tract of forest. In those applications, tree locations are randomly generated (the so-called Poisson forest) and tree characteristics are independently and identically distributed random variables, which are independent of location. For management planning (e.g. Ek and Monserud 1974) and ecological (e.g. Shugart 1984) simulations the generated hypothetical forests, which are used to simulate thinning and growth, consist of only tree characteristics.

For single-tree growth simulation models based on tree locations (e.g. Adlard 1974; Pukkala 1988; Pretzsch 1997) and on geometrical developments of tree crowns (e.g. Ford and Sorrensen 1992; Kellomäki et al. 1999) are developed to simulate the competition among trees. Stoyan and Penttinen (2000, pp. 70-71) reviewed recent applications of forestry growth models and stated that

> … single-tree models can be seen as models of spatio-temporal point processes. They pose the following statistical problems: (i) model validation, i.e. checking that a given single tree model produces realistic artificial forests, (ii) construction of geometrical models for tree crowns, which will be used as marks, (iii) modeling mortality of trees, the main random component in many forests and (iv) developing of models for realistic start configurations based on a statistical analysis of any forest to be used in simulation of its development.

Parallel to the generation of hypothetical forests, point pattern analysis methods are used to study spatial arrangement and to generate locations of trees (e.g. Newnham 1968). Ripley (1976) introduced the *K*-function, which is used to examine whether the configuration of a mapped point pattern is

Poisson, clustered or regular. Tomppo (1986) discussed models and methods for analyzing spatial point patterns of trees in forests. Penttinen et al. (1992) presented statistical methods, based on the marked point processes, for explanatory analysis of interactions among trees.

From a statistical point of view, tree characteristics depend on location and on their marginal distributions, which may follow arbitrary analytical or empirical distributions. There may be isotropic or anisotropic spatial covariances between the characteristics of neighboring trees. Therefore, statistical properties of location and tree characteristics need to be considered simultaneously to generate hypothetical tree populations.

Although there have been attempts to model marginal distributions and the covariance structure of tree characteristics, it is still a difficult task if tree locations and characteristics are to be specified directly for a single point of time. For example, Kokkila et al. (2002) proposed using the Gibbs marked point process to generate a spatial configuration for a given set of tree characteristics. Tree locations and characteristics (marks) were taken into account simultaneously by permuting tree locations within the simulated area. As Griffith (1988, p. 229) pointed out, such a permutation procedure ignored

> …indirect effects associated with the feedback loops of multi-directional dependences… and …the realized surface is not necessarily the outcome of a homogeneous process operating over the planer [2-dimensional] surface in question (stationarity).

The procedure can fail when the desired spatial structure is not within the distribution of spatial structures that can be generated from a given set of independent random variables through permutation.

Forests develop over time via the processes of regeneration (reproduction), growth and mortality (Oliver and Larson 1990, Chap. 6). In this study we formulate models for locations and for these ecological processes and allow these models to interact iteratively. Specifically, the objectives of this study are:

1. To formulate statistical models, which correspond to the ecological processes, for simulating tree locations and characteristics.

2. To generate correlated tree characteristics with empirically estimated cumulative distribution functions.

3. To generate hypothetical forests iteratively with spatio-temporal dependencies.

The computer software SPATE (SPAtio-TEmporal stand simulator) is developed to implement the simulation.

This study belongs to the intersection of applied statistics and quantitative forestry. We apply techniques of statistics to the generation of hypothetical forests that correspond to a given pattern of locations, marginal distributions and covariance structures of tree characteristics. The model specifications are general and the results can be applied to a wide range of forestry and statistical studies (cf. Ripley 1984).

Statistical criteria, i.e. Poisson regression and Q-function analyses, for examining mapped tree patterns are formulated in Chapter 2. These tools facilitate the analysis of mapped point patterns without homogeneity assumptions.

Transformation has been used extensively in forestry for stabilizing variances or for modeling purposes. However, only a few studies have taken into account how transformation affects the correlation structures of tree characteristics. This may be a problem, especially when the underlying distribution does not have a closed form, as in the case of an empirically

estimated density function. We investigate the transformation effects in Chapter 3.

The general formulation of the generalized linear mixed models in Chapters 3 and 4 covers a wide range of stochastic (and deterministic) models that appear in forestry literature. We discuss how the models can be integrated into a global model to iteratively simulate forest development in Chapter 5. We show an application of examining errors caused by commonly used location independent thinning models in Chapter 6.

2. GENERATING INITIAL TREE LOCATIONS

The goals of this chapter are to develop methods for generating point patterns as tree locations, and to formulate statistical measures of point patterns. We consider three types of point patterns, namely, Poisson, clustered and regular (cf. Tomppo 1986), with homogeneous and inhomogeneous intensities. These point patterns are used to generate tree locations for one-step (Chapter 3) or for the initial stands of iteratively generated (Chapter 5) hypothetical forests.

In Section 2.1 we discuss the generation of locations using homogeneous and inhomogeneous Poisson processes. In Section 2.2 we illustrate the generation of clustered locations using the Neyman-Scott process. In Section 2.3 we show how to simulate pairwise interaction processes that display inhibition and local dependency. In Section 2.4 we formulate statistical criteria of spatial patterns and show how to select an appropriate model for a point pattern. In Section 2.5 we examine the longleaf pine mapped stand data with the statistical criteria developed in Section 2.4 and show examples of generated point patterns.

2.1 Poisson Forests

Suppose natural regeneration is underway in a newly harvested forest stand. If there are no significant environmental variations, the locations of regenerated seedlings can be described by the Poisson distribution (de Vries 1986, pp. 371-372).

Let $A \subset \mathbb{R}^2$ be a Borel set (Feller 1971, pp. 113-114) with area (Lebesgue measure) $|A|$. Define $N(A)$ to be a random variable representing the number of tree locations in A. Provided that $N(A)$ is finite whenever $|A| < \infty$, we can think of the collection of random variables

$\{N(A): A \subset \mathbb{R}^2$ is a Borel set$\}$ as a spatial point process, for a rigorous theoretical foundation see van Lieshout (2000, pp. 4-8).

Suppose there is a function $\mu(x) \geq 0$ on $\mathbb{R}^2$. Define $\mu(A) = \int_A \mu(x)\,dx$. A spatial point process is a Poisson process with intensity $\mu(x)$ if

$$P\{N(A) = k\} = \frac{exp(-\mu(A))\mu(A)^k}{k!}, \ k = 0,1,...$$

for all bounded Borel sets $A \subset \mathbb{R}^2$ and counts in disjoint sets are independent (Rényi 1967). If $\mu(x) \equiv \mu$, we have a homogeneous Poisson process. Often this is referred to as the complete spatial randomness (Cressie 1991, p. 586). In this case $\mu(A) = \mu|A|$.

Realizations from a homogeneous Poisson process can be obtained based on the property that, conditionally on $N(A)$, the points are uniformly distributed over A. The shape of A need not be restricted to rectangles. For example, Lewis and Shedler (1979, p. 411) considered the generation of Poisson point patterns in a circular area, and Hsuan (1979) used rejection sampling to generate uniformly distributed random locations on a polygonally shaped area.

If there are factors (such as soil fertility) that affect the spatial distribution of the seedlings, then the intensity may not be constant. In this case the seedling locations can be described by an inhomogeneous Poisson process with a variable intensity function over A (Diggle 1983, p. 52). Suppose $max\,\mu \equiv max\{\mu(x) \mid x \in A\}$, realization of such a process can be generated by the rejection method (Lewis and Shedler 1979, pp. 411-412):

1. Generate $N(A)$ points from a homogeneous Poisson process of intensity $max \, \mu$.
2. Generate $u \sim U[0,1]$ independently. If $u > \mu(x_i)/max \, \mu$, $i \in N(A)$, reject this point.

The result is a realization of an inhomogeneous Poisson process with intensity $\mu(x)$.

2.2 Clustered Forests

For natural regeneration it is customary to leave mother trees in a harvested stand. Consequently we can expect spatial clustering of seedlings around the mother trees. The Poisson cluster process can be used to model this type of clustered spatial point pattern.

An early example is the process described by Neyman and Scott (1958) that was applied to problems of Cosmology. In the case of a forest stand, the 'cluster center' and 'galaxy' in Neyman and Scott's discussion correspond to mother trees and offspring, respectively. This setting corresponds to the seed-tree silvicultural practice in which mother trees are left in a harvested stand to produce seeds and they are removed when seedlings have established (Smith 1962, pp. 421-436). Note that realizations of the Neyman-Scott process consist of the offspring only.

Boundary conditions affect the generation of offspring locations in a bounded area A when mother trees are located near the boundaries. If the boundaries of A are defined by ownership rather than by ecological boundaries, then we have a spatial process that is not influenced by the boundaries. In this case we can generate mother trees and offspring locations in $A^* \subset A$, which is clearly larger than A (Ripley 1981, p. 151). In general, A can be a polygonally shaped area with reasonably regular boundaries.

On the other hand, lakes, ravines, roads, etc. can create ecologically significant boundaries such that the spatial process near boundaries differs from the spatial process elsewhere. This is a difficult problem and methods used to account for such boundary effects (Griffith 1988, Chap. 7) do not reflect the full ecological reality. We simply discard the generated offspring locations that are outside A.

We use the following procedure to generate realizations of the Neyman-Scott process:

1. Generate mother tree locations x_i, $i = 1,\ldots,m$ from a homogeneous Poisson process in A.
2. For each mother tree, generate a Poisson random number to be the number of offspring.
3. Independently locate offspring around the mother tree according to a density function $g(\cdot)$.
4. Retain the generated offspring locations inside A.

Taking an inhomogeneous Poisson process in step 1 will produce an inhomogeneous Neyman-Scott process (Cressie 1991, pp. 663-664). The density function $g(\cdot)$ can be a radially symmetric function such as the normal density function with $\mu = 0$ and $\sigma =$ offspring dispersion distance (Diggle 1983, p. 56). The expected number of mother trees and offspring can be determined based on field observations, or estimated using statistical methods (e.g. van Lieshout and Baddeley 2001, pp. 5-10) and the statistical criteria described in Section 2.4.

2.3 Regular Forests

Saplings may inhibit others from growing in close proximity. This can result in a regular pattern of saplings. Markov point processes provide a flexible inhibition mechanism that preserves local dependencies.

Following Diggle (1983, pp. 63-64), let $f(\cdot)$ be the likelihood function of the process with respect to a Poisson process of unit intensity. Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be any finite set of points in a bounded set $A \subset \mathbb{R}^2$. Then the likelihood function of the process is $f(\mathcal{X}) \exp(-|A|)$. For a fixed $r > 0$, define $x_i, x_j \in A$ to be neighbors if $\|x_i - x_j\| < r$, where $\|x_i - x_j\| = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2}$ is the Euclidean distance between x_i and x_j, and r is the interaction radius. A Markov point process is a Strauss process provided that

$$f(\mathcal{X}) = ab^n c^s, \qquad (2.1)$$

where $a > 0$ is a normalizing constant, n is the number of points in $\mathcal{X}$, $b > 0$, and s is the number of distinct pairs of neighbors in $\mathcal{X}$. Parameter c describes the interaction between pairs of neighboring points. When $c = 1$, we have a Poisson process. If $c = 0$, we have a process in which no two points in $\mathcal{X}$ may be neighbors. When $0 < c < 1$, we have a non-strict inhibition process (Kelly and Ripley 1976).

For a class of pairwise interaction processes such that the interaction between pairs of points depends on the distance between them, process (2.1) can be generalized as

$$f(\mathcal{X}) = a b^n \prod\nolimits_n h(\|x_i - x_j\|), \qquad (2.2)$$

where $h(\cdot) \geq 0$ is bounded and $\prod_n$ is the product over pairs $x_i \neq x_j, i, j = 1, \ldots, n$. For example, if $h(\cdot) = c$ for $\|x_i - x_j\| < r$ and $h(\cdot) = 1$ for $\|x_i - x_j\| \geq r$ we get (2.1).

Denote $x_n = x$ and define $\mathcal{X}^* = \{x_1, \ldots, x_{n-1}\}$. It follows that

$$f(\mathcal{X}) = a b^n \prod_{n-1} h\left(\|x_i - x_j\|\right) \prod_{i=1}^{n-1} h\left(\|x - x_i\|\right).$$

Suppose we know the locations in $\mathcal{X}^*$. Then the conditional density of x is

$$f(\mathcal{X}) \big/ f(\mathcal{X}^*) = b \prod_{i=1}^{n-1} h\left(\|x - x_i\|\right).$$

It follows that (2.2) can be simulated by randomly deleting a point in $\mathcal{X}$ and then adding a new point with a density proportional to $\prod_{i=1}^{n-1} h\left(\|x - x_i\|\right)$. We can then let x_{n-1} play the role of x and repeat the procedure. This iterative process can be shown to lead to a point pattern with density $f(\cdot)$ (Ripley 1977).

Markov point processes can be simulated using the spatial birth-and-death process, or the Metropolis algorithm (Cressie 1991, pp. 679-680). We use the algorithm due to Ripley (1979), which was developed under the spatial birth-and-death process framework. Given an expected number of points, an initial set of points can be generated from a homogeneous Poisson process. A point is chosen at random and replaced by a point drawn from the conditional density given the remaining points. The conditional density is proportional to c^s as in (2.1).

For generating inhomogeneous pairwise interaction point patterns, we can transform a homogeneous pairwise interaction point pattern to correspond to an inhomogeneous intensity function (Jensen and Nielsen 2000, pp. 8-11). Instead of attempting to fit a parametric or non-parametric model to the inhomogeneous intensity interface, we can transform a regular point pattern to an inhomogeneous regular point pattern using estimated empirical cumulative distribution functions from generated and data point patterns. The transformation involves a probability integral transformation

and an inverse transformation (see Section 3.3 for a detailed discussion of implementations). We introduce this approach to generate inhomogeneous regular point patterns since it is much simpler than the methods used by Jensen and Nielsen (2000). We give an example of the transformation in Section 2.5.

2.4 Statistical Criteria for Selecting Patterns of Tree Location

In general the type of point pattern to be simulated is chosen based on prior ecological knowledge, or on the intended use of the simulation results. If the goal is to replicate a particular forest stand, then it may be useful to have statistical guides for making the choices. We present two procedures. In Section 2.4.1 we show how Poisson regression can be used to estimate and test both for the intensity surfaces and for type of point patterns. In Section 2.4.2 we show how measures that are related to the K-function can be used to give a refined analysis of point patterns.

2.4.1 Poisson Regression

Suppose there is a partition $A = A_1 \cup A_2 \cup \ldots \cup A_q$ and let $N(A_i)$ be the number of locations in A_i. This information can be obtained, for example, from a mapped stand. We write $E[N(A_i)] = \lambda(A_i)|A_i|$, where $\lambda(A_i) = \int_{A_i} \mu(\boldsymbol{x})d\boldsymbol{x}\big/|A_i|$ is the average intensity in A_i. $N(A_i)$'s are mutually independent because of the definition of the Poisson process. In general there are no fixed rules as to how to partition A. A common practice is to use a regular grid consisting of rectangles A_i (Diggle 1983, p. 23). In this example we can use a log-linear model (McCullagh and Nelder 1989, pp. 193-200) to estimate $\lambda(A_i)$. In the simplest case it is of the form

$log\left(\lambda\left(A_i\right)\right) = \mathbf{v}_i\boldsymbol{\beta}$, where $\mathbf{v}_i = \left(1, l_i, m_i\right)$ depends on the center $\left(l_i, m_i\right)$ of A_i in the grid and $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \beta_2\right)^T$.

When a log-linear model is used, the expected number of locations in A_i is approximately of the form $E\left[N\left(A_i\right)\right] = exp\left(\mathbf{v}_i^T\boldsymbol{\beta}\right)|A_i|$ for some $\mathbf{v}_i$. If $N\left(A_i\right)$ is Poisson distributed and independent, we can test for an inhomogeneous intensity surface by taking $H_0: \beta_j = 0$ for $j > 1$ and using a likelihood ratio test statistic. However, if $N\left(A_i\right)$ is not Poisson distributed, then the test is only approximately valid (McCullagh and Nelder 1989, p. 200).

Another use of Poisson regression is to examine the under- or over-dispersion of the tree locations. For the Neyman-Scott process we expect more random variability than the Poisson process (i.e. over-dispersion). For the pairwise interaction process we expect less random variability than the Poisson process (i.e. under-dispersion). A simple way to test for over- and under-dispersion is to introduce a scale parameter (McCullagh and Nelder 1989, p. 194). In what follows we assume that $N\left(A_i\right)$'s are independent.

Suppose that $Var\left(N\left(A_i\right)\right) = \sigma^2 E\left[N\left(A_i\right)\right]$, where σ^2 is the dispersion parameter, for the Poisson distribution $\sigma^2 = 1$, if $\sigma^2 > 1$ we have over-dispersion, and if $\sigma^2 < 1$ we have under-dispersion. The dispersion parameter can be estimated using the formula

$$\hat{\sigma}^2 = \sum_{i=1}^{q}\frac{\left(N\left(A_i\right) - \hat{\lambda}\left(A_i\right)|A_i|\right)^2}{\hat{\lambda}\left(A_i\right)|A_i|}\Bigg/\left(q - p\right),$$

where $\hat{\lambda}\left(A_i\right)$ is the estimate of $\lambda\left(A_i\right)$ from the log-linear model and p is the dimension of $\mathbf{v}_i$.

An approximate test for an inhomogeneous intensity surface is obtained as before, but now the estimated dispersion parameter is incorporated. To assess whether $\sigma^2 = 1$, we can use the following approximate procedure. We expect that for large $\lambda(A_i)$, $(q-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{q-p}$ (Rao 1973, pp. 392-393). Let the ν quartile of the χ^2_{q-p} distribution be $w_{\nu, q-p}$. Then

$$P\left(w_{\frac{\nu}{2}, q-p} \leq \frac{(q-p)\hat{\sigma}^2}{\sigma^2} \leq w_{1-\frac{\nu}{2}, q-p}\right) = 1 - \nu.$$

This yields an approximate $(1 - \nu)$ confidence interval for σ^2 of the form

$$P\left(\frac{(q-p)\hat{\sigma}^2}{w_{1-\frac{\nu}{2}, q-p}} \leq \sigma^2 \leq \frac{(q-p)\hat{\sigma}^2}{w_{\frac{\nu}{2}, q-p}}\right) \approx 1 - \nu.$$

If the confidence interval includes 1, we may have a Poisson case. If the lower limit of the confidence interval is > 1, we may have a case of over-dispersion. If the upper limit of the confidence interval is < 1, we may have a case of under-dispersion.

We offer this as an approximate guide only because the true sampling distribution of $\hat{\sigma}^2$ under the Neyman-Scott or the pairwise interaction processes is unknown. Of course the point pattern could be a realization of some other point process as well. Therefore, there is little hope of finding an exact test. However, the purpose of this test is merely to act as a guide for choosing a point pattern generation method from those introduced in Sections 2.1-2.3.

2.4.2 *K*-Function and Related Measures

In forests tree location distributions are affected by many factors (such as soil and competition) and measures based on the homogeneous Poisson assumption may not be adequate. Alternative measures are needed to identify an appropriate point pattern for simulation purposes. In the following we formulate an approach that is also applicable in the case of an inhomogeneous intensity surface.

Consider a disk $U_r(\mathbf{x})$ of radius r centered at any point $\mathbf{x} \in A$. Let the expected number of points (i.e. trees) in the disk be

$E[N(U_r(\mathbf{x}))] = \lambda(U_r(\mathbf{x}))\pi r^2$. Define

$K^*(\mathbf{x},r) = E[N(U_r(\mathbf{x}))|\text{ there is a tree at } \mathbf{x}] - 1$. This is the expected number of other trees in $U_r(\mathbf{x})$ given that there is a tree at the center (Ripley 1977, p. 190). It follows that $Q(\mathbf{x},r) = K^*(\mathbf{x},r)\big/\lambda(U_r(\mathbf{x}))\pi r^2$ is a measure of the effect of the tree at the center on the number of other trees in the disk (cf. Baddeley et al. 2000).

For a homogeneous Poisson process with intensity λ,

$\lambda(U_r(\mathbf{x}))\pi r^2 = \lambda\pi r^2 = K^*(\mathbf{x},r)$, and hence $Q(\mathbf{x},r) = 1$ for all $\mathbf{x}$. The famous *K*-function (Ripley 1976) can be defined as $K(r) = K^*(\mathbf{x},r)\big/\lambda$. As pointed out by Cressie (1991, p. 639), if we pick out a random location $\mathbf{x} \in A$, then we can define $E[Q(\mathbf{x},r)] = Q(r)$, where the expectation is taken with respect to the intensity of $\mathbf{x}$. The expectation does not depend on $\mathbf{x}$, but it may depend on r. For clustered processes with intensity λ we expect that $Q(r) > 1$. For regular point patterns we expect that $Q(r) < 1$.

Continue to consider the case where $Q(r)$ does not depend on $\boldsymbol{x}$.
Define $A(r) = \{\boldsymbol{x} \in A \,|\, U_r(\boldsymbol{x}) \subset A\}$. Then an estimator of the Q-function is

$$\hat{Q}(r) = \frac{1}{N(A(r))} \sum_{x_i \in A(r)} \frac{N(U_r(\boldsymbol{x}_i)) - 1}{\hat{\lambda}(U_r(\boldsymbol{x}_i))\pi r^2},$$

where $\hat{\lambda}(U_r(\boldsymbol{x}_i))$ is the estimated average intensity. It can be estimated
using a parametric model such as the log-linear models discussed in 2.4.1, or
using non- and semi-parametric kernel estimation methods (Baddeley et al.
2000). In the homogeneous Poisson case $\hat{\lambda}(U_r(\boldsymbol{x}_i)) = N(A)/|A|$. The plot
of $\hat{Q}(r)$ against r provides a heuristic estimate of whether tree locations in
a mapped stand display a random, clustered or regular configuration.

Although the Q-function can be a useful tool for analyzing point
patterns, it is sensitive to r and the model for $\hat{\lambda}$. For example, in the log-
linear model case the chosen grid size and the order of the model will affect
the behavior of $\hat{Q}(r)$. This also applies to the non-parametric cases with the
choice of kernel function and bandwidth (cf. Cressie 1991, p. 660). It is
possible that two identical Q-function plots could result from very different
$\hat{\lambda}$'s (Baddeley and Silverman 1984).

2.5 Examples of Forests with Locations

We will analyze the mapped longleaf pine data (Appendix B.1) and
then use the analysis results to simulate tree locations. The mapped stand is
partitioned into a 5×5 grid and the counts in the grid cells are fitted with a
constant and a log-linear models. The results are listed in Table 2.1. The
deviance difference $38.11 \; (= 150.62 - 112.51)$ is greater than
$w_{0.05,\,2} = 5.99$, which indicates that the log-linear model fits significantly

better. The estimated regression coefficients and standard errors indicate that the intensity surface increases with x_1, which suggests inhomogeneity. Because the lower bounds of the confidence interval for $\hat{\sigma}^2$ for both models are greater than 1, there is over-dispersion (clustering) in the point pattern.

Figure 2.1 shows the estimated Q-function and the transformed K-function with variable-width edge correction, i.e. $\hat{L}(r) = \sqrt{\hat{K}(r)/\pi} - r$, where $\hat{K}(r) = |A| \sum_{x_i \in A(r)} (N(U_r(x_i)) - 1)/(N(A(r))N(A))$ (Cressie 1991, p. 616). The Q-function plot indicates a clustered point pattern for longleaf pine. Thus, we can conclude that the longleaf pine mapped stand is an inhomogeneous clustered point process; the same conclusion was drawn by Baddeley et al. (2000). We can see from Figure 2.1 that, as r increases, the Q-function plot decreases steadily towards 1, which is the theoretical value if the intensity surface is homogeneous, and the plot flattens out at around $r = 10$ m. We may take this as an approximate cluster radius.

Although the statistical analysis agrees with qualitative observations (Platt et al. 1988, p. 505) that

> …juveniles and subadults were highly aggregated, occurring in discrete, widely separated clumps…,

the underlying ecological processes that generate the stand are in neither

Table 2.1 List of coefficients (and their standard errors), deviance, estimated dispersion parameter ($\hat{\sigma}^2$) and confidence interval for dispersion parameter (CI) of the fitted intensity surface models of a mapped longleaf pine stand.

Explanatory variable and model statistics	Constant model	Log-linear model
Intercept	–4.22 (.04)	–4.68 (.12)
x_1	—	.0045 (.0008)
x_2	—	–.00034 (.00073)
Deviance	150.62	112.51
$\hat{\sigma}^2$	6.21	5.66
CI	(3.47, 11.80)	(3.16, 10.75)

case the ones we have discussed so far. The reason is that the juvenile and subadult trees are in clumps due to inhibition by the adult trees. In addition the mapped stand data excluded other species, which account for about one quarter of the mapped longleaf pine trees (Platt et al. 1988, p. 493). This means that there might be other undetected point processes interacting with the juvenile and subadult trees (cf. Högmander and Särkkä 1999).

A further investigation was conducted by analyzing the point patterns of juvenile, subadult and adult size classes (Figure B.2). We repeat the Q-function analysis and the results are shown in Figure 2.2 and Table 2.2. The results show that at least two different point processes co-exist in the stand. Therefore, attempting to fit a single clustered point process to the mapped pattern will not yield reasonable parameter estimates.

As an example we generate an inhomogeneous clustered point pattern, which is used as an initial point pattern for generating one-step stands in Chapter 3. We use the intensity surface estimated from adult trees for generating mother tree locations. From the Q-function plot (Figure 2.2) we choose the clustering distance to be *10* m. Judging from the stand map for juveniles there are approximately *10* trees per cluster.
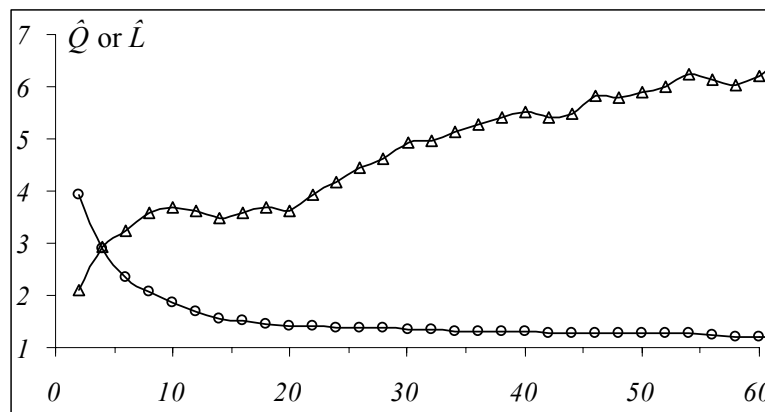
Figure 2.1 Estimated Q-function (circles) and transformed K-function $\hat{L}(r)$ (triangles) of the longleaf pine data. Both functions are computed with an increment of 2 m.
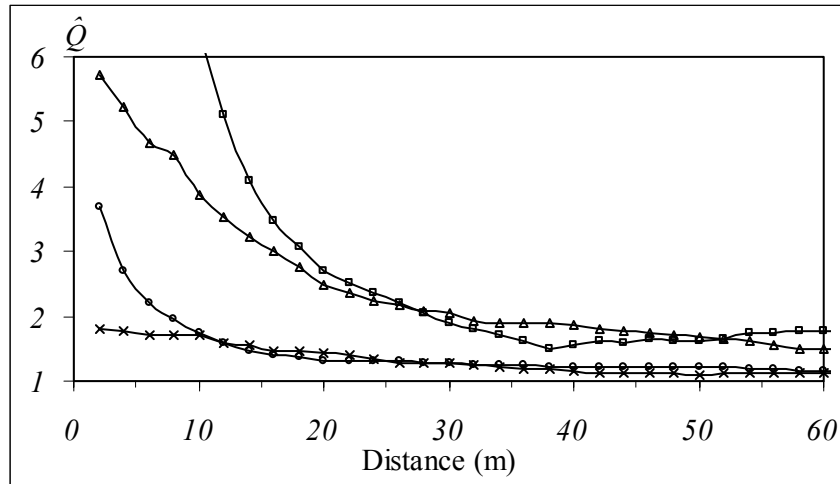
Figure 2.2 Q-function plots for combined (circles), juvenile (squares), subadult (triangles) and adult (crosses) size groups of longleaf pine data. Note that the first five estimated Q-functions for juvenile longleaf pine data are 27.6, 15.2, 10.2, 7.8 and 6.4, which are not plotted.

Table 2.2 List of coefficients (and their standard errors), deviance, estimated dispersion parameter ($\hat{\sigma}^2$) and confidence interval for dispersion parameter (CI) of the fitted intensity surface models of a mapped longleaf pine stand.

Explanatory variable and model statistics	Juvenile	Subadult	Adult
Intercept	−7.809 (.2940)	−6.312 (.240)	−4.418 (.153)
x_1	.0100 (.0155)	.0060 (.0015)	.00060 (.00107)
x_2	.0097 (.0016)	.0012 (.0014)	−.00713 (.00113)
Deviance	169.05	124.89	28.19
$\hat{\sigma}^2$	7.48	6.30	1.30
CI	(4.84, 16.45)	(4.07, 13.85)	(.84, 2.85)
Intensity surface	Inhomogeneous	Inhomogeneous	Inhomogeneous
Point pattern	Clustered	Clustered	Poisson

After a few experiments we choose the number of mother points to be
64 to generate a total number of offspring locations that is close to the actual
number of trees in the longleaf pine stand. For example, the average number
of offspring locations in this example is *587* with *1200* simulation runs. An
example of the generated mother and offspring locations are shown in Figure
2.3.

In fact, the above analyses only deal with tree locations without
considering tree characteristics and environmental factors, such as soil and
relief. Cressie (1991, p. 618) suggested that the longleaf pine point pattern
results from the dynamics of the ecological processes of growth, mortality
and reproduction. Thus, it is preferable to consider point processes and the
dynamics of ecological processes jointly. In the following chapters we
introduce an integrated 'marked point process', which iteratively simulates
locations and tree characteristics (marks).
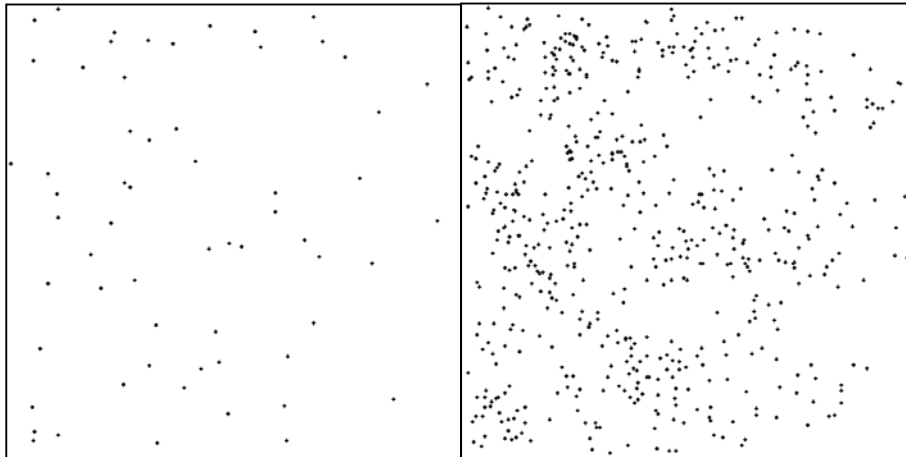
*Figure 2.3 Example point patterns of generated inhomogeneous mother tree
locations (left) and offspring locations (right) using the Neyman-Scott
process.*

As mentioned at the end of Section 2.3, an inhomogeneous regular point pattern can be generated through transformation. For illustration purposes we arbitrarily defined an empirical cumulative distribution function (edf), which assumes that a slope on the x-axis has introduced the inhomogeneity to transform a homogeneous regular point pattern of Strauss process with interaction potential $c = .3$ and interaction radius $r = 10$ m. The generated point patterns and corresponding edf are shown in Figure 2.4.
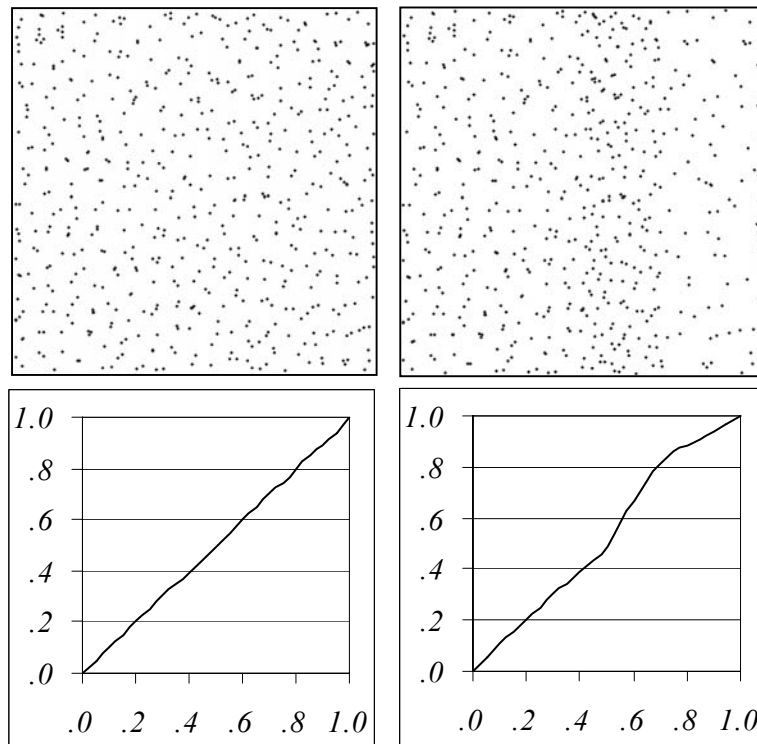
Figure 2.4 Simulated homogeneous regular point pattern (top-left) and its edf (lower-left) versus transformed inhomogeneous regular point pattern (top-right) using an arbitrarily defined edf (lower-right).

3. EMPIRICAL MODELS FOR TREE CHARACTERISTICS

Simplified descriptions of tree populations can be given in terms of means, variances, covariances and marginal distributions of tree characteristics. In general the moments may vary due to environmental differences, genetic differences and their interactions. In this chapter we introduce empirical models for the mean and variance, and show how to use that information to generate tree characteristics that depend on their locations.

We describe the location-scale family of distributions for modeling tree characteristics in Section 3.1. We discuss marginal distributions of tree characteristics in Section 3.2. We show methods to investigate the effect of transformations on correlations in Section 3.3. We formulate mixed linear models for tree characteristics in Section 3.4. We explain the sources of variation via variance components and their statistical properties in Section 3.5. We discuss between-tree and within-tree correlations of the random effects in Sections 3.6. We show methods to examine stand structure in Section 3.7. We show examples of generated tree characteristics with given locations in Section 3.8.

3.1 Location-Scale Representation

Define $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{Y} \in \mathbb{R}^n$, where Y_i is an independent random variable (r.v.), which represents a tree characteristic, with cumulative distribution function (cdf) $F(\cdot)$. We assume that tree characteristics belong to the location-scale family of distributions (Lehmann 1991, pp. 19-22). That is, the marginal distribution of the characteristic Y_i at

location x_i is $F_i(y_i) = P[Y_i \leq y_i]$, which is of the form

$F_i(y_i) = G((y_i - \mu_{Y_i})/\sigma_{Y_i})$, where G is a known distribution with density

$g(\cdot)$ such that $\int_{-\infty}^{\infty} yg(y)\,dy = 0$ and $\int_{-\infty}^{\infty} y^2 g(y)\,dy = 1$. We use the

notation $Y \sim G(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$, where $\boldsymbol{\mu_Y} = (\mu_{Y_1}, \ldots, \mu_{Y_n})^T$ and

$\boldsymbol{\Sigma_Y} = E\left[(\boldsymbol{Y} - \boldsymbol{\mu_Y})(\boldsymbol{Y} - \boldsymbol{\mu_Y})^T \right]$.

In forestry applications tree characteristic measurements are strictly positive. Therefore, we transform the observed values to the logarithmic scale and perform the location-scale transformations, empirical marginal distribution estimations and random variable generations. Then the generated random numbers (tree characteristics) can be transformed back to the measurement scale.

3.2 Marginal Distributions

In general we do not know the true distribution functions of tree characteristics. However, we can define an empirical cumulative distribution function (edf) $F_n(y \mid \boldsymbol{Y}) = \left(\text{Number of } Y_1, \ldots, Y_n \text{ that are } \leq y\right)/n$, for $-\infty \leq y \leq \infty$. We can obtain a continuous version of an edf by joining the midpoints of successive steps of $F_n(y \mid \boldsymbol{Y})$ with straight lines (Dudewicz and Mishra 1988, pp. 197-201). Alternatively, we can fit the midpoints with cubic spline interpolation (Press et al. 1992, pp. 107-110) to obtain a smoothed edf.

A problem may arise in the estimation of edf when cut-off sampling (Särndal et al. 1992, pp. 531-533) is applied. For instance, the smallest measured tree in the Finnish national forest inventory permanent plots is *4.5* cm at breast height (*1.3* m above ground) (Metsäntutkimuslaitos 1985, p. 59). The resulting inventory data consist of trees with diameters between *4.5*

cm and the diameter of the largest sampled tree. This determines the support of the edf (Lehmann 1991, p. 25).

3.3 Transformation of Empirical Distribution with Dependencies

One of the objectives of this research is to generate dependent r.v.'s from the edf of mapped stand data. In general we cannot generate dependent r.v.'s directly from an estimated cdf because we do not know the form of the corresponding joint probability density function. However, we can generate dependent r.v.'s in the normal case and then transform them to match the edf (Lakhan 1981).

Define $X = [X_1, X_2]^T$ and $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $|\rho| < 1$. The procedure for generating dependent r.v.'s is:

1. Generate independent and identically distributed $X_i \sim N(0,1)$, $i = 1,2$ using the Box-Muller method (Box and Muller 1958).
2. Compute $Z = LX \sim N(0,\Sigma)$, where $Z = (Z_1, Z_2)^T$ and L is a lower triangle of the Cholesky decomposition of Σ (Griffith 1988, pp. 227-228).
3. Transform $Y_i = G^{-1}(\Phi(Z_i))$ so that $Y \sim G(0,\Sigma)$, where G^{-1} is an arbitrary inverse cumulative distribution function.

$Y_i = G^{-1}(\Phi(Z_i))$ is a composite function of $U_i = \Phi(Z_i)$, $Z_i \in \mathbb{R}$, $U_i \in [0,1]$ and $\Phi(\cdot)$ is the standard normal cdf (the probability integral transformation) and $Y_i = G^{-1}(U_i)$, $Y_i \in \mathbb{R}$ (the inverse transformation).

A problem arises because a nonlinear transformation alters the covariances in step 3 of the procedure. We will investigate the effects of transformations in three steps: (1) we study a special case where analytical results are available; (2) we use the Taylor-series based approximations to evaluate the transformation effects; and (3) we show how to evaluate the effects using simulation.

Define $Y_i = exp(X_i)$, $i = 1,2$ and let $\mathbf{X} \sim N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$.

Using the characteristic function of the normal distribution (Rao 1973, p. 519), one can show that $E[Y_i] = exp\left(\mu + \frac{1}{2}\sigma^2\right)$, $E[Y_i^2] = exp\left(2\mu + 2\sigma^2\right)$, and $E[Y_1 Y_2] = exp\left(2\mu + \sigma^2 (1 + \rho)\right)$. Therefore,

$$Corr(Y_1, Y_2) = \frac{exp\left(\sigma^2 (1 + \rho)\right) - exp\left(\sigma^2\right)}{exp\left(2\sigma^2\right) - exp\left(\sigma^2\right)}. \tag{3.1}$$

In the lognormal case the explicit forms of the expectations are known. The explicit forms may not be available for other distributions, but approximations to the moments of functions of r.v.'s can be derived using the Taylor-series based approximations (Dudewicz and Mishra 1988, pp. 263-264).

Using a second order Taylor-series we get the approximation

$$Corr(Y_1, Y_2) \approx \frac{f'(x)\rho + \frac{1}{2}f''(x)^2 \rho^2}{f'(x) + \frac{1}{2}f''(x)^2}, \tag{3.2}$$

where $f(x)$ is the lognormal transformation function (see Appendix A for the derivations). Although using a higher order Taylor-series approximation can yield a more precise approximation, the formula expands drastically. For our purposes a second order approximation balances the precision and the computational costs. Let $f(X) = exp(X)$, then an approximated formula for $Corr(Y_1, Y_2)$ corresponding to (3.1) is

$$Corr(f(X_1), f(X_2)) \approx \frac{\rho + \frac{1}{2}\sigma^2 \rho^2}{1 + \frac{1}{2}\sigma^2}.$$

Figure 3.1 illustrates the graphs of $|\rho| \le .9$ with $\sigma = .25$, $.50$, $.75$ and 1. In the case of an affine transformation, correlations do not change and this is shown by the reference line in Figure 3.1. We can see from the graphs that the effects of the logarithmic transformation on correlations is limited when $\sigma < .5$ and $|\rho| \le .4$. For instance, when $\sigma = .25$, $\rho = -.4$, $-.2$, $.2$, and $.4$, the corresponding $\hat{\rho}$ are $-.383$, $-.193$, $.195$ and $.393$, respectively. In practice, the range $[.2,.4]$ might be the most relevant in forestry applications. We see that in this case the transformed correlations are approximately 2% smaller than the original correlations.

The above findings have direct implementations in forestry since we expect that at the logarithmic scale, tree characteristics typically have a
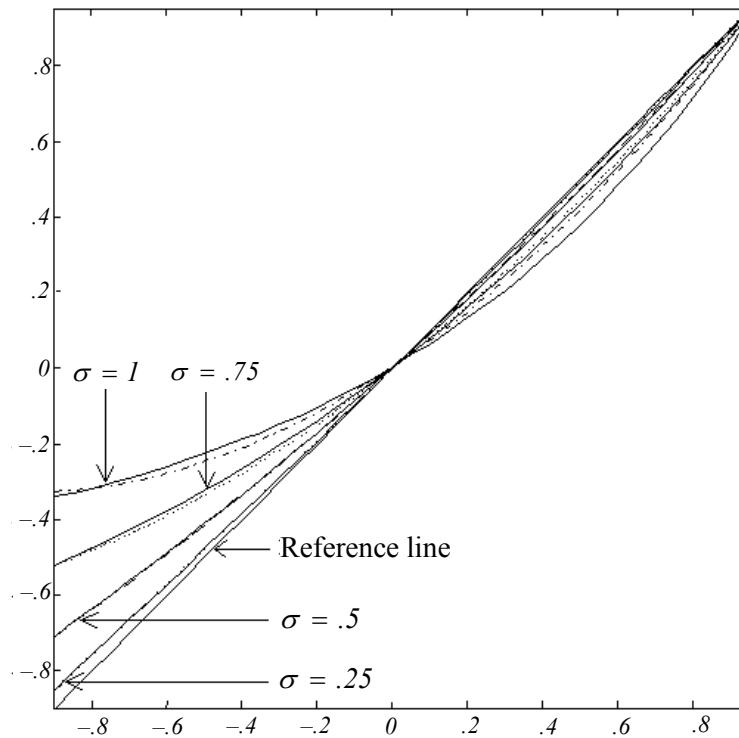
Figure 3.1 Graphs of analytically computed (solid lines) and approximated (dot or dash lines) correlations.

standard deviation less than $.5$. For example, let X be a r.v. that represents tree height and assume that a forest has a median tree height of 20 m and one standard deviation is 5 m, and that the logarithm of tree height is

$Y = log(X) \sim N(\mu_Y, \sigma_Y^2)$. Then $exp(\mu_Y)\sqrt{exp(2\sigma_Y^2) - exp(\sigma_Y^2)} = 5$.

Taking $\mu_Y = log(20)$, we get $\sigma_X = .24$. In practice the variations of tree measurements may be even smaller than this.

Let $\boldsymbol{X} = [X_1, X_2]^T$ as earlier, but with $\mu_X = 0$ and $\sigma_X = 1$. Define

$Y_i = f(X_i) = W^{-1}(\Phi(X_i))$, and $W^{-1}(t) = \delta(-log(1 - \Phi(X_i)))^{\varphi^{-1}}$ is the inverse cdf of the Weibull distribution $W(t) = 1 - exp(-(t/\delta)^\varphi)$,

$t \in [0, +\infty)$, $\delta, \varphi \in [0, +\infty)$ (Johnson and Kotz 1970, pp. 250-252). We can apply (3.2) to calculate the approximated $Corr(Y_1, Y_2)$ with various δ, φ, and ρ (Appendix A).

In general the accuracy of the Taylor-series based approximations may depend on the order of the series and $f(x)$. We may investigate the accuracy via simulation experiments to compute the exact correlation between Y_i for various δ and φ. As an example, approximated and simulated correlations for $\delta = 1$ and $\varphi = 2$ are listed in Table 3.1. The results show that the transformation of r.v.'s from the normal to the Weibull distributions does affect correlations. However, the effects are not crucial when compared with the potential errors associated with the assumptions and simplifications for modeling tree characteristics. Notice that the transformation effects vary with some other combinations of δ and φ.

The methods discussed above can be extended to evaluate the effect of inverse transformations on correlations for an edf. A restriction of the Taylor-series based approximation is that the inverse cdf (and composition functions) must be differentiable. Simulation, on the other hand, provides the

exact consequence of the transformation effects given that the sample size is sufficiently large. For example, we need to make one million simulation runs to obtain one standard error of approximately *.001* for $\hat{\rho}$ (Afifi and Azen 1979, pp. 140-141).

Another problem may occur in step 3 of the transformation procedure when we generate tree characteristics at the logarithmic scale, and we need to transform them back to the measurement scale. We cannot directly use exponential transformation for two reasons. First, the joint distribution will no longer be normal, and it may be highly skewed after having been manipulated by the growth, mortality and reproduction processes (Chapter 4). Thus the inter-relations between normal and lognormal r.v.'s (Patil et al. 1984, pp. 98-100) do not apply. Second, the marginal distribution of the

Table 3.1 The approximated and simulated correlations of the Weibull r.v.'s.

Actual Correlation	Correlation of Taylor-series based approximation	Simulated Correlation
−.9	−.859	−.855
−.8	−.765	−.761
−.7	−.671	−.668
−.6	−.577	−.573
−.5	−.482	−.480
−.4	−.386	−.382
−.3	−.291	−.287
−.2	−.194	−.193
−.1	−.097	−.095
.0	.000	.002
.1	.098	.100
.2	.196	.199
.3	.295	.300
.4	.394	.395
.5	.494	.493
.6	.594	.593
.7	.695	.695
.8	.796	.797
.9	.898	.898

transformed random variables should follow a distribution that is observed in real forests. Therefore, without taking into account the joint distribution of generated tree characteristics, exponential transformation does not produce correct marginal distributions that correspond to tree characteristics in real forests.

To account for the two aspects above, we can use the composite function as in step 3, but replace both analytical cdf with edf. For example, if we take an exponential transformation of a generated log-diameter of 4, we get a transformed diameter of 54.6 cm. However, if we take an edf transformation for the same log-diameter, we get a transformed diameter of 45.8 cm (Figure 3.2).

The edf of the generated tree characteristics is computed from all generated values. For example, the total number of generated log-diameters is the expected number of trees multiplied by the number of simulation runs. In principle we can sort the log-diameters and find the corresponding percentiles. However, this will require considerable computer resources. Alternatively, we can divide the support into intervals that are chosen to balance the numerical precision needed and required computer resources.
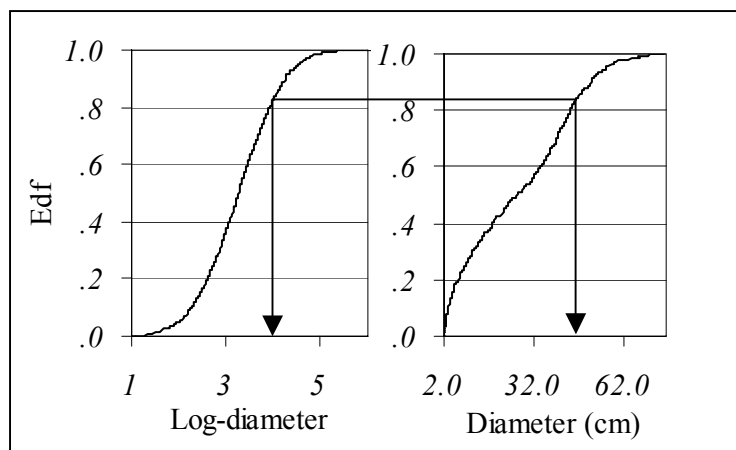
Figure 3.2 Illustration of transformation using edf. Arrows show the corresponding log-diameter and diameter in cm.

3.4 Mixed Linear Models for Tree Characteristics

Assuming that tree characteristics are linear combinations of fixed effects and random effects, we can write a mixed linear model for one tree characteristic of the form (Searle 1971, pp. 384-385)

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \alpha\boldsymbol{\tau} + \gamma\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \tag{3.3}$$

where $\boldsymbol{Y} = (Y(\boldsymbol{x_1}),\ldots,Y(\boldsymbol{x_n}))^T$ is a column vector of a tree characteristic for n trees, $Y(\boldsymbol{x_i})$ represents a tree characteristic of the i th tree at location $\boldsymbol{x_i}$, $\boldsymbol{X} = (X_{jk})$, $j = 1,\ldots,n$, $k = 1,\ldots,p$, is a $n \times p$ design matrix, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^T$ is a column vector of regression coefficients, $\boldsymbol{\tau} = (\tau_1,\ldots\tau_n)^T$ is a random vector of initial heterogeneity, $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_n)^T$ is a random vector of environmental effects, α and γ are parameters that represent the scales for, respectively, $\boldsymbol{\tau}$ and $\boldsymbol{\xi}$, and $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_n)^T$ is the error term representing the lack-of-fit of the model. We assume that the random components follow the normal distribution. To simplify the notations, we use $Y(\boldsymbol{x_i})$ and Y_i interchangeably.

The elements X_{jk} of the design matrix can be other measured tree characteristics, site indices, soil characteristics, and geographical information such as elevation (e.g. Lappi and Bailey 1988; Hasenauer et al. 1998). Competition indices derived from locations and tree characteristics can also be included (Pukkala 1988).

To complete the model formulation, we assume

$$E[\boldsymbol{\tau}] = \boldsymbol{0}, \quad Cov(\boldsymbol{\tau}) = \boldsymbol{\Sigma}_\tau = \boldsymbol{I},$$

$$E[\boldsymbol{\xi}] = \boldsymbol{0}, \quad Cov(\boldsymbol{\xi}) = \boldsymbol{\Sigma}_\xi = \left(Cov(\xi_i, \xi_j)\right)$$

$$\text{with } Var(\xi_i) = 1,$$

$$E[\boldsymbol{\varepsilon}] = \boldsymbol{0}, \quad Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \boldsymbol{I},$$

where $\boldsymbol{\Sigma}_\xi$ is a symmetric and positive definite matrix, and

$$E[\boldsymbol{Y}] = \boldsymbol{X\beta}, \quad Cov(\boldsymbol{Y}) = \alpha^2 \boldsymbol{I} + \gamma^2 \boldsymbol{\Sigma}_\xi + \sigma_\varepsilon^2 \boldsymbol{I}.$$

When there is more than one tree characteristic, we can write (3.3) in its multivariate version (cf. Rao 1973, p. 544)

$$\underline{\boldsymbol{Y}} = \underline{\boldsymbol{X}}\,\underline{\boldsymbol{\beta}} + \boldsymbol{\alpha} \otimes \boldsymbol{\tau} + \boldsymbol{\gamma} \otimes \boldsymbol{\xi} + \underline{\boldsymbol{\varepsilon}}, \tag{3.4}$$

where $\underline{\boldsymbol{Y}} = \left(\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_m^T\right)^T$,

$$\underline{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 & \cdots & \boldsymbol{0} \\ \vdots & \cdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{X}_m \end{bmatrix},$$

$\underline{\boldsymbol{\beta}} = \left(\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_m^T\right)^T$ is a matrix of coefficients, and $\otimes$ denotes the Kronecker product. For example, $\boldsymbol{\alpha} \otimes \boldsymbol{\tau} = \left(\alpha_1 \boldsymbol{\tau}^T, \ldots, \alpha_m \boldsymbol{\tau}^T\right)^T$. Finally, $\underline{\boldsymbol{\varepsilon}} = \left(\boldsymbol{\varepsilon}_1^T, \ldots, \boldsymbol{\varepsilon}_m^T\right)^T$ and $Cov(\underline{\boldsymbol{\varepsilon}}) = \boldsymbol{\Sigma}_\varepsilon \otimes \boldsymbol{I}$, where $\boldsymbol{\Sigma}_\varepsilon = \left(Cov(\varepsilon_i, \varepsilon_j)\right)$, $i, j = 1, \ldots, m$, represents the dependencies between characteristics of the same tree. The covariance of tree characteristics in (3.4) is expressed by

$$Cov(\underline{\boldsymbol{Y}}) = \boldsymbol{\alpha\alpha}^T \otimes \boldsymbol{I} + \boldsymbol{\gamma\gamma}^T \otimes \boldsymbol{\Sigma}_\xi + \boldsymbol{\Sigma}_\varepsilon \otimes \boldsymbol{I}. \tag{3.5}$$

Notice that (3.4) reduces to the univariate case if $m = 1$. If $\boldsymbol{\tau} = \boldsymbol{0}$ and $\boldsymbol{\xi} = \boldsymbol{0}$, (3.4) reduces to the so-called seemingly unrelated regressions (Zellner 1962). We do not constrain the matrices $\boldsymbol{X}_i$ to be equal, and therefore, different explanatory variables can be used for different characteristics. When the same set of predictors is used for all tree characteristics, the design matrix has the form $\underline{\boldsymbol{X}} = \boldsymbol{I} \otimes \boldsymbol{X}$.

The covariance parameterization of (3.5) accounts for spatial autoregressive models (Griffith 1988, pp. 82-106), geo-statistical models (Keitt et al. 2002, pp. 617-618) and space-time models (Cressie 1991, pp. 449-452). The choice of various models depends on the sources of variation and correlation between and among trees (see Section 3.5), on how the modeler thinks (Cressie 1991, p. 408) and on the available data (Rennolls 1997). Alternatively, criteria (such as the predictive residual sum of squares) are used to select from simple to complicated models (Gelfand et al. 1998). Note that parameters α and γ are tuning instruments that allow us to scale the source of variation based on the ecological knowledge and the data measured in real forests.

3.5 Sources of Variation

Benjamin and Hardwick (1986, p. 765) wrote

> …environmental-induced variation in plant relative growth rate must be due to variation between plants either in the availability of resources or to variation in the ability of different plants to translate those resources into biomass.

We can think that the initial heterogeneity (τ) represents the ability of different plants to utilize the available resources and the environmental effects (ξ) represent the availability of resources.

In forestry applications the net effects of τ and ξ are often estimated as one variance component (Gregoire 1987). Nevertheless, it is reasonable to allow separate random effects so that we can use the existing biological information (such as genetic variation) in simulations. However, without additional assumptions we are not able to separate the variance components relating to τ and ε.

The initial heterogeneity is a highly simplified form of the genetic source of variation (Namkoong 1981, pp. 3-6). We can think that it consists of: (1) characteristics that mother trees may transmit to their offspring; and (2) the interactions between alleles that influence the quality of the offspring. Not knowing the genetic variation, we may utilize the available information (e.g. the narrow sense heritability (Zobel and Jackson 1995, p. 35)) to get a lower bound and eventually come up with an educated guess of the proportion of variation that can be attributed to genetic sources.

The environmental effects (ξ) may be a function of: (1) available resources for a tree when there are neighboring trees sharing the available resources; (2) the distance between a tree and its neighbors; or (3) the τ of a tree and its neighbors. We expect ξ to vary from tree to tree since microenvironment (such as growing space, shading conditions, soil fertility, etc.) varies for trees at different locations.

In principle interactions between initial heterogeneity and environment effects may change the performance of trees differently in different environments. However, Zobel and Jackson (1995, pp. 45-46) suggested that the interaction is relatively small unless the differences between the environments are very large. Our models will not include such interaction.

The random component ε represents the residual variation due to the failure of simplified model assumptions concerning $X\boldsymbol{\beta}$ and the joint distributions of $\boldsymbol{\tau}$ and $\boldsymbol{\xi}$ (i.e. $\varepsilon = Y - E[Y | \boldsymbol{\tau}, \boldsymbol{\xi}]$, where $E[Y | \boldsymbol{\tau}, \boldsymbol{\xi}] = \boldsymbol{\mu}_Y + \alpha\boldsymbol{\tau} + \gamma\boldsymbol{\xi}$).

3.6 Between- and Within-Tree Correlations

The between-tree correlations are represented by the off-diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_\xi$. Cressie (1991 pp. 20-23) suggested that $\boldsymbol{\Sigma}_\xi$ is generally not known and must be parameterized for estimation. Commonly assumed models for covariance structure include moving average and autoregressive models, and a distance-decaying exponential family of functions (Richardson et al. 1992).

A possible parametric correlation function can be of the form $corr(\xi_i, \xi_j) = exp(\theta_1 + \theta_2 \|x_i - x_j\|)$, where $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$, $\theta_1, \theta_2 \le 0$ (cf. Cook and Pocock 1983, pp. 365-366). Because $Var(\xi_i) = 1$, the proposed correlation function is also a covariance function and we get $\hat{\boldsymbol{\Sigma}}_\xi(\hat{\boldsymbol{\theta}}) = exp(\hat{\theta}_1 + \hat{\theta}_2 \|x_i - x_j\|)$. This covariance function is positive definite in $\mathbb{R}^3$, and thus in $\mathbb{R}^1$ and $\mathbb{R}^2$, since it satisfies $c'(0) < 0$ and $c''(\|x_i - x_j\|) - \|x_i - x_j\| c'''(\|x_i - x_j\|) \ge 0$, where $c'(\cdot)$ is the first derivative, etc. (Christakos 1984, p. 256). In the case $i = j$, $exp(\theta_1)$ is the so-called nugget effect, which represents the correlation of two trees when the distance between them decreases to zero. Notice that although the covariance function is positive definite, some combinations of θ's may still produce non-positive definite covariance matrices in some configurations of $\{x_1, \ldots, x_n\}$.

The within-tree correlations are represented by $\boldsymbol{\alpha\alpha}^T$ and the diagonal elements of $\boldsymbol{\Sigma}_\xi$ and $\boldsymbol{\Sigma}_\varepsilon$. The initial heterogeneity may affect different tree characteristics in different ways. For m tree characteristics we have a total of $m(m-1)/2$ combinations of $\alpha_i\alpha_j$, where $i > j$, $i,j = 1,\dots,m$, when τ differs for the i th and the j th tree characteristics. In other words, $\alpha_i\alpha_j$ introduce interactions between tree characteristics.

If the initial heterogeneity effects on different tree characteristics are the same, then $\boldsymbol{\alpha\alpha}^T$ reduces to $\alpha^2 \boldsymbol{11}^T$, where $\boldsymbol{1}$ is a n-dimensional column vector with all elements 1. The same argument applies to the environmental effects. Assuming homoscedasticity, we can express the within-tree correlations by $\alpha_i\alpha_j + \gamma_i\gamma_j + \sigma_{\varepsilon_i\varepsilon_j}$. Similar to the case for the between-tree correlations, we can think that $\alpha_i\alpha_j$ and $\gamma_i\gamma_j$ represent the interactions between tree characteristics.

3.7 Stand Structure Evaluation

The goal of this study is to generate hypothetical forests with spatio-temporal dependencies. Therefore, we need to corroborate the simulation results with mapped forest stand data (Swartzman and Kaluzny 1987, pp. 209-215). In forestry literature, mean diameter, basal area and diameter frequency distribution are used to describe and compare stand structures. Because of the transformation method we implemented (Section 3.3), these descriptive statistics computed from simulated stands are guaranteed to match those from mapped stand data. In fact, these descriptive statistics do not show the discrepancies between stand structures of random and correlated tree characteristics since they are location independent statistics.

To examine locations and tree characteristics simultaneously, we can use regression analysis and empirical correlation plot (Ripley 1981, p. 51) to summarize spatial patterns of trees. Regression analysis using competition indices as explanatory variables shows the scale of local competition effects. Empirical correlation plots show the correlation structure of tree characteristics given their locations.

By simulating a large number of hypothetical stands we can construct Monte Carlo distributions of the regression coefficients and empirical correlations. We can compute regression and empirical correlation coefficients from the mapped stand data and compare these with Monte Carlo distributions of simulated coefficients to calculate empirical p-values. The empirical p-values show the likelihood of the models and input parameter values meeting the objective of mimicking the stand structure of real forests.

A small p-value, say *.01*, indicates that the modeled real forest is a rare event among the simulated forests. Because we are using models to predict the behavior of stand development processes, a p-value of *.5* does not automatically mean that we have correct models and input parameter values. This is because a wrong model together with a specific set of parameter values may accidentally produce a similar stand structure as in the mapped stand data. We can reparameterize the models and calibrate the model parameters according to the hypothesis (or knowledge) of stand development processes. However, mapped stand data may contain very limited information for us to do so.

3.8 Examples of One-Step Hypothetical Forest

A common practice in forestry is to fit the Weibull distribution
(Section 3.3) to the diameter frequency distribution (Bailey and Dell 1973).
For the longleaf pine data the parameter estimates are $\delta = 33.8$ and
$\varphi = 1.04$. Notice that the fitted Weibull distribution (Figure 3.3) does not
match well with the longleaf pine data since the longleaf pine diameter
frequency distribution exhibits a multi-modal distribution. Alternatively, we
use cubic spline to fit the edf of longleaf pine diameters. The smallest and
largest measured diameters in the mapped longleaf pine data are,
respectively, *2* cm and *76* cm. This is the support of the edf.

The transformation effects of using the Weibull distribution and edf
are examined through simulations. The results (Figure 3.4) show that the
simulated correlations using edf are about 5% smaller than the original
correlation at the *.1 - .4* interval. The transformation effects from the Weibull
distribution at the same interval are 2.5 times higher. When the original
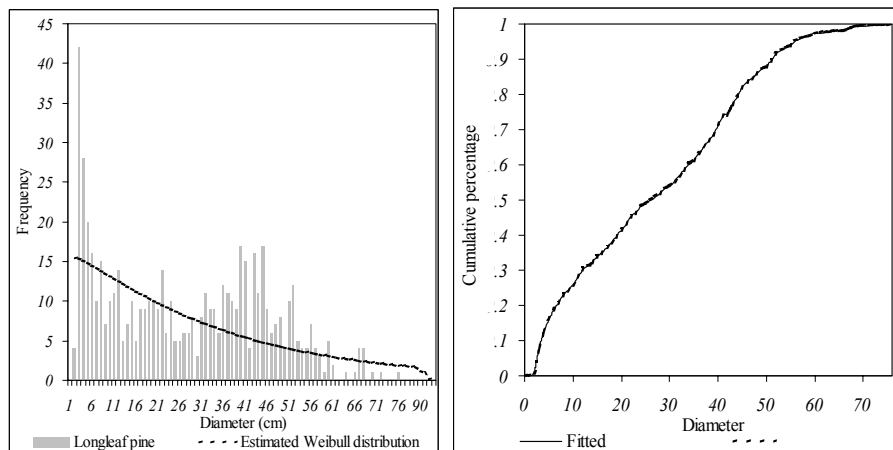
*Figure 3.3 Diameter frequency and fitted Weibull distributions (left) and the
edf and its fitted cumulative distribution function using cubic spline
interpolation method (right) of longleaf pine diameters.*

correlation is *.7* , the transformation effect using edf is *3%* smaller. This suggests that if we want to introduce autocorrelation of *.7* (as in the residuals of log-diameters), then we need to use *.72* as the original correlation for the edf, or *.75* for the Weibull distribution to carry out the transformation.
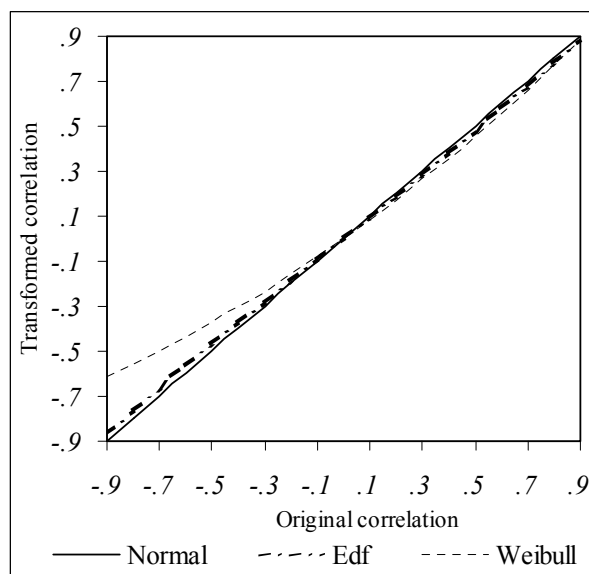
Figure 3.4 Transformation effects using edf and Weibull distributions.

Table 3.2 Regression coefficient estimates of competition indices for the longleaf pine diameter at the logarithm scale.

Explanatory Variable	Coefficient	Standard Error
Intercept	*2.52*	*.19*
Density	*-39.90*	*2.67*
Average distance	*.13*	*.02*

Mean square error = *.63* R-square = *.37*

In this example, we generate tree characteristics (diameters) for the locations that were generated in Section 2.5. Distance dependent competition indices (i.e. density of trees in $U_r(x)$ and average distances from neighboring trees to a target tree) are used as explanatory variables to obtain the estimated effects of neighboring trees. Regression analysis results (Table 3.2) indicate that the clustering of trees has negative effects on log-diameters, and that a tree has a larger log-diameter when its neighboring trees are further away.

We use the method-of-moment method to estimate α^2, γ^2 and σ_ε^2. Let $\hat{e} = Y - X\beta$ and $\hat{\Sigma}_\xi(\hat{\theta}) = exp(\hat{\theta}_1 + \hat{\theta}_2 \|x_i - x_j\|)$. Setting $E[\hat{e}\hat{e}^T] = \hat{\Sigma}_\xi(\hat{\theta})$, we can apply the least squares principles to estimate θ and γ^2 (Searle 1971, p. 450; Richardson et al. 1992, pp. 541-543). For example, a set of estimated parameter values are $\hat{\theta}_1 = -.14$, $\hat{\theta}_2 = -.85$ and $\hat{\gamma}^2 = 1.74$. This leads to a negative estimate for $\alpha^2 + \sigma_\varepsilon^2$.

As suggested by Searle (1971, pp. 407-408), the negative variance component estimate is an indication that either $\gamma^2 = 0$, or we have a wrong model. In fact, if we consider that there are three cohorts in the stand (see Section 2.5), we would need three estimates of $\hat{\gamma}$ for each cohort. Also Nelder (1954) showed that if the intra-cohort correlation is less than inter-cohort correlation, then a negative covariance component estimate can appear. Indeed the empirical correlation for juveniles is *.55* (within a neighboring distance of *8* m), which is higher than the overall empirical correlation of *.43*.

We generate four types of one-step hypothetical forests: (1) mean fixed effect with uncorrelated random components (MFUR); (2) mean fixed effect with correlated random components (MFCR); (3) trend fixed effect with uncorrelated random components (TFUR); and (4) trend fixed effect with correlated random effects (TFCR). The estimated parameters for generating the initial tree characteristics are listed in Table 3.3. The total number of simulation runs is *1200*, which is determined so that there are sufficient observations in the tails of the simulated distributions for the 95% Monte Carlo confidence interval. Note that we introduce environmental effects to generate the spatially correlated diameters, although the above analysis shows no environmental effects. The simulated mean diameter, basal area (sum over *4* ha area) and regression coefficients are listed in Table 3.4, and diameter frequency distributions are shown in Figure 3.5.

As expected, the coefficients for density and average distance indices for the stands generated with trend fixed effects (TFUR and TFCR) and the coefficients for the mapped stand are not significantly different. The empirical correlation plots (Figure 3.5) show that the simulated MFCR and TFCR stands have spatial structures corresponding to the specified

Table 3.3 Initial tree characteristics parameter estimates for generating one-step hypothetical stands.

Fixed effect	Parameter estimate
Mean	$\beta_0 = 2.92$
Trend	$\beta_0 = 2.52$, $\beta_1 = -39.9$ and $\beta_2 = .13$

Random effect	Parameter estimate
Uncorrelated	$\sigma_\varepsilon = 1.0$
Correlated	$\gamma = .78$, $\theta_1 = -.13$ and $\theta_2 = -.22$

Symbols: β_0 is intercept, β_1 and β_2 are coefficients for density and average distance competition indices, respectively, σ_ε is residual standard deviation, γ is environmental effect parameter, and θ_1 and θ_2 are parameters for the covariance function.

correlation function. However, they do not match the stand structure of longleaf pine.

When experimenting with various sets of $\hat{\boldsymbol{\theta}}$ to match the estimated correlations to the mapped stand correlation structure, non-positive definite covariance matrices were occasionally detected. Because of this we allowed the estimated correlations to decrease more rapidly than the empirical estimate correlations. Notice that the short distance negative correlations that appear in the MFUR and TFUR empirical correlation plots are the so-called spurious autocorrelations since the covariance matrices for generating log-diameters were diagonal, and the log-diameters were randomly allocated to the locations (Haining et al. 1983, p. 253).

Table 3.4 Mean diameter (cm), basal area (m^2) and regression coefficient estimates of mapped stand and simulated stands. The values in ( ) are empirical p-values.

Stand type	Mean diameter	Basal area	Explanatory variable		
			Intercept	Density	Average distance
Mapped stand	26.89	193.75	2.52	−39.90	.13
MFUR	26.77 (.57)	193.08 (.53)	2.90 (.20)	−.1 (.00)	.0 (1.0)
MFCR	26.80 (.51)	193.79 (.50)	2.90 (.14)	−.09 (.00)	.0 (1.0)
TFUR	27.01 (.48)	193.38 (.49)	2.69 (.27)	−42.81 (.81)	.12 (.69)
TFCR	27.00 (.49)	193.47 (.51)	2.67 (.35)	−44.68 (.78)	.1 (.59)
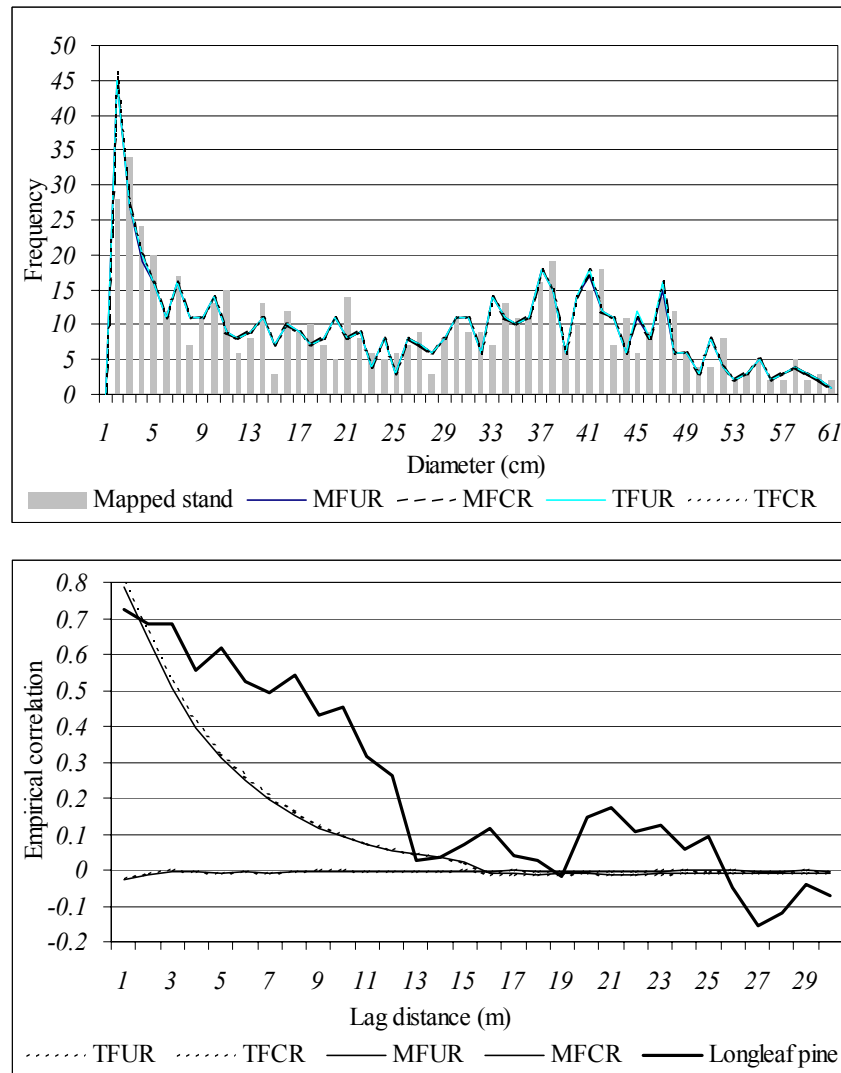
Figure 3.5 Diameter frequency distributions (top) and empirical correlation plots (bottom) of simulated forests and the longleaf pine mapped stand. The abbreviations are: TF for trend fixed effect; MF for mean fixed effect; CR for correlated random effect; and UR for uncorrelated random effect.

The negative variance component estimates raise the question of whether the model and parameter values for simulating the spatially correlated diameters are appropriate. Since we only have *583* trees in the longleaf pine data, there is not sufficient data to further analyze the spatial structure. Furthermore, the simulated forests do not take into account the ecological processes that produced the longleaf pine forest. Therefore, it is unlikely that the one-step approach will yield a stand structure that matches all aspects of the ecologically produced structure.

In fact, generating correlated tree characteristics on an inhomogeneous plane (such as forest land) via modeled covariance function are extremely difficult since correlations among trees result from simultaneous spatio-temporal processes. In the following chapter we formulate growth, mortality and reproduction models, which are structured to introduce spatial and temporal dependencies iteratively. The iterative simulation enables us to simulate growth, mortality, and reproduction processes simultaneously while working with each process individually. This is illustrated with two examples in Chapter 5.

4. GROWTH PHASE

Spatio-temporal dependencies in forests result from growth, mortality and reproduction processes over time. A natural way to reproduce the dependencies is to generate an initial forest and let the forest undergo a growth phase and reach the desired state of forest development (Goreaud et al. 1997, p. 169). In the following sections we formulate generalized linear mixed models for growth (4.1) and mortality (4.2) processes, and a model of relative intensity for the reproduction process (4.3). Examples follow at the end of each section.

4.1 Growth

Let t be the index of iterations (or period of time) and $n(t)$ be the number of trees at the end of the t th iteration. The number of trees may differ at different iterations due to mortality and reproduction. In order not to complicate the notation unnecessarily, we will re-index tree index number at each iteration. Re-indexing does not alter tree locations.

Model (3.4) becomes a yield model (cf. Vanclay 1994, pp. 106-108) with the following modifications

$$\underline{Y}^t = \underline{\mu}_Y^{t-1} + \underline{X}^t \underline{\beta}^t + \alpha^t \otimes \tau^t + \gamma^t \otimes \xi^t + \underline{\varepsilon}^t. \qquad (4.1)$$

We discussed the case when $t = 0$ in Chapter 3. When $t > 0$,
$\underline{X}^t = diag\left(\left[X_i^t, Y_i^{t-1}\right]\right)$, $i = 1,\ldots,m$. We assume that the variance components are independent between iterations except that τ remains the same for each individual tree.

Because the modeling and simulation are carried out at the logarithm scale, we can rewrite (4.1) as

$$\underline{Y}^t - \underline{\mu}_Y^{t-1} = \underline{X}^t \underline{\beta}^t + \boldsymbol{\alpha}^t \otimes \boldsymbol{\tau}^t + \boldsymbol{\gamma}^t \otimes \boldsymbol{\xi}^t + \underline{\varepsilon}^t. \qquad (4.2)$$

It follows that the left-hand side of (4.2) is the relative growth from time $t-1$ to t, and that the right-hand side represents the increment (Rao 1973, pp. 212-213). In forestry this corresponds to current (or periodic) annual increment, which depends on t.

Model (4.2) is similar to the general forms of growth equations that appear in forestry literature. For example, Zeide (1993, p. 604) showed that most of the growth equations could be simplified and linearized as

$$log\left(Y_i^t\right) = \beta_0 + \beta_1 \, log\left(Y_i^{t-1}\right) + \beta_2 \, log\left(a\right), \qquad (4.3)$$

where Y_i^t is the increment of tree characteristics (diameter, height or volume), β_0 and $\beta_1 > 0$, $\beta_2 < 0$ and a is age. An alternative formulation of (4.3) is to replace $log\left(a\right)$ with a. Note that notations in (4.3) are modified from the source reference to follow the notation conventions used in this thesis.

In general tree age is difficult to determine, especially in uneven-aged (-sized) forests (Schreuder et al. 1993, pp. 262-263). The restriction on β_2 limits the applicability of (4.3). For example, mean diameter increment increases with age for young longleaf pine trees (Platt et al. 1988, p. 502). In addition, there are only two explanatory variables in (4.3), but we may include random effects and explanatory variables in (4.2) other than past yield and age.

Growth is treated in model (4.2) as a spatio-temporal process in two aspects. Firstly, spatial component $\underline{\xi}^t$ and within tree variation $\underline{\varepsilon}^t$ are subject to the changes of $n(t)$, whereas temporal component $\underline{\tau}^t$ remains the same throughout the simulation. Parameters α^t and γ^t balance the magnitude and interactions between initial heterogeneity and environmental effects. Secondly, tree characteristics at time $t - 1$ are included in the design matrix at time t. This propagates the history of fixed and random effects in the system. Hence, we will build up spatio-temporal dependencies as the iterations go on.

4.1.1 Source of Growth Variation

Variation in the Mean Due to Soil Characteristics. Tree growth depends on many factors, but soil characteristics play an important overall role (Kozlowski 1971, p126-139). If soil and geographical information is available, we can use this information to model the mean and generate $\underline{Y}^t$ depending on tree locations. If we do not have explicit knowledge of the soil characteristics, we can think that the variations due to different soil characteristics are captured in $\underline{\xi}^t$. The magnitude of the variations is given by γ^t. However, we may still include available explanatory variables that are related to soil characteristics in the design matrix.

Competition for Light, Nutrients and Water. Trees react to the changes of growing resources and we refer to the reactions as competition (cf. Mou et al. 1993, p. 2180). We can treat competition as fixed effects by including stand characteristics (e.g. stand density) and competition indices in the design matrix. Alternatively, we can consider competition between trees as random effects that are described by $\underline{\Sigma}_{\xi}^t$. In this case γ^t corresponds to the

magnitude of competition effects. When γ^t increases due to the shortage of growing resources, α^t may decrease until trees can react to the environmental changes (Zobel and Jackson 1995, p. 37).

Temporal Variation Due to Temperature and Precipitation. Temporal climatic variations play an important role in tree growth and earlier growth significantly affects current growth. In (4.2) we can introduce the variations in the mean due to temperature and precipitation by including climatic variables and indices in the design matrix. In this case τ^t will play the role of the random component in time-series models (cf. Jordan and Lockaby 1990) and α^t represents the magnitude of climatic variations. We exclude the cases of the extreme temperature and precipitation conditions (e.g. storms), which would kill trees. The effects of extreme weather conditions on trees are discussed in Section 4.2. The effects of temperature and precipitation on tree growth are difficult to separate from other environmental effects. We can think that the temperature and precipitation conditions affect ξ^t as a multiplier (Phipps 1979) and that γ^t is determined with all other environmental factors fixed. However, we are not able to fix any part of environmental effects for estimation purposes. Therefore, γ^t represents the magnitude of mixed climatic and other environmental effects.

4.1.2 Example of Growth Modeling

There are no growth increment data for the mapped longleaf pine data. Therefore, we use external data sources to study location dependent tree growth patterns. The relative growth of Norway spruce in Alkkianvuori plot 8 and 9 data (Appendix B.2) is analyzed using the diameters (measured in 1989) at the logarithmic scale (log-diameters) and competition indices listed in Table 4.1, with $r = 10$ m, as explanatory variables. To avoid edge effects,

trees within a *10* m zone around the borders are excluded after competition indices have been calculated.

Regression analyses suggest that distance dependent indices do not significantly explain differences in growth rates, but the tree characteristic dependent competition indices do (Table 4.2). The results indicate that the relative growth decreases as log-diameters increase and when there are more neighboring large trees within *10* m, and that the relative growth increases when target trees are larger than the neighboring trees.

Note that these explanatory variables have captured most of the variations in the residuals. Therefore, we can conclude that the variation in the residuals of relative growth has been explained by the competition indices and the past yield. The small amount of variation may be due to the initial heterogeneity and/or unpredictable variations.

The silvicultural practice for Alkkianvuori plot 8 was selection thinning, which resulted in an uneven-sized stand structure. The regression analysis is repeated with plot 9 data, which was an even-sized stand that

Table 4.1 Competition indices and their computational formulas for the tree located at x_i.

Competition Index	Computational formula
Density	$C_{\lambda_i} = N(U_r(x_i))/|U_r(x_i)|$
Average-distance	$C_{a_i} = \sum_{j=1}^{N(U_r(x_i))} \|x_i - x_j\|/(N(U_r(x_i)) - 1), \ x_j \in U_r(x_i)$
Inverse distance	$C_{I_i} = \sum_{j=1}^{N(U_r(x_i))} 1/\|x_i - x_j\|, \ x_j \in U_r(x_i), Y(x_i) < Y(x_j)$
Mark-sum	$C_{s_i} = \sum_{j=1}^{N(U_r(x_i))} Y(x_j), x_j \in U_r(x_i)$
Mark-difference	$C_{d_i} = \sum_{j=1}^{N(U_r(x_i))} Y(x_i) - Y(x_j), x_j \in U_r(x_i)$

resulted from low thinning. The aim is to examine how different stand structures affect competition pattern.

Interestingly, the mark-sum index rather than mark-difference becomes significant for the trees in an even-sized stand (Table 4.2). This suggests that the relative growth is higher when there are more large trees in the neighborhood. A possible interpretation is that the mark-sum index reflects the availability of growth resources; in other words, the larger the mark-sum index the more growth resources are available. In this case the mark-sum index is behaving in a similar way to the fertility index (Pukkala 1989, p. 104). Unfortunately without site condition information (e.g. soil map) and a larger data set, we will not be able to verify the interpretation.

Table 4.2 Estimated coefficients of growth increment regression analysis for spruce trees in Alkkianvuori data with centered explanatory variables. Y^{t-1} corresponds to the log-diameter in 1989.

Alkkianvuori plot 8				
Explanatory variable	*Min.*	*Max.*	*Coefficient*	*Standard error*
Intercept	*1.00*	*1.00*	*.144*	*.001*
Y^{t-1}	*−1.03*	*.98*	*−.008*	*.003*
Inverse distance	*−4.42*	*6.72*	*−.004*	*.001*
Mark-difference	*−.12*	*.15*	*.823*	*.066*

Root MSE = *.01*, Adjusted R-Sq = *.98*

Alkkianvuori plot 9				
Explanatory variables	*Min.*	*Max.*	*Coefficient*	*Standard error*
Intercept	*1.00*	*1.00*	*.141*	*.003*
Y^{t-1}	*−.69*	*.71*	*−.040*	*.010*
Inverse distance	*−2.12*	*2.79*	*−.029*	*.003*
Mark-sum	*−.46*	*.59*	*.037*	*.013*

Root MSE = *.02*, Adjusted R-Sq = *.85*

For comparison purposes, the mean growth increments (in cm) of longleaf pine and Norway spruce under different silvicultural treatments are plotted in Figure 4.1. We can see that, for diameter < 20 cm, both tree species follow a similar growth pattern. Notice that Norway spruce trees had a growth period of *10* years, while longleaf pine was *4* years. The size (and age) of the longleaf pine trees are larger than the Norway spruce trees, which means that we do not know the effects of past yield and competition on relative growth of large longleaf pine trees. Nevertheless, for simulation purposes we can take the estimated regression coefficients from the Norway spruce stands as initial guesses and calibrate them to mimic the longleaf pine growth pattern.

A problem may arise when we apply those parameter estimates in simulating growth: the marginal distributions of initial tree characteristics (i.e. Y^0) follow the normal settings of model (3.4) in Section 3.4, but we do not know the marginal distributions for the log-diameters of mapped Norway
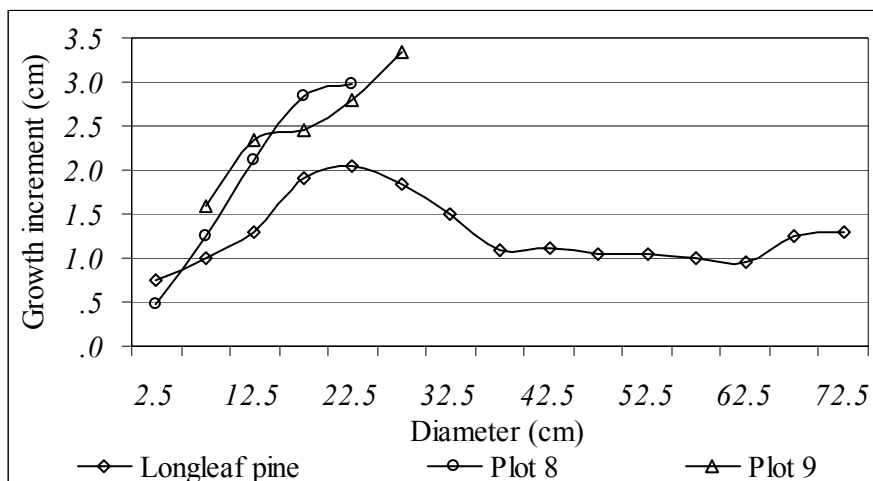

Figure 4.1 Mean diameter growth increment (cm) by 5 cm size class for Longleaf pine and Norway spruce (Alkkianvouri plot 8 and 9). The longleaf pine diameter increments (1979-1983) are interpreted from Figure 6 of Platt et al. (1988, p. 503).

spruce trees. In the cases when the log-diameters of mapped Norway spruce trees are marginally normally distributed, we can calibrate the parameter estimates for simulating longleaf pine growth. For example, the approximate p-values of the Kolmogorov-Smirnov test for normality are $> .15$ for the log-diameters from Alkkianvuori plots 8 and 9. Therefore, the parameter estimates are applicable in the normal settings of the initial characteristic model. However, we may need to carry out transformations before estimating the parameters if the empirical distribution of Norway spruce log-diameters deviates from the normal case.

4.2 Mortality

Mortality in forests may arise from aging, competition, forest damage and silvicultural practices. There are also many deaths occurring in forests that cannot be satisfactorily explained. In the following, we will distinguish tree death due to poor growth as 'regular mortality', which is due to competition and aging during stand development, and 'irregular mortality', which is due to forest damage, silvicultural treatments and unexplained causes (Oliver and Larson 1990, p. 213).

In the previous section we assume that the growth increment (i.e. $Y^t - \mu_Y^{t-1}$) is normally distributed. When simulating growth, negative relative growth can occur if the fixed effects are negative (which indicates trees are severely stressed by competition), or initial heterogeneity is at the lower tail of the standard normal distribution, or both. For simulating regular mortality we can simply kill-off (remove) trees in Y^t that have a negative relative growth. However, this method may not be appropriate for species (e.g. Norway spruce) that tolerate shade (Vanclay 1994, pp. 176-178).

Logistic regression is often used in forestry to model the probability of irregular mortality. There are two main problems in modeling mortality, namely, data availability and difficulties in identifying the cause of tree mortality. These problems make the mortality modeling difficult and increase the uncertainty of the model predictions. However, a mortality model that incorporates understandings of the ecological processes that affect mortality in forests has an increased likelihood to perform well. A series of studies to test the assumptions and applicability of the model are required to examine the model performance (Hamilton 1990).

Define $\mathbf{Y}^t = \left(Y_1^t, \dots, Y_{n(t)}^t \right)^T$ that represents tree mortality, and let Y_i^t be a Bernoulli random variable which takes value 1 if tree i dies at the t th period and 0 otherwise. Define $\pi_i^t = E\left[Y_i^t \mid X_i^t \right] = P\left(Y_i^t = 1 \mid X_i^t \right)$, where $\mathbf{X}_i^t = \left(x_{i1}, \dots, x_{ip} \right)$ is the i th row of the design matrix that consists of p explanatory variables. We can write a mixed logistic regression model of the form (cf. McCullagh and Nelder 1989, pp. 432-437)

$$logit\left(\pi_i^t \right) = \mathbf{X}_i^t \boldsymbol{\beta}^t + \phi^t \tau_i + \kappa^t \zeta_i^t, \qquad (4.4)$$

where τ_i is the initial heterogeneity, $\zeta_i^t \sim N(0, 1)$, $Cov\left(\zeta_i^t, \zeta_j^t \right) \neq 0$, represents the environmental effects, and $\phi^t, \kappa^t \in \mathbb{R}$ are the parameters of the corresponding random effects (cf. Lesaffre and Spiessens 2001, p. 327). The estimation of the parameters can be achieved using, for example, the empirical Bayes analysis (Carlin and Louis 1996, pp. 330-348).

In general competition indices and site variables can be included in $\mathbf{X}_i^t$. If mortality probability is higher near the stand edge (see for example, Peltola and Kellomäki 1993) we can either include an indicator variable in

X_i^t to flag edge-trees or take the distance of a tree to the stand edge, where higher mortality probabilities occurs, as an explanatory variable.

We can think that ζ_i^t describes the spatial variation of mortality at different t and that κ^t signifies the levels of damage and occurrence frequency. For example, rare catastrophic hurricanes can destroy forests but the damage may be restricted to particularly wind-prone sites, whereas windstorms cause frequent tree falls of a few trees per ha per year in much smaller impact areas (Waring and Schlesinger 1985, pp. 213-217).

We can use the rejection-sampling method to kill-off trees in generated forests. Given a uniform random number u, $Y_i^t = 1$, if $u \leq \pi_i^t$, and $Y_i^t = 0$ otherwise. Once the dead trees are identified, we remove them from the simulated tree population and re-index tree numbers.

4.2.1 Cause of Mortality

Mortality Due to Self-thinning. The term 'self-thinning' is used to describe the density-dependent regular mortality that occurs in crowded even-aged forests. We can model this mortality by $\Sigma_\zeta^t = \left(Cov\left(\zeta_i^t, \zeta_j^t\right) \right)$, $i \neq j$, $i, j = 1, \ldots, n(t)$, where $Cov\left(\zeta_i^t, \zeta_j^t\right)$ may be a parameterized covariance function of the stand structure, or of the pattern determined by the causal agents. The initial heterogeneity determines the competitiveness of a tree. If a tree has 'good' initial heterogeneity, then we can expect this tree to survive from competition. On the other hand, if the tree is under environmental stress, then it may die along with neighboring trees. We can think that covariances between ζ^t describe the inhomogeneity of self-thinning rate over the forest area. The role of κ^t is to indicate the scale of self-thinning. The assumptions about self-thinning may apply to even-aged

monoculture forests, but not in the uneven-aged and mixed forests (Vanclay 1994, pp. 175-176).

Mortality Due to Forest Damage. Abiotic forest damage agents (such as storms) often cause fatal mechanical injuries (such as stem breakage). Usually trees are first damaged by abiotic causes and later killed by biotic agents (such as insects). The infected trees may initiate epidemics in a forest. In this case, the number of trees infected and killed in a particular area will be higher than surrounding areas (Smith 1970, pp. 243-246). We can use ζ_i^t to represent the correlated irregular mortality due to forest damage that happens at the t th iteration and κ^t to adjust the degree of the damage. We may include explanatory variables that are closely linked to the causal agents in the model. For example, Kellomäki and Peltola (1998, p. 77) showed that snow and wind damage was closely related to the ratio between diameter and height. If the irregular mortality occurs as a chance event, we can use τ_i to represent the vigor of the tree and the trees with 'bad' initial heterogeneity will die.

Mortality Due to Silvicultural Practices. Silvicultural practices (such as thinning) are regularly executed in managed forests. There are some 'rules of thumb' as to how to carry out thinning. For example, Davis (1966, pp. 42-43) gives a rule to calculate the desired between tree spacing in terms of tree diameter and number of trees per acre from a stocking guide. Alternatively, tree diameter and density can be used as explanatory variables in the logistic regression. In the case when different thinning methods are used on forests that are sufficiently inhomogeneous (the so-called free thinning method) (Smith 1962, pp. 89-90), we can use τ_i^t to represent the decision of the field workers, and ϕ^t to represent the degree of thinning they carry out. We may

observe different harvesting patterns in the forests, which can be represented by Σ_ζ. The magnitude of the pattern is then given by κ^t.

4.2.2 Example of Mortality Modeling

We use the mapped longleaf pine data in this example. For a complete mortality analysis, we ought to have the disturbance data (e.g. storm strength and duration, harvesting path, etc.), geographical variation data, other tree characteristics (e.g. vigor status), surface vegetation and other species information. However, we have very limited information from the longleaf pine data to model the mortality probability properly. Therefore, the following analysis should be seen as a preliminary examination of the longleaf pine mortality based on the competition indices that are computed from the mapped stand data.

According to Platt et al. (1988, pp. 501-502), the annual mortality of longleaf pine varied by diameter size classes (Figure 4.4). Yearly mortality was correlated with tree size and, partly, with the annual mortality of trees in the large size classes. They noted that lightning strikes and windfalls accounted for at least *54*% and *31*%, respectively, of all death among trees
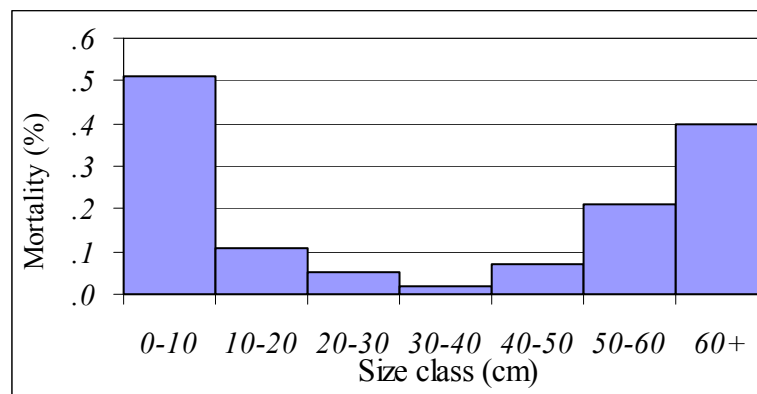
Figure 4.2 Observed mortality, by size class, in the mapped longleaf pine stand during 1979-1987.

larger than *30* cm, and that mortality increased progressively as diameter increased. The mortality of juvenile trees was mainly due to competitions from neighboring sub-adult and adult trees (see also Grace and Platt 1995). The annual mortality rate of juvenile trees was about *4-5%* in the Wade Tract mapped forest.

Based on the above observations, we analyze the mortality of longleaf pine using logistic regression with competition indices (Table 4.1) and log-diameter as explanatory variables. To examine the possible edge effects on the parameter estimation, trees in $A(r)$ (Section 2.4) with $r = 5$, *15*, *25* and *35* m are used to compute competition indices. The mortality rate of the mapped stand and the four variable-edge stands, by size classes, are listed in Table 4.4. Notice that the mortality rates at $r = 35$ m have the largest change, indicating that the edge effect may affect the mortality modeling. This is reflected in the estimated logistic regression coefficients (Table 4.3). Otherwise, the estimated coefficients of full data sets (the 'No edge' entries in Table 4.4) and reduced data sets are rather consistent with the compatible models at different neighborhood radii. We see this as an indication that edge effects are negligible for $r \le 25$ m.

Table 4.3 Mortality rates (number of trees) of full and reduced data sets by radius and size classes.

Radius for reduced data set	Juvenile	Subadult	Adult
5 m	51.3% (145)	7.7% (143)	8.6% (243)
15 m	50.8% (128)	8.2% (135)	8.4% (191)
25 m	52.3% (109)	8.9% (124)	9.1% (143)
35 m	56.8% (88)	10.3% (107)	6.1% (115)
Full data set	51.0% (155)	8.3% (157)	10.3% (271)

Table 4.4 Estimated logistic regression coefficients (standard error) with
centered competition indices, by size classes, for the longleaf pine data.

Juvenile

Radius	Explanatory variable	
	Intercept	*Mark-difference*
5 m	.05 (.17)	–.29 (.09)
No edge	.05 (.17)	–.27 (.09)
15 m	.07 (.18)	–.10 (.02)
No edge	.05 (.20)	–.10 (.02)
25 m	.05 (.17)	–.03 (.01)
No edge	.11 (.20)	–.03 (.01)

Radius	Intercept	Inverse distance
35 m	.05 (.17)	.24 (.02)

Radius	Intercept	Log-diameter	Density	Mark-sum
No edge	.48 (.26)	–11.30 (4.83)	1012.7 (419.8)	.12 (.06)

Subadult

Radius	Explanatory variable	
	Intercept	*Log-diameter*
5 m	–2.82 (.40)	–3.25 (1.10)
No edge	–2.93 (.44)	–3.34 (1.20)
15 m	–2.82 (.40)	–3.25 (1.10)
No edge	–2.85 (.44)	–3.33 (1.20)
25 m	–2.82 (.40)	–3.25 (1.10)
No edge	–2.72 (.43)	–3.17 (1.19)
35 m	–2.82 (.40)	–3.25 (1.10)
No edge	–2.53 (.42)	–3.08 (1.18)

Adult

Radius	Explanatory variable	
	Intercept	*Log-diameter*
5 m	–2.54 (.26)	4.93 (1.09)
No edge	–2.72 (.30)	4.72 (1.23)
15 m	–2.54 (.26)	4.93 (1.09)
No edge	–2.82 (.36)	5.14 (1.46)
35 m	–2.54 (.26)	4.93 (1.09)
No edge	–3.12 (.52)	4.67 (2.09)

Radius	Intercept	Log-diameter	Density
25 m (1)	–2.77 (.30)	4.26 (1.15)	–132.45 (47.64)
No edge	–3.23 (.52)	4.71 (1.81)	–138.96 (79.87)

Radius	Intercept	Log-diameter	Mark-sum
25 m (2)	–2.74 (.30)	4.58 (1.14)	–.017 (.006)
No edge	–3.16 (.50)	5.27 (1.79)	–.023 (.010)

For juveniles, the mark-difference index is significant for $r \leq 25$ m. The mortality probability increases with increasing surrounding large trees, but it decreases when the target tree is larger than the surrounding trees. The effects of this index are similar in both full and reduced data sets. Grace and Platt (1995, p. 102) suggested that the effects of large trees on juvenile mortality might extend up to 15 m. Although we come to a similar result in our analysis, the age (and size) of juveniles in Grace and Platt were much younger than those we analyzed.

For subadults, the mortality probability is explained by the size of trees (i.e. the mortality probability decreases with increase in log-diameters). This result is similar to that reported in Rathbun and Cressie (1994, p. 1172). However, they used a neighborhood distance of $r = 50$ m and a 50 m wide guard region around the borders of the mapped stand to avoid edge effects. We can see from Table 4.5 that the effects of size (log-diameter) on subadult mortality are similar despite the neighborhood conditions.

For adult trees, larger trees have a higher mortality probability. At $r = 25$ m, higher density or mark-sum indices would reduce the mortality probability. Although density is a location dependent index and mark-sum is a location and tree characteristic dependent index, they seem to have similar effects on adult tree mortality.

Rathbun and Cressie (1994, p. 1172) concluded that mortality within adults was clustered and that an adult tree had a higher mortality probability if one or more of its neighbors had died. We recompute the competition indices using only adult trees for $r = 25$. Logistic regression results show that the mortality probability is affected by the size (log-diameter) only. This indicates that the adult mortality, when all trees are included in computing competition indices, is affected by both the adults and the younger generations, which contradicts the findings in Rathbun and Cressie. In fact, large subadult trees can reach a height of 20-25 m and the average height for

adult trees is about *30* m. Thus, they can provide similar kinds of protection to adult trees from storm and lightning damage.

Platt et al. (1988, p. 502) observed that the mortality of juvenile and subadult trees was correlated with the mortality of large trees. The correlation between mortalities of different size classes may be explained by environmental effects. Although we can assume a covariance structure (or models, see the discussion in Section 3.6) for modeling such correlation, we do not have sufficient information, or long periods of records on individual tree mortality, for empirically estimating the parameters.

The coefficients in Table 4.5 are estimated from the diameters and locations of longleaf pine trees in 1979. Since we do not know the time (year) when mortality occurred, the computed competition indices may include trees that died right after 1979. These trees would have much less effect on juvenile mortality than the ones that survived until 1987.

4.3 Reproduction

Reproduction, which includes regeneration and recruitment, normally occurs when open spaces are made available due to mortality. New trees are those that have out-competed other individuals from undergrowth. The number of reproduced new trees depends on factors such as regeneration method used, site conditions, climate conditions, etc. In addition, conditions that are suitable for regeneration may not be suitable for recruitment (Coates 2002). The complicated relationships among these factors make the modeling and simulation of reproduction difficult.

Nevertheless, we can use information such as field observations to construct a simplified model for simulation purposes. For example, the density of naturally regenerated new trees is low near large mother trees (or old trees) due to lack of growth resources. The density of new trees increases as the influence of mother trees decreases, and then density decreases as the

maximum seed dispersal distance is reached (T. Pukkala, personal communication, March 2002).

There are three steps for simulating reproduction. First, we need to know the number of new trees to be generated. The number of new trees can be either obtained from field measurements or estimated from models (see for example, Kellomäki et al. 1987). Second, we need to generate the location of new trees given the locations and characteristics of old trees. Third, we need to generate new tree characteristics. We can use the mixed linear model (Chapter 3) to generate characteristics for new trees, given the locations and characteristics of old trees. A remaining question is how to generate new tree locations at the second step.

4.3.1 Generating New Tree Locations

Our goal is to find a method to generate locations of new trees that is simple to implement with limited data and information. An important consideration of the methodological development is the balance between forestry and statistical modeling. In forestry the influence of old trees on reproduction has been seen as competition effects, which are often modeled using the competitive influence-zone principle (Bella 1971), or the ecological field theory (Kuuluvainen and Pukkala 1989). These approaches attempt to model the mechanism of competition, which arises in our simulation through iteratively growth and mortality processes.

In statistical modeling, marked point processes have been used to model and simulate similar directional influences of old tree locations influence on new tree locations, but not the other way around (Högmander and Särkkä 1999). The disadvantage of this approach is that the model can produce unpredictable outcomes if the process exhibits attractions (Särkkä 1993, p. 21), which is the case in reproduction.

Rathbun and Cressie (1994) used a non-stationary Cox process to model the recruits of juvenile longleaf pine trees given the fixed locations of subadult and adult trees. They used distance to model interactions between juvenile and neighboring trees and a Gaussian random field to represent the effects of unknown or random sources of environmental variation. This approach is close to what we are searching for, but the model specification and parameter estimation are not suitable for simulation purposes.

Inspired by these studies, we propose a simpler approach to model the relative intensity of new tree locations within concentric rings around old trees. Suppose $A = \bigcup_{ij} A_{ij} \cup A_0$, $j = 0,\ldots,q$, where q is the number of concentric rings centered at x_i, $A_0 = A \setminus \bigcup_{ij} A_{ij}$, and let $N(A_{ij})$ be the expected number of new trees that are in the j th ring around the i th old tree (Figure 4.3). Note that A_{i0} represents tree bole area and $N(A_{i0}) = 0$. In general the shapes of A_{ij} are not restricted. We have

$N(A_{ij}) \sim Po(\lambda(A_{ij})|A_{ij}|)$, where $\lambda(A_{ij}) = \int_{A_{ij}} \mu(x)\,dx \big/ |A_{ij}|$ (see Section 2.4.1).
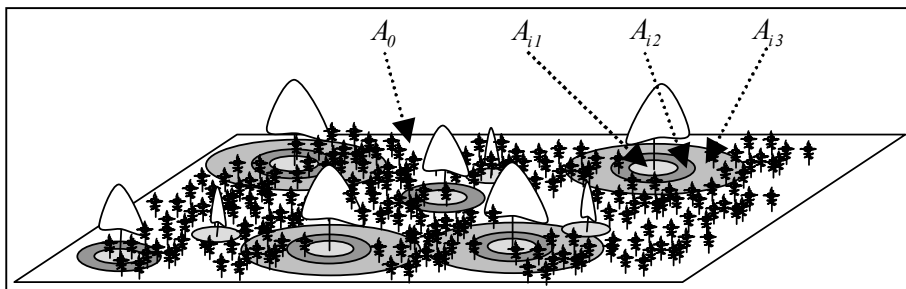

Figure 4.3 Hypothetical conditions of concentric rings around old trees.

Define $\lambda(A_0)$ to be the background intensity (i.e. the intensity surface for A_0 without the influence of old trees) that represents the environmental effects on new tree locations. For a homogeneous case $\lambda(A_0) = N(A_0)/|A_0|$. For an inhomogeneous case we can use the partitioning method introduced in Section 2.4.1 to obtain an estimated inhomogeneous intensity surface (see also Rathbun and Cressie 1994, p. 1166).

We can model the relative intensity $\lambda(A_{ij})/\lambda(A_0)$ of A_{ij} with

$$log\left(\lambda\left(A_{ij}\right)/\lambda\left(A_0\right)\right) = X_i\,\boldsymbol{\beta} + \delta_j\,, \qquad (4.5)$$

where $X_i = \left(1, X_{i1}, \ldots, X_{ip}\right)$ is the i th row of design matrix, $\boldsymbol{\beta} = \left(\beta_0, \ldots, \beta_p\right)^T$ and δ_j is the j th ring effect. The parameters $\boldsymbol{\beta}$ and δ_j can be estimated from the concentric rings using Poisson regression. This approach introduces modeling error in the intersection areas of A_{ij}. However, we can use either a rule to allocate the intersection areas (see Ford and Sorrensen 1992, pp. 376-387), or a model to summarize the effects of old trees on the intersection areas (Figure 4.4). Following Kuuluvainen and Pukkala (1989), we assume a multiplicative model for the intensities of the intersection areas. For example, suppose the relative intensity of A_{12} and
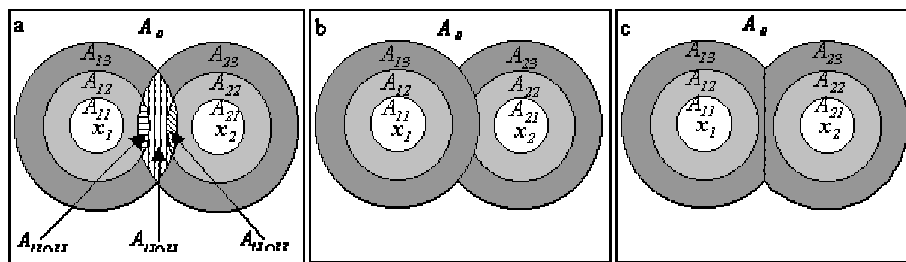

Figure 4.4 Hypothetical cases of average intensities in three ring-overlapping conditions: intersection (a), dominant (b) and shared (c).

A_{23} are $exp(X_1\beta + \delta_2)$ and $exp(X_2\beta + \delta_3)$, respectively. Then the relative intensity of $A_{12\cap23}$ is $exp((X_1 + X_2)\beta + \delta_2 + \delta_3)$.

A disadvantage of this modeling approach is that the parameters need to be estimated from non-intersection rings. If the data for parameter estimation consist only a few non-intersection rings, partially intersection rings may have to be used to estimate the parameters, which may introduce, hopefully negligible, bias in the estimates. However, the advantage of this modeling approach is that both attraction and inhibition point processes can be modeled with the same method.

The procedure to generate new tree locations is similar to the procedure for an inhomogeneous Poisson process, except that points are added instead of thinned. We propose the following procedure:

1. Generate a Poisson number n, with intensity λ, to be the number of new trees to be located. Set $i = 0$.
2. Generate x and compute $\mu(x)$, which is equal to intensity in the area where x is located.
3. Generate $u \sim U[0,1]$. If $x \notin A_{i0}$ and $u \leq \mu(x)/\lambda$, accept this point and set $i = i + 1$. Repeat steps 2 and 3 until $i = n$.

After the locations are generated, we can use (3.4) to generate tree characteristics. We assume that the reproduced trees are not yet competing with each other, but their characteristics depend on the surrounding old trees. Therefore, the competition indices in Table 4.1 need a further condition that x_j's are the locations of existing trees. Since we are generating $Y(x_i)$, the mark-difference index cannot be used.

4.3.2 Example of Reproduction Modeling

As discussed previously, we assume a log-linear model for the relative intensity of non-intersection rings and a multiplicative model for the intersection ring areas. Following Platt et al. (1988, p. 508), we determine a

three-ring model with equal ring widths of *3* m. For each ring, the number of juvenile trees that fall into the rings and the log-diameter of the old tree are recorded. This applies to the partially intersection rings. Juvenile trees not falling into any ring or intersection areas belong to the background area. Then $\boldsymbol{\beta}$ and δ_j's can be estimated simultaneously by classifying background area (A_0) and rings (A_{ij}) as factors using Poisson regression (see for example, Crawley 1993, pp. 226-264). In this example, we consider a homogeneous $\lambda(A_0)$ due to lack of information and a few partially intersection ring data are included in the analysis.

The results (Table 4.5) show that juvenile locations are inhibited around old trees for a radius of up to *6* m. Then they have a higher probability of establishment up to *9* m around the old trees when compared to the background area. Notice that the coefficient for log-diameter is negative, which means that the intensity is even lower when the juvenile trees are near larger trees. The analysis agrees with the field observations made by Grace and Platt (1995, p. 104) that

> …cohorts of seedlings entering the old-growth stand on the Wade Tract are most likely to be reduced in numbers, with patches of surviving juveniles located in areas where interference is reduced.

Table 4.5 Estimated parameter values for 3-ring effects using Poisson regression.

Explanatory variable and ring effect	Coefficient
Background intensity	−5.15
Log-diameter	−.28
δ_1	−1.19
δ_2	−.54
δ_3	.50

Similar to the procedure for generating initial tree characteristics, we can compute competition indices using the adult and subadult trees, and then estimate parameter values of the mixed linear model (3.4) for newly recruited juvenile trees. The estimated regression coefficients (Table 4.6) indicate that the size of juvenile trees decreases when the density and inverse distance indices increases. Similar to the explanations in Section 4.1.2, the mark-sum index represents the site conditions and tree size increases as the mark-sum increases.

Notice that in the generated stands it is not necessary for the size of old trees to be larger than the new trees. In the longleaf pine case, the threshold diameter values for distinguishing size classes are arbitrarily defined. Therefore, the lower tail of the distribution of the existing trees may overlap with the upper tail of the distribution of new trees.

It should be noted that when using the reproduction model for simulation purposes, the population growth rate should match the recruitment rate at different stages of stand development. Furthermore, the population growth and mortality rates need to be balanced so that the forest structure remains feasible. For example, Platt et al. (1987, p. 514) suggested that, at a low disturbance rate, a longleaf pine population size could reach an upper bound where there would be no space left for recruitment. In this case the expected number of new trees is zero.

Table 4.6 Estimated regression coefficients for juvenile longleaf pine trees with the competition indices computed by including subadult and adult trees within a 15 m radius.

Explanatory variable	Min.	Max.	Coefficient	Standard Error
Intercept	1.00	1.00	1.46	.03
Density	−.03	.02	-17.68	5.26
Inversed distance	−2.76	7.26	−.09	.02
Mark-sum	−32.47	49.65	.02	.004

Mean square error = .41, Adjusted R-Square = .26.

5. STRUCTURAL SPATIO-TEMPORAL ITERATION

The purpose of this chapter is to show how to structure the models introduced in the earlier chapters to iteratively generate hypothetical forests with spatio-temporal dependencies. We explain how to structure models to iteratively simulate forest development in Section 5.1. We describe a computer program, SPATE, that implements the iterative simulation in Section 5.2. We show two examples of iteratively generated hypothetical forests using SPATE in Section 5.3.

5.1 Structured Iterative Simulation

In the models for growth, mortality and reproduction processes, tree characteristic distributions are conditioned on tree locations and characteristics of the previous iteration. These models are conditionally independent of each other. This allows us to decompose the complicated spatio-temporal processes into a set of simpler processes, which can be represented using generalized linear mixed models. Then they can be structured to become a global model (Cressie 1991, p. 618).

At the initial iteration, simulated stands can consist of seedlings, saplings and/or trees. We will call them initial stands and they correspond to time $t = 0$ (Section 4.1). The simulated initial stands may be the end products of one-step simulations, or they can be a starting configuration for iterative simulations.

If mapped stands or inventoried sample plot data are available, we can use the statistical criteria introduced in Section 2.5 to choose a point process (Sections 2.1-2.3) for generating tree locations for the initial stand. Tree characteristics can be modeled using locations and available explanatory variables from the data. In most cases mapped stand data are collected from

mature stands (see for example, Oderwald et al. 1980). If we use these data for modeling, then the state of the generated 'initial' stands is already in maturity and the ecological processes that produced the mapped forest are ignored. This can affect the modeling of growth and mortality significantly. For example, Varmola (1996, p. 286) wrote

> …the location of a tree in a stand is determined by the genetic properties of the tree, microsite variation in the soil, the spacing resulting from the regeneration method used and the interspecific competition between trees.

Thus, we may observe a smooth growth pattern from aggregated data (Shugart 1984, p. 49-52), but there will be a large variation in the growth of individual trees (see also a cautionary remark by Kozlowski 1971, p. 38). In other words, we may have an incorrect spatio-temporal dependency model for the initial stands if we ignore (or do not know) how trees have developed over time. As shown in the examples in Section 3.8, the simulated stands may display the introduced correlation structure, but may not match the stand structure as a whole.

A drawback of modeling dependency via covariance functions is that the assumptions of stationarity (and isotropy) are needed to enable modeling of the spatial process by its mean and covariance matrix (cf. Griffith 1988, p. 17-19). Arbia (1989, p. 49) remarked that stationarity is a useful property of a stochastic process, but often unrealistic. Bennett (1979, p. 489) suggested that a non-stationary process could be treated as if it was stationary so that a representative model could be achieved to give a good description of the process. However, the outcomes of simulations using stationary models will inherit the stationarity properties, which conflict with real forests. We can take advantage of iterative simulation so that the simulated forests are no longer stationary and isotropic, although the processes in the global model are.

If we have sufficient forest inventory data, we may try a long sequence of iterations and hope that the initial stand structure will not influence the final stand structure, as in the Markov process. However, the initial stand structure is often measured at the stem exclusion or stand reinitiation stages (Oliver and Larson 1990, pp. 142-157), or it may not be known. Thus, we may still not be sure of generating the correct spatio-temporal dependencies. However, we can validate the assumptions and calibrate the models during the iteration simulations. This allows us to study the effects of individual processes in the global model and to reproduce the desired dependencies.

We can use iterative simulation to replicate the ecological processes so that new trees appear successively after major disturbances (Kellomäki et al. 1987). After we have generated an initial stand, we can iteratively grow and kill-off generated trees, and add new trees through reproduction. This allows us to simulate forest stands at any stage of forest development, and to introduce spatio-temporal dependencies to match the stand structure of real forests. Examples of hypothetical stands using iterative simulation are given in Section 5.3.

5.2 Simulation Using SPATE

SPAtial and TEmporal stand simulator (SPATE) has been developed to implement the methods. SPATE was written in standard Fortran 90 programming language. The structure of SPATE is shown in Figure 5.1. Boxes on the right-hand side display the required input parameters, which may come from empirical data analysis, prior knowledge or assumptions, or a combination of the above. The current version of SPATE simulates locations and diameters in a square area. A DOS version of SPATE and technical details are available at
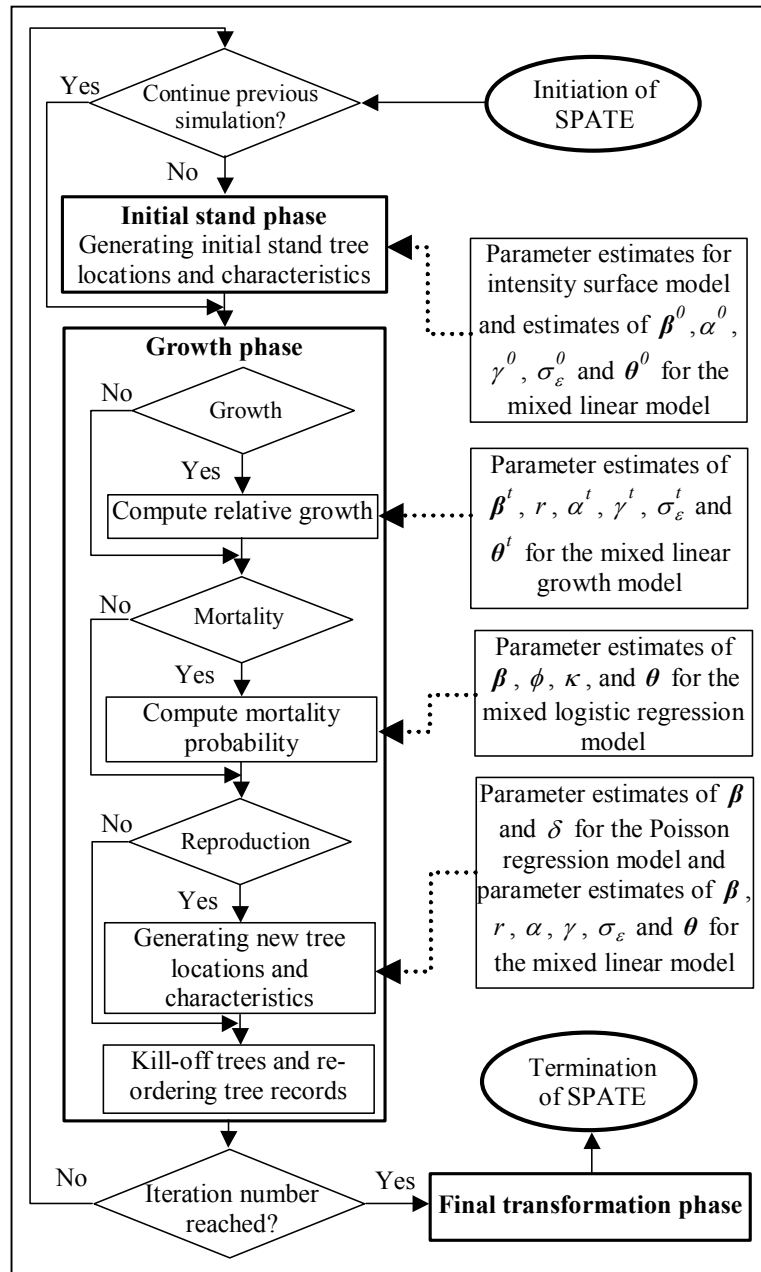
http://www.joensuu.fi/statistics/lin/spate.html

Figure 5.1 Flow chart of the procedures of and required information for the stand (SPATE). Note that $\boldsymbol{\beta}$ has different meanings for different processes in the chart.

SPATE carries out simulations in three phases: initial stand, growth, and final transformation. The outcome of the initial stand phase is either a one-step stand (e.g. the hypothetical forests of Section 3.8) or a stand for further growth. In the latter case SPATE proceeds iteratively with growth, mortality and reproduction. In both cases SPATE will write the generated log-diameters, and x- and y-coordinates to a file for each simulation run. The first two phases are repeated until the number of simulation runs is reached.

In the final transformation phase SPATE transforms the generated log-diameters to the measurement scale (e.g. cm) using cubic spline interpolations. The interpolations require empirical cumulative distribution functions (edf) from the generated log-diameters and data. SPATE can either compute edf from the generated stands and measured data, or take computed edf from external files. The edf in external files can come from earlier simulations, or they can be approximated from analytical cumulative distribution function. The transformed diameters and coordinates are written to another file. The transformation is repeated until the number of simulation runs is reached. At this point SPATE has completed the simulation.

For evaluating the simulated stand structure, a subroutine is included to compute stand structure statistics from tree records in the transformed stand files. The statistics to be computed are discussed in Section 3.7.

5.3 Examples of Iteratively Generated Hypothetical Forests

In this section we demonstrate how to use SPATE to simulate hypothetical forests. The goal is to reproduce the stand structure of the longleaf pine mapped stand (Appendix B.1). We have concluded in Section 2.5 that point pattern of the longleaf pine stand exhibits inhomogeneity. We use inhomogeneous background intensities for reproduction processes to recreate the inhomogeneity.

In general the design of a simulation strategy for carrying out iterative simulations should be based on the observed process in real forests. The number of iterations corresponds to a specific time frame. In practice it may be determined by the intervals between successive forest inventories. In any case the chosen number of iterations should correspond to the available information and assumptions made. At the end of iterative growth phase, the cdf of generated tree characteristics should match the cdf of mapped stand data. The assumptions made for the processes in the early iterations may affect the later iterations.

Inspecting the histogram of longleaf pine diameters (Figure 3.3) and the stand maps of adults, subadults and juveniles (Figure B.2), we identify cohorts with mean log-diameters that correspond to the three size classes. We assume that diameters of each cohort follow a lognormal distribution. In addition, young and mature trees are identified within cohorts. We assume that: (1) new generations are inhibited by existing mature trees; (2) young trees have higher probabilities to be reproduced in certain areas in the simulated stands; and (3) later cohorts have smaller log-diameters than the previous cohort.

We show two strategies for simulating longleaf pine stand. The first strategy (the artificial strategy) is to generate tree locations by artificially superimposing cohorts without considering growth and mortality. The second strategy (the ecological strategy) is to mimic ecological processes to generate hypothetical stands.

5.3.1 Artificial Strategy

The parameter values for background and relative intensity models are listed in Table 5.1. The distribution of the tree characteristic (i.e. log-diameter) is shown in Table 5.2 and log-diameters are randomly assigned to

the generated locations. The assumed inhomogeneous intensity surfaces are shown in Figure 5.2.

   Initial Iteration. There are two steps needed to generate the adults. The first step is to generate an initial stand (the first cohort) with a homogeneous Poisson point pattern and randomly assigned log-diameters to the locations.

Table 5.1 Input model parameter values of initial stand and reproduction location point processes for artificial strategy.

Cohort	Initial iteration
1	Homogeneous Poisson point pattern
	First iteration
2	Reproduction location model: Background intensity: $log\left(\lambda\left(A_0\right)\right) = -6.5 - .02x_1$ Relative intensity: $log\left(\lambda\left(A_{ij}\right)/\lambda\left(A_0\right)\right) = -.3Y_i^{t-1} + \delta_j$, for $Y_i^{t-1} > 3.4$, $\delta_1 = -1.0$ (*15* m) and $\delta_2 = -.5$ (*5* m)
	Second iteration
3	Reproduction location model: Background intensity: $log\left(\lambda\left(A_0\right)\right) = -10.2 + .08x_1 - .001x_2$ $\qquad -.0004x_1^2 + .000046x_1x_2 - .00002x_2^2$ Relative intensity: $log\left(\lambda\left(A_{ij}\right)/\lambda\left(A_0\right)\right) = -.5Y_i^{t-1} + \delta_j$, for $Y_i^{t-1} > 3.4$, $\delta_1 = -1.5$ (*15* m) and $\delta_2 = -1.5$ (*5* m)
	Third iteration
4	Reproduction location model: Background intensity: $log\left(\lambda\left(A_0\right)\right) = -8.0 + .047x_1 - .035x_2$ $\qquad -.0003x_1^2 + .0006x_1x_2 - .0004x_2^2$ Relative intensity: $log\left(\lambda\left(A_{ij}\right)/\lambda\left(A_0\right)\right) = -.5Y_i^{t-1} + \delta_j$, for $Y_i^{t-1} > 3.4$, $\delta_1 = -1.5$ (*10* m) and $\delta_2 = -.5$ (*5* m)

Table 5.2 Input parameter values for tree characteristics (log-diameter) in the reproduction processes for the artificial strategy.

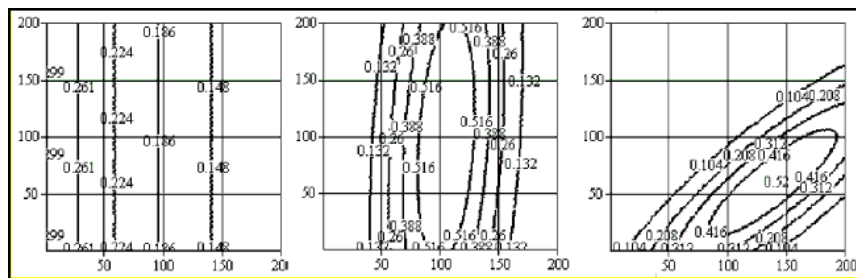| Cohort	Target		Iterative reproduction processes
	Log-diameter distribution	*No. of trees (4 ha)*	*Log-diameter Distribution*
		Initial iteration	
1	$N(4.1,.01)$	100	$N(4.1,.01)$
		First iteration	
1	$N(4.1,.01)$	100	
2	$N(3.75,.0025)$	100	$N(3.75,.0025)$
		Second iteration	
1	$N(4.1,.01)$	100	
2	$N(3.75,.0025)$	100	
3	$N(3.0,.004)$	150	$N(3.0,.004)$
		Third iteration	
1	$N(4.1,.01)$	100	
2	$N(3.75,.0025)$	100	
3	$N(3.0,.004)$	150	
4	$N(1.5,.009)$	233	$N(1.5,.009)$

Figure 5.2 Inhomogeneous background intensity surfaces on 200×200 m areas for generating new tree locations of the second (left), third (middle) and fourth (right) cohorts of the artificial strategy.

First Iteration. The next step is to generate the second cohort of trees that have smaller mean log-diameters than the first cohort, and that are inhibited by the trees of the first cohort.

Second Iteration. The second cohort is reproduced on the left-hand side of the simulated stands, which creates space for subadult and juvenile reproductions on the right-hand side of the simulated stands. The subadults (the third cohort) are reproduced with the inhibition effects from adult trees and mostly distributed near the middle of the stand. The combination of inhibition effects and the assumed inhomogeneous background intensity surface forces the subadults to be located in clumps as in the mapped stand.

Third Iteration. Finally, the juveniles (the fourth cohort) are generated near the lower-right corner of the simulated stand. Notice that the juveniles are in the top right hand side corner of the mapped longleaf pine stand. We assume that only trees with log-diameter > *3.4* have inhibition effects. This will allow juveniles to be located near the young subadult trees, but away from mature trees in the first three cohorts.

5.3.2 Ecological Strategy

For this strategy, we apply growth, mortality and reproduction processes iteratively by considering the observed longleaf pine stand dynamics. Platt et al. (1988, p. 506; p. 517) suggested

> …spatial structure of this [longleaf pine] population consisted of a
> mosaic of discrete clumps of juveniles and subadults superimposed
> upon a background matrix of widely spaced adults… and
> …recruitment within this population thus appears to occur
> primarily within open spaces created by the death of large trees.

Therefore, we first generate an adult cohort and kill-off trees in this cohort iteratively to create space for recruitment. The input parameter values for the ecological strategy are listed in Table 5.3. Table 5.4 lists the target number

of trees and the log-diameter distributions of each cohort. Log-diameters are randomly assigned to locations.

Initial Iteration. Trees in the first cohort are generated at the initial stand phase with an inhomogeneous Poisson pattern. The strategy is to allocate this cohort on the right side of the stand to create space for recruitment when they are killed-off.

First Iteration. In the first iteration about *30*% of the first cohort trees are killed-off to create openings for later cohorts. The second cohort (adult) trees are generated with a homogeneous background intensity surface and they are inhibited by the first cohort trees. The inhibition effects extend to *35* m around the first cohort trees so that sufficiently large spaces will be created later.

Second Iteration. To produce the 'background matrix of adult trees', we generate more adult trees (the third cohort) in the second iteration with an inhomogeneous background intensity surface, which force these trees to be located on the left-hand side of the simulated stands. Because the third cohort trees are densely populated, recruitment of subadult (fourth) and juvenile (fifth) cohorts is inhibited from this area.

Third Iteration. About *35*% of the remaining first cohort trees are killed-off to create more openings for juveniles (the fifth cohort). Subadults (the fourth cohort) are generated in this iteration. The simulated background intensity surface for subadults is similar to that for the third cohort in the artificial strategy. Subadult trees are inhibited by the mature trees in the first three cohorts, and they are expected to be located in the space created in the first iteration.

Table 5.3 Input model parameter values for initial stand and iterative
processes of the ecological strategy.

Cohort	Initial iteration
1	Inhomogeneous Poisson point process model for initial point pattern: $log\left(\mu\left(\boldsymbol{x}_i\right)\right) = -10.0 + .02x_{1i} - .005x_{2i}$
	First iteration
1	Mortality model: $logit\left(\pi_i\right) = -.65 + 100.0C_{\lambda_i}$, $r = 20$ m
2	Reproduction location model: Background intensity: $log\left(\lambda\left(A_0\right)\right) = -6.5$ Relative intensity: $log\left(\lambda\left(A_{ij}\right)/\lambda\left(A_0\right)\right) = -.3Y_i^{t-1} + \delta_j$, for $Y_i^{t-1} > 3.6$, $\delta_1 = -2.5$ (*30* m) and $\delta_2 = -1.5$ (*5* m)
	Second iteration
3	Reproduction location model: Background intensity: $log\left(\lambda\left(A_0\right)\right) = -6.5 - .02x_1$ Relative intensity: $log\left(\lambda\left(A_{ij}\right)/\lambda\left(A_0\right)\right) = -.5Y_i^{t-1} + \delta_j$, for $Y_i^{t-1} > 3.6$, $\delta_1 = -2.0$ (*5* m) and $\delta_2 = -1.5$ (*2* m)
	Third iteration
1	Mortality model: $logit\left(\pi_i\right) = -.65 - 200.0C_{\lambda_i}$, $r = 20$ m
4	Reproduction location model: Background intensity: $log\left(\lambda\left(A_0\right)\right) = -17.0 + .13x_1 + .005x_2$ $-.00045x_1^2 + .000026\,x_1x_2 - .000045x_2^2$ Relative intensity: $log\left(\lambda\left(A_{ij}\right)/\lambda\left(A_0\right)\right) = -.5Y_i^{t-1} + \delta_j$, for $Y_i^{t-1} > 3.6$, $\delta_1 = -1.5$ (*20* m) and $\delta_2 = -.5$ (*5* m)
	Fourth iteration
1	Mortality model: $logit\left(\pi_i\right) = 10.0$
3	Mortality model: $logit\left(\pi_i\right) = 50.0C_{\lambda_{ii}}$, $r = 10$ m
4	Growth model: $Y_i = 1.0 + .5Y_i^{t-1} + .025C_{s_i} + \varepsilon_i$, $\sigma_\varepsilon = .18$ Mortality model: $logit\left(\pi_i\right) = -.1C_{d_i}$, $r = 20$ m

Fourth Iteration. We assume that something (e.g. a storm) killed all the largest trees (the first cohort) and half of the third and fourth cohort trees in the fourth iteration. The second cohort trees are been protected by the third cohort trees. For the adult trees, the density competition index is included in the mortality model. For subadult tree, the mortality probability is determined by the mark-difference index. We select these competition indices and initial parameters based on the mortality modeling results in Section 4.3. During the fourth iteration, the subadult trees are grown by *50*% of the size from the previous iteration. We assume that trees that have a higher past yield and mark-sum competition index values grow better. The fifth cohort, which consists of juveniles, was reproduced in this iteration. The assumed inhomogeneous background intensity was similar to the one that was used to generate locations for the fourth cohort (subadults). However, the inhibition distance is extended to *45* m for the juveniles and *25* m for subadults. We used log-diameter of *3.6* as the cut-off value for mature trees to introduce inhibition effects. Therefore, subadults and juveniles might be inhibited by the same trees, but the inhibition effects are different. This forces the juveniles to be more concentrated than subadults in the same areas created by the mortality of the first cohort trees.

Table 5.3 (Continued)

Cohort	Fourth iteration
5	Reproduction location model: Background intensity: $log\left(\lambda\left(A_0\right)\right) = -18.0 + .145x_1 - .0025x_2$ $\qquad -.00045x_1^2 + .000026x_1x_2 - .000045x_2^2$ Relative intensity: $log\left(\lambda\left(A_{ij}\right)/\lambda\left(A_0\right)\right) = -.5Y_i^{t-1} + \delta_j$, for $Y_i^{t-1} > 3.6$, $\qquad \delta_1 = -2.5$ (*40* m) and $\delta_2 = -1.5$ (*5* m)

Table 5.4 Target number of trees and distributions for cohorts of the ecological strategy.

Cohort	Target		Iterative processes		
	Log-diameter distribution	No. trees (4 ha)	Growth	Mortality	Reproduction log-diameter Distribution
			Initial iteration		
1	N (4.4,.0001)	20			
			First iteration		
1	N (4.4,.0001)	13		.35	
2	N (4.0,.01)	90			N (4.0,.01)
			Second iteration		
1	N (4.4,.0001)	13			
2	N (4.0,.01)	90			
3	N (3.7,.01)	330			N (3.7,.01)
			Third iteration		
1	N (4.4,.0001)	9		.35	
2	N (4.0,.01)	90			
3	N (3.7,.01)	330			
4	N (2.0,.01)	340			N (2.0,.01)
			Fourth iteration		
1	N (4.4,.0001)	0		1.0	
2	N (4.0,.01)	90			
3	N (3.7,.01)	165		.5	
4	N (3.0,.055)	170	1.0	.5	
5	N (1.7,.1225)	160			N (1.7,.1225)

5.3.3 Simulation Results and Discussion

We carried out *1200* simulation runs for each strategy. The stand structure statistics discussed in Section 3.7 were used to evaluate simulated stand structure. The execution times for both strategies are listed in Table 5.5. The output stand files took about *70* megabytes hard disk space to store.

The non-spatial stand structure statistics and regression analyses results are listed in Table 5.6. The 95% Monte Carlo confidence intervals of the empirical correlation coefficients of the simulated stands and the empirical correlation coefficients computed from the mapped stand data are plotted in Figure 5.3. Examples of simulated stands are shown in Figure 5.4.

Table 5.5 Execution times for the examples in Section 5.3.

Type of computer	Artificial strategy	Ecological strategy
Pentium II (*200* MHz)	*140 minutes*	*230 minutes*
Pentium 4 (*1.75* GHz)	*38 minutes*	*65 minutes*
SGI Origin 2000[†]	*23 minutes*	*38 minutes*

[†] SGI Origin 2000 is a UNIX mainframe computer at the Scientific Computing Ltd., Helsinki, Finland, and technical specifications can be found at http://www.csc.fi/metacomputer/cedar.html.en

Table 5.6 Mean diameter, basal area of the mapped longleaf pine stand and simulated stands and their empirical p-values.

Variable	Mapped stand	Artificial strategy (p-value)	Ecological strategy (p-value)
Mean diameter (cm)	*26.89*	*26.69 (.61)*	*26.68 (.60)*
Basal area (m^2)	*193.75*	*192.48 (.53)*	*194.19 (.48)*
Intercept	*2.92*	*2.89 (.72)*	*2.89 (.75)*
Density	*−24.43*	*−25.32 (.57)*	*−25.59 (.57)*
Inverse distance	*−.09*	*.00 (.05)*	*−.03 (.12)*
Mark-difference	*.07*	*.08 (.06)*	*.07 (.41)*

Similar to the results in Section 3.8, the mean diameters and basal areas of the simulated stands agree well with the mapped stand. However, the spatial stand structure statistics show that the resulting stand structures of the strategies differ in other respects. The regression analysis results suggest that density and inverse distance indices computed using the artificial strategy do not conform well to the mapped stand data, although they are within the approximated confidence intervals. The ecological strategy that follows the observed stand dynamics produces better results than the artificial strategy. This is also shown in Figure 5.3 where the empirical correlations of ecological strategy stands match better, on average, with the longleaf pine data.

We expected that the coefficients computed from mapped stand data would be close to the median of the distribution of the coefficients computed from simulated stands. However, there is discrepancy between the expectations and the simulation results. Possibly, the inhomogeneity that we
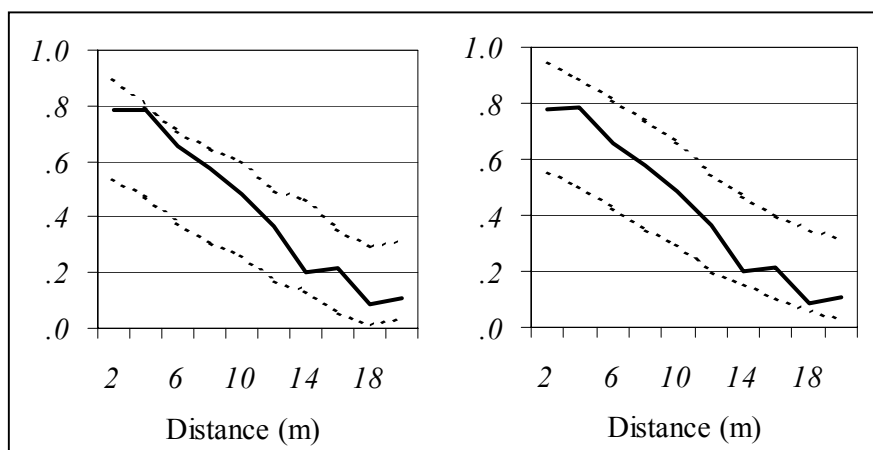

Figure 5.3 Empirical correlation plot of the mapped stand (solid line), lower and upper bounds of 95% Monte Carlo confidence intervals (dashed lines) of the simulated stands from artificial strategy (left) and ecological strategy (right).

introduced may deviate from the inhomogeneity on the mapped stand, which may have been caused by the interaction between longleaf pine and other species, by the annual ground fires and storm damages, and by salvage operations (Platt et al. 1988). Further experimentation with SPATE may reveal whether the shortcoming is due to inherent limitations of SPATE (and the global model), or whether an alternative strategy will achieve the expectations.

During the parameter calibration we noticed that a small change in one of the input parameter values could have a significant impact on the whole system. Thus, the ecological strategy required substantially more effort to calibrate the input parameter values compared with the artificial strategy. However, the parameter calibration process allowed us to validate the input parameter values against the observed ecological phenomena.
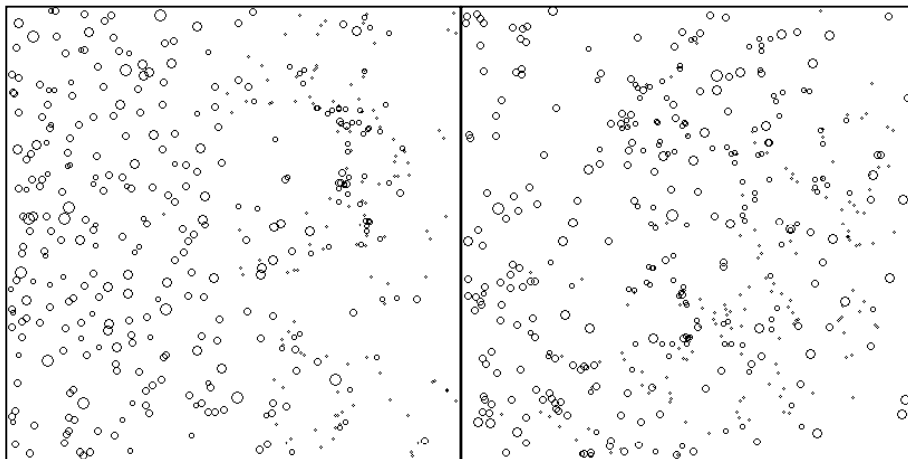

Figure 5.4 Examples of transformed stands of artificial strategy (left) and ecological strategy (right).

6. AN APPLICATION

In this chapter we demonstrate a potential use of the iterative hypothetical forest generation procedure. In conventional forest management planning, models that depend on stand characteristics (e.g. dominant height and number of trees per ha) are used to simulate tree populations and to project future yields. Location independent thinning models are often used to select trees for intermediate thinnings (see for example, Gadow and Hui 1999, pp. 53-60). However, in real forests field workers often select trees to be harvested based on local stand conditions (such as density). Location independent models do not take into account such factors. Thus, models may not match with reality. We can examine the modeling errors by comparing two thinning rules, namely, the size class and the neighborhood dependent rules. The criterion for thinning is basal area (cf. Isomäki and Niemistö 1983).

In Section 6.1 we describe a logistic regression model for thinning probabilities. In Section 6.2 we explain the size class dependent thinning rule. In Section 6.3 we explain the neighborhood dependent thinning rule and associated thinning strategies. In Section 6.4 we explain the methods for estimating thinning probability. In Section 6.5 we show the growth models for simulating growth after thinning. In Section 6.6 we examine the errors caused by the size class dependent thinning rule through simulations.

A program called 'SPATE_thinning' was developed to carry out the thinning simulation. The technical details of the thinning software are available at the same Web address for SPATE. SPATE_thinning carries out thinnings on the transformed stands (diameters in cm) and produce thinned stands (log-diameters). The thinned stands can be readmitted to SPATE for growth simulation.

6.1 Logistic Regression Model for Thinning Probability

Consider a stand $A \subset \mathbb{R}^2$. Let s be the number of thinning classes. In our application they correspond to size classes defined in terms of diameter. Let T_i be the set of trees in the ith thinning class, $i = 1, \ldots, s$, and let $n(T_i)$ be the number of trees in T_i. Let r_i be the neighborhood radius relevant for class i. Define ω_{ij} as the thinning probability of a tree located at $\boldsymbol{x}_{ij}$, $j \in T_i$. We assume a logistic model for the thinning probability, i.e.

$$logit(\omega_{ij}) = \beta_{i0} + \beta_{i1}\left(\lambda\left(U_{r_i}(\boldsymbol{x}_{ij})\right) - \lambda(A)\right),$$

where β_{i0} represents the level of thinning and β_{i1} determines the effect of local density compared to stand density, $\lambda(A)$ is the stand density, and $\lambda\left(U_{r_i}(\boldsymbol{x}_{ij})\right) = N\left(U_{r_i}(\boldsymbol{x}_{ij})\right)/\left|U_{r_i}(\boldsymbol{x}_{ij})\right|$. This is a highly simplified model, which may not reproduce all aspects of real thinning practices. However, it is sufficient for illustration purposes.

We assume that the areas outside stand edges have recently been harvested and no edge correction is imposed on the density calculation. In addition, ingrowth and mortality caused by thinning operation are not considered.

6.2 Size Class Dependent Thinning Rule

The size class dependent thinning rule (SCD) states that trees in each thinning class are thinned uniformly. For each thinning class a constant thinning probability is determined by the thinning rate, which corresponds to the thinning probability model with β_{i0} only. For trees in T_i, if $\omega_{ij} \geq u \sim U[0,1]$ then the tree located at $\boldsymbol{x}_{ij}$ is thinned.

Size class dependent thinning is commonly used in forest management planning studies. This rule does not depend on tree locations. For example, if two trees are in the same thinning class then they have the same thinning probability even though one tree is under competition stress caused by surrounding trees and the other is standing alone (T. Pukkala, personal communication, March 2003).

6.3 Neighborhood Dependent Thinning Rule

For the neighborhood dependent thinning rule (ND), the thinning probability is determined by β_{i0} and β_{i1}, which takes into account the relative density $\lambda\left(U_{r_i}\left(\boldsymbol{x}_{ij}\right)\right) - \lambda(A)$. Neighborhood dependent thinnings can be carried out in a similar way as in the previous section. However, trees that have higher relative intensities are more likely to be thinned. Obviously, the number of trees that are thinned depends on the model parameters and it may differ from SCD.

To mimic the selection process of thinning inspectors, we include three strategies, namely, DEterministic thinning (NDDE), Stochastic thinning with Fixed thinning probabilities (NDSF), and Stochastic thinning with Adjustable thinning probabilities (NDSA). To implement these thinning strategies a Poisson random number p_i is generated, based on the thinning rate of T_i and $n(T_i)$, to be the number of trees to be thinned. Thinnings begin from the largest thinning class and move toward smaller classes.

Deterministic Thinning Strategy. For NDDE, ω_{ij} is ranked in descending order, and the first p_i trees are thinned. This thinning strategy is deterministic because there is no random mechanism in the selection of trees to be thinned (cf. Pukkala and Miina 1998). We can think that NDDE

mimics the decision making of a thinning inspector who is able to rank thinning probabilities of trees based on the local density.

Stochastic Thinning Strategy with Fixed Thinning Probability. By chance, a tree with the highest thinning probability may not be selected at the first inspection. The aim of NDSF is to mimic a process in which (inexperienced) thinning inspectors may accidentally mark trees that have smaller thinning probabilities. To implement this thinning strategy at each thinning run, the tree that has the largest thinning probability is identified and checked if $\omega_{ij} \geq u \sim U[0,1]$. If $\omega_{ij} \geq u$, then this tree is marked to be thinned and the next thinning run begins. Otherwise, the tree with the second largest thinning probability is found and a check is made if $\omega_{ij} \geq u$. The above checking process continues until a tree is selected or $j = n(T_i)$. This check-and-thin procedure is repeated p_i times.

Stochastic Thinning Strategy with Adjustable Thinning Probability. In principle NDSA is similar to NDSF except that after a tree is thinned, thinning probabilities for trees in $U_{r_i}(x_{ij})$ are recomputed. This strategy reflects a thinning inspector's ability to picture the surroundings of a target tree with respect to marked trees (which will be removed later) within the neighborhood defined by r_i. Notice that in NDDE the number of trees to be thinned is fixed, but the number may vary in NDSF and NDSA. In an extreme case, no tree is thinned, but on average the number of trees to be thinned should be the same.

6.4 Thinning Probability Estimation

We arbitrarily determine a hypothetical curve (the dotted line in Figure 6.1), which represents an ideal regularly all-sized diameter frequency distribution in a managed forest for the longleaf pine data (cf. Lähde et al. 2001). Judging from Figure 6.1, simulated thinnings will be applied to the longleaf pine mapped stand for trees with diameter larger than *30* cm to reduce the amount of adult trees to a desired level. Three thinning classes, namely, *30 - 45* cm, *45 - 55* cm and *> 55* cm, are identified. The proportion of the number of trees to be thinned and neighborhood radius for these thinning classes are shown in Table 6.1.

In general β_{i0} and β_{i1} might be estimated from empirical thinning data. However, we do not have such data. Therefore, we make educated guesses for the parameter values. The rationale is that the estimated thinning probabilities should cover a reasonable range of relative intensities obtained
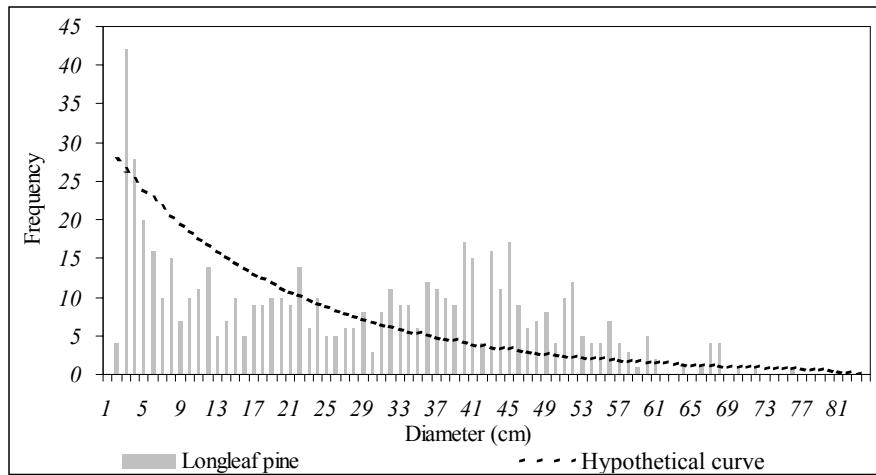

Figure 6.1 Histogram of longleaf pine diameters (cm) and a hypothetical curve that represents an ideal regularly all-sized diameter frequency distribution.

from the longleaf pine data (Table 6.2). We choose $\beta_{i0} = .5$ and, since β_{i1} discriminates the thinning probability, three levels of β_{i1}, namely, *50, 100* and *200*, for each thinning classes. For example, the thinning probabilities of the three levels of β_{i1} for the thinning class with diameter > 55 cm are shown in Figure 6.2.

6.5 Tree Growth after Thinnings

Platt et al. (1988, p. 502) suggested that during a *4* years period the growth increment was approximately *10%* of the past period for trees < 30 cm in diameter, and diameter increment decreased sharply among larger trees. Because young trees in the simulated stands are in clumps, we will consider that the growth of these trees is similar to the even-sized Norway spruce growth pattern (Table 4.3).

Table 6.1 Thinning rate and neighborhood radius for the selected thinning classes.

Thinning class	Thinning rate	Neighborhood radius (m)
30 - 45 cm	.55	10
45 - 55 cm	.65	20
> 55 cm	.55	25

Table 6.2 Minimum, maximum and mean of relative density for the selected thinning classes.

Thinning class	Min.	Max.	Mean
30 - 45 cm	−.015	.040	−.001
45 - 55 cm	−.013	.011	−.005
> 55 cm	−.013	.011	−.004

For a growth period of *10* years the growth model for young trees is

$$Y_i^1 = Y_i^0 + .2 + .05Y_i^0 - .001C_{I_i} + .0005C_{S_i} + \varepsilon_i \text{ with } \sigma_\varepsilon = .02 \text{ and a}$$

neighborhood radius of $r = 6$ m. Note that Y_i^0 are log-diameters in the
thinned stands. The growth model for mature trees (i.e. diameter larger than
30 cm) is $Y_i^1 = Y_i^0 + .01 - .005C_{I_i} + \varepsilon_i$ with $\sigma_\varepsilon = .002$ and a neighborhood
radius of $r = 30$ m.

6.6 Results and Discussions

Size class dependent (SCD) and neighborhood (ND) thinning rules are
applied to the *1200* simulated stands for the ecological strategy (Section
6.5.3). Then we grow trees in the thinned stands using SPATE. The removal
and growth of trees, in terms of basal area (m^2), of each simulated stand (*4*
ha) are recorded. The resulting mean removal, growth and correlation
between removal and growth are listed in Table 6.3. Note that removal
differs for SCD and ND thinning rules because trees have a fixed thinning
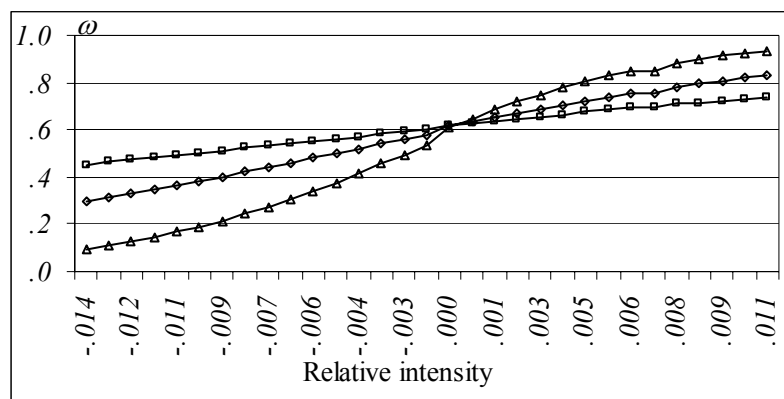

Figure 6.2 Thinning probabilities of $\beta_{i1} = 50$ *(squares),* $\beta_{i1} = 100$
(diamonds) and $\beta_{i1} = 200$ *(triangles) plotted against various relative
intensities for the largest thinning class (diameter > 55 cm).*

probability with SCD thinning rule, but the thinning probability depends on the relative intensity of a target tree with ND thinning rule (strategies).

Comparing the results of SCD and ND thinning rules, we notice that there are three errors made by applying the SCD thinning rule:

1. Variance of simulated removals is under-estimated.

2. Growth increment is over-estimated.

3. Removal and growth are either uncorrelated or weakly positively correlated when the SCD rule is applied, but they are weakly negatively correlated when the ND thinning rule is applied.

Table 6.3 Simulated removal and growth of basal area (m^2) over 4 ha thinned stands. The mean, standard deviation (Std) and correlation between removal and growth are computed over 1200 simulation runs.

Thinning	Removal		Growth		Correlation
strategy	Mean	Std	Mean	Std	(p-value)
$\beta_1 = 50$					
SCD	100.66	9.00	14.25	1.01	−.008 (.794)
NDDE	99.87	11.26	14.19	1.01	−.093 (.001)
NDSF	99.85	11.26	14.19	1.01	−.091 (.002)
NDSA	100.03	11.25	14.19	1.01	−.094 (.001)
$\beta_1 = 100$					
SCD	100.69	8.99	14.25	1.00	.019 (.505)
NDDE	99.65	11.41	14.20	1.02	−.117 (.000)
NDSF	99.64	11.41	14.20	1.02	−.117 (.000)
NDSA	99.82	11.39	14.20	1.02	−.121 (.000)
$\beta_1 = 200$					
SCD	100.57	8.76	14.24	1.01	.012 (.668)
NDDE	99.89	11.35	14.19	1.01	−.061 (.034)
NDSF	99.81	11.30	14.20	1.01	−.061 (.035)
NDSA	99.84	11.29	14.20	1.01	−.057 (.048)

Symbol: β_1 is the coefficient for relative intensity; SCD stands for Size Class Dependent thinning; NDDE stand for Neighborhood Dependent rule (ND) with DEterministic strategy; NDSF stand for Stochastic strategy with Fixed thinning probability; and NDSA stand for Stochastic strategy with Adjustable thinning probability.

Thus, by applying SCD thinning simulations forest owners or policy makers may be mislead into thinking that a heavier thinning may yield better growth afterwards, and thus to overestimate the future yield. Instead, ND thinning simulations reveal that the mean removal is smaller and has a higher variance, and that the future growth is not so high. The implication of the weakly negative correlation between removal and growth is that if a forest owner decides to carry out a heavy thinning to get more cash income now, then the future income will be lower than expected, or the owner will have to wait longer for the trees to grow.

Mäkelä et al. (2000, p. 296) pointed out

…methods of decision-making and analysis in forest management are gradually moving toward a more general, causal-oriented approach.

We have demonstrated that the model specifications are general and the iterative procedure is causal-oriented. In principle modules for other potential applications can be integrated in SPATE for more complex simulation studies.

APPENDIX A: TAYLOR-SERIES APPROXIMATION FOR CORRELATIONS

Define $X = [X_1, X_2]^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = [\mu, \mu]^T$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{bmatrix}$, $|\rho| \leq 1$. The second order Taylor-series based approximation of the function of random variable X is

$$f(X) \cong f(\mu) + f'(\mu)(X - \mu) + \tfrac{1}{2} f''(\mu)(X - \mu)^2.$$

Using the characteristic function of the bivariate normal distribution, we can show that

$$E\left[X_1^2 X_2\right] = E\left[X_1 X_2^2\right] = 2\mu\sigma^2 \rho + \mu\sigma^2 + \mu^3,$$

and that

$$E\left[X_1^2 X_2^2\right] = \sigma^4 + 2\sigma^4 \rho^2 + 2\mu^2\sigma^2 + 4\mu^2\sigma^2 \rho + \mu^4.$$

Therefore, the approximate moments of $f(X)$ are

$$E[f(X_i)] \cong f(\mu) + \tfrac{1}{2} f''(\mu)\sigma^2,$$

$$E\left[f(X_i)^2\right] \cong f(\mu)^2 + \sigma^2 f(\mu) f''(\mu) + \sigma^2 f'(\mu)^2 + \tfrac{3}{4}\sigma^4 f''(\mu)^2, \quad i = 1, 2,$$

and

$$E[f(X_1) f(X_2)] \cong f(\mu)^2 + \sigma^2 f(\mu) f''(\mu) + \sigma^2 \rho f'(\mu)^2 + \tfrac{1}{4}\sigma^4 f''(\mu)^2 + \tfrac{1}{2}\sigma^4 \rho^2 f''(\mu)^2.$$

Define $f(\mu) = G(\Phi(\mu))$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function (cdf), and $G(\Phi(\mu)) = \delta(-log(1 - \Phi(\mu)))^{\varphi^{-1}}$ is the inverse cdf of the Weibull

distribution $W(t) = 1 - exp\left(-(t/\delta)^{\varphi}\right)$, $t \in [0, +\infty)$, $\delta, \varphi \in [0, +\infty)$. Notice that we use G in replace of W^{-1} (see Section 3.3) to simplify the notations below.

In the case of $\mu = 0$ and $\sigma = 1$, we know that $\Phi(0) = .5$, $\Phi'(0) = 1/\sqrt{2\pi}$ and $\Phi''(0) = 0$. Then the equations for the $f(0)$ and its first two derivatives are

$$f(0) = G(\Phi(0)) = \delta\left(-log(.5)\right)^{\varphi^{-1}},$$

$$f'(0) = G'(\Phi(0))\Phi'(0) = -2\delta\frac{\left(-log(.5)\right)^{\varphi^{-1}}}{\sqrt{2\pi}\varphi\left(log(.5)\right)}$$

and

$$f''(0) = G''(\Phi(0))\Phi'(0)\Phi'(0) + G'(\Phi(0))\Phi''(0)$$
$$= \frac{\delta}{2\pi}\left(\frac{4\left(-log(.5)\right)^{\varphi^{-1}}}{\varphi^2\left(log(.5)\right)^2} - \frac{4\left(-log(.5)\right)^{\varphi^{-1}}}{\varphi\left(log(.5)\right)} - \frac{4\left(-log(.5)\right)^{\varphi^{-1}}}{\varphi\left(log(.5)\right)^2}\right).$$

Therefore, the variances of the transformed random variables are

$$Var(Y_i) \cong f'(0)^2 + \tfrac{1}{2}f''(0)^2,$$

and their covariance is

$$Cov(Y_1, Y_2) = f'(0)^2\rho + \tfrac{1}{2}f''(0)^2\rho^2.$$

Finally, we can calculate the correlation with respect to δ and φ as follows

$$Corr(Y_1, Y_2) = \frac{f'(0)^2\rho + \tfrac{1}{2}f''(0)^2\rho^2}{f'(0)^2 + \tfrac{1}{2}f''(0)^2}.$$

APPENDIX B: DATA SETS

B.1 Longleaf Pine Mapped Stand Data

The mapped longleaf pine (*Pinus palustris*) data (Figure B.1) was downloaded (June 2000) from http://www.stat.uga.edu/faculty/RATHBUN/ and was posted on the Web by Dr. Stephen L. Rathbun at the Department of Statistics, University of Georgia. The dataset consists of *583* trees in a *4* ha area (200×200 m) with locations (x- and y-coordinates in meters), diameter at breast height (cm) and mortality. For analysis, the trees were separated into three size classes: juvenile (diameter $\leq$ *10* cm); subadult (*10* cm $<$ diameter $\leq$ *30* cm); and adult (diameter $>$ *30* cm) (Figure B.2).

The data were collected from Wade Tract, southern Georgia, which is an old-growth longleaf pine forest. Wade Tract is part of a plantation established in the early 1800s, which has not been cleared or used for
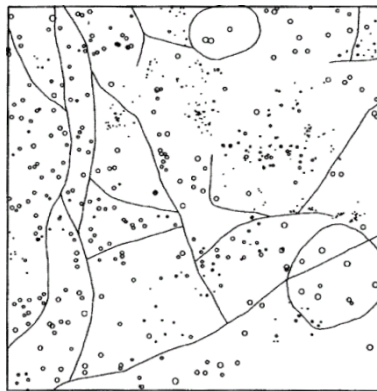

Figure B.1 Longleaf pine mapped stand (Rathbun and Cressie, 1994). The size of the circles is proportional to diameters and the lines indicate salvage paths in 1979. Reproduced with permission from The Journal of the American Statistical Association. Copyright 1994 by the American Statistical Association. All rights reserved.

agriculture, but selective logging has possibly been conducted. A study area of *39.4* ha within the Wade Tract was surveyed in 1979. All trees with diameter *> 2* cm were tagged and measured. Salvaging of pines damaged by lightning and windthrow were carried out in the 1960s and the 1970s; the salvage paths are indicated by the solid lines in Figure B.1. Stand characteristics and population dynamics were described in Platt et al. (1988). Other than longleaf pine, there were *17* additional species surveyed (about one quarter of the total mapped trees), but the locations and measurements are not available.

B.2 Alkkianvuori Experimental Plot Data

These experimental plot data have been made available to this study by Drs. Olavi Laiho (Parkano Research Station) and Erkki Lähde (Vantaa Research Center), Finnish Forest Research Institute. The Alkkianvuori stand (Parkano, Finland) was naturally regenerated in the 1940s with seed trees felled in 1960-61. The stand is a mixed Norway spruce (*Picea abies*) stand. Norway spruce dominated the stand until the first thinning in 1985-86. To examine the effects of low- and selection thinning methods, experimental
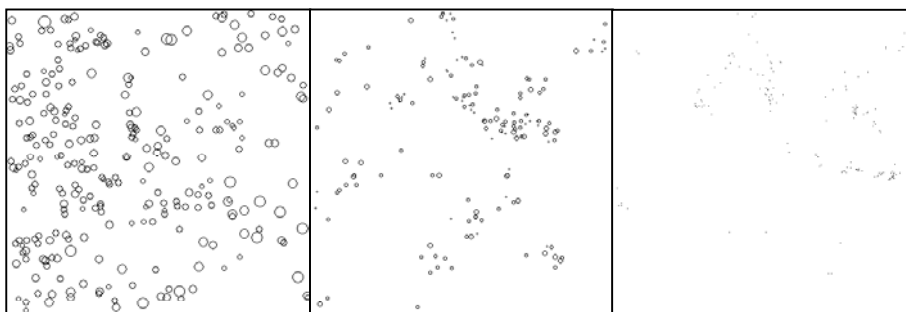

Figure B.2. Longleaf pine stand maps of the adult (left), subadult (center) and juvenile (right) size classes. Note that the circles are proportional to tree diameter and the ratio is adjusted in different maps.

plots (50×50 m) were established in 1990. Tree coordinates, relative relief, tree species, tree crown area, mortality and diameter were recorded.

Tree diameters were measured in 1990, 1994 (before treatment) and 1998. Alkkianvouri plot 8 and 9 data are used in Chapter 4. Selection thinning was applied to plot 8 and low thinning was applied to plot 9 in 1994 (Figure C.3). A buffer zone of *5* m wide around the mapped area received the same treatment, but trees were not measured.
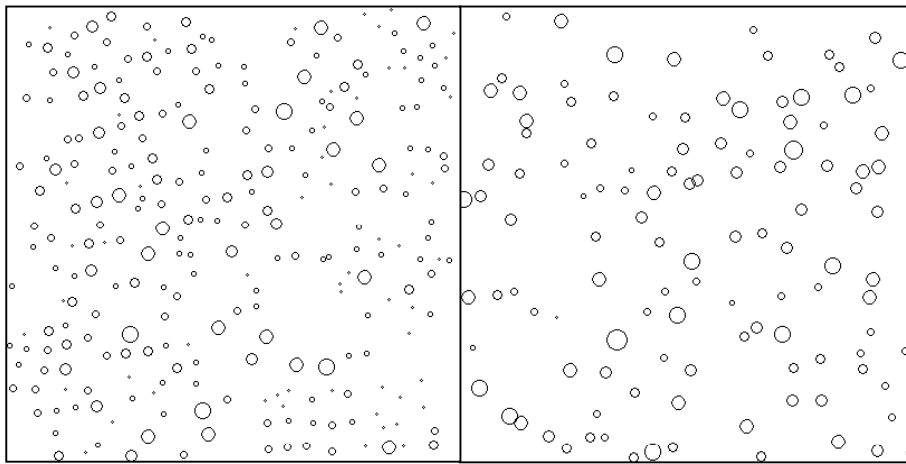

Figure B.3 Maps of Alkkianvuori experimental plot 8 (left) and 9 (right) for Norway spruce trees which have survived in 1998. The circles are proportional to diameters in 1990.

REFERENCES

Adlard, P. G. (1974), "Development of an empirical competition model for individual trees within a stand," in *Growth models for tree and stand simulation*, ed. J. Fries, Uppsala, Sweden: Royal College of Forestry, pp. 22-37.

Afifi, A. A., and Azen, S. P. (1979), Statistical Analysis - A computer oriented approach (2nd ed.), New York, NY: Academic Press, Inc.

Arbia, G. (1989), Spatial data configuration in statistical analysis of regional economic and related problems, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000), "Non- and semi-parametric estimation of interaction in inhomogeneous point patterns," *Statistica Neerlandica*, 54, 329-350.

Baddeley, A. J., and Silverman, B. W. (1984), "A cautionary example on the use of second-order methods for analyzing point patterns," *Biometrics*, 40, 1089-1093.

Bailey, R. L., and Dell, T. R. (1973), "Quantifying diameter distributions with the Weibull function," *Forest Science*, 19, 97-104.

Bella, I. E. (1971), "A new competition model for individual trees," *Forest Science*, 17, 364-372.

Benjamin, L. R., and Hardwick, R. C. (1986), "Sources of variation and measures of variability in even-aged stands of plants," *Annals of Botany*, 58, 757-778.

Bennett, R. J. (1979) Spatial time series, London, UK: Pion Limited.

Box, G. E. P., and Muller, M. E. (1958), "A note on the generation of normal deviates," *Annals of Mathematical Statistics*, 29, 610-611.

Carlin, D. P., and Louis, T. A. (1996), Bayes and empirical Bayes methods for data analysis, London, UK: Chapman & Hall.

Christakos, G. (1984), "On the problem of permissible covariance and variogram models," *Water Resources Research*, 20, 251-265.

Coates, K. D. (2002), "Tree recruitment in gaps of various size, clearcuts and undisturbed mixed forest of interior British Columbia, Canada," *Forest Ecology and Management*, 155, 387-398.

Cook, D. G., and Pocock, S. J. (1983), "Multiple regression in geographical mortality studies, with allowance for spatially correlated errors," *Biometrics*, 39, 361-371.

Crawley, M. J. (1993), GLIM for Ecologists, Oxford, UK: Blackwell Scientific Publications.

Cressie, N. (1991), Statistics for spatial data, New York, NY: John Wiley & Sons, Inc.

Davis, K. P. (1966), Forest management: Regulation and Valuation (2nd ed.), New York, NY: McGraw-Hill Book Company.

Diggle, P. J. (1983), Statistical analysis of spatial point patterns, London, UK: Academic Press Inc.

Dunster, J., and Dunster, K. (eds.) (1996), Dictionary of natural resource management, Oxon, UK: CAB International.

Dudewicz, E. J., and Mishra, S. N. (1988), Modern mathematical statistics, New York, NY: John Wiley & Sons, Inc.

Ek, A. R., and Monserud, R. A. (1974), FOREST: A computer model for simulating the growth and reproduction of mixed-species forest stands, Report R2635, Madison, WI: School of Natural Resources, University of Wisconsin-Madison.

Feller, W. (1971), An introduction to probability theory and its applications (Vol. 2, 2nd ed.), New York, NY: John Wiley & Sons, Inc.

Ford, E. D., and Sorrensen, K. A. (1992), "Theory and models of inter-plant competition as a spatial process," in *Individual-based models and approaches in ecology: populations, communities, and ecosystems*, eds. D. L. DeAngelis, and L. J. Gross, New York, NY: Routledge, Chapman and Hall, Inc., pp. 393-407.

Gadow, K., and Hui, G. (1999), Modelling Forest Development, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Gelfand, A. E., Ghosh, S. K., Knight, J. R., and Sirmans, C. F. (1998), "Spatio-temporal modeling of residential sales data," *Journal of Business & Economic Statistics*, 16, 312-321.

Goreaud, F., Courbaud, B., and Collinet, F. (1997), "Spatial structure analysis applied to modelling of forest dynamics: a few examples," in *Empirical and process-based models for forest tree and stand growth simulation*, eds. A. Amaro and M. Tomé, Lisbon, Portugal: Edições Salamandra, Lda., pp. 155-172.

Grace, S. L., and Platt, W. J. (1995), "Neighborhood effects on juveniles in an old-growth stand of longleaf pine, *Pinus palustris*," *Oikos*, 77, 99-105.

Gregoire, T. G. (1987), "Generalized error structure for forestry yield models," *Forest Science*, 33, 423-444.

—— (1993), "Estimating of forest growth from successive surveys," *Forest Ecology and Management*, 56, 267-278.

Griffith, D. A. (1988), Advanced spatial statistics, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Haining, R., Griffith, D. A., and Bennett, R. (1983), "Simulating two-dimension autocorrelated surface," *Geographical Analysis*, 13, 248-255.

Hamilton, D. A. (1990), "Extending the range of applicability of an individual tree mortality model," *Canadian Journal of Forest Research*, 20, 1212-1218.

Hasenauer, H., Monserud, R. A., and Gregoire, T. G. (1998), "Using simultaneous regression techniques with individual-tree growth models," *Forest Science*, 44, 87-95.

Hsuan, F. (1979), "Generating uniform polygonal random pairs," *Applied Statistics*, 28, 170-172.

Högmander, H., and Särkkä, A. (1999), "Multitype spatial point patterns with hierarchical interactions," *Biometrics*, 55, 1051-1058.

Isomäki, A., and Niemistö, P. (1983), Koealapuuston harvennusvalinta tietokone-ohjelman avulla (Abstract: The selection of trees in thinning experiments: a computer method), Folia Forestalia 557, Helsinki, Finland: Metsäntutkimuslaitos.

Jensen, E. B. V., and Nielsen, L. S. (2000), Inhomogeneous spatial point processes, unpublished technical report, Aarhus, Denmark: Laboratory for Computational Stochastics, University of Aarhus.

Johnson, N. L., and Kotz, S. (1970), Continuous univariate distributions-1, New York, NY: John Willey & Sons, Inc.

Jordan, D. N., and Lockaby, B. G. (1990), "Time series modelling of relationships between climate and long-term radial growth of loblolly pine," *Canadian Journal of Forest Research*, 20, 738-742.

Keitt, T. H., Bjørnstad, O. N., Dixon, P. M., and Citron-Pousty, S. (2002), "Accounting for spatial pattern with modeling organism-environment interactions," *Ecography*, 25, 616-625.

Kellomäki, S., Hänninen, H., Kolström, T., Kotisaari, A., and Pukkala, T. (1987), "A tentative model for describing the effects of some regenerative processes on the properties of natural seedling stands," *Silva Fennica*, 21, 1-10.

Kellomäki, S., Ikonen, V., Pletola., H., and Kolström, T. (1999), "Modelling the structural growth of Scots pine with implications for wood quality," *Ecological Modelling*, 122, 117-134.

Kellomäki, S., and Peltola, H. (eds.) (1998), Silvicultural strategies for predicting damage to forests from wind, fire and snow, Silva Carelica 73, Joensuu, Finland: Faculty of Forestry, University of Joensuu.

Kelly, F. P., and Ripley, B. D. (1976), "A note on Strauss's model for clustering," *Biometrika*, 63, 357-60.

Kokkila, T., Mäkelä, A., and Nikinmaa, E. (2002), "A method for generating stand structures using Gibbs marked point process," *Silva Fennica*, 36, 65-277.

Kozlowski, T. T. (1971), Growth and development of trees (Vol. 2), New York, NY: Academic Press.

Kuuluvainen, T., and Pukkala, T. (1989), "Effect of Scots pine seed trees on the density of ground vegetation and tree seedlings," *Silva Fennica*, 23, 159-167.

Lakhan, V. C. (1981), "Generating autocorrelated pseudo-random numbers with specific distributions," *Journal of Statistical Computer Simulation*, 12, 303-309.

Lappi, J., and Bailey, R. L. (1988), "A height prediction model with random stand and tree parameters: An alternative to traditional site index methods," *Forest Science*, 34, 907-927.

Lesaffre, E., and Spiessens, B. (2001), "On the effect of the number of quadrature points in a logistic random-effects model: an example," *Applied Statistics*, 50, 325-335.

Lehmann, E. L. (1991), Theory of point estimation, Belmont, CA: Wadsworth, Inc.

van Lieshout, M. N. M. (2000), Markov point processes and their applications, London, UK: Imperial College Press.

van Lieshout, M. N. M., and Baddeley, A. J. (2001), "Extrapolating and interpolating spatial patterns," Technical Report PNA-R0117, Amsterdam, The Netherlands: National Research Institute for Mathematics and Computer Science.

Lewis, P. A. W., and Shedler, G. S. (1979), "Simulation of non-homogeneous Poisson processes by thinning," *Naval Research Logistics Quarterly*, 26, 403-413.

Lähde, E., Laiho, O., and Norokorpi, Y. (2001), "Structure transformation and volume increment in Norway spruce-dominated forests following contrasting silvicultural treatments," *Forest Ecology and Management*, 151, 133-138.

Mackisack, M. S., and Wood, G. B. (1990), "Simulating the forest and the point-sampling process as an aid in designing forest inventories," *Forest Ecology and Management*, 36, 79-103.

McCullagh, P., and Nelder, J. (1989), Generalized linear models (2nd ed.), London, UK: Chapman and Hall Ltd.

Metsäntutkimuslaitos (1985), Valtakunnan metsien 8. inventointi pysyvien koealojen kenttätyön ohjeet, Helsinki, Finland: Author.

Mou, P., Mitchell, R. J., and Jones, R. H. (1993), "Ecological field theory model: a mechanistic approach to simulate plant-plant interactions in southeastern forest ecosystems," *Canadian Journal of Forest Research*, 23, 2180-2193.

Mäkelä, A., Landsberg, J., Ek, A., Burk, T., Ter-Mikaelian, M., Ågren, G., Oliver, C., and Puttonen, P. (2000), "Process-based models for forest ecosystem management: current state of the art and challenges for practical implementation," *Tree Physiology*, 20, 289-298.

Namkoong, G. (1981), Introduction to quantitative genetics in forestry, London, UK: Castle House Publications.

Nelder, J. A. (1954), "The interpretation of negative components of variance," *Biometrika*, 51, 544-548.

Newnham, R. M. (1968), The generation of artificial population of points (spatial pattern) on a plane, Information Report FMR-X-10, Ottawa, Canada: Forest Management Institute.

Neyman, J., and Scott, E. L. (1958), "Statistical approach to problems of cosmology," *Journal of the Royal Statistical Society*, *Ser. B*, XX, 1-28.

Oderwald, R. G., Stuart, W. B., and Farrar, K. D. (1980), "The forest model file: A mapped stand library at Virginia Tech," *Forest Science*, 26, 193-194.

Oliver, C., and Larson, B. (1990), Forest stand dynamics, New York, NY: John Wiley & Sons, Inc.

Palley, M. N., and O'Regan, W. G. (1961), "A computer technique for the study of forest sampling methods," *Forest Science*, 7, 282-294.

Patil, G. P., Boswell, M. T., and Ratnaparkhi, M. V. (1984), Dictionary and classified bibliography of statistical distributions in scientific work, Volume 2: Continuous univariate models, Burtonsville, MD: International Co-operative Publishing House.

Peltola, H., and Kellomäki, S. (1993), "A mechanistic model for calculating windthrow and stem breakage of Scots pine at stand edge," *Silva Fennica*, 27, 99-111.

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

Phipps, R. L. (1979), "Simulation of wetlands forest vegetation dynamics," *Ecological Modelling*, 7, 257-288.

Platt, W. J., Evans, G. W., and Rathbun, S. L. (1988), "The population dynamics of a long-lived conifer (*Pinus palustris*)," *The American Naturalist*, 131, 491-525.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992), Numerical recipes in FORTRAN: the art of scientific computing (2nd ed.), New York, NY: Cambridge University Press.

Pretzsch, H. (1997), "Analysis and modelling of spatial stand structures. Methodological considerations based on mixed beech-larch stands in Lower Saxony," *Forest Ecology and Management*, 97, 237-253.

Pukkala, T. (1987), "Simulation model for natural regeneration of *Pinus sylvestris*, *Picea abies*, *Betula pendula* and *Betula pubescens*," *Silva Fennica*, 21, 37-53.

——(1988), Studies on the effect of spatial distribution of trees on the diameter growth of Scots pine, Ph.D. dissertation, University of Joensuu, Joensuu, Finland.

——(1989), "Prediction of tree diameter and height in a Scots pine stand as a function of the spatial pattern of trees," *Silva Fennica*, 23, 83-99.

Pukkala, T., and Miina, J. (1998), "Tree-selection algorithms for optimizing thinning using a distance-dependent growth model," *Canadian Journal of Forest Research*, 28, 693-702.

Rao, C. (1973), Linear statistical inference and its applications, New York, NY: John Wiley & Sons, Inc.

Rathbun, S. L., and Cressie, N. (1994), "A space-time survival point process for a longleaf pine forest in southern Georgia," *Journal of the American Statistical Association*, 89, 1164-1174.

Rennolls, K. (1997), "Data requirements for forest modelling," in *Empirical and process-based models for forest tree and stand growth simulation*, eds. A. Amaro and M. Tomé, Lisbon, Portugal: Edições Salamandra, Lda., pp. 11-22.

Rényi, A. (1967), "Remarks on the Poisson process," *Studia Scientiarum Mathematicarum Hungarica*, 2, 119-123.

Richardson, S., Guihenneuc, C., and Lasserre, V. (1992), "Spatial linear models with autocorrelated error structure," *The Statistician*, 41, 539-557.

Ripley, B. D. (1976), "The second order analysis of stationary point processes," *Journal of Applied Probability*, 13, 255-266.

—— (1977), "Modelling spatial patterns," *Journal of the Royal Statistics Society*, Ser. B, 39, 172-192.

—— (1979), "Simulating spatial patterns: Dependent samples from a multivariate density," *Applied Statistics*, 28, 109-112.

—— (1981), Spatial statistics, New York, NY: John Wiley and Sons, Inc.

—— (1984), "Present position and potential developments: Some personal views statistics in the natural sciences," *Journal of the Royal Statistics Society*, Ser. A, 147, 340-348.

Schreuder, H. T., Gregorie, T., and Wood, G. (1993), Sampling methods for multipurpose forest inventory, New York, NY: John Willey & Sons, Inc.

Searle, S. (1971), Linear models, New York, NY: John Wiley & Sons, Inc.

Shugart, H. H. (1984), A theory of forest dynamics, New York, NY: Springer-Verlag New York, Inc.

Smith, D. M. (1962), The practice of silviculture, New York, NY: John Wiley and Sons, Inc.

Smith, W. H. (1970), Tree pathology, New York, NY: Academic Press.

Stoyan, D., and Penttinen, A. (2000), "Recent applications of point process methods in forestry statistics," *Statistical Science*, 15, 61-78.

Swartzman, G. L., and Kaluzny, S. P. (1987), Ecological simulation primer, New York, NY: Macmillan Publishing Company.

Särkkä, A. (1993), "Pseudo-likelihood approach for pair potential estimation of Gibbs processes," Ph.D. dissertation, University of Jyväskylä, Dept. of Mathematics and Statistics, Jyväskylä, Finland.

Särndal, C. E., Swensson, B., and Wretman, J. (1992), Model assisted survey sampling, New York, NY: Springer-Verlag New York, Inc.

Tomppo, E. (1986), Models and methods for analyzing spatial patterns of trees, Communicationes Instituti Forestalis Fenniae 138, Helsinki, Finland: Finnish Forest Research Institute.

Vanclay, J. K. (1994), Modelling forest growth and yield - Applications to mixed tropical Forests, Oxon, UK: CAB International.

Varmola, M. (1996), "Initial density and precommercial thinning affects the development of Scots pine cultivations: A simulation model for combining quantity and quality," in *Modelling regeneration success and early growth of forest stands*, eds. J. P. Skovsgaard, and V. K. Johannsen, Copenhagen, Denmark: Danish Forest and Landscape Research Institute, pp. 285-297.

de Vries, P. G. (1986), Sampling theory for forest inventory: A teach-yourself course, Berlin, Germany: Springer-Verlag.

Waring, R. H., and Schlesinger, W. H. (1985), Forest ecosystems: concepts and management, Orlando, FA: Academic Press, Inc.

Wright, J. W. (1976), Introduction to forest genetics, New York, NY: Academic Press.

Zeide, B. (1993), "Analysis of growth equations," *Forest Science*, 39, 594-616.

Zellner, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregations bias," *Journal of the American Statistical Association*, 57, 348-368.

Zobel, B. J., and Jackson, B. J. (1995), Genetics of weed production, Berlin, Germany: Springer-Verlag.

JOENSUUN YLIOPISTON YHTEISKUNTATIETEELLISIÄ JULKAISUJA
UNIVERSITY OF JOENSUU, PUBLICATIONS IN SOCIAL SCIENCES

1. Hautamäki, Jarkko. Peruskoululaisten loogisen ajattelun mittaamisesta ja esiintymisestä. English abstract: The Measurement and Distribution of Piagetian Stages of Thinking in Finnish Comprehensive School. Joensuu 1984. 157 s.

2. Nyblom, Jukka. Comparisons of Tests for the Presence of Random Walk Coefficients in Two Simple Linear Models. Joensuu 1984. 18 p.

3. Vartiainen, Perttu. Maantieteen konstituoitumisesta ihmistieteenä. English abstract: On the Constitution of Geography as a Human Science. Joensuu 1984. 590 s.

4. Saariluoma, Pertti. Essays on the Foundations of Cognitive Psychology. Joensuu 1985. 129 p.

5. Eskelinen, Heikki. Sijainti ja tuotantorakenne riippuvuuksineen. Tutkimuksia alue-taloudellisesta integraatiosta. English abstract: Interdependencies in Location and Production Structure. Joensuu 1985. 73 s.

6. Hilpelä, Jyrki. Filosofia kritiikkinä. English abstract: Philosophy as Criticism. Joensuu 1986. 205 s.

7. Karjalainen, Pauli. Geodiversity as a Lived World: On the Geography of Existence. Joensuu 1986. 190 p.

8. Paasi, Anssi. Neljä maakuntaa. Maantieteellinen tutkimus aluetietoisuuden kehittymisestä. English abstract: Four Provinces in Finland. A Geographical Study of the Development of Regional Consciousness. Joensuu 1986. 345 s.

9. Paasi, Anssi. The Institutionalization of Regions. Theory and Comparative Case Studies. Joensuu 1986. 36 p.

10. Räty, Hannu. Uhka vai uhri. Tutkimus mielisairauteen asennoitumisesta. English abstract: Threat or Victim. A Study of Attitudes toward Mental Illness. Joensuu 1987. 142 s.

11. Sabour, M'hammed. Homo Academicus Arabicus. Joensuu 1988. 329 p.

12. Rannikko, Pertti. Metsätyö-pienviljelykylä. Tutkimus erään yhdyskuntatyypin noususta ja tuhosta. Joensuu 1989. 114 s.

13. Ahponen, Pirkkoliisa. Kulttuuripolitiikka ja sen representaatiot. Tutkimus yle-vän maallistumisesta. Joensuu 1991. 311 s.

14. Sabour, M'hammed (ed.). Liberté, Egalité, Fraternité. Bicentenaire de la Grande Révolution Française. Joensuu 1992. 246 s.

15. Häyrynen, Yrjö-Paavo & Perho, Hannu & Kuittinen, Matti & Silvonen, Jussi. Ilmapiirit, kentät ja kulttuurit. Suomen korkeakoulutus 1973-1989. Joensuu 1992. 151 s.

16. Turunen, Panu. Der unbefangene Blick: Mach, Husserl und die ursprüngliche Anschauung. Joensuu 1993. 169 s.

17. Koski, Leena. Tieteen tahtomana, yliopiston tekemänä. Yliopiston sisäiset symboliset järjestykset. Joensuu 1993. 201 s.

18. Pirttilä, Ilkka. Me ja maailman mallit. Tiedonsosiologian ydintä etsimässä. Joensuu 1993. 244 s.

19. Mjelde, Liv. Apprenticeship. From Practice to Theory and Back Again. Joensuu 1993. 308 p.

20. Leiman, Mikael. Integrating the Vygotskian Theory of Sign-mediated Activity and the British Object Relations Theory. Joensuu 1994. 70 p.

21. Puuronen, Vesa. Nuoret maailmansa tuottajina? Tutkimus nuorisososiologian metodologiasta ja nuorten keskusteluista. Joensuu 1995. 186 s.

22. Niemeläinen, Heikki. Essays on Labor Supply, Wage Formation and the Trade Union Fee. Joensuu 1995. 97 p.

23. Hölttä, Seppo. Towards Self-Regulative University. Joensuu 1995. 278 p.

24. Kortelainen, Jarmo. Tehdasyhdyskunta talouden ja ympäristötietoisuuden murrosvaiheissa. Joensuu 1996. 155 s.

25. Hong, Jianzhong. Development of Verbal and Non-verbal Thinking. A Review and Re-analysis of Studies in Children of Six Nationalitites on the Qinghai Plateau. Joensuu 1996. 159 s.

26. Kekäle, Jouni. Leadership Cultures in Academic Department. Joensuu 1997. 253 p.

27. Tammi, Timo. Essays on the Rationality of Experimentation in Economics. The Case of Preference Reversal Research. Joensuu 1997. 194 p.

28. Moisseinen, Eija. Contingent Valuation. The Case of the Saimaa Seal. Joensuu 1997. 208 p.

29. Jolkkonen, Arja. Paikan vaihto. Tutkimus irtisanomisuhan alaisten naisten työ-markkinastrategioista ja paikallisista työmarkkinoista. Joensuu 1998. 328 s.

30. Oksa, Jukka. Syrjäkylä muutoksessa. Joensuu 1998. 300 s.

31. Räty, Tarmo. Efficient Facets, Influential Observations and Revisions of the Best Practice Boundary for DEA. Joensuu 1998. 141 p.

32. Ilmonen, Kari. Kulttuuri alueen käyttövarana. Kulttuuripolitiikan hallitsevat ja marginaaliset diskurssit Keski-Pohjanmaalla. Joensuu 1998. 287 s.

33. Varis, Eira. Syrjäkylien murros Venäjän Karjalassa ja Unkarissa. Tutkimus post-sosialistisen maaseudun restrukturaatiosta ja resurssiyhdyskuntien selviytymisestä. Joensuu 1998. 286 s.

34. Vähämöttönen, Timo. Reframing Career Counselling in Terms of Counsellor-Client Negotiations. An Interpretive Study of Career Counselling Concepts and Practice. Joensuu 1998. 183 p.

35. Komulainen, Katri. Kotihiiriä ja ihmisiä. Retorinen minä naisten koulutusta koskevissa elämänkertomuksissa. Joensuu 1998. 312 s.

36. Pyy, Ilkka. Paikallisvaltion loppu? Hyvinvointivaltion murros syrjäisessä kunnassa. Joensuu 1998. 287 s.

37. Susag, Chris. Finnish American Ethnicity as Measured by Collective Self Esteem. Joensuu 1999. 182 p.

38. Aaltonen, Elli. Sosiaalitoimen johtaminen - byrokratiaa vai tulosjohtamista? Tutkimus sosiaalitoimen johtamisesta Suomessa 1950-luvulta 1990-luvun puoliväliin. English abstract: Management within Municipal Social Administration - Bureaucracy or Managing for Results? A Study of the Management of Municipal Social Administration in Finland from the 1950's to the mid 1990's. Joensuu 1999. 201 s.

39. Korhonen, Merja. Isyyden muutos. Keski-ikäisten miesten lapsuuskokemukset ja oma vanhemmuus. The Change in Fatherhood. The Childhood Experiences of Middle-aged Men and Their Own Parenthood. Joensuu 1999. 292 s.

40. Tedre, Silva. Hoivan sanattomat sopimukset. Tutkimus vanhusten kotipalvelun työntekijöiden työstä. English abstract: The Unspoken Contracts in Social Care. A Case Study of Paid Workers in Finnish Municipal Home Help Services for the Elderly. Joensuu 1999. 188 s.

41. Kurvinen, Arja. Tilinteon aika. Tutkimus pankista työnsä menettävien naisten identiteetin uudelleen arvioinneista. Joensuu 1999. 357 s.

42. Koivuluhta, Merja. Ammatti-intressit ja ura. Pohjoiskarjalaisen ammattikoulutetun ikäluokan seuranta 1975 – 1991. English summary: Vocational Interests and Career. A Follow-up of a Vocationally Trained North Karelian Age Cohort from 1975 – 1991. Joensuu 1999. 381 s.

43. Westman, Anna Liisa. Under the Northern Lights. The Reflection of Gender on the Career of Women Managers in Finnish Municipalities. Tiivistelmä/Finnish abstract: Revontulten alla. Sukupuolen heijastumia naisten uraan ja työhön Suomen kunnissa. Tutkimus naisten urakehityksestä ja työstä johtajina Suomen kunnissa. Joensuu 2000. 460 p.

44. Puhakka, Antero. Hoitoonohjaus hoitoon ohjatun silmin – subjektiivinen totuus ja sen tulkinta. English abstract: Referral to Treatment from the Point of View of the Referred: Interpreting Subjective Truth. Joensuu 2000. 264 s.

45. Puustinen, Raimo. Sign, Symptom and Interpretation. Medical Consultation in the Light of Mikhail Bakhtin's Semiotics. Tiivistelmä/Finnish abstract: Merkki, oire ja tulkinta – lääkärin vastaanotto Mihail Bahtinin semiotiikan valossa. Joensuu 2000. 79 p.

46. Silvonen, Jussi. "Tässä muuten hienossa näytelmässä." – Työhön–ryhmät työttömän roolimuutoksen tukijoina laadullisen analyysin valossa. English abstract: "In this otherwise so great play …". JOBS groups in Finland as supporter of the change in the job-seeker role of the unemployed. Joensuu 2000. 267 s.

47. Komonen, Katja. Koulutusyhteiskunnan marginaalissa? Ammatillisen koulutuksen keskeyttäneiden nuorten yhteiskunnallinen osallisuus. English summary: In the Margin of an Education Society? The Social Inclusion of Adolescents who Drop out of Vocational Education. Joensuu 2001. 304 s.
48. Leskinen, Pekka. Statistical methods for measuring preferences. Joensuu 2001. 113 p.
49. Motola Michael. Social Habitus and the Impact of Socio-Cutural Factors on Vocational Behavior of Adolescents. A Cross-National Comparison of French and Finnish Students. Tiivistelmä/Finnish abstract: Sosiaalinen habitus ja sosiokulttuuristen tekijöiden vaikutus nuorten ammatilliseen käyttäytymiseen. Suomalaisten ja ranskalaisten opiskelijoiden vertailu. Joensuu 2001. 327 p.
50. Lehtola, Ilkka: Palvelujen muutos ja syrjäkyläläisten arki. English abstract: Change of Services and Everyday Life in a Remote Village. Joensuu 2001. 238 s.
51. Vapaavuori, Jukka. Dialoginen muutos ryhmäkeskusteluissa: työyhteisön jumi konsultaation kohteena. English abstract: Dialogical change in staff group discussions: the 'snag' as target of consultation. Joensuu 2001. 214 s.
52. Kuittinen, Matti. Defensiivinen käyttäytyminen yhteistyön ja kommunikaation esteenä. Pienyrityksen tapaustutkimus. English abstract: Organisational defensive routines as obstacles of learning and communication: Case of a Family Company in Crisis. Joensuu 2001. 185 s.
53. Karttunen, Sari: Taiteilijan määrittely: refleksiivisyyden esteitä ja edellytyksiä taidepoliittisessa tutkimuksessa. English abstract: Defining an Artist: Obstacles and Requisites for Reflexivity in Art-Political Research. Joensuu 2002. 105 s.
54. Saqib, Omar Farooq. Currency Crises in Emerging Economies: Essays in International Macroeconomics. Joensuu 2002. 145 + VIII p.
55. Oker-Blom, Max. Bakom rätten. En rättsekonomisk argumentationsmodell. Joensuu 2002. 311 s.
56. Sireni, Maarit. Tilansa tekijät. Tutkimus emännyydestä ja maatilasta naisen paikkana. English abstract: Gendered Identities and Place. A Study of Female Farmers. Joensuu 2002. 320 s.
57. Häyrynen, Simo. Kansakunnan ja maakunnan välissä. Pohjois-Karjalan alueyhteisön rakentuminen kulttuuripoliittisen edunvalvonnan keskus/periferia –suhteessa. English abstract: Between the Nation and Province. Constructing the Regional Community of North Karelia in the Centre/Periphery Relation of Cultural Policy. Joensuu 2002. 314 s.
58. Kasanen, Kati. Lasten kykykäsitykset koulussa. English abstract: School Children's Notions of Ability. Joensuu 2003. 165 s.

59. Gasso, Chaltu Deressa. Breaking the Silence: Conceptions of Adolescents on Sexual Health Issues in Modernising Development Society: The Case of Nakamte High School, Ethiopia. Tiivistelmä/Finnish abstract: Hiljaisuutta murtamassa. Nuorten käsitykset seksuaalisuudesta ja ehkäisystä modernisoituvassa kehitysmaassa: tapaustutkimus Nakamten lukiosta, Etiopiasta. Joensuu 2003, 188 s.

60. Huusko, Liisa. Sopiva tiimi. English abstract: A Suitable Team. Joensuu 2003. 302 s.

61. Moore, Erja. Pitkä opintie. Aikuisiällä suoritettu yliopistotutkinto ja koulutuksellisen elämänkulun muutos. English abstract: Long Study Paths. Studying for a Degree at Adult Age and Change in Educational Life Course. Joensuu 2003. 212 s.

62. Kerkkänen, Paavo. Huumorintaju ja terveys itäsuomalaisten poliisien työssä 1995 - 1998. English abstract: Sense of Humor and Health of Finnish Policemen 1995 -1998. Joensuu 2003, 190 s.

63. Hurskainen, Aimo. Hallintopäätöksen perustelemisesta. English summary: On the Importance of the Presentation of Reasons. Joensuu 2003, 374 s.

64. Lin, Chijien. Generating Forest Stands with Spatio-Temporal Dependencies. Tiivistelmä/Finnish abstract: Tila-aika –riippuvien metsiköiden generointi.中文摘要：用时空相关性模拟森林林相. Joensuu 2003. 123 p.