AN ABSTRACT OF THE DISSERTATION OF

<u>Jeff D. Hamann</u> for the degree of <u>Doctor of Philosophy</u> in
<u>Forest Engineering</u> presented on <u>July 14, 2008</u>.
Title: <u>Optimizing the Primary Forest Products Supply Chain:</u>
<u>A Multi-Objective Heuristic Approach</u>

Abstract approved: _____

Kevin Boston

This thesis is a collection of four submitted manuscripts that present
methods to assist forest ecosystem service managers wanting to develop oper-
ational sampling, monitoring, and production plans for a set of specific quan-
tifiable ecosystem services, which are formulated as a series of general multi-
objective optimization problems. The problems are solved using a heuris-
tic solution technique to determine the best trade-off, efficient, or Pareto
frontiers, among the potentially competing and possibly non-commensurate
objectives, with the intention that the decision maker(s) will select and im-
plement a single plan from the Pareto frontier.

The first manuscript presents the general formulation and solution frame-
work, and demonstrates the method with a problem that has five objectives.
The method demonstrates that Pareto frontiers for problems with unknown
inputs, many competing objectives, and complex constraints can be analyzed
using simple search rules.

The second manuscript examines design-based estimation and model-based prediction methods to obtain guesses of unknown inputs, and the resulting outputs, for operational production plans. The results indicate that model-based prediction methods, using simple correlation models, provide benefits by reducing production uncertainties, and thus offer substantial cost savings, or increases in net revenue, when comparison to traditional design-based methods.

The third manuscript approximates the Pareto frontier between the maximum information content (i.e. entropy) and the minumum cost for a forest sample, where the results from the sample will be used for many objectives (e.g. prediction, simulation, and optimization). The results depend on the definition of the sample design, but follow similar patterns for all 36 sample designs examined.

Finally, the fourth manuscript presents an examination of the Pareto frontier for an operational harvest schedule, using the sample that contains the maximum information content, and the objectives for the operation must satisfy multiple internal and external customers (i.e. production, financial, environmental, logistics, and marketing).

By including additional information (i.e. spatial correlation) in the prediction, simulation, and optimization process, these manuscripts demonstrate substantial potential increases in financial objectives (i.e. maximize net revenue, minimize costs), environmental objectives (i.e. maximize unharvested area), materials management objectives (i.e. minimize product degredation), information objectives (i.e. maximum entopy sampling) as well as provide a framework for the objective examination of complex forest ecosystem supply chain problems with multiple objectives.

Optimizing the Primary Forest Products Supply Chain:

A Multi-Objective Heuristic Approach

by

Jeff D. Hamann

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented July 14, 2008
Commencement June 2009

Doctor of Philosophy dissertation of <u>Jeff D. Hamann</u> presented on <u>July 14, 2008</u>

APPROVED:

_____

Major Professor, representing Forest Engineering

_____

Head of the Department of Forest Engineering

_____

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

_____

Jeff D. Hamann, Author

ACKNOWLEDGMENTS

DEDICATION

For Ron and Carole.

Without you, I would not exist.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

1. INTRODUCTION

A broad definition of forestry is a combination of biological, quantitative, managerial and social sciences applied to the management and conservation of forest resources (Helms, 1998). The practice of forestry involves the intervention of forest conditions for the sake of producing some desired societal, biological, or economic objectives (Gregoire, 2002). Often, these objectives require quantitative information which are often diverse, incomplete, and frequently dynamic. The task of acquiring and assessing the status of the forest biome, which typically falls under the chore of forest inventory, can include assessments of woody biomass, flora, fauna, water, and atmospheric gases (Gregoire, 2002). The production of these goods and services, which is the role of forest engineering, requires outputs from forest inventory (i.e. predictions), capital inputs (i.e. finance), and has been the subject of numerous optimization studies. As the objectives of these two fields, the measurement of their associated metrics (i.e. water quality, air quality, wood fiber, recreation, and economic value) and their optimization (e.g. maximize water quality, minimize prediction error, maximize information content, minimize cost, maximize revenue, minimize waste, minimize bias, maximize historical range of variability, minimize environmental degradation) become more prevalent in society, often these two objectives appear to be in conflict or competition, and so, our need to understand the trade-offs among them becomes paramount.

The need to address these issues is large indeed, and to accomplish this, these manuscripts present a general framework and solution method for environments with multiple objectives, complex constraints, and diverse solution spaces. The manuscripts are couched within a supply chain logistics man-

agement framework, presented by Riopel et al. (2005) where the authors provide a framework for business logistics decision-making by classifying logistics decisions and highlight the relevant linkages among them, focusing on the precedence of the relationships and how each decision influences others. For these manuscripts, the specific focus is on the communication, information, and logistics networks within the primary forest products supply chain. Specifically, these manuscripts develop a general framework to formulate, examine, and solve complex problems in probability sampling, optimization, and decision making using a population of individual trees within a bounded polygon, sets of complex constraints with multiple objectives, and a heuristic solution technique.

In Chapter One, I present the formulation for and give an example of a small primary forest products supply chain, where the data are individual point locations that represent volume supplies, which in this case, represent eight individual and arbitrary stems. The "sample" data are taken from Table 12.1 in Isaaks and Srivastava (1989). The manuscript presents a formulation and solution framework for a general non-linear, mixed-integer assignment problem, capable of handling single or multiple objectives, datum with multiple supports (point and area input data), and complex constraints.

In Chapter Two, I examine the differences between traditional design-based estimation of the volume (i.e. biomass) in a finite harvest area, as is commonly performed in forest supply chain planning, against model-based predictions of the same attribute using simple spatial correlation models to describe relationship between proximity and similarity among observations at the sample locations. The chapter presents an examination of sampling using non-traditional sample location patterns, sample intensity, and the application of spatial correlation models, and their influence on both traditional

design-based and model-based guesses of the total volume in a harvest polygon. This chapter answers the question regarding which method of inference is more appropriate for precisely managing forest ecosystem services, specifically, the planning for and the optimization of the primary forest products supply chain where practitioners require multiple inputs from potentially disparate data sources, and when little or no a priori information is available.

In Chapter Three, I determine the optimal spatial sample schemes, as defined by de Gruijter et al. (2006), from a population of sample designs, where the interest is in optimization of a single sample from which the results will be used for multiple purposes. Since each intended use of the data may yield a different optimal sample set depending on the criteria (i.e. estimation, prediction, optimization, simulation, or some combination thereof), the problem is to examine the trade-offs between the maximum information content and the minimum sample cost. The problem is formulated as a bi-objective problem where the objective is to determine the set of samples for a population of sample designs that maximize entropy (Shewry and Wynn, 1987) and minimize the sampling cost, defined as a traveling salesman problem (Lawler et al., 1985). This technique combines methods used in in large scale environmental and hydrological network monitoring, which is often subject to intense budgetary scrutiny, and the well-established traveling salesman problem, which is often used in transportation planning. The resulting samples can be used for any variety of, or multiple, purposes (i.e. D-optimal experimental design), from which one sample design is selected as the sample design that is used to generate the inputs into a multi-objective/multi-criteria decision making framework, presented in Chapter One.

In Chapter Four, I combine the previous three manuscripts, to construct a simplified ecosystem services supply chain planning and optimization prob-

lem. The practitioner is required to satisfy a demand vector, maintain an inventory, maximize net revenue, and maximize the area left unharvested. The method allows the practitioner to examine the optimum trade-off curves, or Pareto frontiers, among various management plans, for a variety of operational metrics, and provide them with a rational to select a single plan for implementation.

To accomplish the numerous goals of these manuscripts, which like much of forestry, is really a collection of concepts from a wide variety of disciplines, I present the framework, problems, and solution techniques using the frameworks of their respective fields. I use definitions that are consistent with classical design-based sampling texts (Cochran, 1977; Knottnerus, 2003; Tillé, 2006), model-based geostatistics texts (Cressie, 1993; Wackernagel, 1998), information theory (Shannon, 1948), combinatorial optimization (Lawler et al., 1985), multi-objective optimization (Steuer, 1986), graph theory (Diestel, 2005), inventory theory (Hillier and Lieberman, 1995; Porteus, 2002), and evolutionary algorithms (Eiben and Smith, 2003). I have attempted to use notation consistent with these texts, and have translated the notation for formulations linking the fields when required.

Finally, this collection of manuscripts completes the initial work set forth by researchers at Oregon State University, to develop a framework for the study, management, and optimization of the primary forest products supply chain. To provide these tools, these manuscripts make original contributions to the following subject areas within forest information science:

- Multiple-objective optimization

    - A general framework for non-linear mixed-integer multi-objective NP-hard planning problems with deterministic and stochastic elements;

- The use of Multi-Objective Evolutionary Algorithms (MOEA), in forest engineering, forest sampling, and forest operations for examination of the trade-offs among potentially competing and possibly non-commensurate objectives;

- Basic research regarding evolution strategy parameter values and their influence on the performance of the search of the solution space; and

- A completely enumerated non-trivial example problem which can be used to examine and compare multi-objective solution search techniques.

- Optimal data acquisition for forest planning problems

  - Examination of design-based and model-based inference for the estimation or prediction of missing values which are then used for planning, simulation, and optimization, of forest operations;

  - The inclusion of simple spatial correlation structures to assess forest product availability in an operational environment;

  - The examination of various spatial correlation models, from which more precise predictions of forest products can be obtained;

  - The application of entropy (Shannon, 1948) in forest sampling where the goal is to obtain data from a finite region for multiple purposes; and

  - Optimal experimental design sampling, expressed as a sample data acquisition problem for supply chain optimization activities.

- Precision production planning in forest environments

– A framework that can combine potentially disparate forest operations (e.g. production and log yard inventory) into a single problem where the practitioner can examine the consequences of various decisions objectively;

– A framework that provides ability to combine deterministic, stochastic inputs and outputs (i.e. probability maximization problems, single and joint-chance constraint problems), and general nonlinear mixed-integer multi-objective planning problems into a single framework;

– Examination and communication of the Pareto set for high dimensional solution spaces (i.e. many objectives), which are common in forest engineering; and

– Model-based sampling can be used to derive the inputs needed to formulate problems that include stochastic elements such as chance-constraints and probability objectives

BIBLIOGRAPHY

Cochran, W. G. (1977). *Sampling techniques.* John Wiley & Sons, New York. 428 p.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* John Wiley and Sons, Inc., New York, New York, revised edition edition.

de Gruijter, J. J., Brus, D. J., Bierkins, M. F. P., and Knotters, M. (2006). *Sampling for Natural Resource Monitoring.* Springer-Verlag.

Diestel, R. (2005). *Graph Theory (Graduate Texts in Mathematics).* Springer.

Eiben, A. E. and Smith, J. E. (2003). *Introduction to Evolutionary Computing.* SpringerVerlag.

Gregoire, T. (2002). Forestry. In El-Shaarawi, A. H. and Piegorsch, W. W., editors, *Encyclopedia of Environmetrics*, volume 2. John Wiley and Sons.

Helms, J. A., editor (1998). *The Dictionary of Forestry.* Society of American Foresters, Bethesda, MD.

Hillier, F. S. and Lieberman, G. J. (1995). *Introduction to Operations Research.* McGraw-Hill, Inc., sixth edition.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics.* Oxford University Press, Inc., New York, New York.

Knottnerus, P. (2003). *Sample Survey Theory: some Pythagorean perspectives.* Springer-Verlag.

Lawler, E. L., Lenstra, J. K., Kan, A. H. G. R., and Shmoys, D. B., editors (1985). *The Traveling Salesman Problem.* John Wiley, New York.

Porteus, E. L. (2002). *Foundations of Stoahstic Inventory Theory.* Stanford University Press.

Riopel, D., Langevin, A., and Campbell, J. F. (2005). The network of logistics decisions. In Langevin, A. and Riopel, D., editors, *Logistical Systems:Design and Optimization*, number 2 in GERAD, pages 1–38. Springer, New York, New York.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423.

Shewry, M. and Wynn, H. (1987). Maximum entropy sampling. *Journal of Applied Statistics*, 46:165–170.

Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation and Application.* John Wiley, New York.

Tillé, Y. (2006). *Sampling Algorithms.* Springer.

Wackernagel, H. (1998). *Multivariate Geostatistics*, chapter 1. Springer, Berlin, Germany, 2nd edition.

2. A GENERALIZED FRAMEWORK FOR OPTIMIZING MULTI-OBJECTIVE FOREST SUPPLY CHAIN OPERATIONS

2.1 ABSTRACT

We present a formulation and heuristic solution technique to determine the Pareto frontier for multi-objective mixed-integer non-linear spatially correlated forest products supply chain planning problems. The Pareto Archiving Evolutionary Strategy (PAES), presented by Knowles and Corne (2000), is applied to a small operational forest production planning problem with the following objectives: 1) obtain the unbiased minimum variance stem volume estimates for unmeasured stems, 2) minimize constraint violations, 3) maximize daily production subject to demand constraints, 4) minimize the reserve stem volume subject to certification constraints, and 5) minimize soil disturbance. The resulting Pareto set contains the values of the objective functions, daily cutting schedules, and the associated path a ground-based harvester must travel in order to implement the plans that maximize or minimize the objectives while meeting the constraints.

The example problem contains 68,584,320 possible solutions of which 465 are global Pareto optimal solutions. Heuristic performance is measured and discussed by examination of the mutation rate at which new Pareto optimal solutions were discovered and the proportion at which mutated candidate solutions entered the Pareto archive. Candidate solution mutation probability rates where limited to 0.01, 0.05, and 0.1 mutations per candidate solution, each of which were applied to populations candidate solutions of 10, 50 and 100 individuals respectively.

All mutation rates for the population size of 10 candidate solutions yielded

the slowest rate of discovery or "learning rate". The discovery of new Pareto optimal solutions was strongly influenced by the mutation rate for the population of 50 and 100 candidate solutions. The best discovery rate of new Pareto solutions was obtained using the largest population size (100) and the highest mutation rate (0.1).

The framework presented here supports precision production planning in forest environments by 1) providing a statement of how the decision maker obtained the inputs upon which, 2) any selected decision, is considered optimal, given a set of inputs, objectives, and constraints, 3) documents the attainment of the *best* possible set of decision alternatives from which the decision maker can select, and 4) facilitates rapid examination of the trade-offs among potentially competing and possibly non-commensurate objectives when solution time is critical as might be the case in general decision making "roundtable" type environments.

2.2   INTRODUCTION

Strategic, tactical and operational plans for forest environments should 1) explicitly state how the decision maker obtained the inputs upon which, 2) any selected decision for a particular formulation, is considered optimal, given a set of inputs, objectives, and constraints, and 3) document the framework that generated the *best* possible set of decision alternatives.

To accomplish these goals, deference should be given to imputation and solution search methods, regardless of the underlying model formulation, that can 1) produce the *best* (unbiased minimum prediction variance) predictions for fixed, but unknown target parameters (e.g. biomass, stems, species abundance), 2) handle problems with multiple potentially competing and possibly

non-commensurate objectives, which are common in complex supply chain optimization problems in forest environments, 3) detect and communicate to the practitioner, unique features of the solution space for further exploration and exploitation (Ducheyne et al., 2006, 2004), and 4) require as few runs as possible, preferably a single run, which allows a practitioner to select and implement, an optimal plan within a time-sensitive environment.

To address these criteria, we present a framework for practitioners of precision forest planning that can 1) be formulated, implemented, and modified using simple formulation, search, and analysis rules, 2) obtain the best set of trade-offs among multiple objectives when the analysis time is limited, and 3) further refine the optimal set as time permits.

The format of the paper is to first introduce the nomenclature and in Section 2.3, present the formulation for the framework. In Section 2.4, a small non-linear mixed-integer multi-objective operational forest supply chain optimization problem is presented. Section 2.5 briefly presents the heuristic developed by (Knowles and Corne, 2000) and describes the parameters used to explore the optimal trade-offs or efficient frontier for the example problem. Our results and a discussion of the heuristic's performance and a short examination of some Pareto optimal precision production plans are graphically generated in Section 2.6 followed by our conclusions, which are presented in Section 2.7.

Nomenclature

Unknown values and random variables are denoted using lower and upper case Greek symbols. Vectors and matrices are represented in bold.

Problem Formulation

$\text{argmin}\,\mathbf{f}$	The arguments that achieve the global minimum of $\mathbf{f}$
d	The total number of objectives
D_i	The total distance traveled by a harvester on day i
D'	The total distance traveled
f_i	An individual objective function
$\mathbf{f}$	A vector of d objective functions
g	Generation index
G	Maximum number of generations
g_k	An inequality constraint
$\mathbf{g}$	A vector of K inequality constraints
h_m	An equality constraint
$\mathbf{h}$	A vector of M equality constraints
i	Period number
j	Stem number in daily cutting schedule
k	Inequality constraint index
m	Equality constraint index
K	The total number of inequality constraints
λ_g	The number of mutations (candidate solutions) at generation g
μ_g	The Pareto optimal set at generation g
$|\mu_g|$	The size of the Pareto optimal set at generation g
M	The total number of equality constraints
$\min\mathbf{f}$	The global minimum of $\mathbf{f}$
N	The total number of observation locations
N_i	The number of stems to be harvested on day i
ψ	A scalar penalty function
r	The length of the fixed and known values of the input data
v_i	The volume of stem i
V_i	The volume harvested in period i
x_i	A fixed and known scalar input
ν	The number of fixed but unknown variables in the inputs
ξ_i	A fixed but unknown scalar input
$\mathbf{x}$	A vector of fixed and known inputs of length r
ξ	A vector of fixed but unknown inputs of length ν
$\hat{\xi}$	A vector of predicted fixed but unknown inputs

Operators and Miscellaneous Notation

$\mathbf{0}$	A vector of zeros
$\mathbb{R}$	The real numbers

$\mathbb{C}$	The complex numbers
$\mathbf{f}_a \prec \mathbf{f}_b$	Component-wise less-than vector inequality
$\exists, \in, \subset$	exists, in, and subset
$\vee, \wedge, \neg$	logical `OR`, `AND`, and `NEGATION` operators
$\mathbb{E}[\cdot], \mathbb{V}[\cdot], \mathbb{C}[\cdot]$	Expected value, variance, and covariance operator
$\mathbb{G}_k[\cdot]$	Heaviside operator for inequality k (Osyczka, 2002)
$\mathbf{C}', \mathbf{c}'$	transpose of matrix $\mathbf{C}$ or vector $\mathbf{c}$
$\mathbf{C}^{-1}$	Inverse of $\mathbf{C}$
$\mathcal{F}$	The feasible region
$\mathcal{S}$	The search space $\subset \mathbb{R}^{r+\nu}$

2.3   PROBLEM FORMULATIONS

Kuhn and Tucker (1951) introduced a vector-valued objective function in mathematical programming called a vector maximum problem, and derived the optimality conditions for efficient solutions (Coello Coello, 2002). Since precision production planning in forest environments is often considered an optimization task with more than one objective, the goal here is the minimization of a general non-linear constrained deterministic d-objective problem ($d \geq 1$), which can be expressed using the general formulation (Karush, 1939; Kuhn and Tucker, 1951; Osyczka, 2002; Zhang, 2003):

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x} \in \mathcal{S}} \{ \mathbf{f}(\mathbf{x}) \in \mathbb{R}^d | \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0} \} \tag{2.1}$$

where the goal is to obtain the arguments $\mathbf{x}^*$ that yield the minimum $\mathbf{f}$, a vector of d-objectives, subject to $\mathbf{g}$, a vector of K inequality constraints, and $\mathbf{h}$, a vector of M equality constraints and the fitness of any candidate solution $\mathbf{x}$ can be evaluated directly from the functions $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$.

The feasible region, $\mathcal{F}$ is a subset of the entire search space $\mathcal{S} \subset \mathbb{R}^r$

and $\mathcal{F} \subseteq \mathcal{S}$. For an inequality constraint that satisfies $g_k(\mathbf{x}) = 0$, then the inequality k is active at $\mathbf{x}$. All equality constraints $h_m(\mathbf{x})$ are considered active at all values of $\mathcal{F}$, regardless of the value of $\mathbf{x}$ (Bishop, 2006).

For problems with potentially competing and possibly non-commensurate objectives (i.e. $d > 1$), an objective vector $\mathbf{f}_a$ is said to *dominate* another objective vector $\mathbf{f}_b$, denoted by $\mathbf{f}_a \prec \mathbf{f}_b$ (component-wise), if and only if:

$$\mathbf{f}_{a,i} \leq \mathbf{f}_{b,i} \; \forall \, i \in \{1, \dots, d\} \; \wedge \; \mathbf{f}_{a,j} < \mathbf{f}_{b,j} \; \exists \, j \in \{1, \dots, d\} \qquad (2.2)$$

In words, an objective vector is called *non-dominated* if there are no other objective vectors that can increase the value of any one of the d objective functions without decreasing the value of another of the d objective functions. The set of all *non-dominated* solutions is called the *Pareto set*, *Pareto front* or *efficient frontier* (Eiben and Smith, 2003).

When all of the inputs are known without uncertainty, Equations (2.1) and (2.2) can be used to examine the trade-offs among Pareto optimal policies which satisfy the goals of a decision maker (Deb, 2001). More often than not, the inputs for forest planning problems contain a combination of r fixed and known values (e.g. supply and delivery locations or tree measurements) and ν estimates for fixed, but unknown values (e.g. yield estimates or demand forecasts), which can be expressed by a $r + \nu$ partitioned vector:

$$(\mathbf{x}, \xi) = (x_1, x_2, \dots, x_r, \xi_{r+1}, \xi_{r+2}, \dots, \xi_{r+\nu}) \qquad (2.3)$$

where the values $x_1, x_2, \dots, x_r$ are the fixed and known inputs and ξ_{r+1}, ξ_{r+2}, $\dots$, $\xi_{r+\nu}$ are random variables, used to represent a guess for the last ν fixed, but unknown inputs, given the first r fixed and known values of the data. When an optimization problem contains fixed, but unknown values, represented by a random vector for some subset of the inputs, the resulting

formulation is a stochastic optimization problem (Gen and Cheng, 1997).

A stochastic optimization problem can be expressed by letting $\xi = (\xi_{r+1}, \xi_{r+2}, \ldots, \xi_{r+\nu})$ be the random vector of fixed, but unknown inputs, with the joint probability mass function ϕ, defined by $\mathbf{x}$:

$$\phi_{\mathbf{x}}(\xi) = \phi_{x_1,\ldots,x_r}(\xi_{r+1}, \ldots, \xi_{r+\nu}) \tag{2.4}$$

Since the maximization or minimization of a random vector or random function itself is meaningless, Equation (2.1), now a non-linear constrained stochastic d-objective ($d \geq 1$) minimization problem:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^{r+\nu}}{\operatorname{argmin}} \{ \mathbb{E}[\mathbf{f}(\mathbf{x}, \xi)] \in \mathbb{R}^d \,|\, \mathbb{E}[\mathbf{g}(\mathbf{x}, \xi)] \geq \mathbf{0}, \mathbb{E}[\mathbf{h}(\mathbf{x}, \xi)] = \mathbf{0} \} \tag{2.5}$$

can be solved using a deterministic-substitution formulation (Marti, 2005) where the goal is to obtain risk-neutral (e.g. expected value) optimal criterion, and $\mathbb{E}[\mathbf{f}(\mathbf{x}, \xi)]$, $\mathbb{E}[\mathbf{g}(\mathbf{x}, \xi)]$, and $\mathbb{E}[\mathbf{h}(\mathbf{x}, \xi)]$ are estimated from $\mathbf{x}$ and $\phi_{\mathbf{x}}(\xi)$ using *a priori*, sample, and structural information since $\mathbf{x}$ contains *all known* information available to the decision maker. The available information includes inputs such as product yields, forest conditions, spatial relationships, current management policies, desired objectives, and constraints, and *any* functions thereof.

The resulting expected values for the functions in Equation (2.5) are:

$$\mathbb{E}[f_i(\mathbf{x}, \xi)] = \sum_{\xi_{r+\nu}} \cdots \sum_{\xi_{r+2}} \sum_{\xi_{r+1}} f_i(\mathbf{x}, \xi)\phi_{\mathbf{x}}(\xi) \quad \forall \ i = 1, \ldots, d \tag{2.6}$$

$$\mathbb{E}[g_k(\mathbf{x}, \xi)] = \sum_{\xi_{r+\nu}} \cdots \sum_{\xi_{r+2}} \sum_{\xi_{r+1}} g_k(\mathbf{x}, \xi)\phi_{\mathbf{x}}(\xi) \quad \forall \ k = 1, \ldots, K \tag{2.7}$$

$$\mathbb{E}[h_m(\mathbf{x}, \xi)] = \sum_{\xi_{r+\nu}} \cdots \sum_{\xi_{r+2}} \sum_{\xi_{r+1}} h_m(\mathbf{x}, \xi)\phi_{\mathbf{x}}(\xi) \quad \forall \ m = 1, \ldots, M \tag{2.8}$$

Here we assume that the expected values can be estimated without bias and so, to conserve notation, we will express the general formulation as:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^{r+\nu}}{\operatorname{argmin}} \{ \mathbf{f}(\mathbf{x}, \hat{\boldsymbol{\xi}}) \in \mathbb{R}^d | \mathbf{g}(\mathbf{x}, \hat{\boldsymbol{\xi}}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}, \hat{\boldsymbol{\xi}}) = \mathbf{0} \} \qquad (2.9)$$

where $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$ are now defined as statistical or econometric models of the data (Konishi and Kitagawa, 2008), which can also be viewed as optimization models themselves (White, 1999).

We now present a small multi-objective non-linear mixed-integer operational supply chain optimization problem that contains fixed and known inputs $\mathbf{x}$, and fixed, but unknown inputs $\boldsymbol{\xi}$.

2.4   AN EXAMPLE

The goal is to determine the best plan (i.e. a set of daily cutting schedules) for a three day harvest operation where a finite population of stems $\mathcal{T}$, within a harvest polygon $\mathcal{A}$, are assigned either a harvest day-order combination, or a selected to remain in $\mathcal{A}$ (i.e no-harvest). The polygon $\mathcal{A}$ is defined by the rectangle with lower-left and upper-right corners of $(55, 125)$ and $(80, 145)$ distance units, respectively for a total surface area of 500 distance units squared.

The dataset for this problem, presented in Table 2.1, contains eight stems ($N = 8$), with seven of the stems having perfectly measured volumes ($\mathcal{T}^v$) *and* perfectly known positions ($\mathcal{T}^{\mathbf{u}}$). The eighth stem, contains a *known* position, and a fixed but *unknown* volume, and so we do not know the total $V_{\mathcal{A}} \equiv \sum_{i=1}^{N} \mathcal{T}_i^v$ or the spatial mean $\overline{V}_{\mathcal{A}} \equiv \lambda(A)^{-1} V_{\mathcal{A}}$, where $\lambda(\mathcal{A})$ is a function that returns the surface area of the polygon $\mathcal{A}$.

Stem, i	Location, $\mathcal{T}^{\mathbf{u}}$	Volume, $\mathcal{T}^v$	Day, d	Sequence, s
1	(61,139)	477	d_1	s_1
2	(63,140)	696	d_2	s_2
3	(64,129)	227	d_3	s_3
4	(68,128)	646	d_4	s_4
5	(71,140)	606	d_5	s_5
6	(73,141)	791	d_6	s_6
7	(75,128)	783	d_7	s_7
8	(65,137)	v_8	d_8	s_8

Table. 2.1: Data from Table 12.1 in Isaaks and Srivastava (1989). The variable v_8 represents the fixed but unknown value for the unobserved volume associated with the eighth stem. The column vectors d and s represent the decision variables for which day and order to harvest stem i.

The objectives are to 1) obtain the *best* (i.e. unbiased minimum prediction variance) stem volume predictions for unmeasured stems, 2) minimize constraint violations, 3) maximize daily production subject to demand constraints, 4) minimize the residual volume while meeting retention tree requirements, 5) minimize soil disturbance, and 6) provides to the practitioner the *best* set of possible choices for implementation.

The constraint requirements are such that 1) in order to meet contractual obligations, the delivered volumes meet or exceed 1100, 800, and 1200 volume units, for each of the three days, 2) the soil disturbance is minimized in order to maintain the soil resource as defined in the Montreal Protocol (Block et al., 2002), and 3) the plan meet certification standards by maintaining at least 300 volume units (i.e. reserve stems) in $\mathcal{A}$ (Stringer, 2006).

2.4.1 Formulation

The problem can be expressed using the general formulation:

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^{r+\nu}} \{ \mathbf{f}(\mathbf{x}, \hat{\xi}) \in \mathbb{R}^d | \mathbf{g}(\mathbf{x}, \hat{\xi}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}, \hat{\xi}) = \mathbf{0} \} \qquad (2.10)$$

where $\hat{\xi}$ are predictions from the best unbiased minimum variance predictors for ξ, determined from $\mathbf{x}$ and the probability mass function $\phi_{\mathbf{x}}(\xi)$, defined by the decision maker.

Since the temporal scale of the operation is sufficiently short so that repeated measurements of the volume for stem eight yield the same value (e.g. no growth or measurement error), the volume for stem eight $(\mathcal{T}_8^v)$ is the only fixed, but unknown input in the vector ξ, in which the prediction will be denoted as $\hat{\xi}$.

The variables representing harvest day and harvest sequence are the decision variables, which are known and fixed inputs as they are dictated by the decision maker. A decision to harvest a stem on day four represents a decision to set aside the stem as a residual (e.g. green tree retention or legacy stem), thus the decision variables, determined by the decision maker, are the elements of $\mathbf{x}$, a $(6 \times 8) - 1$ row vector representing the fixed and known inputs and ξ representing $\mathcal{T}_8^v$, the last $(\nu = 1)$ element which is a fixed, but unknown input. To reduce the notation, we denote T_i^v as v_i, $T_i^{\mathbf{u}}$ as $\mathbf{u}_i$, T_i^d as d_i and T_i^s as s_i.

2.4.1.1   Objective 1: Unbiased minimum variance predictions for volume

The first objective is to predict the *actual* stem volume, v_8, which can be accomplished by adding *structural information* to the formulation in order to obtain $\hat{v}_8$ from the best unbiased minimum variance estimator (Cressie, 1993).

Here, we assume stem volumes are point realizations from a random field $V_\mathcal{A}$ over a finite area, $\mathcal{A}$, which accounts for both local irregularities (e.g. randomness) and a structured aspect (e.g. large scale tendencies) and the that for any distance $|\mathbf{h}| = |\mathbf{u}_i - \mathbf{u}_j|$, the distribution of the random variables $V(\mathbf{u}_1), V(\mathbf{u}_2), \ldots, V(\mathbf{u}_k)$ is the same as $V(\mathbf{u}_1 + \mathbf{h}), V(\mathbf{u}_2 + \mathbf{h}), \ldots, V(\mathbf{u}_k + \mathbf{h})$ for the first two moments (i.e. constant mean and covariance) (Reed and Burkhart, 1985). These assumptions, known as second-order stationarity, are critical in determining the optimal weights so that the prediction at some unsampled location $\mathbf{u}_0$, over a region $\mathcal{A}$, is unbiased ($E[\hat{V}_\mathcal{A} - V_\mathcal{A}] = 0$) and the error or *prediction* variance ($Var[\hat{V}_\mathcal{A} - V_\mathcal{A}]$) is minimum (Armstrong, 1998).

These assumptions and procedures are often used to describe spatial and temporal processes in forest inventory (Hudak et al., 2002; Wallerman et al., 2002; Tuominen et al., 2003), forest operations optimization (Hof and Bevers, 1998; Mandallaz, 2000; Hamann and Boston, 2007), and heuristics (Hyoun-Jin et al., 2005).

Using these assumptions, the additional structural information included in the formulation can be defined as:

$$\hat{v}_8 = \sum_{i=1}^{7} \omega_i v(\mathbf{u}_i) \tag{2.11}$$

where the predicted stem volume for stem eight, $\hat{v}_8$, is a weighted linear combination of the nearby seven stem volumes and the distances that separate

them, $\mathbf{u}_i$ is the position of the i-th nearby stem, $v(\mathbf{u}_i)$ is the i-th realization of the regionalized variable $V_{\mathcal{A}}(\mathbf{u})$ at location $\mathbf{u}_i$, and ω_i is the weight applied to the observed value at position $\mathbf{u}_i$ (Isaaks and Srivastava, 1989).

The first objective then becomes to select the weights that yield the *best* (i.e. unbiased minimum prediction variance) prediction for v_8, subject to the unbiasedness constraint ($\sum_{i=1}^{n} \omega_i = 1$), which can be expressed as a set of $n + 1$ equality constraints using Equation (2.5):

$$h_i(\mathbf{x}, \hat{\xi}) \equiv \sum_{j=1}^{n} \omega_j \hat{C}_{ij} + \lambda - \hat{C}_{i,8} = 0 \text{ for } i = 1, \ldots, N \quad (2.12)$$

$$h_{N+1}(\mathbf{x}, \hat{\xi}) \equiv \sum_{i=1}^{n} \omega_i - 1 = 0 \quad (2.13)$$

where $N = 7$, the number of observed stem volumes, ω_i are the kriging weights and λ is the Lagrangian parameter, required to convert the constrained minimization problem into an unconstrained problem for this objective (Boyd and Vandenberghe, 2004; Bishop, 2006), and maintains unbiasedness when the weights sum to one. Here, the spatial covariance model, $\hat{C}_{ij}$, defined by Isaaks and Srivastava (1989), is an exponential function: $\hat{C}_{ij} = 10e^{-0.3 \cdot |\mathbf{h}|}$ where $|\mathbf{h}|$ is the distance separating stem i and j.

Alternatively, the system of $n+1$ kriging equations can be expressed more compactly in matrix notation as:

$$\begin{bmatrix} \hat{C}_{11} & \ldots & \hat{C}_{1,n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \hat{C}_{n1} & \ldots & \hat{C}_{n,n} & 1 \\ 1 & \ldots & 1 & 0 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \\ \lambda \end{bmatrix} - \begin{bmatrix} \hat{C}_{1,8} \\ \vdots \\ \hat{C}_{n,8} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad (2.14)$$

or more simply:

$$\mathbf{C}\omega - \mathbf{D} = \mathbf{0} \quad (2.15)$$

where $\boldsymbol{\omega}$ is the vector of weights, $\mathbf{C}$ is the covariance matrix of the observations and $\mathbf{D}$ is the vector of the covariances at the points themselves, that is, $\mathbf{C}(\mathbf{h}) = 0$ where $\mathbf{h}$ is the distance between sample points, or in this case stems, and $\mathbf{0}$ is a $n + 1$ vector of zeros. This generates a system of $n + 1$ equations that can be easily solved for $\mathbf{C}^{-1}$ to obtain the weights $\boldsymbol{\omega}$:

$$\boldsymbol{\omega} = \mathbf{C}^{-1}\mathbf{D} \tag{2.16}$$

provided $\mathbf{C}$ is a positive definite function ($\mathbf{x}'\mathbf{A}\mathbf{x} > \mathbf{0}; \mathbf{A}, \mathbf{x} \in \mathbb{C}^n$), where the resulting values for $\boldsymbol{\omega}$ produce the unbiased predictions with the minimum prediction variance (Cressie, 1993).

2.4.1.2 Objective 2: Minimize constraint violations

Ideally, all production plans must not violate *any* constraints, or should a particular plan be infeasible, constraint violations should be minimal and for infeasible solutions, the decision maker should know the magnitude of the constraint violations for a given candidate solution.

To ensure that an objective vector, for any given candidate solution, is as close as possible to, or is within the feasible region, a constraint violation function, defined by (Osyczka, 2002), is included as the first objective function:

$$f_1(\mathbf{x}, \hat{\xi}) \equiv \psi \equiv \sum_{m=1}^{M} h_m(\mathbf{x}, \hat{\xi})^2 + \sum_{k=1}^{K} \mathbb{G}_k[g_k(\mathbf{x}, \hat{\xi})]^2 \tag{2.17}$$

where $\mathbb{G}_k$ is defined as:

$$\mathbb{G}_k = \begin{cases} 0 & \text{for } g_k(\mathbf{x}, \hat{\boldsymbol{\xi}}) \geq 0 \\ 1 & \text{for } g_k(\mathbf{x}, \hat{\boldsymbol{\xi}}) < 0 \end{cases} \tag{2.18}$$

so that when the predicted value of the first objective, $f_1(\mathbf{x}, \hat{\boldsymbol{\xi}})$, equals zero, the candidate solution yields an objective vector $\mathbf{f}(\mathbf{x}, \hat{\boldsymbol{\xi}})$ within, or on the boundary of the feasible region $\mathcal{F}$.

2.4.1.3 Objective 3: Maximize production

The predicted value for the volume produced during each day can be defined as a set of equality constraints:

$$h_1(\mathbf{x}, \hat{\boldsymbol{\xi}}) \equiv \sum_{j=1}^{N_1} \hat{v}_{1,j} - \hat{V}_1 = 0 \tag{2.19}$$

$$h_2(\mathbf{x}, \hat{\boldsymbol{\xi}}) \equiv \sum_{j=1}^{N_2} \hat{v}_{2,j} - \hat{V}_2 = 0 \tag{2.20}$$

$$h_3(\mathbf{x}, \hat{\boldsymbol{\xi}}) \equiv \sum_{j=1}^{N_3} \hat{v}_{3,j} - \hat{V}_3 = 0 \tag{2.21}$$

where $\hat{v}_{i,j}$ is the predicted volume of stem i, harvested on day j, and N_j is the number of stems harvested on day j. These equality constraints are constructed as "accounting variables" since the values are required for the formulation of both objective functions and inequality constraints.

The objective to maximize daily production, can then be expressed as three separate objectives:

$$f_2(\mathbf{x}, \hat{\xi}) \equiv -\hat{V}_1 \tag{2.22}$$

$$f_3(\mathbf{x}, \hat{\xi}) \equiv -\hat{V}_2 \tag{2.23}$$

$$f_4(\mathbf{x}, \hat{\xi}) \equiv -\hat{V}_3 \tag{2.24}$$

and in order to meet the demand constraints listed above, a set of inequality constraints is included:

$$g_1(\mathbf{x}, \hat{\xi}) \equiv \hat{V}_1 - 1100.0 \geq 0 \tag{2.25}$$

$$g_2(\mathbf{x}, \hat{\xi}) \equiv \hat{V}_2 - 800.0 \geq 0 \tag{2.26}$$

$$g_3(\mathbf{x}, \hat{\xi}) \equiv \hat{V}_3 - 1200.0 \geq 0 \tag{2.27}$$

2.4.1.4 Objective 4: Minimize residual standing volume

Another common goal in forest planning is to maximize the unharvested area, or maximize the green-tree retention, which here we describe as unharvested volume. Thus, another goal is then to minimize the residual volume in $\mathcal{A}$, another equality constraint is needed to account for the stem volume assigned to the residual volume, is defined as:

$$h_h(\mathbf{x}, \hat{\xi}) \equiv \sum_{j=1}^{N_4} v_{4,j} - \hat{V}_4 = 0 \tag{2.28}$$

and to minimize the residual volume, an objective is included as:

$$f_5(\mathbf{x}, \hat{\xi}) \equiv \hat{V}_4 \tag{2.29}$$

Finally, one additional inequality constraint is required to ensure that at least 300 volume units remain within $\mathcal{A}$ upon completion of the operation:

$$g_4(\mathbf{x}, \hat{\xi}) \equiv \hat{V}_4 - 300.0 \geq 0 \tag{2.30}$$

2.4.1.5 Objective 5: Minimize soil disturbance

Many methods exist to reduce soil degradation during forest operations such as seasonal constraints, limiting the number of machine passes, or maximum ground pressure constraints. To minimize soil disturbance, we included an objective function to minimize the total soil contact, expressed as a *tour* that a single machine would take to visit each harvested stem, for a given sequence, over the planning horizon. This minimization of the tour distance is more commonly known as a traveling salesman problem (TSP) has been used widely in single objective (Lin and Kernighan, 1973; Lawler et al., 1985) and multi-objective planning problems (Yan et al., 2003; Michalewicz and Fogel, 2004).

Beginning with the first stem in the sequence for each day, the distance traveled, each day, is the distance required to travel a path to each stem in the cut sequence for the day, which can be expressed as a set of equality constraints.

The daily travel distance is defined as:

$$D_i = \sum_{j=1}^{N_i} \sqrt{(x_{i,j} - x_{i,j+1})^2 + (y_{i,j} - y_{i,j+1})^2} \ \ i = 1, 2, 3, \ \ j = 1, \ldots, N_i - 1$$

$$\tag{2.31}$$

plus the distance required to travel from the last stem, on day i, N_i, to the first stem of the next day:

$$D_i = D_i + \sqrt{(x_{i,N_i} - x_{i+1,1})^2 + (y_{i,N_i} - y_{i+1,1})^2} \quad i = 1, 2 \qquad (2.32)$$

where x and y are the easting and northing of the each stem in the daily cutting schedule, respectively. Since all locations are known without error, the expected value of the length of any path is a constant, and so the objective to minimize the total machine travel distance is simply:

$$f_6(\mathbf{x}, \hat{\xi}) \equiv D' \equiv \sum_{i=1}^{3} D_i \quad i = 1, 2, 3 \qquad (2.33)$$

and for periods with no stems harvested, the distance from the previously ending position to the next day's starting position is used.

2.4.1.6   Complete Formulation

The complete formulation of the problem is:

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^{r+\nu}} \{ \mathbf{f}(\mathbf{x}, \hat{\xi}) \in \mathbb{R}^d | \mathbf{g}(\mathbf{x}, \hat{\xi}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}, \hat{\xi}) = \mathbf{0} \} \qquad (2.34)$$

where the elements of the objective vector, $\mathbf{f}$, are:

$$\mathbf{f}(\mathbf{x}, \hat{\xi}) = (\psi, -\hat{V}_1, -\hat{V}_2, -\hat{V}_3, \hat{V}_4, D') \qquad (2.35)$$

subject to ($K = 4$) inequality constraints, $\mathbf{g}$:

$$g_1(\mathbf{x}, \hat{\xi}) \equiv \quad \hat{V}_1 - 1100.0 \quad \geq 0 \tag{2.36}$$

$$g_2(\mathbf{x}, \hat{\xi}) \equiv \quad \hat{V}_2 - 800.0 \quad \geq 0 \tag{2.37}$$

$$g_3(\mathbf{x}, \hat{\xi}) \equiv \quad \hat{V}_3 - 1200.0 \quad \geq 0 \tag{2.38}$$

$$g_4(\mathbf{x}, \hat{\xi}) \equiv \quad \hat{V}_4 - 300.0 \quad \geq 0 \tag{2.39}$$

and the $(M = 13)$ equality constraints, $\mathbf{h}$:

$$h_i(\mathbf{x}, \hat{\xi}) \equiv \quad \sum_{j=1}^{n} \omega_j \hat{C}_{i,j} + \lambda - \hat{C}_{i,8} \quad = 0 \text{ for } i = 1, \ldots, N \tag{2.40}$$

$$h_{N+1}(\mathbf{x}, \hat{\xi}) \equiv \quad \sum_{i=1}^{n} \omega_i - 1 \quad = 0 \tag{2.41}$$

$$h_{N+2}(\mathbf{x}, \hat{\xi}) \equiv \quad \sum_{j=1}^{N_1} v_{1,j} - \hat{V}_1 \quad = 0 \tag{2.42}$$

$$h_{N+3}(\mathbf{x}, \hat{\xi}) \equiv \quad \sum_{j=1}^{N_2} v_{2,j} - \hat{V}_2 \quad = 0 \tag{2.43}$$

$$h_{N+4}(\mathbf{x}, \hat{\xi}) \equiv \quad \sum_{j=1}^{N_3} v_{3,j} - \hat{V}_3 \quad = 0 \tag{2.44}$$

$$h_{N+5}(\mathbf{x}, \hat{\xi}) \equiv \quad \sum_{j=1}^{N_4} v_{4,j} - \hat{V}_4 \quad = 0 \tag{2.45}$$

2.5 HEURISTIC DESCRIPTION

While exact solution methods for spatially explicit planning formulations have been presented for problems with two objectives, examples of exact solution methods for formulations in higher dimensions $(d \geq 2)$ are limited and many techniques are incapable of handling complex Pareto frontiers (Tóth et al., 2006). For multiple objective formulations, heuristic methods such as Tabu Search (TS) (Hansen, 1997), Simulated Annealing (SA) (Deusen, 1999; Nam and Park, 2000), Genetic Algorithms (GA), (Ducheyne et al., 2004), and Evolutionary Algorithms (Stewart et al., 2004; Ducheyne et al., 2006) where $d \geq 2$, while rare in forest planning, show promise.

The class of heuristics known as Evolutionary Algorithms (EA), inspired from Darwinian evolution, selects, mutates and promotes candidate solutions based on competition, fitness, and reproductive success (Eiben and Smith, 2003; Falcão and Borges, 2001). The major difference between GA and EA results from the use of the mutation operator. In GA, the crossover operator is used to promote both diversification and intensification, whereas EAs rely on the mutation operator alone to change mutation parameters depending on a variety of metrics such as convergence (Igel et al., 2007), stopping (Laumanns et al., 2002), diversification (Farhang-Mehr and Azarm, 2002), and optimality (Deb et al., 2007) criteria.

Several comprehensive reviews of multi-objective evolutionary algorithms are available from Coello Coello (2002), Tan et al. (2005), and Deb (2001) and all suggest that evolutionary algorithms have many advantages for high dimensional vector valued problems with potentially highly disconnected, non-uniformly distributed, and concave Pareto frontiers. In forest planning however, examination of EA for single objective (Falcão and Borges, 2001) and multi-objective (Ducheyne et al., 2004) formulations are rare. To examine the applicability, behavior, and performance of a generalized multi-objective evolutionary algorithm (MOEA), the evolutionary algorithm presented by (Knowles and Corne, 2000) was used to solve the formulation in Section 2.5.

The $(\mu + \lambda)$-Pareto Archive Evolution Strategy $((\mu + \lambda)$-PAES), requires no assumptions about the decision maker or their preferences in the form of weights or scaling factors to reduce the solution space to a scalar function (Steuer, 1985). The method, presented in Algorithm 1, is an elitest strategy, which guarantees convergence (Rudolph and Agapie, 2000), is simple to program, generates λ mutations at each generation, maintains a set of μ non-dominated solutions throughout the search, and has been shown to work

well for multitude of Pareto front types (Tan et al., 2005).

The initial population of λ candidate solutions are generated randomly and evaluated using the formulations presented in Section 2.3. The set of initial λ candidate solutions are then added to the Pareto archive (i.e. $\mu = \lambda$). A scalar rank is then assigned to each candidate solution, where the rank is defined as the number of $\mu + \lambda$ individuals that dominate the candidate solution. When there are no solutions that dominate the candidate solution, (i.e. non-dominated) a rank of 0 is assigned to the candidate solution. All candidate solutions with $rank > 0$ are then removed from the archive (i.e. dominated). The remaining non-dominated (μ) candidate solutions, which define the current Pareto frontier, are then selected for promotion and further mutation.

Selection for promotion is performed using tournament selection where a set of h candidate solutions enter a tournament and a single winner is selected with probability p (Dumitrescu et al., 2000). We used deterministic binary tournaments ($h = 2, p = 1$), where two solutions are selected with uniform probability from the Pareto archive (i.e. uniform, $P(Sel = 1) = 1/|\mu|$), and one of two candidate solutions is selected, with a probability of one, to be copied into the population buffer (λ) for promotion and mutation. The number of individual objectives that dominated the tournament for each candidate solution determined the winner with probability of one. In the event of a tie (i.e. weakly non-dominated), the first contestant was deemed the winner. The process of selection, promotion, and mutation is performed until some stopping criteria is met. For this manuscript, we used a deterministic stopping criteria defined by the number of generations (Deb, 2001).

Since a portion of the *input data* for the problem is a constant regardless of the values of the decision variables needed to evaluate a specific candidate

solution, the decision vector included only variables for the day and sequence combination:

$$\mathbf{x}_c = \langle d_1, \ldots, d_i, \ldots, d_8, s_1, \ldots, s_i, \ldots, s_8, \sigma \rangle \tag{2.46}$$

where $d_1, \ldots, d_i, \ldots, d_8$ represents the harvest/reserve day in which to assign stem i, and $s_1, \ldots, s_i, \ldots, s_8$ represents the sequence on which harvest day the stem is to be harvested on day d_i.

The last entry in the candidate solution, σ represents the probability of mutation for an individual candidate solution, which in this example, was examined for three different values (0.01,0.05,0.1). The evolution strategy parameters are presented in Table 2.2.

To examine the behavior and performance of the heuristic, 1000 replications were performed for each of the three mutation rates and population sizes listed in Table 2.2 as is suggested by Bartz-Beielstein (2006) and Bang-Jensen et al. (2007).

2.6   RESULTS AND DISCUSSION

The results and discussion presented here contains a brief presentation of the geostatistical method used to predict the volume for stem eight, $\hat{v}_8$, followed by a discussion of the behavior and performance of the heuristic. Finally, we present a detailed discussion of a *single* non-dominated solution as would be the case in many operational organizations, where only one Pareto optimal solution that meets the multiple objectives of the decision maker is selected for implementation.

Data: $\mathbf{x}, \hat{\xi}, \mathbf{f}(\mathbf{x}, \hat{\xi}), \mathbf{g}(\mathbf{x}, \hat{\xi}), \mathbf{h}(\mathbf{x}, \hat{\xi})$
Data: field data, yield estimates,objective and constraint formulae,
      software
Result: $(\mu+\lambda)$-PAES Multi-Objective Evolutionary Algorithm
      (Knowles and Corne, 2000)
$g \leftarrow 1$;
$\mu_g = \emptyset$;
$\lambda_g \leftarrow GenerateRandom()$;
$Evaluate(\lambda_g)$;
$\mu_g \leftarrow UpdateParetoArchive(\lambda_g)$;
for $g \leftarrow 2$ **to** G **do**
    $\lambda_g \leftarrow SelectCandidatesFromArchive(\mu_{g-1})$;
    $\lambda_g \leftarrow Mutate(\lambda_g)$;
    $Evaluate(\lambda_g)$;
    $\mu_g \leftarrow UpdateParetoArchive(\lambda_g)$;
end
$ExportArchive(\mu_G)$;

Algorithm 1: Pareto archiving multi-objective evolutionary algorithm.

Parameter	Specification
Representation	Real/Integer-valued vector
Population size	10, 50, and 100 Candidate solutions
Generations	1000
Recombination	None
Mutation	1%, 5%, and 10%
Parent Selection	Uniform random
Tournament type	Binary deterministic (k=2,p=1)
Survivor Selection	$(\mu + \lambda)$-PAES
Specialty	None
Stopping Criteria	1000 generations

Table. 2.2: Evolution strategy parameters.

2.6.1 Optimal prediction of spatially correlated inputs

Since the first objective was to obtain the *best* input (i.e. unbiased minimum error variance prediction) for the missing stem volume, regardless of the solution methodology, the system of $n+1$ ordinary kriging equations was solved independently (Cressie, 1993). It should be noted here that ordinary kriging is normally used for situations where the regionalized variable of interest is continuous with an unknown mean (Isaaks and Srivastava, 1989). Rather than create a realistically complex prediction situation, ordinary kriging was used to reduce formulation complexity while remaining consistent with the support of the decision variables (i.e. individual tree or point support). The resulting unbiased minimum variance prediction for the volume of stem eight was $\hat{\xi} \equiv \hat{v}_8 = 592.729$ volume units. The accompanying minimum prediction variance was 8.956 volume units squared, which, while not used here, can be used for chance constraints and probability maximization formulations (Hof et al., 1992).

The ability to determine the best unbiased minimum variance predictions and solve the optimization formulation simultaneously has both advantages and disadvantages. One advantage is that both the prediction method for missing observations and the solution method for the optimization problem are connected in both documentation via formulation and the computer procedures used to solve them. Should the outputs agree with results from independently predicted values, this should give credence to the method. For more realistic problems, the decision maker can refine the formulation by adding imputation metrics for missing values, (Little and Rubin, 2002), more complex kriging methods for both finite populations (Hoef, 2002), infinite populations (Goovaerts, 1997), chance-constraints (Hof et al., 1992),

and probability objectives (Hof and Bevers, 1994, 1998).

The disadvantage of this method is that, if the decision variables, or functions thereof, are independent of the fixed, but unknown input values, then solving the prediction problem and the optimization problem simultaneously with each run, unnecessarily increases the solution time for a given analysis and creates a situation where the predictions for the unknown inputs between heuristic runs may not be consistent, which is undesirable when alternatives are compared among heuristic outputs or several runs (Bartz-Beielstein, 2006).

2.6.2   Heuristic behavior and performance

While the mutation rate and population size parameters are typically the variables of interest in evolutionary computation research (Bartz-Beielstein, 2006), we fixed the mutation rate and population sizes for each run to 1) minimize potentially confounding factors that might arise from dynamically modifying the mutation rate and population sizes, and 2) because, for a given population size, there are endless methods in which the mutation rate could have been modified. To examine the effectiveness of the heuristic, we now briefly discuss the heuristic's ability to search $\mathcal{S}$, which contained 465 global Pareto optimal solutions (see Appendix A), by examining the size of the Pareto archive and the rate at which non-dominated solutions are ejected from the Pareto archive as a result of mutation, selection pressure, and the "learning rate".

2.6.2.1   Pareto frontier approximation

Figure 2.1 shows the the size of the Pareto archive (e.g. non-dominated solutions) at the end of every fiftieth generation for the three mutation rates ($\sigma \in \{0.01, 0.05, 0.1\}$) and population sizes ($\lambda \in \{10, 50, 100\}$). This value represents the number of solutions that have dominated all previous μ non-dominated and λ mutated candidate solutions. The lowest population buffer size ($\lambda = 10$) yielded the lowest non-dominated solution discovery rates evident by lowest slopes of the mean number of discovered non-dominated solutions. Interestingly, the lowest population size and mutation rate also produced local optima as is evident by the number of runs that resulted in Pareto sets with more solutions than the global set suggesting that this combination was insufficient for determining the actual frontier. The number of Pareto solutions discovered increased with the mutation rate, for all three population sizes, suggesting that the rate at which the number of solutions were added to the archive was constant over the run.

The rate at which new non-dominated solutions entered the Pareto archive was approximately constant (i.e. linear) until the population of candidate solutions reached 50 with a mutation rate of 0.01. On average, a decrease in the "discovery rate" is apparent near the middle of a run of 1000 generations as is evident by the change from a linear to a curvilinear form in higher population sizes and mutation rates. For the highest mutation rate, the discovery of the Pareto optimal solutions increased, and then again, slowed quickly by the end of the first quarter of the runs. The variance of the number of non-dominated solutions discovered, decreased as the population size and mutation rates increased. The variation decreased most rapidly for the higher mutation rates near the end of the runs, indicative of convergence. These patterns suggests

Number of Pareto Optimal Solutions
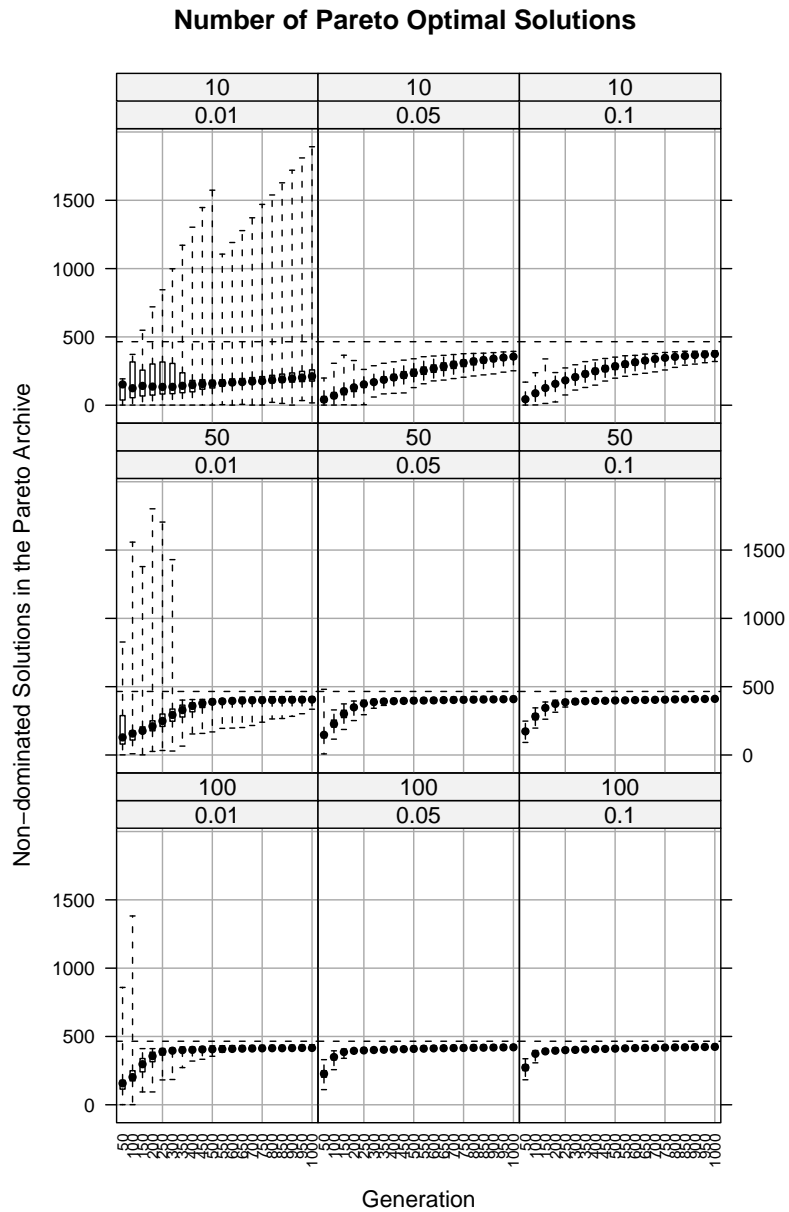
Figure. 2.1: Box and whisker plot of the number of non-dominated Pareto optimal solutions for three candiate solution population sizes $\lambda \in \{10, 50, 100\}$ and the three mutation rates $\sigma \in \{0.01, 0.05, 0.1\}$ every 50th generation. The dashed line represents the size of the global optimal Pareto set of 465 solutions.

that the mutation rate, when combined with the population size, has more influence on the rate at which the search "learns" new non-dominated solutions than does the size of the population alone. This is evident in the pattern of curvature within the population sizes. For small population sizes, the slope of the median number of non-dominated solutions was constant while the slope becomes more curvilinear as both the population and the mutation rate increase.

2.6.2.2 Mutation Efficiency

Since Pareto candidate solutions were selected for promotion with equal probability, at each generation, we examined the ratio at which the non-dominated solutions were dominated by new mutations (Figure 2.2). We define the mutation efficiency as the ratio of the total number of non-dominated solutions originating from the archive over the number of non-dominated solutions from both the candidate population and the existing archive. A higher value is indicative that the search is not generating dominating mutations and so, the progress is slow. A low value indicates that the mutation process is generating solutions that displace non-dominated solutions in the Pareto archive. The ratio is similar to the "progress ratio", as presented by (Tan et al., 2005), but is not a moving average nor was it used as a stopping criteria. More comprehensive stopping criteria for multi-objective heuristics are presented in the machine learning and evolutionary algorithm literature (Mansour and Schain, 1997; Farhang-Mehr and Azarm, 2002; Zhang, 2003).

For all population sizes and mutation rates, the search effectiveness appeared to slow early in the search suggesting that a change in the mutation

**Proportion of Non–dominated Pareto Optimal Solutions
Originating from within the Pareto Archive**
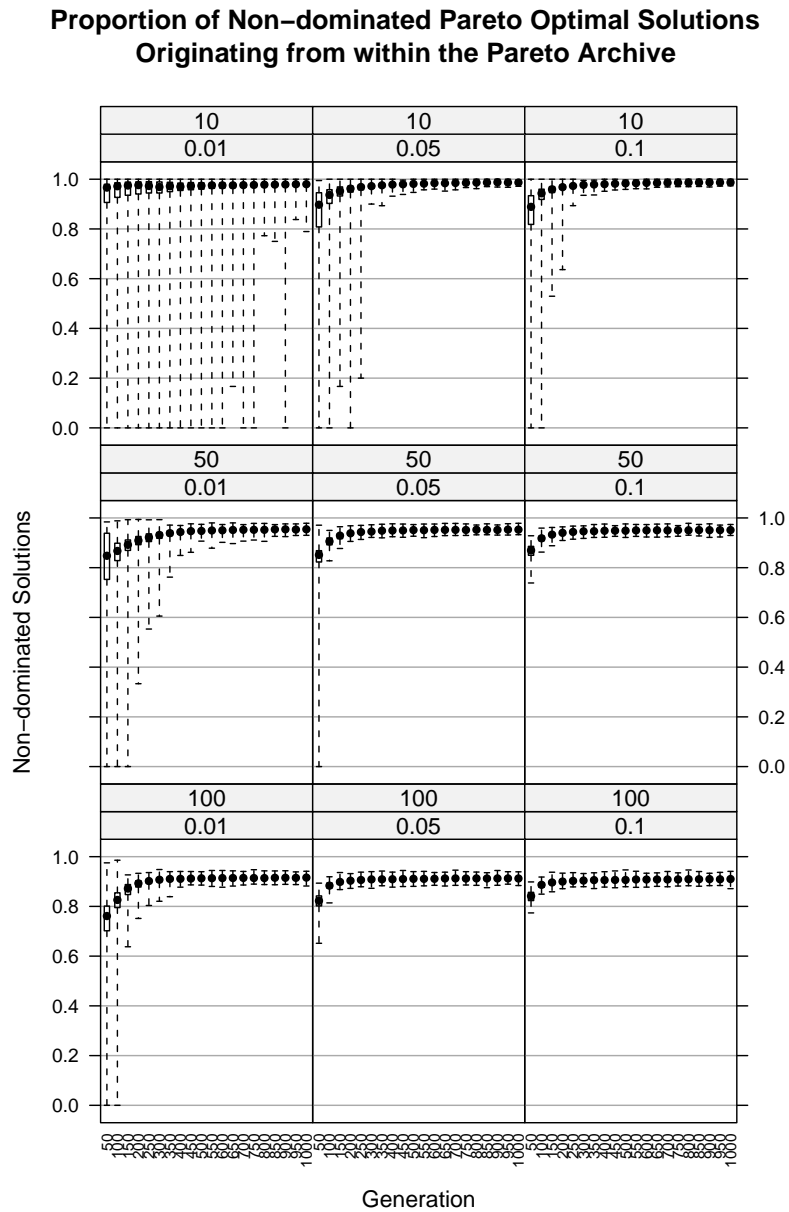
Figure. 2.2: Box and whisker plot of the mutation efficiency for three candidate solution population sizes $\lambda \in \{10, 50, 100\}$ and the three mutation rates $\sigma \in \{0.01, 0.05, 0.1\}$ every 50th generation. A flat slope means that few mutations have displaced non-dominated solutions in the Pareto archive which indicates the search progress is slow.

strategy may be beneficial. An asymptote near unity was pronounced for all nine combinations of population size and mutation rate. For the smallest population size (i.e. 10 candidate solutions), all three mutation probability rates yielded a higher number of function evaluations than did larger populations. The larger population sizes however, resulted faster convergence rates (i.e. more overall mutations) and thus ultimately more exhaustive searches.

Since the goal here was a general examination of the $(\mu + \lambda)$-PAES and not metrics such as the mutation operator, the plot in Figure 2.2 reveals that the mutation operator effectiveness can be examined over time, and used as means of selecting mutation strategies (Deb, 2001), determining candidate population sizes (Coello Coello and Toscano Pulido, 2001), or as a stopping criteria (Tan et al., 2005).

2.6.3 Optimal operational plans

To examine a specific set of Pareto optimal solutions, as would be the case in an operational environment, the run with the highest number of discovered non-dominated solutions was selected for further examination. Beginning with a matrix plot, as suggested by Deb (2001), no discernible single plan appeared uniquely sufficient to be declared an "obvious" choice for implementation (Figure 2.3).

The selected run contained 408 solutions (87.742 percent of the global Pareto optimal set), where each solution represents a different optimal operational plan. It is important to note again that a *single* run of the heuristic generated the 408 best possible plans from which any single plan from the Pareto optimal set yields an optimal solution subject to the time constraints of the analysis. This allows the decision maker to have a documented analysis

Matrix Plot of Pareto Archive

Figure. 2.3: Matrix plot of the Pareto optimal set.

of the trade-offs that does not require direct assignment of weights or preferences as do exact solution methods thus reducing potential anthropocentric biases without having to completely enumerate the solution space.

Figure 2.3 reveals a variety of patterns among the Pareto set for this small problem. A relationship among the daily volume production values can be observed for the top left block of scatter plot matrix. The Pareto frontier is visible as a diagonal boundary among the three daily production volumes for the three days, $\hat{V}_1$, $\hat{V}_2$, and $\hat{V}_3$, respectively. The distribution of the residual volumes, $(\hat{V}_4)$ is minimized, for a given solution, as represented by the vertical cluster on the left (lower end of the variable), demonstrating that the set of solutions yields a residual volume that is as low as possible, while still meeting the target of 300 volume units over $\mathcal{A}$. A pattern in the objective to minimize the total distance is less pronounced, but there does appear to be a weak relationship between the residual volume and the distance a machine must travel in the Pareto set. For larger datasets and problems, a stronger relationship between the distance and production may be more pronounced.

A query of the selected Pareto set was performed to determine how the different plans maximized or minimized the various objectives. There were 2, 5, and 3 plans that maximized daily production, for each of the three production periods, respectively and 1 plan that minimized machine travel distance (see Table 2.6.3). There were 371 plans that minimized the residual volume at it's lowest possible level while remaining feasible. These solutions have been removed from Table 2.6.3 for brevity. All 408 plans did not violate any of the constraints ($\psi_i = 0 \ \forall \ i \in |\lambda_G|$).

Finally, in order to further understand the trade-offs among the various optimal plans, the machine paths for the three solutions that maximized

Table. 2.3: Pareto Optimal Plans for each of the objectives. Note that $\mathcal{T}_1^v$ is the residual volume in all but one plan since it is the only one that can meet the residual volume objectives and meet the constraints by itself. The solutions that minimize the residual volume are removed for brevity. The format for the cutting plan is four sets of stems representing the paritions (i.e. knapsacks) to which each stem is assigned, $\hat{\xi} \equiv \hat{v}_8 = 592.729$, and $\psi_i(\mathbf{x}, \hat{\xi}) = 0 \, \forall i \in \mu_G$. The plans followed by the †symbol, are plotted in Figure 2.4.

Pareto Plan	$\mathbf{x}$	$\widehat{V}_1(\mathbf{x}, \hat{\xi})$	$\widehat{V}_2(\mathbf{x}, \hat{\xi})$	$\widehat{V}_3(\mathbf{x}, \hat{\xi})$	$\widehat{V}_4(\mathbf{x}, \hat{\xi})$	Tour Distance
1†	(7,2,6),(5,3),(4,8),(1)	**2270**	833	1239	477	55.9
2	(7,6,2),(8,3),(4,5),(1)	**2270**	820	1252	477	51.4
292†	(2,5),(6,7,3),(4,8),(1)	1302	**1801**	1239	477	48.0
308	(2,8),(3,7,6),(5,4),(1)	1289	**1801**	1252	477	50.5
324	(4,5),(6,7,3),(8,2),(1)	1252	**1801**	1289	477	50.5
340	(8,4),(3,7,6),(5,2),(1)	1239	**1801**	1302	477	48.0
356	(8,5),(6,7,3),(4,2),(1)	1199	**1801**	1342	477	50.3
339†	(5,4),(3,8),(2,6,7),(1)	1252	820	**2270**	477	51.4
355	(8,4),(3,5),(6,2,7),(1)	1239	833	**2270**	477	55.9
371	(5,8),(3,4),(7,6,2),(1)	1199	873	**2270**	477	49.1
398†	(5,6),(2,8),(4,7),(1,3)	1397	1289	1429	704	**32.4**
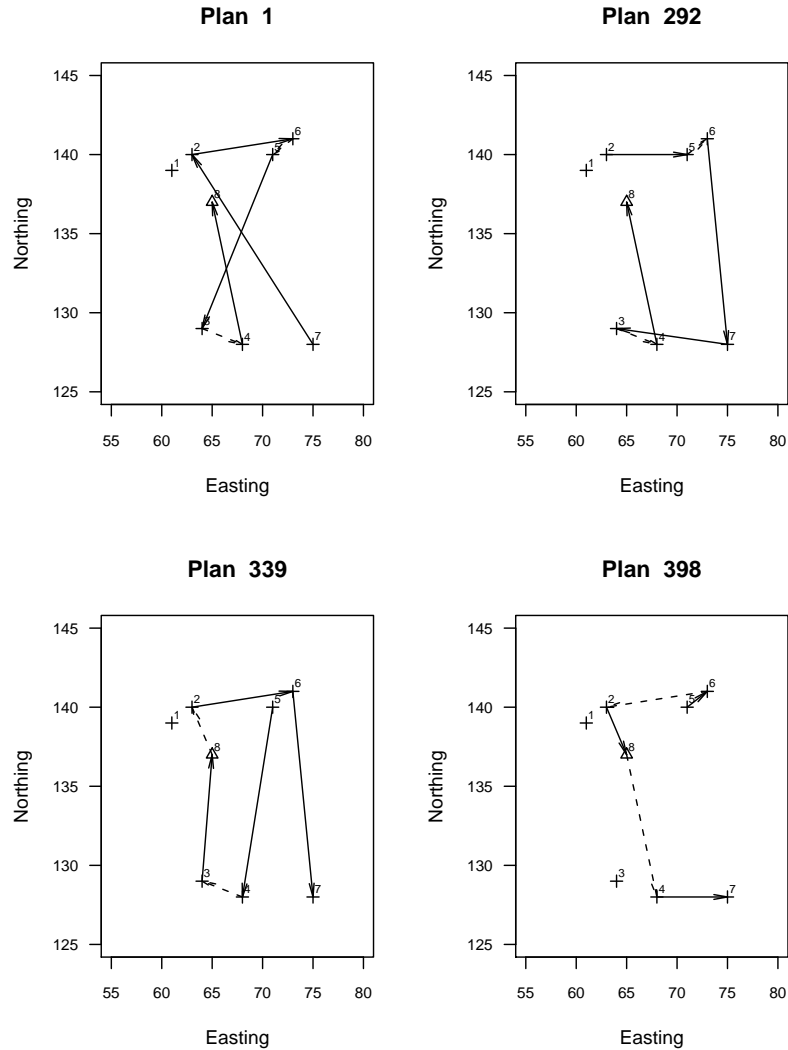
Figure. 2.4: The three machine routes presented in the upper left, upper right, and lower left maximize $\hat{V}_1$, $\hat{V}_2$, and $\hat{V}_3$, or daily delivered volume objectives for days one, two, and three respectively. The one tour that minimizes the soil disturbance is presented in the lower right. Stems to be harvested on the same day are connected with solid arrows and the dashed arrow is the path the machine takes at the end of the period to reposition for the next day. Stems not on the machine's path are reserve stems. The triangle represents the only fixed, but unknown stem volume in the dataset, which was predicted using ordinary kriging.

daily volume production and one that minimized the distance objective were plotted (Figure 2.4). For each of the objectives, the solution that was first in the partially ordered set was selected for plotting. The routes for the plans that maximized the daily production for the three days are plotted in the upper half and lower left in Figure 2.4. The plots of each route are represented by solid and dashed lines where the solid lines represent the daily route among the stem locations and the dashed lines represent the path the harvester must take to move from the last stem of the day to the first stem of the next.

The plan that minimized the soil disturbance (i.e. total soil contact) regardless of the daily production volume, is plotted in the lower right of Figure 2.4. This solution had the shortest travel distance over all solutions while also meeting the constraints. For a majority of the plans, stem one was marked as the only reserve stem as it was the only stem that had enough volume to meet the reserve volume constraint by itself while meeting the other constraints as well.

2.7 CONCLUSIONS

The framework and heuristic presented here can be used to formulate and solve complex non-linear mixed-integer supply chain planning and optimization problems with potentially competing and non-commensurate multiple objectives which are common in strategic, tactical, and operational forest planning. The objectives and constraints can be constructed using simple formulas that may contain commonly used logical operators and can include both deterministic and stochastic components. By using a Pareto archive evolution strategy, the resulting set of non-dominated Pareto optimal solu-

tions, can 1) be reported at any point during a solution process, 2) be refined by reloading the Pareto archive and continuing the search, 3) be used to modify mutation strategies, 4) be used determine the stopping criteria, 5) be guaranteed to converge given sufficient running time, and 6) facilitate an understanding of the relationships (e.g. elasticity) among the potential decisions and their inputs.

The features of this method are beneficial for practitioners who might be unfamiliar with the minutiae of a particular problem formulation, solution method, and constraint set. The framework allows the decision maker to examine the consequences of optimal alternatives, for various datum supports (i.e. tree models, stand tables, or forest models), with a single run of the heuristic without assigning anthropocentric weights or preferences as is common in other multi-objective methods. The method shows promise in situations where a quick examination of the trade-offs is required, possibly to be refined later, as might be the case in "roundtable" policy analysis type environments, fast paced production environments, or highly variable decision spaces.

As the focus was on the basic formulation and the search procedure, a detailed examination of the influence of the mutation and selection operators on solution quality and times were not reported. The solution time for each of the 4500 runs was less than one minute, which was considerably faster than the 341.39 hours required to completely enumerate $\mathcal{S}$. The need to examine mutation operators, selection operators, and methods to visualize Pareto frontiers in forest planning problems is large indeed, as there are many complex relationships among input data, decision variables, and policies defined by $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$ and is the subject of future research.

BIBLIOGRAPHY

Armstrong, M. (1998). *Basic Linear Geostatistics.* Springer.

Bang-Jensen, J., Chiarandini, M., Goegebeur, Y., and Jørgensen, B. (2007). Mixed models for the analysis of local search components. In Stützle, T., Birattari, M., and Hoos, H. H., editors, *SLS*, volume 4638 of *Lecture Notes in Computer Science*, pages 91–105. Springer.

Bartz-Beielstein, T. (2006). *Experimental Research in Evolutionary Computation: The New Experimentalism (Natural Computing Series).* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Block, R., Van Rees, K. C. J., and Pennock, D. J. (2002). Quantifying Harvesting Impacts using Soil Compaction and Disturbance Regimes at a Landscape Scale. *Soil Sci Soc Am J*, 66(5):1669–1676.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization.* Cambridge University Press, Cambridge, UK.

Coello Coello, C. A. (2002). Evolutionary multi-objective optimization: A critical review. In Sarker, R., Mohammadian, M., and Yao, X., editors, *Evolutionary Optimization*, chapter 5. Kluwer Academic Publishers, Norwell, MA, USA.

Coello Coello, C. A. and Toscano Pulido, G. (2001). A Micro-Genetic Algorithm for Multiobjective Optimization. In Zitzler, E., Deb, K., Thiele, L., Coello, C. A. C., and Corne, D., editors, *First International Conference on Evolutionary Multi-Criterion Optimization*, pages 126–140. Springer-Verlag. Lecture Notes in Computer Science No. 1993.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, chapter 2. John Wiley and Sons, Inc., New York, New York, revised edition edition. 900.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms.* John Wiley & Sons, Chichester, UK. ISBN 0-471-87339-X.

Deb, K., Tewari, R., Dixit, M., and Dutta, J. (2007). Finding Trade-off Solutions Close to KKT Points Using Evolutionary Multi-Objective Optimization. KanGAL Report 2007006, Indian Institute of Technology Kanpur, Kanpur, PIN 208016, India.

Deusen, P. C. V. (1999). Multiple solution harvest scheduling. *Silva Fennica*, 33(3):207–216.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2004). Even flow scheduling problems in forest management. In Coello, C. A. C. and Lamont, G. B., editors, *Applications of Multi-Objective Evolutionary Algorithms*, pages 701–726. World Scientific, Singapore.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2006). A spatial approach to forest-management optimization: linking gis and multiple objective genetic algorithms. *International Journal of Geographical Information Science*, 20(8):917–928.

Dumitrescu, D., Lazzerini, B., Jain, L., and Dumitrescu, A. (2000). *Evolutionary Computation*. CRC Press.

Eiben, A. E. and Smith, J. E. (2003). *Introduction to Evolutionary Computing*. SpringerVerlag.

Falcão, A. O. and Borges, J. G. (2001). Designing an Evolution Program for Solving Integer Forest Management Scheduling Models: An Application in Portugal. *Forest Science*, 47(2):158–168.

Farhang-Mehr, A. and Azarm, S. (2002). Diversity assessment of pareto optimal solution sets: An entropy approach. In *Congress on Evolutionary Computation (CEC'2002)*, volume 1, pages 723–728, Piscataway, New Jersey. IEEE Service Center.

Gen, M. and Cheng, R. (1997). *Genetic Algorithms and Engineering Design*. John Wiley and Sons, Inc., New York.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics Series. Oxford University Press, New York, New York.

Hamann, J. D. and Boston, K. (2007). Using spatial statistics to improve the primary forestry supply chain. In Reynolds, K. M., Thompson, A., Shannon, M., Köhl, M., Ray, D., and Rennolls, K., editors, *Sustainable Forestry: from Monitoring and Modelling to Knowledge Management & Policy Science*, chapter 18, pages 285–297. CABI International.

Hansen, M. P. (1997). Tabu Search in Multiobjective Optimisation : MOTS. In *Proceedings of the 13th International Conference on Multiple Criteria Decision Making (MCDM'97)*, Cape Town, South Africa.

Hoef, J. M. V. (2002). Sampling and geostatistics for spatial data. *Ecoscience*, 9(2):152–161.

Hof, J. and Bevers, M. (1994). Optimal timber harvest scheduling with spatailly degined sediment objectives. *Canadian Journal of Forest Research*, 30(1494-1500).

Hof, J. G. and Bevers, M. (1998). *Spatial Optimization for Managed Ecosystems*. Columbia University Press, New York.

Hof, J. G., Kent, B. M., and Pickens, J. B. (1992). Chance constraints and chance maximization with random yield coefficients in renewable resource optimization. *Forest Science*, 38:305–323(19).

Hudak, A. T., Lefsky, M. A., Cohen, W. B., and Berterretche, M. (2002). Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, 82(2-3):397–416.

Hyoun-Jin, S., Kyung-Joon, C., Jae-Eung, O., and Je-Seon, R. (2005). Kriging with meta-heuristic methods for optimal design to reduce the noise of the engine cooling fan. *JSME international journal. Series C, Mechanical systems, machine elements and manufacturing*, 48(3):363–370.

Igel, C., Hansen, N., and Roth, S. (2007). Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15(1):1–28.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, Inc., New York, New York.

Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. Master's thesis, Department of Mathematics, University of Chicago, Chicago, Ill.

Knowles, J. D. and Corne, D. W. (2000). Approximating the nondominated front using the pareto archived evolution strategy. *Evolutionary Computation*, 8(2):149–172.

Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In Neyman, J., editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1950*, pages 481–492, Berkeley, California. University of California Press. Reprinted in: *Readings in Mathematical Economics, Vol 1, Value Theory*, (P. Newman, ed.), The Johns Hopkins University Press, Baltimore, 1968, pp. 3–14.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002). Combining convergence and diversity in evolutionary multiobjective optimization. *Evol. Comput.*, 10(3):263–282.

Lawler, E. L., Lenstra, J. K., Kan, A. H. G. R., and Shmoys, D. B., editors (1985). *The Traveling Salesman Problem.* John Wiley, New York.

Lin, S. and Kernighan, B. W. (1973). An effective heuristic for the traveling salesman problem. *Operations Research*, 21:498–516.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* John Wiley and Sons, Inc.

Mandallaz, D. (2000). Estimation of the spatial covariance in Universal Kriging: Application to forest inventory. *Environmental and Ecological Statistics*, 7:263–284.

Mansour, Y. and Schain, M. (1997). Learning with maximum-entropy distributions. In *Computational Learing Theory*, pages 201–210.

Marti, K. (2005). *Stochastic Optimization Methods.* Springer.

Michalewicz, Z. and Fogel, D. B. (2004). *How to Solve It: Modern Heruistics*, chapter 5, page 554. Springer Verlag, 2 edition.

Nam, D. and Park, C. (2000). Multiobjective simulated annealing: A comparative study to evolutionary algorithms. *International Journal of Fuzzy Systems*, 2(2):87–97.

Osyczka, A. (2002). *Evolutionary Algorithms for Single and Multicriteria Design Optimization.* Physica Verlag, Germany. ISBN 3-7908-1418-0.

Reed, D. D. and Burkhart, H. E. (1985). Spatial autocorrelation of individual tree charactersitics in loblolly pine stands. *Forest Science*, 31(3):575–587.

Rudolph, G. and Agapie, A. (2000). Convergence Properties of Some Multi-Objective Evolutionary Algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, volume 2, pages 1010–1016, Piscataway, New Jersey. IEEE Press.

Steuer, R. (1985). *Multiple Criteria Optimization: Theory, Computation and Application.* John Wiley & Sons, New York, NY.

Stewart, T. J., Janssen, R., and van Herwijnen, M. (2004). A genetic algorithm approach to multiobjective land use planning. *Comput. Oper. Res.*, 31(14):2293–2313.

Stringer, C. (2006). Forest certification and changing global commodity chains. *J Econ Geogr*, 6(5):701–722.

Tan, K., Khor, E., and Lee, T. (2005). *Multiobjective Evolutionary Algorithms and Applications*. Springer.

Tóth, S., McDill, M. E., and Rebain, S. (2006). Finding the efficient frontier of a bi-criteria, spatially explicit, harvest scheduling problem. *Forest Science*, 52(1):93–107.

Tuominen, S., Fish, S., and Poso, S. (2003). Combining remote sensing, data from earlier inventories and geostatistical interpolation in multisource forest inventory. *Canadian Journal of Forest Research*, 33:624–634.

Wallerman, J., Joyce, S., Vencataswamy, C. P., and Olsson, H. (2002). Prediction of forest stem volume using kriging adapted to detected edges. *Canadian Journal of Forest Research*, 32:509–518.

White, H. (1999). *Asymptotic Theory for Econometricians*. Acandemic Press.

Yan, Z., Zhang, L., Kang, L., and Lin, G. (2003). A New MOEA for Multiobjective TSP and Its Convergence Property Analysis. In Fonseca, C. M., Fleming, P. J., Zitzler, E., Deb, K., and Thiele, L., editors, *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pages 342–354, Faro, Portugal. Springer. Lecture Notes in Computer Science. Volume 2632.

Zhang, B.-T. (2003). *A unified Bayesian framework for evolutionary learning and optimization*, pages 393–412. Springer-Verlag New York, Inc., New York, NY, USA.

3. AN EXAMINATION OF DESIGN-BASED AND MODEL-BASED SAMPLING FOR PRIMARY FOREST PRODUCTS SUPPLY CHAIN PLANNING AND OPTIMIZATION

3.1 ABSTRACT

We compare and discuss the performance of design-based estimation and model-based prediction on the primary forest products supply chain, specifically, the communication and information network system as defined by Riopel et al. (2005). Specifically, we examine and compare the characteristics and performance of design-based estimation and model-based prediction for guessing the total volume within a finite area (8.24 ha) for four sampling location patterns (random placement, rectangular, non-aligned rectangular, and hexagonal grids), four spatial correlation models, for nine target sample location counts $M \in \{10, 15, 20, 25, 30, 35, 40, 45, 50\}$.

For each combination of sample size, sample location pattern, and assumed spatial model, we generated 1000 replications, with and without edge-bias correction, from which we examine the design-based estimator developed by Horvitz and Thompson (1952). For model-based predictions, we used ordinary kriging (Cressie, 1993a), where the spatial correlation was defined using either a one parameter linear (LIN), a one-parameter spherical (SPH), or a two parameter exponential (EXP) variogram model.

We present the results of the estimation or prediction bias, a computed standard error, and finally the root mean squared error (RMSE), using box-and-whisker plots over the target sample sizes. Visually, performance gains resulting from non-edge-corrected model-based predictions were negligible. However, edge-corrected model-based predictors were superior, in every case,

to both edge-corrected design-based estimations and model-based predictions indicating that significant gains in sampling efficiency and the utility of the output data can be obtained by the inclusion of simple spatial correlation models and edge-correction, regardless of the selected spatial correlation model selected.

3.2 INTRODUCTION

Decision support and optimization models for precision forest operations require inputs such as the density of stems (i.e. stems ha^{-1}), volume density (i.e. m^3ha^{-1}), and mean stem volume (i.e. m^3), within a finite area, in order to determine a set of potential plans that minimize costs (Clark et al., 2000), maximize revenue (Vera et al., 2003), minimize environmental deterioration (Gjedtjernet, 1995), minimize portfolio risk (Reeves and Haight, 2000) or some combination thereof (Connaughton and Fight, 1984; Ducheyne et al., 2006). When the exact values of the inputs are fixed, but unknown, guesses of their values (i.e. global estimates or local predictions) and associated properties (i.e. distributions, moments, and correlations) can be obtained by both design-based estimates or model-based predictions (Gregoire, 1998).

Design-based estimators are well established and many studies examine the *influence* of plot size (Gambill et al., 1985), plot spacing (Chapman, 1982), and plot density (Lohr, 1999) on sampling costs (Mawson and Mack, 1982) for simple and complex (Oderwald and Jones, 1992; Mandallaz and Ye, 1999) sampling schemes (de Gruijter et al., 2006). These studies are often used to justify a particular sample intensity to minimize costs for a pre-specified level of precision (i.e. confidence interval half-width). In contrast, model-based prediction methods, such as geostatistics (Goovaerts, 1997) have

been used to examine structural relationships (e.g. spatial correlation) among tree size (Samra et al., 1989), volume (Holmgren and Thuresson, 1997), site index (Hock et al., 1993), cork production (Montes et al., 2005), and value (King, 2000).

While design-based methods are considered standard in many forest inventory texts (Avery and Burkhart, 2003; Husch et al., 2003), model-based methods can be more effective at describing phenomenon at operational scales as they 1) can be used to obtain values for an arbitrary datum support (i.e. point, block, or volume), 2) allow for the inclusion of some measure of dissimilarity versus proximity among sample locations (i.e. spatial correlation), 3) facilitate cross-validation to assess model assumptions, and 4) can directly incorporate measurement error directly into the prediction process (Wackernagel, 1998). One particularly useful feature of model-based inference, which addresses a key notion in forest operations that phenomenon (i.e. product abundance, costs, productivity rates), located more closely in space, or time, are more related than more distant phenomenon (Karlsson et al., 2003; Weintraub and Murray, 2006; Montes et al., 2005; Dimov et al., 2005) and that these relationships can be used to enhance the decision making process in the primary forest products supply chain (Hamann and Boston, 2007).

To examine the influence of spatial structure on precision production planning, we examine the influence of spatial correlation on estimations (i.e. design-based) and predictions (i.e. model-based) of the total volume in a finite operational area. We use definitions that are consistent with classical design-based sampling texts (Cochran, 1977) and model-based geostatistics texts (Wackernagel, 1998). In Section 3.3, we present the formulations for design-based estimation (Section 3.3.1) and model-based predictions (Section 3.3.2). Section 3.4 presents a description of the population and the simulation

process used to examine various sampling schemes as defined by de Gruijter et al. (2006). Section 3.5 presents the results and a discussion of our findings, followed by our conclusions in Section 3.6.

Nomenclature

Harvest Unit Attributes

$\mathcal{A}$	A harvest unit, stand, or general operational polygon
$\lambda(\mathcal{A})$	A function that returns the surface area of a polygon $\mathcal{A}$
N	The number of stems in $\mathcal{A}$
M	The number of sampling locations in $\mathcal{A}$
$V_{\mathcal{A}}$	The total volume in $\mathcal{A}$
$V_{\mathcal{A}}$	Intrinsic regionalized variable over $\mathcal{A}$ with point support

Tree Attributes

$\mathcal{T}$	Set of N stems in $\mathcal{A}$ that define the population and universe of inference
$\mathcal{T}_k$	Stem k in a population of N stems
$\mathcal{T}_k^{\mathbf{u}}$	Position of $\mathcal{T}_k$ $\mathbf{u} \in \mathbb{R}^2$
$\mathcal{T}_k^d$	Outside bark stem diameter at 1.3 m for $\mathcal{T}_k$ (i.e. DBH), in mm
$\mathcal{T}_k^h$	Total height for $\mathcal{T}_k$, in m
$\mathcal{T}_k^v$	Total volume for $\mathcal{T}_k$ ($\mathcal{T}_k^v = f(\mathcal{T}_k)$), in m^3
$\mathcal{T}_k^p$	The detection polygon for $\mathcal{T}_k$
$\mathcal{T}_k^{\pi_j}$	The inclusion probability for $\mathcal{T}_k$ as a function of $p_j(\,\cdot\,)$
$\mathcal{U}$	The set of unique labels for each $\mathcal{T}$

Plot Attributes

$\mathcal{P}$	Set of M sampling locations in $\mathcal{A}$
$\mathcal{P}_\ell$	Plot ℓ in a sample of M locations
$\mathcal{P}^{\mathbf{u}}$	Position of $\mathcal{P}_\ell$ $\mathbf{u} \in \mathbf{R}^2$
$\mathcal{P}_\ell^v$	A realization of a regionalized random process V, for the total volume of plot $\mathcal{P}_\ell$
$\mathcal{W}$	The set of unique labels for each $\mathcal{P}$

Estimator Formulation

$\widehat{V}_{\mathcal{A}}^{i,m,p}$	The estimator for V_A
i	Design-based or model-based inference

	method index ($i \in \{HT, OK\}$)
m	Spatial correlation model ($m \in \{\emptyset, L, S, E\}, \emptyset = $ none)
p	Sampling pattern index $p \in \{RND, REG, NON, HEX\}$
$\mathbf{I}_{k,\ell}(D, \theta_1, \theta_2)$	An $N \times M$ indicator vector or matrix for sample inclusion of $\mathcal{T}_k$ at sample location $\mathcal{P}_\ell$
$p(\,\cdot\,)$	A sample design function that assigns a probability of inclusion, π, to each element of $\mathcal{T}$
D	Minimum $\mathcal{T}^d$ to be included in fixed area plot or variable radius plot
θ_1	Fixed area plot radius when $\mathcal{T}^d \leq D$
θ_2	Angle gauge constant for variable radius plots when $\mathcal{T}^d > D$
$d_{k,\ell}$	$N \times M$ distance matrix from $\mathcal{T}_k$ to $\mathcal{P}_\ell$
$\mathbf{h}$	Lag or distance between $\mathcal{P}_i^{\mathbf{u}}$ and $\mathcal{P}_j^{\mathbf{u}}$

Operators and Miscellaneous Notation

$\mathbf{0}$	A vector of zeros
$\mathbb{N}, \mathbb{R}, \mathbb{C}$	The set of natural, real, and complex numbers
$\forall, \exists, \nexists, \in, \notin \subset$	for all, exists, not exists, in, not in, and subset
$\vee, \wedge, \neg$	logical OR, AND, and NEGATION operators
$\mathbb{E}[\,\cdot\,], \mathbb{V}[\,\cdot\,], \mathbb{C}[\,\cdot\,]$	Expected value, variance, and covariance operator
$\mathbf{C}', \mathbf{c}'$	transpose of matrix $\mathbf{C}$ or vector $\mathbf{c}$
$\mathbf{C}^{-1}$	Inverse of $\mathbf{C}$
α	Coordinates for a location $(\alpha_1, \alpha_2) \in \mathcal{A}$

3.3   METHODS

Let $\mathcal{A} \subset \mathbb{R}^2$, represent an operational harvest polygon with a finite boundary, where the surface area, measured in hectares, is $\lambda(\mathcal{A})$. The polygon $\mathcal{A}$ contains a finite population of N stems denoted by the set $\mathcal{T}$. Each stem $\mathcal{T}_k$ is defined by a unique label and $\mathcal{U} = \{1, \ldots, k, \ldots, N\}$ is the set of labels that is used to uniquely identify each stem. The attributes for each stem includes a position $\mathcal{T}_k^{\mathbf{u}} \in \mathbb{R}^2$, a diameter $\mathcal{T}_k^d$, and volume $\mathcal{T}_k^v$ and are assumed to be

known or measured without error. To reduce the notation, we shall write $\mathcal{T} \cap \mathcal{A}$ to represent all stems within the boundary of $\mathcal{A}$.

Let $\mathcal{P}$ represent a set of M measurement locations placed within $\mathcal{A}$, where each measurement location (i.e. point/plot) $\mathcal{P}_\ell$ is defined by a unique label and $\mathcal{W} = \{1, \ldots, \ell, \ldots, M\}$ is the set of labels that are used to uniquely identify each sampling location. For this examination, each sampling location contains a position vector $\mathcal{P}_\ell^{\mathbf{u}} \in \mathbb{R}^2$ and a volume density observation $\mathcal{P}_\ell^v$ (i.e. m^3 ha^{-1}) that are measured without error. Again, to reduce the notation, we shall write $\ell \in \mathcal{W}$ or $\mathcal{A}_{\ell \in \mathcal{W}}$ to represent all sampling locations in $\mathcal{A}$.

The target parameter, which here is the fixed, but unknown total volume in $\mathcal{A}$, denoted by $V_\mathcal{A}$, expressed as the function of interest as:

$$V_\mathcal{A} = f(\mathcal{T}_1^v, \ldots, \mathcal{T}_k^v, \ldots, \mathcal{T}_N^v) = \sum_{k \in U} \mathcal{T}_k^v \tag{3.1}$$

with the spatial mean over $\mathcal{A}$, denoted by $\overline{V}_\mathcal{A}$, defined as:

$$\overline{V}_\mathcal{A} = \lambda(\mathcal{A})^{-1} f(\mathcal{T}_1^v, \ldots, \mathcal{T}_k^v, \ldots, \mathcal{T}_N^v) = \lambda(\mathcal{A})^{-1} \sum_{k \in U} \mathcal{T}_k^v \tag{3.2}$$

with the variance of the total volume expressed as:

$$\sigma_{V_\mathcal{A}}^2 = \frac{1}{N} \sum_{k \in U} (\mathcal{T}_k^v - \overline{V}_\mathcal{A})^2 = \frac{1}{2N^2} \sum_{k \in U} \sum_{\ell \in U} (\mathcal{T}_k^v - \mathcal{T}_\ell^v)^2 \tag{3.3}$$

The sampling unit is the individual stem $\mathcal{T}_k$. All sample units that are observed, taken collectively, are referred to as a sample (Tillé, 2006). A function $p(\,\cdot\,)$, commonly referred to as a sample design or plot design in forest sampling, assigns an inclusion probability to each stem $\mathcal{T}_k$ for each unique combination of $\mathcal{T}$ and $p(\,\cdot\,)$ (de Gruijter et al., 2006; Tillé, 2006). Common definitions of $p(\,\cdot\,)$ include "split-plot designs" (i.e. $p(\,\cdot\,) = f(D, \theta_1, \theta_2)$) where $\mathcal{T}$ is stratified into two populations based on some diameter D, so that when

$T_k^d \leq D$, stems are included on a fixed-area plot, and when $T_k^d > D$, stems are included on a variable-radius or "prism" plot, where θ_1 defines the fixed-area plot radius, and θ_2 is an angle gauge constant, or basal area factor, that defines the radius of the circular inclusion probability polygon associated with stem $\mathcal{T}_k$, when $T_k^d \geq D$.

Our goal is to obtain the *best* guess of $V_\mathcal{A}$ and our scope of inference, and thus our understanding of the current conditions in $\mathcal{A}$, for a single sample in time, is a function of the unique combination of $\mathcal{T}$, $\mathcal{P}$, and $p(\,\cdot\,)$ (Gregoire, 1998; Knottnerus, 2003; Tillé, 2006). In the next two sections, we present two methods to obtain a guess of $V_\mathcal{A}$, which will then be used as an input for a precision supply chain decision support framework (Riopel et al., 2005).

The first method of guessing is the classic design-based estimator (i.e. estimation) of the population total, presented by Horvitz and Thompson (1952), denoted $V_\mathcal{A}^{HT}$, and is most commonly used to make inference, not about a particular stand, but about a larger population of similar polygons (Gregoire, 1998). The second method of guessing the total volume in A is a model-based inference method which can be used to obtain predictions for various datum support levels (i.e. point, block, and volume datum), is the classic geostatistical predictor known as Kriging, denoted $V_\mathcal{A}^{OK}$. The method is not used to make inference about a larger population of similar polygons as is the case in estimation, but to make predictions for a specific population (i.e. $\mathcal{T} \cap \mathcal{A}$) (Mandallaz, 2000).

Since we wish to examine the influence of these two inference types (i.e. traditional sampling, or estimation ($V_\mathcal{A}^{HT}$) versus model-based sampling (i.e. spatial sampling, or prediction ($V_\mathcal{A}^{OK}$)) within a precision supply chain framework, specifically the communication and information network (Riopel et al., 2005), our interest is specifically on the performance (i.e. bias and variance)

of the two methods of inference, and so we compare the two, examine their applicability to predict volume density within $\mathcal{A}$ to refine our guess, over $\mathcal{A}$, to enhance the input data for complex decision making.

3.3.1 Design-Based Estimation

A sample can be represented by the column vector $\mathbf{s} = (s_1, \ldots, s_k, \ldots, s_N)' \in \{0, 1\}^N \,\forall\, k \in U$, where $s_k = 1$ if stem k is in the sample, and $s_k = 0$ if it is not (Knottnerus, 2003). An empty sample is denoted by $\mathbf{s} = (0, \ldots, 0, \ldots, 0)'$ and a census is defined by $\mathbf{s} = (1, \ldots, 1, \ldots, 1)'$ (Tillé, 2006).

The well known unbiased minimum variance estimator for Equation (3.1), developed by Horvitz and Thompson (1952), is:

$$\widehat{V}_{\mathcal{A}}^{HT} \equiv \sum_{i=1}^{n} \frac{\mathcal{T}_i^v}{\pi_i} \equiv \sum_{j=1}^{N} \frac{\mathcal{T}_j^v s_j}{\pi_j} \tag{3.4}$$

where n is the sample size, which for now will considered fixed, but need not be (Gregoire, 1998; Mandallaz and Ye, 1999; Tillé, 2006), and $\boldsymbol{\pi}$ is the vector of N inclusion probabilities, defined below.

The associated variance of its sampling error is expressed as:

$$Var_{HT}[\widehat{V}_{\mathcal{A}}^{HT}] = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\pi_{ij} - \pi_i \pi_j \right) \frac{\mathcal{T}_i^v \mathcal{T}_j^v}{\pi_i \pi_j} \tag{3.5}$$

which can be estimated by (Knottnerus, 2003):

$$\widehat{Var}_{HT}[\widehat{V}_{\mathcal{A}}^{HT}] = \sum_{i=1}^{N} \sum_{j=1}^{N} s_i s_j \left(1 - \frac{\pi_i \pi_j}{\pi_{ij}} \right) \frac{\mathcal{T}_i^v}{\pi_i} \frac{\mathcal{T}_j^v}{\pi_j} \tag{3.6}$$

where π_i, π_j, and π_{ij} are the first and second-order inclusion probabilities for stem i and j and $\pi_{ij} \geq 0$. It should be noted that when $\mathcal{T}_k^\alpha \propto \pi_k \,\forall\, k \in$

$\mathcal{U}$, where α is some variable measured on stem $\mathcal{T}_k$, then $Var[\hat{V}_{\mathcal{A}}^{HT}] = 0$ (Knottnerus, 2003; Tillé, 2006).

The first-order probability of inclusion, is expressed as:

$$\pi_k = \frac{\lambda(\mathcal{A} \cap \mathcal{T}_k^p)}{\lambda(\mathcal{A})} \quad \forall k \in \mathcal{U} \tag{3.7}$$

and is normally defined in an operational environment by the intersection of $\mathcal{T}_k^p$, the non-edge corrected inclusion polygon, commonly defined as a circle of fixed radius. To correct for edge-bias, the first-order inclusion probability polygon is defined as the intersection of $\mathcal{T}_k^p$ and the boundary of $\mathcal{A}$ (Mandallaz and Ye, 1999). Here, we assume a sample can include a single sample location (i.e. $M = 1$), and so the second order inclusion probabilities are equal to zero (Gregoire, 1998).

At each sampling location $\mathcal{P}_\ell^{\mathbf{u}}$, stems are selected into P_ℓ, by the following indicator function:

$$\mathbf{I}_{k,\ell}(D, \theta_1, \theta_2) = \begin{cases} 1 & \mathcal{T}_k^d \leq D \wedge d_{k,l} \leq \theta_1 \\ 1 & \mathcal{T}_k^d > D \wedge d_{k,l} \leq F(\theta_2) \\ 0; & \text{otherwise} \end{cases} \tag{3.8}$$

where D is the breast height (1.3 m) diameter for $\mathcal{T}_k$ to be included on either a fixed radius plot of radius θ_1, or a variable radius plot with a limiting distance, expressed as a function of θ_2 for $\mathcal{T}_k$ and $p(\cdot)$ alone (i.e. $F(\theta_2) = 2500 \cdot 2 \sin \theta_2/2$ (Husch et al., 2003)). When $M = 1$, the result is a N binary vector $\mathbf{s}$ and when $M > 1$, the result is an $N \times M$ indicator matrix $\mathbf{S}$.

Now, let the probability of inclusion be $p_k = P(\mathbf{I}_{k,\ell}(D, \theta_1, \theta_2) = 1)$, which coincides with π_k when $M = 1$, a single plot (Gregoire, 1998). Since π_k is a function of $\mathcal{T}_k$ and $p(\cdot)$ alone, and not the location of plot $\mathcal{P}_\ell$, the position of $\mathcal{P}_\ell$ has no effect on π_k $\forall k \in U$ when $\mathcal{T}_k$ is close to the edge of $\mathcal{A}$ (Gregoire, 1998; Mandallaz, 2000; Mandallaz and Ye, 1999).

The design-based estimator for $V_\mathcal{A}$, presented by Horvitz and Thompson (1952), can now be expressed as:

$$\widehat{V}_\ell^{HT} = \sum_{k \in U} \frac{\mathcal{T}_k^v}{p_k} \mathbf{I}_{k,\ell}(D, \theta_1, \theta_2) \ \forall \ \ell \in \mathcal{W} \tag{3.9}$$

As noted by Gregoire (1998), using $\widehat{V}_\ell^{HT}$ to estimate the total volume in $\mathcal{A}$ at each $\mathcal{P}_\ell$, yields M *independent* estimates of $V_\mathcal{A}^{HT}$, which can then be used to estimate the total volume in $\mathcal{A}$:

$$\widehat{V}_\mathcal{A}^{HT} = M^{-1} \sum_{\ell=1}^{M} \widehat{V}_\ell^{HT} \tag{3.10}$$

Since the $\widehat{V}_\ell^{HT}$ are considered independent and identically distributed (i.e. no correlation among attributes at different sampling locations), the estimated variance for the estimated total volume, over $\mathcal{A}$, can be expressed as (Gregoire, 1998):

$$Var[\widehat{V}_\mathcal{A}^{HT}] = M^{-1} Var[\widehat{V}_\ell^{HT}] \ \ \forall \ell \in 1, \ldots, M \tag{3.11}$$

The estimator developed by Horvitz and Thompson (1952), considered the "gold-standard" estimator, is a linear combination of the data elements (i.e. $\mathcal{T}, \mathcal{P}$, and $p(\cdot)$), weighted by the inverse of their inclusion probabilities, is standard in forest sampling theory (Gregoire, 1998; Mandallaz and Ye, 1999; Cooper, 2006). The estimator can be used to examine any number of equal and unequal design-based probability sampling design functions $p(\cdot)$, and thus entire *sampling schemes* (e.g. (de Gruijter et al., 2006, see)) for sampling with or without replacement using algorithms that rely on post-sample permutations (Knottnerus, 2003) or super-population models (Tillé, 2006).

While the estimator developed by Horvitz and Thompson (1952) is not without caveats (Knottnerus, 2003; Tillé, 2006), we now present a geostatistical method to predict $V_\mathcal{A}$ from $\mathcal{P}$ assuming some correlation structure among $\mathcal{P}$, within $\mathcal{A}$, assuming $\mathcal{T}$ is the result of a complex random process (Gregoire, 1998).

3.3.2 Model-Based Prediction

Now, let $V(\alpha)$ be a *regionalized variable or function* defined by a random process on the domain $\mathcal{A}$, for point sample support, where $V(\alpha)$ produces a density (i.e. $\mathrm{m}^3\ \mathrm{ha}^{-1}$) at some point location $\alpha \in \mathcal{A}$ (Armstrong, 1998; Wackernagel, 1998; de Gruijter et al., 2006).

Given that π is a function of $\mathcal{T}$ and $p(\,\cdot\,)$ only, the local density is equal to the true density, $V(\alpha)$, by construction:

$$V_\mathcal{A} \overset{\text{Horvitz and Thompson (1952)}}{=} \sum_{k \in U} \frac{\mathcal{T}_k^v s_k}{\pi_k} \overset{\text{Mandallaz and Ye (1999)}}{=} \int_\mathcal{A} V(\alpha)\, d\alpha = \overline{V}_\mathcal{A} \cdot \lambda(\mathcal{A})$$

$$(3.12)$$

where the local density of $V(\alpha)$ at any location α, is also the estimator developed by Horvitz and Thompson (1952) and $\overline{V}_\mathcal{A}$ is the spatial mean or density (i.e. $\mathrm{m}^3\ \mathrm{ha}^{-1}$) of $V_\mathcal{A}$ over $\mathcal{A}$. This construction translates the problem of estimating $V_\mathcal{A}$ from the finite population $\mathcal{T}$, as is common in forest sampling, to estimating integrals of $V(\alpha)$ over spatial domains of $\mathcal{A}$ (i.e. stands, treatment polygons, harvest areas), which are commonly used for inputs into forest planning models (Weintraub and Murray, 2006), which can be done using model-based prediction methods such as mixed-effects models (Robinson and Wykoff, 2004) or geostatistics (Mandallaz and Ye,

1999; Mandallaz, 2000).

Geostatistics commonly uses measures that relate proximity and similarity (Wackernagel, 1998). One such measure is one half the square of the difference between two values, expressed as:

$$\gamma(i,j) = \frac{(\mathcal{P}_i^v - \mathcal{P}_j^v)^2}{2} \tag{3.13}$$

where $\mathcal{P}_i^v, \mathcal{P}_j^v$; $i,j \in \mathcal{W}$ are realizations of $V(\alpha)$ at $P_i^{\mathbf{u}}$ and $P_j^{\mathbf{u}}$ which, for all pairs of data in a sample, when plotted, forms a cloud of points known as a *variogram* (Isaaks and Srivastava, 1989; Wackernagel, 1998).

For large datasets, the *variogram cloud* can be non-informative and so, curve fitting techniques are used to fit the data to an *experimental variogram*:

$$\gamma(\mathbf{h}_k) = \frac{1}{2 \cdot n_c} \sum_{i=1}^{n_c} (P_{\alpha+\mathbf{h}}^v - P_\alpha^v)^2 \ \ \forall \, \mathbf{h} \in \mathbf{h}_k \tag{3.14}$$

where P_α^v are realizations of $V(\alpha)$ at location $\mathcal{P}^\alpha$, k is the distance that can be associated with one of the distance classes, $\mathbf{h}$, and $\mathbf{h}_k$ is a k vector of distance classes whose orientation is the same up to a given tolerance angle (Wackernagel, 1998).

Since it may not be possible to obtain sufficiently large samples to obtain adequate fits for Equation (3.14), *theoretical variograms* (Wackernagel, 1998), defined as:

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathbb{E}[(V(\alpha + \mathbf{h}) - V(\alpha))^2] \tag{3.15}$$

which require that the expectations of the first two moments (i.e. mean and covariance), expressed as:

$$\mathbb{E}[V(\alpha + \mathbf{h}) - V(\alpha)] \;\; = \;\; 0, \tag{3.16}$$

$$\mathbb{V}[V(\alpha + \mathbf{h}) - V(\alpha)] \;\; = \;\; 2\gamma(\mathbf{h}) \tag{3.17}$$

are invariant to translation which assumes that the mean of the $\mathbf{h}$, or the drift, is supposed to be zero, and invariant for any translation in $\mathcal{A}$, and that the variance has a finite value, depending on the length and orientation, but not the position of $\mathbf{h}$ within $\mathcal{A}$ (Cressie, 1985; Dimov et al., 2005).

A variety of theoretical variogram models can be fit using the classical estimator, described by Cressie (1993b):

$$\widehat{\gamma}(\mathbf{h}) = \frac{1}{2 \cdot N_h} \sum_{i=1}^{N_h} (V(P^{\mathbf{u}+\mathbf{h}}) - V(P^{\mathbf{u}}))^2 \;\; h = h(1), h(2), \ldots \tag{3.18}$$

where $V(P^{\mathbf{u}+\mathbf{h}}) = \mathcal{P}^v$ at $P^{\mathbf{u}+\mathbf{h}}$, $V(P^{\mathbf{u}}) = \mathcal{P}^v$ at $P^{\mathbf{u}}$, $\mathbf{h} = \mathcal{P}_j^{\mathbf{u}} - \mathcal{P}_i^{\mathbf{u}}$, and N_h is the number of lag differences defined by distance classes (Wackernagel, 1998). Using this general relationship, geostatistical methods can predict values of $V(\alpha)$ at an unsampled location $\alpha \in \mathcal{A}$ as well as provide an estimate of the uncertainty of the prediction at location α.

One such geostatistical prediction method, known as ordinary kriging (Isaaks and Srivastava, 1989), uses fitted theoretical variogram models is like the estimator developed by Horvitz and Thompson (1952), which is also is a linear combination, or weighted average of the data $\mathcal{P}$, and be used to predict the realization of the volume density (i.e. m^3 ha^{-1}), at some unsampled location $\mathcal{P}_\alpha$.

The prediction for the realization of $V(\alpha)$ at location α, can be expressed

as:

$$\mathcal{P}_\alpha^{\hat{v}} = \sum_{\ell=1}^{M} \omega_\ell V(\mathcal{P}_\ell^{\mathbf{u}}) \tag{3.19}$$

where $\mathcal{P}_\alpha^{\hat{v}}$ is the predicted response of $V(\alpha)$, at some unsampled location $\mathcal{P}_\alpha$, from the weights ω_ℓ and the M plots, or realizations of $V(\alpha)$, $\mathcal{P}_\ell^v$ within $\mathcal{A}$, from the sample locations $\mathcal{P}_\ell^{\mathbf{u}}$.

Ordinary kriging, like many statistical problems (e.g. ordinary least squares), is a constrained multi-criteria minimization problem (Gentle et al., 1997) that produces a predictor that is unbiased ($E[\hat{V}_\mathcal{A} - V_\mathcal{A}] = 0$) and the error or *prediction* variance ($Var[\hat{V}_\mathcal{A} - V_\mathcal{A}]$) is minimum (Armstrong, 1998).

The system of ordinary kriging equations can be expressed compactly using matrix notion:

$$\begin{bmatrix} \widehat{\mathbf{C}}_{1,1} & \dots & \widehat{\mathbf{C}}_{1,\ell} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \widehat{\mathbf{C}}_{\ell,1} & \dots & \widehat{\mathbf{C}}_{\ell,\ell} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_\ell \\ \lambda \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{C}}_{1,\alpha} \\ \vdots \\ \widehat{\mathbf{C}}_{\ell,\alpha} \\ 1 \end{bmatrix} \tag{3.20}$$

or more simply:

$$\mathbf{C}\boldsymbol{\omega} = \mathbf{D} \tag{3.21}$$

where $\boldsymbol{\omega}$ is a vector of the kriging weights and the Lagrangian multiplier, λ, which converts the constrained minimization problem into an unconstrained minimization problem (Bishop, 2006), $\mathbf{C}$ is the covariance matrix of the observations, from the relationship $\gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \mathbf{C}(\mathbf{h})$ (Wackernagel, 1998), and $\mathbf{D}$ is the vector of the covariances at the points themselves, that is, $\gamma(\mathbf{h}) = \mathbf{0}$ (Isaaks and Srivastava, 1989).

The resulting system of $M + 1$ equations can be easily solved for $\mathbf{C}^{-1}$ (Press et al., 2007), resulting in the kriging weights $\boldsymbol{\omega}$:

$$\boldsymbol{\omega} = \mathbf{C}^{-1}\mathbf{D} \qquad (3.22)$$

where the resulting values for $\boldsymbol{\omega}$ produce the unbiased estimates with the minimum prediction variance, provided the candidate model $\mathbf{C}$ is consistent with a positive definite covariance matrix ($\mathbf{C} \equiv \mathbf{x}'\mathbf{A}\mathbf{x} > \mathbf{0}; \mathbf{A}, \mathbf{x} \in \mathbb{C}^n$) (Isaaks and Srivastava, 1989; de Gruijter et al., 2006).

The predicted volume density (i.e. $\mathrm{m}^3\,\mathrm{ha}^{-1}$) at some unsampled location, $\mathcal{P}_\alpha^{\hat{v}}$, can then be calculated by:

$$\mathcal{P}_\alpha^{\hat{v}} = \sum_{\ell=1}^{M} \omega_i P_\ell^v \qquad (3.23)$$

with the associated minimized residual prediction variance, expressed as:

$$\hat{\sigma}_r^2 = \hat{\sigma}^2 - \sum_{\ell=1}^{M} \omega_\ell \mathbf{C}_{\ell,\alpha} + \lambda \qquad (3.24)$$

where $\hat{\sigma}^2$ is the "sill" or covariance value for $|\mathbf{h}| = 0$ and λ is a Lagrangian parameter.

To obtain $\widehat{V}_\mathcal{A}$ and the associated $\widehat{Var}[\widehat{V}_\mathcal{A}]$, a grid of prediction locations representing possible supply locations (i.e. stem locations, turn locations, or landing locations), are placed within $\mathcal{A}$ and the predicted local volume density, and the associated prediction variance, are computed for each location and then combined to obtain $\widehat{V}_\mathcal{A}$ and $\widehat{Var}[\widehat{V}_\mathcal{A}]$ using block kriging equations (Journel and Huijbregts, 1978; Kim and Baafi, 1984).

As is the case with the estimator developed by Horvitz and Thompson (1952), model-based prediction is also not without its caveats (Isaaks and Srivastava, 1989; Gregoire, 1998; Mandallaz, 2000). For now however, we

assume $V(\alpha)$ is well defined and there are an infinite number of possible sample and prediction locations in $\mathcal{A}$ (i.e. $\alpha \in \mathbb{R}^2$) that follow the relationship presented in Equation (3.12).

3.4 DATA AND SIMULATIONS

To compare these two methods for guessing $V_{\mathcal{A}}$, a bounded 8.24 ha polygon, denoted $\mathcal{A}$, was located on the Oregon State University McDonald-Dunn Research Forest (Figure 3.1). The boundary for $\mathcal{A}$ and the stem positions, $\mathcal{T}^{\mathbf{u}} \, \forall \, k \in \mathcal{U}$, for all stems over 15 cm, were recorded using standard survey methods (Davis et al., 1981). Attributes collected for each $\mathcal{T}_k$ included the species, $\mathcal{T}_k^s$, the diameter at breast height $\mathcal{T}_k^d$, the total height, $\mathcal{T}_k^h$, the height to live crown $\mathcal{T}_k^c$, the number of pieces in the main stem, $\mathcal{T}_k^n$, and surface characteristics for each piece $\mathcal{T}_k^q \, \forall \, q \in (1, \cdots, \mathcal{T}_k^n)$. We assume measurement error is non-existent.

A script was created to determine the volume for each stem, $\mathcal{T}_k^v$, where all stems were merchandised into log lengths of 4 m, to a 2 cm top, using the taper equation presented by Kozak et al. (1968). If the stem could not be cut into a round number of standard log lengths, any remaining stem length was bucked into a short log no less than 2 m. Stems were merchandised with a stump height of 0.3 m and each log included 0.2 m of trim. For each log in the stem, the starting and ending height of the log, the nominal length and the actual length (nominal length plus trim), the small and large end diameters, and the Smalian volume was recorded. For this examination, the Smalian volume from each of the logs were then tallied and assigned to $\mathcal{T}_k^v$.

Only Douglas-fir (*psudeotsuga mensezii* (Mirb. Franco)) stems were used resulting in a final count of 2053 stems with a spatial mean density $\overline{V}_{\mathcal{A}}$ of

249.17 stems ha^{-1}. The total volume for the study area $V_{\mathcal{A}}$, was 7560.07 m^3 yielding an average stem volume of 3.68 m^3 and an mean volume density of 917.57 m^3 ha^{-1}.

The N vector of first-order inclusion probabilities, π, were computed using the "tree-concentric method" (Schreuder et al., 1993) for a single sample design, $p(60, 10, 10)$. This design means that a "plot design" is placed at the sample location so that stems less than 60 cm are included in a fixed-area plot, with a radius of 10 m and stems over 60 cm are sampled using a basal area factor equal to 10 m^{2}ha^{-1}(i.e. $\theta_2 = 0.030303°$ (Husch et al., 2003)). To correct for edge-bias, the polygon defining the probability of detection for $\mathcal{T}_k$, $\mathcal{T}_k^p$, was intersected with the edge of $\mathcal{A}$, as described by Beers (1966), to obtain the edge-corrected weights for the observations.

Once the sample locations were established, stems were selected for inclusion into the sample using Equation (3.8), and both the corrected and uncorrected expansion factors (i.e. $1/\pi$) were recorded and the results archived.

The theoretical spatial correlation models were defined by either a one-parameter isotropic linear model (LIN):

$$\gamma(\mathbf{h}) = \begin{cases} \frac{h}{a} & 0 \leq h \leq a \\ 1 & h > a \end{cases} \tag{3.25}$$

a one parameter isotropic spherical (SPH) variogram model:

$$\gamma(\mathbf{h}) = \begin{cases} \frac{3}{2}\frac{|\mathbf{h}|}{a} - \frac{1}{2}\frac{|\mathbf{h}|^3}{a} & |\mathbf{h}| < a \\ 0 & |\mathbf{h}| \geq a \end{cases} \tag{3.26}$$

or an isotropic one-parameter exponential (EXP) variogram model:

$$\gamma(\mathbf{h}) = 1 - \exp\left(\frac{-|\mathbf{h}|}{a}\right) \tag{3.27}$$

where $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$ (Goovaerts, 1997), and a represents an unknown,

Random

Regular Systematic

Non–aligned Systematic

Hexagonal Systematic
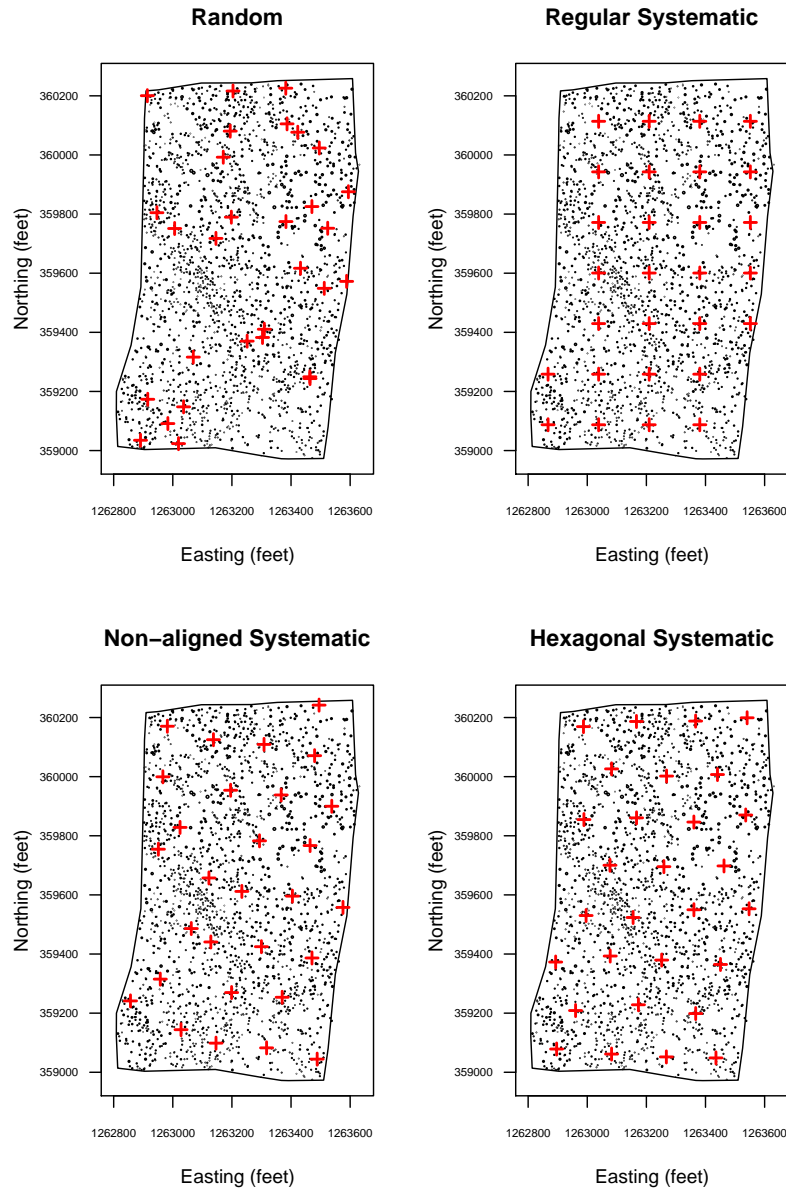
Figure. 3.1: Plots of $\mathcal{T}^u_{k\in\mathcal{U}}$ and the plot densities for 1 replication of sample locations for random (RND), systematic (REG), non-aligned systematic (NON), and hexagonal systematic (HEX) samples from $\mathcal{A}$.

but fixed population parameter specific to each model (Cressie, 1993a). For simplification, we did not record the fit statistics for the estimated variogram parameters.

These particular models were used as they represent a cross-section of the possible peculiarities encountered when fitting variogram models such as might arise in a production sampling environment (i.e. exponential is non-linear in the parameters but differentiable everywhere, the spherical is non-linear and non-differentiable for some parameters, and the linear increases without bound, and so no covariogram exists) (Cressie, 1985).

For each replication, the theoretical variogram models were fit with a lag distance class width of 100 m from zero to 800 m and the range was fixed at 800 m as it was assumed that there would be no correlation among plot volumes past 800 m and to ensure the fitting algorithm converged. As suggested by Cressie (1993a), the variogram models were fit using weighted least squares where the weights for the observations were $\frac{N_j}{h_j^2}$ (Pebesma, 2004).

For the model-based predictions, a 12 m$\times$20 m grid of prediction locations (i.e. possible production locations), was used to compute $V_{\mathcal{A}}^{OK,m,p}$. The average value of $V_{\mathcal{A}}^{OK,m,p}$, denoted $\overline{V}_{\mathcal{A}}^{OK,m,p}$, was obtained using block-kriging (Goovaerts, 1997), by averaging the variogram from each sample location in $\mathcal{A}$ to estimate the coefficients of a point-to-block ordinary kriging system (Montes et al., 2005). Once the value of $\overline{V}_{\mathcal{A}}^{OK,m,p}$ was obtained, $\widehat{V}_{\mathcal{A}}^{OK,m,p}$ was computed as the product of $\lambda(\mathcal{A}) \cdot \overline{V}_{\mathcal{A}}^{OK,m,p}$. The resulting block-kriging prediction variance was then also used to compute the standard error of the predicted total as $\sqrt{\lambda(\mathcal{A})^2 \cdot \widehat{Var}[\widehat{V}_{\mathcal{A}}^{OK,m,p}]}$.

We generated 1000 samples and computed $\widehat{V}_{\mathcal{A}}^{i,m,p}$, the bias ($\beta^{i,m,p} = \widehat{V}_{\mathcal{A}}^{i,m,p} - V_{\mathcal{A}}$), the standard error ($\sqrt{\widehat{Var}[\widehat{V}_{\mathcal{A}}^{i,m,p}]}$), the mean squared error ($MSE^{i,m,p} = \frac{\sum \beta^2}{n}$), and root mean squared error ($RMSE^{i,m,p} = \sqrt{MSE^{i,m,p}}$)

where n is the number of observations in each combination of sample pattern, for each combination of $i \in \{HT, OK\}$, $m \in \{\emptyset, LIN, SPH, EXP\}$, and $p \in \{RND, REG, NON, HEX\}$ for $n \in \{10, 15, 20, 25, 30, 35, 40, 45, 50\}$ sample locations within $\mathcal{A}$ (see Figure 3.1).

All calculations were performed using R (Ihaka and Gentleman, 1996) with the sample locations obtained using the `spsample` function in the `sp` package (Pebesma and Bivand, 2005), and the variogram estimation and kriging prediction functions in the `gstat` package (Pebesma, 2004). Finally, box-and-whisker plots of the bias and standard errors were generated using the `lattice` package (Sarkar, 2008).

3.5 RESULTS AND DISCUSSION

To examine the results, we first visually inspected box-and-whisker plots of the bias and standard error, for $\widehat{V}_{\mathcal{A}}^{i,m,p}$, and then plotted RMSE. Figures 3.2 and 3.3 display the results of the simulations where the design-based estimator $\widehat{V}_{\mathcal{A}}^{HT,\emptyset,p}$ is displayed in the first column and the $\widehat{V}_{\mathcal{A}}^{OK,EXP,p}$, $\widehat{V}_{\mathcal{A}}^{OK,SPH,p}$, and $\widehat{V}_{\mathcal{A}}^{OK,LIN,p}$ model-based predictions are displayed in the second, third and fourth columns, respectively. The rows correspond to the sample location placement patterns for the random, hexagonal, non-aligned grid, and the regularly spaced grid patterns, respectively.

For many of the replications, the actual number of plots installed in $\mathcal{A}$ did not match the target sample size as a result of the randomization process, and so the number of samples in each target sample size class is different. This produced plots with a large number samples that deviated slightly from the target sample sizes. For brevity, phenomenon where the target sample size matched the actual sample size will be discussed. It should be noted the
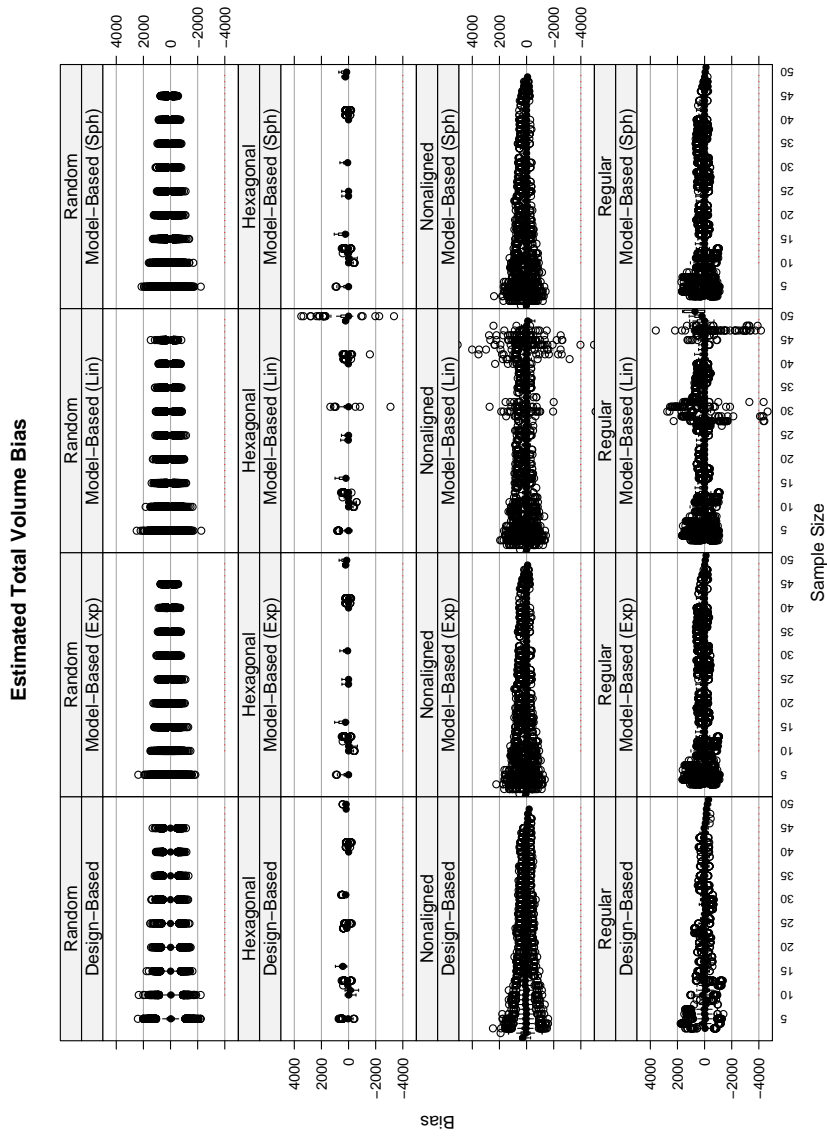
Figure. 3.2: Box-and-Whisker plot of the bias for $\widehat{V}_A$ for design-based and model-based methods using four different sample location patterns (random, hexagonal, non-aligned, and regular grids) and four different spatial correlation models (none, exponential, linear, and spherical variograms). The dashed line placed at the bottom of each panel, represents the range of the target sample sizes.

70
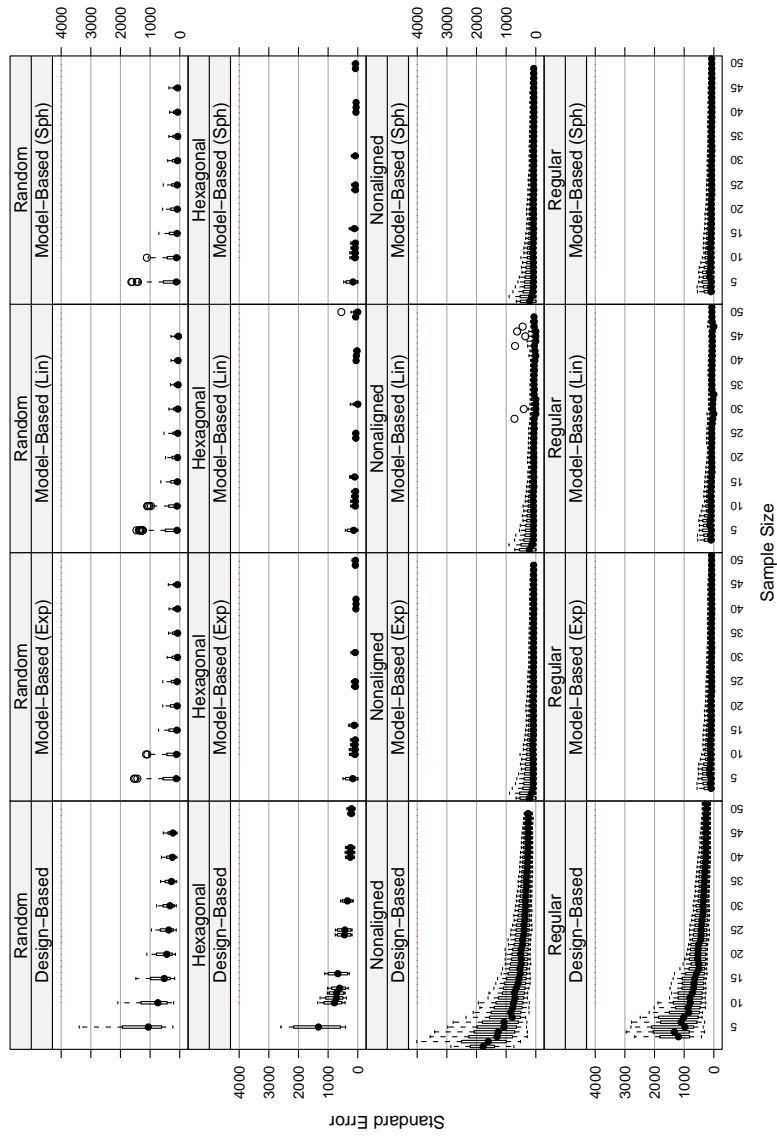
Figure. 3.3: Box-and-Whisker plot of the calculated standard error for $\widehat{V}_A$ for design-based and model-based methods using four different sample location patterns (random, hexagonal, non-aligned, and regular grids) and four different spatial correlation models (none, exponential, linear, and spherical variograms). The dashed line placed at the top of each panel, represents the range of the target sample sizes.

pattern did not change with the removal of these data, or if the corrected and uncorrected inclusion probabilities were segregated, and so here, the discussion is limited to visual inspection of the patterns, regardless of edge-correction. Discussion of edge-correction is presented in Section 3.5.3.

3.5.1 Design-Based Estimates

Distributions of $\widehat{V}_{\mathcal{A}}^{i,m,p}$ and $SE[\widehat{V}_{\mathcal{A}}^{i,m,p}]$ for the design-based estimates were consistent with asymptotic theory for increasing sample sizes in that the range in variation for the estimates decreased with increasing sample size as is evident in the first column in Figures 3.2 and 3.3. The randomly located design-based estimator bias distribution was wider than the non-aligned and regularly spaced estimates. Both the non-aligned and regularly spaced sample locations yielded a slight negative bias in the largest sample sizes. The hexagonal sample pattern, yielded consistent distributions over the limited range of obtained sample sizes, but did not decline as sample size increased. The performance of the non-aligned seemed superior as there were fewer "bumps" between the edge-corrected and non-corrected results and clearly fewer deviations from a smooth curve than the hexagonal and regularly spaced sample locations (e.g. current practice).

3.5.2 Model-Based Predictions

Distributions of $\widehat{V}_{\mathcal{A}}^{i,m,p}$ and $SE[\widehat{V}_{\mathcal{A}}^{i,m,p}]$ for the model-based estimates were also consistent with asymptotic theory for increasing sample sizes in that the range in variation for the estimates decreased with increasing sample size as

is evident in the second through fourth columns in Figures 3.2 and 3.3.

Interestingly, there were larger deviations in the distribution of the bias for the linear variogram model for the hexagonal, non-aligned, or regular sampling patterns. The results are clearly visible around 25 and 30 sample locations and again between 45 and 50 sample locations. This pattern is not present in the other (i.e. exponential and spherical) variogram models.

The computed standard errors for model-based predictions, at each sample size, was notably smaller than the standard errors of design-based estimates for all but the completely random sampling patterns. This supports the notion that with the inclusion of a simple one-parameter linear, exponential, or spherical variogram model, the size of the standard error can be considerably less than the traditional design-based estimator. By examination of Figures 3.2 and 3.3, the practitioner can ascertain that the number of plots needed to achieve the same results is considerably less than the number of plots required for design-based sampling.

3.5.3   Mean Square Error

To more thoroughly examine the performance of each method, we plotted the RMSE $\widehat{V}_{\mathcal{A}}^{i,m,p}$, for both the edge-corrected (solid line) and uncorrected (dashed-line) results (Figure 3.4). The patterns in the RMSE for both were similar; the non edge-corrected results showing worse (i.e. higher RMSE) values than the edge-correct results, regardless of the assumed spatial correlation model and sampling pattern.

The pattern for the design-based results are commensurate with sample theory in that for the randomly located sample types, the RMSE decreases as the sample size increases for all four sample location patterns. The patterns
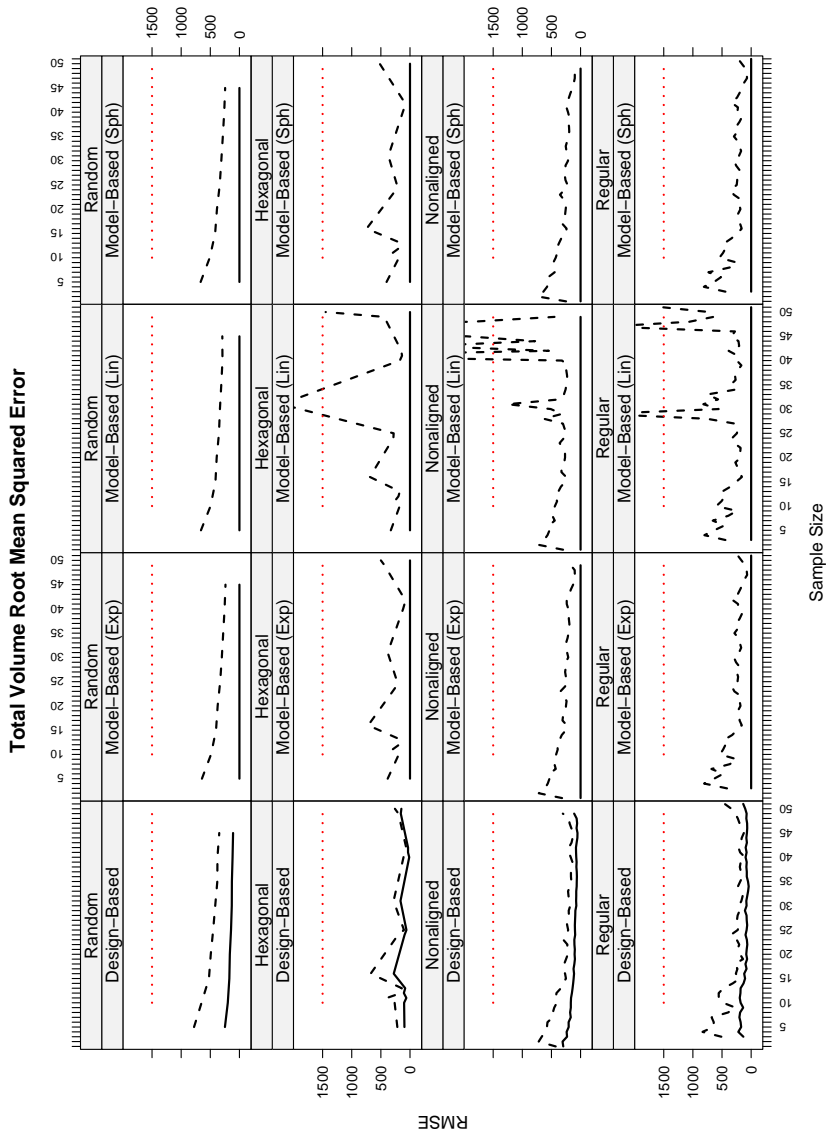
Figure. 3.4: Root Mean Squared Error for the estimation (design-based) and predictions (model-based) of the V_A for four different sample location patterns and three different covariance models (exponential, linear, and spherical variograms) for $\mathcal{A}$. The dashed, solid, and dotted lines represent the non-edge corrected results, the results with edge-correction, and the range of the targeted sample sizes, M, respectively.

in the uncorrected results are more pronounced in that the smoothness of the curve is less for non-random sample location patterns.

Interestingly, the pattern between design and model-based inference is similar regardless of the variogram model in that the bias is negligible across the range of the target sample sizes. While the large deviations in the linear variogram model-based inference, the pattern is prevalent across all sample location patterns, with the exception of the randomly placed samples. In short, performance of the non-edge corrected RMSE roughly the same between design-based and model-based methods, with edge-corrected model-based predictors notably superior to the design-based estimator.

3.6   CONCLUSIONS

For this study, we simulated operational field sampling and examined the consequences of using classical design-based estimations and model-based predictions, for a variety of sample patterns, sampling intensities, and spatial models, in order to guess the total volume in bounded area $\mathcal{A}$, as is often the case with pre-harvest survey samples. Here, the only knowledge available to the sampler was the boundary of $\mathcal{A}$ itself, and so we tested the effects of sampling using various assumptions of the spatial correlation structure, as would be the case in operational sampling environments.

While the general accuracy of model-based predictions depends on a number factors such as 1) the number of sample locations, 2) the quality of the data at each sample location, 3) the positions of the samples within $\mathcal{A}$, 4) the distances between the sample locations and the position of interest, and 5) the spatial continuity of the target parameter, we used a naive approach (e.g. no *a priori* information, no transformations, no covariates, no multi-stage sam-

ple schemes, and non-informative sampling) and found that the model-based predictions were superior, to the design-based estimates for operational environments. The technique provides a method to predict the volume density at unsampled locations within $\mathcal{A}$, without having to make inference about a larger population with similar characteristics, which is typically the goal in forest operations.

From our results, sampling densities, and thus the positions of sample locations, can be critical in obtaining predictions for operational situations as was apparent in the larger sample sizes for the linear variogram model (i.e. larger biases and RMSE values), and should model-based predictions be used for precision production planning, maintaining sufficient diversity among the distances between sample locations is key to obtaining estimates of the variogram model used for prediction, regardless of the model selected (i.e. linear, spherical, or exponential). In short, non-aligned and regular sampling patterns, provided the best benefits (anti-conservative confidence intervals and better sample size flexibility) regardless of inference method or variogram model.

Finally, additional information can be incorporated into the sampling process, more sophisticated variogram models can be examined, covariates can be included, and the trade-offs between sample precision, sample information content, and sampling costs can be explored, thus generating further enhancements to the sampling and prediction process, the outputs of which are ultimately used to in operational, tactical, and strategic planning and optimization models.

BIBLIOGRAPHY

Armstrong, M. (1998). *Basic Linear Geostatistics.* Springer.

Avery, T. E. and Burkhart, H. E. (2003). *Forest measurements, 5th Ed.* McGraw-Hill.

Beers, T. W. (1966). The direct correction for boundary-line slopover in horizontal point sampling. Research Progress Report 224, Purdue University, Agricultural Experiment Station, Lafayette, Indiana.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Chapman, R. C. (1982). Plot spacing in systematic sampling. *Journal of Forestry*, 80:409–419.

Clark, M. M., Meller, R. D., McDonald, T. P., and Ting, C. C. (2000). A new harvest operation cost model to evaluate forest harvest layout alternatives. *Annals of Operations Research*, 95(1):115–129.

Cochran, W. G. (1977). *Sampling techniques.* John Wiley & Sons, New York. 428 p.

Connaughton, K. P. and Fight, R. D. (1984). Applying Trade-Off Analaysis to National Forest Planning. *Journal of Forestry*, 82(11):680–683.

Cooper, C. (2006). Sampling and variance estimation on continous domains. *Environmetrics*, 17:539–553.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17(5):563–586.

Cressie, N. A. C. (1993a). *Statistics for Spatial Data*, chapter 2. John Wiley and Sons, Inc., New York, New York, revised edition edition. 900.

Cressie, N. A. C. (1993b). *Statistics for Spatial Data.* John Wiley and Sons, Inc., New York, New York, revised edition edition.

Davis, R. E., Foote, F. S., Anderson, J. M., and Mikhail, E. M. (1981). *Surveying, theory and practice.* McGraw-Hill, Inc.

de Gruijter, J. J., Brus, D. J., Bierkins, M. F. P., and Knotters, M. (2006). *Sampling for Natural Resource Monitoring.* Springer-Verlag.

Dimov, L. D., Chambers, J. L., and Lockhart, B. R. (2005). Spatial continuity of tree attributes in bottomland hardwood forests in the southeastern united states. *Forest Science*, 51(6):532–540.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2006). A spatial approach to forest-management optimization: linking gis and multiple objective genetic algorithms. *International Journal of Geographical Information Science*, 20(8):917–928.

Gambill, C. W., Harry V. Wiant, J., and Yandle, D. O. (1985). Optimum plot size and baf. *Forest Science*, 31(3):587–594.

Gentle, J. E., Narula, S. C., and Valliant, R. L. (1997). Multicriteria optimization in sampling design. In Ghosh, S., Schucany, W. R., and Smith, W. B., editors, *Statistics of Quality*, volume 153 of *Statistics, textbooks, and monographs*, pages 411–425. Marcel Dekker, Inc.

Gjedtjernet, A. M. F. (1995). Forest operations and environmental protection. *Water, Air, & Soil Pollution*, 82(1):35–41.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation.* Applied Geostatistics Series. Oxford University Press, New York, New York.

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28:1429–1447.

Hamann, J. D. and Boston, K. (2007). Using spatial statistics to improve the primary forestry supply chain. In Reynolds, K. M., Thompson, A., Shannon, M., Köhl, M., Ray, D., and Rennolls, K., editors, *Sustainable Forestry: from Monitoring and Modelling to Knowledge Management & Policy Science*, chapter 18, pages 285–297. CABI International.

Hock, B., Payn, T., and Shirley, J. (1993). Using a geographic information system and geostatistics to estimate site index of *Pinus radiata* for kaingaroa forest. *New Zealnad Journal of Forest Science*, 23:264–277.

Holmgren, P. and Thuresson, T. (1997). Applying Objectively Estimated and Spatially Continuous Forest Parameters in Tactical PLanning to Obtain Dynamic Treatment Units. *Forest Science*, 43(3):317–326.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statisical Assocaition*, 47.

Husch, B., Beers, T., and Kershaw Jr., J. (2003). *Forest Mensuration*. John Wiley & Sons, Inc., Hoboken. 443 p.

Ihaka, R. and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, Inc., New York, New York.

Journel, A. and Huijbregts, C. J. (1978). *Mining Geostatistics*, chapter 5, page 600. Academic Press, Inc., London.

Karlsson, J., Ronnqvist, M., and Bergstrom, J. (2003). Short-term harvest planning including scheduling of harvest crews. *International Transactions in Operational Research*, 10(19):413–431.

Kim, Y. C. and Baafi, E. Y. (1984). Combining local kirging variances for short-term mine planning. In Verly, G., David, M., Journel, A., and Marechal, A., editors, *Geostistics for natural resources characterization*, volume 122 of *Series C: Mathematical and Physical Sciences*, pages 185–199, P.O. Box 17, 3300 AA Dorrecht, Holland. NATO Advanced Study Institute on Geostistics for natural resources characterization, Reidel Publishing Company.

King, S. L. (2000). Sequential gaussian simulation vs. simulated annealing for locating pockets of high-value commercial trees in pennsylvania. *Annals of Operations Research*, 95(1-4):117–203.

Knottnerus, P. (2003). *Sample Survey Theory: some Pythagorean perspectives*. Springer-Verlag.

Kozak, A., Munro, D. D., and Smith, J. (1968). Taper functions and thier application in forest inventory. *Forestry Chronical*, 45:278–283.

Lohr, S. L. (1999). *Sampling: Design and Analysis*, chapter 8. Brooks/Cole Publishing Company.

Mandallaz, D. (2000). Estimation of the spatial covariance in Universal Kriging: Application to forest inventory. *Environmental and Ecological Statistics*, 7:263–284.

Mandallaz, D. and Ye, R. (1999). Forest inventory with optimal two-phase, two-stage sampling schemes based on the anticipated variance. *Candian Journal of Forest Research*, 29:1691–1708.

Mawson, J. C. and Mack, R. J. (1982). Inventory costs on small forests. *Journal of Forestry*, 80(3):165–167.

Montes, F., Hernández, M. J., and Canellas, I. (2005). A geostastical approach to cork production sampling estimation in *Quercus suber* forests. *Canadian Journal of Forest Research*, 35:2787–2796.

Oderwald, R. G. and Jones, E. (1992). Sample sizes for point, double sampling. *Canadian Journal of Forest Research*, 22:980–983.

Pebesma, E. J. (2004). Multivariable geostatistics in s: the gstat package. *Computers & Geosciences*, 30:683–691.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK) and New York, 3rd edition.

Reeves, L. H. and Haight, R. G. (2000). Timber harvest scheduling with price uncertainty using markowitz portfolio optimization. *Annals of Operations Research*, 95:229–250.

Riopel, D., Langevin, A., and Campbell, J. F. (2005). The network of logistics decisions. In Langevin, A. and Riopel, D., editors, *Logistical Systems:Design and Optimization*, number 2 in GERAD, pages 1–38. Springer, New York, New York.

Robinson, A. P. and Wykoff, W. R. (2004). Imputing missing height measures using a mixed-effects modeling strategy. *Canadian Journal of Forest Research*, 34:2492–2500.

Samra, J., Gill, H., and Bhatia, V. (1989). Spatial stochastic modelling of growth and forest resource evaluation. *Forest Science*, 35:663–676.

Sarkar, D. (2008). *lattice: Lattice Graphics*. R package version 0.17-4.

Schreuder, H. T., Gregoire, T. G., and Wood, G. B. (1993). *Sampling methods for multiresource forest inventory*. John Wiley & Sons, Inc. 446 p.

Tillé, Y. (2006). *Sampling Algorithms*. Springer.

Vera, J. R., Weintraub, A., Koenig, M., Bravo, G., Guignard, M., and Barahona, F. (2003). A lagrangian relaxation approach for a machinery location problem in forest harvesting. *Pesquisa Operacional*, 23(1):111–128.

Wackernagel, H. (1998). *Multivariate Geostatistics*, chapter 1. Springer, Berlin, Germany, 2nd edition.

Weintraub, A. and Murray, A. T. (2006). Review of combinatorial problems induced by spatial forest harvesting planning. *Discrete Appl. Math.*, 154(5):867–879.

4. OPTIMAL SPATIAL SAMPLING TO MAXIMIZE SAMPLE INFORMATION CONTENT FOR USE IN FOREST SUPPLY CHAIN PLANNING AND OPTIMIZATION

4.1 ABSTRACT

We determined the Pareto frontier between the information content and the sampling cost by combining the constrained maximum entropy sampling problem (CMESP) Lee (1998) with the classic traveling salesman problem (Lawler et al., 1985) for a discrete random spatial field. To solve the bi-objective problem (i.e. maximize sample information content, minimize sample cost) we used a multi-objective evolutionary algorithm, presented by Knowles and Corne (1999), to solve the general d-objective non-linear mixed integer problem.

Our results obtain the set of best possible trade-off curves (i.e. Pareto frontier) for a population of 36 general "split-plot" sample designs that stratifies a finite population of stems, using a critical diameter, to be selected into a fixed-area plot and a variable radius plot. Our results agree somewhat with those of Zeide (1980), in that larger plots farther apart yield higher information/cost ratios than do smaller, more densely spaced plots.

The resulting Pareto curves, which are equivalent to the D_n-optimal experimental design problem for a finite cost, can then be used to examine a variety of questions (i.e. experiments) about the population of stems, design-based inference, model-based inference, simulation, and optimization problems within a multi-criteria/multi-objective decision making framework so that the uncertainty associated with the unsampled locations is minimized.

4.2   INTRODUCTION

In forest operations, which typically focuses on small spatial and temporal scales (e.g. time-motion studies, daily/weekly crew-assignments, tactical, and operational planning), many well established studies examine the trade-off between sampling costs and plot size (Zeide, 1980), basal area factor (BAF) (Gambill et al., 1985), plot density (Chapman, 1982) for simple and complex (Oderwald and Jones, 1992; Mandallaz and Ye, 1999) sampling *schemes* (de Gruijter et al., 2006) for both design-based estimation and model-based prediction where the goal is to guess a target parameter (i.e. total volume) within a finite region (i.e. harvest or treatment polygon) (Gregoire, 1998).

Conversely, in the environmental and hydrological monitoring literature, where the focus is to minimize data acquisition costs at larger scales (e.g. geographic regions, watersheds, or basins), emphasis is often placed on installing (Bueso et al., 1998), expanding (Le et al., 1993), or contracting (Wu and Zidek, 1992) networks of monitoring stations (i.e. populations of sampling locations), for single and multiple objectives (Trujillo-Ventura and Ellis, 1991), often assuming spatial correlation information (Shewry and Wynn, 1987), where the goal is to maintain or maximize the information content (i.e. entropy) from the selected sample locations, as the cost of traveling to, installing, maintaining, and collecting data at monitoring stations is minimized (Caselton and Hussain, 1980; Lee, 1998).

Entropy, a measure of the uncertainty contained in the variables (Le and Zidek, 2006), is a general approach to sampling which is based on information theory (Shannon, 1948). The measure, originally formalized in statistical mechanics by Boltzmann, and made popular in statistics by Blackwell (1951),

has been applied to a wide variety of disciplines including computational biology (Abbas and Holmes, 2004), data mining (Kononenko and Kukar, 2007), econometrics (Theil, 1971), experimental design (Lee, 1998), hydrological monitoring networks (Caselton and Hussain, 1980), machine learning (Mansour and Schain, 1997), probability sampling (Chen et al., 1994), sampling for single (Bueso et al., 1998) and multiple (Bueso et al., 1999) variables and observations, species diversity assessment (Gill and Joanes, 1979), and convergence criteria for single and multiple objective optimization (Farhang-Mehr and Azarm, 2002).

We use entropy for the following reason: In the collection of forest samples, in which the configuration of the sample location network and data observed can be considered part of the supply chain "information and communications channel" (Caselton and Hussain, 1980), there are often multiple, yet implicit, purposes for collection of field samples were the optimization for one objective may be suboptimal for another (Gentle et al., 1997). Often, forest samples originally used for timber volume assessment (Zeide, 1980), are then used for environmental monitoring (Le et al., 1993), or potentially used for growth and yield modeling (Stage, 1973), or for the prediction of ancillary and possibly spatially correlated variables to estimate product abundance over a finite region (Montes et al., 2005). Here, we assume the original purpose of the survey is to predict the total volume V_A from a single observation (i.e. survey), and that additional information (e.g. correlation structure or ancillary variables such as production estimates) will be used for other purposes such as ecosystem assessment (Stevens, Jr., 1997), activity planning (Hamann and Boston, 2007), simulation (Davis et al., 2001), or for the optimization of multiple outputs (Bettinger et al., 1999), as is often the case in operational environments.

Rather than examine the curves between the between plot travel-time on a rectangular lattice and the resulting probability sampling variance, as is traditionally done in the forest sampling literature (Zeide, 1980; Chapman, 1982; Gambill et al., 1985), we directly examine the Pareto frontier between the maximum average information content (i.e. entropy), for sampling to obtain a guess of $V_{\mathcal{A}}$ using probability sampling, at the minimum cost required to obtain the sample, for a general sample design, $p(\,\cdot\,)$, as is common in environmental sampling (i.e. Shewry and Wynn (1987)). While the primary objective to obtain the set of possible samples that are optimal for predicting $V_{\mathcal{A}}$, regardless of our inference (i.e. design-based versus model-based inference (Gregoire, 1998)), we also wish to use the resulting sample to make predictions (i.e. inputs) for simulation, optimization, and we wish to minimize our uncertainty associated with those results as well.

Assuming we have complete information about the formulation of the optimization problem formulation (i.e. exactly known values for inputs and exactly known values for outputs). This method facilitates the examination of any number of potential models used to determine the status of $\mathcal{A}$ for a variety of purposes. Our purpose is to obtain the samples that provide the maximum information content, for the least cost, when the goal is both to obtain precise samples, but also to formulate and solve non-linear constrained possibly stochastic d-objective $(d \geq 1)$ minimization problems (Marti, 2005) using the resulting estimates or predictions of the total volume in $\mathcal{A}$, denoted $V_{\mathcal{A}}$.

We explain our rational using definitions that are consistent with classical design-based sampling (Cochran, 1977; Tillé, 2006), information theory (Shannon, 1948), combinatorial optimization (Lawler et al., 1985), and graph theory (Diestel, 2005). In Section 4.3, we present the notation, the objec-

tives, and the complete formulation. Section 4.5 presents a description of the population of stems, sample locations (i.e. plots) and the population of sample designs which are common in pre-harvest surveys. A description of the heuristic used to determine the determine the Pareto frontier is then presented in Section 4.4, followed by the results and a discussion in Section 4.6. Finally, we present our conclusions in Section 4.7.

4.3 PROBLEM FORMULATION

Let $\mathcal{A} \subset \mathbb{R}^2$, represent an operational harvest polygon with a finite boundary, where the surface area, measured in hectares, is denoted $\lambda(\mathcal{A})$. The polygon $\mathcal{A}$ contains a finite population of N stems denoted by the set $\mathcal{T}$. Each stem $\mathcal{T}_k$ is defined by a unique label and $\mathcal{U} = \{1, \ldots, k, \ldots, N\}$ is the set of labels that is used to uniquely identify each stem. The attributes for each stem includes a position $\mathcal{T}_k^{\mathbf{u}} \in \mathbb{R}^2$, a diameter $\mathcal{T}_k^d$, and volume $\mathcal{T}_k^v$. To reduce the notation, we shall write $\mathcal{T} \in \mathcal{A}$ to represent all stems in $\mathcal{A}$.

Let $\mathcal{P}$ represent a set of M potential measurement locations placed within $\mathcal{A}$, where each measurement location (i.e. point/plot) $\mathcal{P}_\ell$ is defined by a unique label and $\mathcal{W} = \{1, \ldots, \ell, \ldots, M\}$ is the set of labels that are used to uniquely identify each sampling location. For this examination, each sampling location contains a fixed position vector $\mathcal{P}_\ell^{\mathbf{u}} \in \mathbb{R}^2$ and a volume density observation $\mathcal{P}_\ell^v$ (i.e. m^3 ha^{-1}) that is measured without error. Again, to reduce the notation, we shall write $\ell \in \mathcal{W}$ or $\mathcal{A}_{\ell \in \mathcal{W}}$ to represent all sampling locations in $\mathcal{A}$.

The sampling unit is the sample location (i.e. plot) $\mathcal{P}_\ell$. All sample units that are observed, taken collectively, are referred to as a sample, and the sample size is defined as the number of sample locations visited during the

survey. A general sample design function $p(\,\cdot\,)$ (i.e. plot design), assigns an inclusion probability to each stem $\mathcal{T}_k$. The function $p(\,\cdot\,)$ is typically used to select elements of $\mathcal{T}$ into the sample, at sample location ℓ, using an indicator function (Gregoire, 1998; Mandallaz and Ye, 1999).

The decision variable, design space, or sample space, is represented by the vector $\mathbf{s} = (s_1, \ldots, s_\ell, \ldots, s_M) \in \{0,1\}^M \; \forall \; \ell \in M$, where $s_\ell = 1$ if sample location ℓ is included in the sample, and $s_\ell = 0$ if it is not (Knottnerus, 2003). An empty sample is denoted by $\mathbf{s} = (0, \ldots, 0, \ldots, 0)$ and a census of all M possible sample locations is defined by $\mathbf{s} = (1, \ldots, 1, \ldots, 1)$ (Tillé, 2006). The total number of plots included in the sample (i.e. sample size) is $\sum_{\ell=1}^{M} s_\ell = n$, which is also known as a cardinality constraint (Ko et al., 1995; Tillé, 2006). The cardinality of the sample space, $\mathcal{S}$ thus the total number of possible samples, regardless of any cardinality constraints (i.e. sample size constraints) can be expressed as $2^M = \sum_{i=1}^{M} \binom{M}{i}$.

While the main goal is to obtain the Pareto frontiers for the maximum information content versus cost, the initial purpose (i.e. first objective) of a forest sample is often to obtain the *best* guess of the target parameter which, in this case, is the fixed, but unknown total volume (i.e. m^3) in $\mathcal{A}$, denoted by $V_\mathcal{A}$, expressed as the function of interest:

$$V_\mathcal{A} = f(\mathcal{T}_1^v, \ldots, \mathcal{T}_k^v, \ldots, \mathcal{T}_N^v) = \sum_{k \in U} \mathcal{T}_k^v \qquad (4.1)$$

with the spatial mean (i.e. m^{3}ha^{-1}) over $\mathcal{A}$, denoted by $\overline{V}_\mathcal{A}$, defined as:

$$\overline{V}_\mathcal{A} = \lambda(\mathcal{A})^{-1} f(\mathcal{T}_1^v, \ldots, \mathcal{T}_k^v, \ldots, \mathcal{T}_N^v) = \lambda(\mathcal{A})^{-1} \sum_{k \in U} \mathcal{T}_k^v = \lambda(\mathcal{A})^{-1} \cdot V_\mathcal{A} \quad (4.2)$$

and the variance of the total volume expressed as:

$$\sigma^2_{V_{\mathcal{A}}} = \frac{1}{N} \sum_{k \in U} (\mathcal{T}^v_k - \overline{V}_{\mathcal{A}})^2 = \frac{1}{2N^2} \sum_{k \in U} \sum_{\ell \in U} (\mathcal{T}^v_k - \mathcal{T}^v_\ell)^2 \qquad (4.3)$$

which can be estimated (i.e. $\widehat{V}_{\mathcal{A}}$ and $\widehat{Var}[\widehat{V}_{\mathcal{A}}]$) by a set of n sample locations placed within $\mathcal{A}$ (Gregoire, 1998; Mandallaz and Ye, 1999)).

The scope of inference, and thus the understanding of the current conditions within $\mathcal{A}$, for a single sample in time, is a function of the unique combination of $\mathcal{T}$, $\mathcal{P}$, and $p(\cdot)$ (i.e. triplet) (Knottnerus, 2003; Tillé, 2006; Gregoire, 1998).

Ultimately, our goal is to determine which sample locations within $\mathcal{A}$, yield the highest average information content (i.e. the most *valuable*) using the information content metric, entropy, for a given sampling cost, which represents the willingness to pay for the information. In other words, before we may harvest $\mathcal{A}$ using the optimal "plan", thus optimizing our production objectives, we first need to optimize our harvest of the potential information from $\mathcal{A}$ where our choices to detecting phenomenon within $\mathcal{A}$ are limited (i.e. finite population of sample design functions, $p(\cdot)$) as is the case in typical operations where sampling budgets are often arbitrarily constrained based on the perceived value of $\mathcal{A}$ by the decision maker.

4.3.1 First Objective: Constrained Maximum Entropy Sampling Problem

The maximum entropy sampling problem (MESP) (Ko et al., 1995), which is also equivalent to the D_n-*optimal experimental design* problem (Boyd and Vandenberghe, 2004), maximizes the average information content, or minimizes the uncertainty of a discrete spatial random field $V_{\mathcal{A}}$. The problem is

as follows: Given a set $\mathcal{P}$ of M sample locations, called the design space, a set $\mathcal{F} \in \mathcal{P}$ of f "forced" sample locations, and a design size n, such that $f \leq n \leq m$, the objective is to choose a set $\mathcal{S}$ of n sample locations satisfying $\mathcal{F} \in \mathcal{S} \in \mathcal{P}$, called the *feasible sample*, such that observations taken at these sample locations will be as *valuable* as possible and minimize the uncertainty associated with the unsampled locations (Ko et al., 1995). The constrained maximum entropy sampling problem (CMESP), is similar to the MESP, with K additional constraints (Lee, 1998):

$$\mathbf{s}^* = \underset{\mathbf{s} \in \mathcal{S}}{\operatorname{argmin}} \left\{ -ln(det(\mathbf{C}[\mathbf{s}, \mathbf{s}])) \in \mathbb{R} \Big| \sum_{j \in N} a_{ij} x_j \leq b_i \ i \in K; \sum_{j \in N} x_j = n \right\} \quad (4.4)$$

where $ln[\cdot]$ is the natural logarithm, $det[\cdot]$ is the determinant operator, $\mathbf{C}[\mathbf{s}, \mathbf{s}]$ is the principal sub-matrix of $C[\mathcal{S}]$ with row and column entries indexed by s, n is the cardinality constraint (i.e. sample size), subject to the K constraints, and the decision variable, $\mathbf{s}$ is represented by the column vector $\mathbf{s} = (s_1, \ldots, s_k, \ldots, s_M)' \in \{0, 1\}^M \ \forall \ k \in U$, where $s_k = 1$ if sample location k is included in the sample, and $s_k = 0$ if it is not (Knottnerus, 2003). An empty sample is denoted by $\mathbf{s} = (0, \ldots, 0, \ldots, 0)'$ and a census is defined by $\mathbf{s} = (1, \ldots, 1, \ldots, 1)'$ (Tillé, 2006).

To determine the entropy for a particular sample $\mathbf{s}$, from the M sample locations, we implement the algorithm developed by Ko et al. (1995), where the exact entropy associated with any n-element subset $\mathcal{S}$ of M (i.e. cardinality constraint) is the logarithm of the determinant of the $n \times n$ principal sub-matrix $\mathbf{A}[\mathcal{S}]$ with row and column indices in $\mathcal{S}$ where $\mathbf{s} \in \mathcal{S}$. The matrix $\mathbf{A}[\mathcal{S}]$ defines the correlation structure, in this case $\mathbf{C}$ is the covariance matrix, among the population of possible sample designs, as is similar to the method described by Bueso et al. (1998), who also used entropy based optimal spatial

sampling designs with a spatial correlation structure in order to determine the optimal environmental monitoring network.

4.3.2 Second Objective: Traveling Salesman Problem

Assuming the sampler is free to move between any two of the given M potential sample locations, the total cost to collect observations (i.e. tree measurements) and move among the n selected sample locations, is equivalent to the famous traveling salesman problem (TSP) (Lawler et al., 1985; Michalewicz and Fogel, 2004). The TSP, which is equivalent to finding the minimum length Hamiltonian cycle of a graph, $\mathcal{G}$ (Okabe et al., 2000), is defined as follows: Given the M potential sample locations (i.e. cities or vertices), a $M \times M$ distance matrix $\mathbf{D} = d(i, j)$, can be represented by a complete weighted graph $\mathcal{G} = (M, E, c)$, where E represents the edges (i.e. travel paths) connecting the M potential sample locations, and c is a function that assigns to each edge $(i, j) \in E$, a vector $(c_{i,j}^1, \ldots, c_{i,j}^K)$, where each element $c_{i,j}^k$ corresponds to a vector of K costs between sample location i and j (Paquete et al., 2004). The problem can be expressed mathematically as:

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x} \in \mathcal{S}} \left\{ \sum_{ij} c_{ij} x_{ij} \in \mathbb{R} \,\middle|\, x_{ij} \geq 0; \sum_i x_{ij} = 1 \;\; \forall \, j, \sum_j x_{ij} = 1 \;\; \forall \, i \right\} \quad (4.5)$$

where c_{ij} is defined as the cost to move between sample location i and j, represented by $x_{ij} = 1$ to move along that edge, otherwise $x_{ij} = 0$. Here, the length of the cost vector for each edge, is one (i.e. $K = 1$), representing only the cost of travel (i.e. distance) and since a sampler can move between any two sample locations, $\mathcal{G}$ represents a Euclidean Distance Matrix ($\mathbb{EDM} \equiv \mathbf{D} \in \mathbb{R}^2$) with $M(M-1)/2$ entries (Dattorro, 2005).

Like the CMESP, the TSP is a well known and well studied $\mathcal{NP} - hard$ problem, where the number of possible solutions representing all the tours, or permutations, for each set of selected sample locations is $n!$ (Ehrgott, 2000; Yan et al., 2003). The TSP has been used to solve vehicle routing, computer wiring, sheet cutting, job sequencing, and data clustering problems (Lawler et al., 1985; Michalewicz and Fogel, 2004). Variations on the TSP have been used to determine the minimum spanning tree (Weintraub and Navon, 1976), the longest inter-node distance, routing problems with multiple vehicles, vehicles with capacity constraints, and tours for deliveries with time-windows, all of which are important in primary forest supply chains (Michalewicz and Fogel, 2004).

4.3.3   Complete Formulation

The Pareto frontier between the maximum average information content of the sample (CMESP) and the cost to sample $\mathcal{A}$ (TSP), can be determined by the minimization of a general non-linear constrained deterministic d-objective problem ($d \geq 1$), expressed using the general formulation (Karush, 1939; Kuhn and Tucker, 1951; Osyczka, 2002; Coello Coello, 2002):

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{S}}{\operatorname{argmin}} \left\{ \mathbf{f}(\mathbf{x}) \in \mathbb{R}^d \middle| \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0} \right\} \tag{4.6}$$

where the goal is to obtain the arguments $\mathbf{x}^*$ that yield the minimum $\mathbf{f}$, a vector of d-objectives, subject to $\mathbf{g}$, a vector of K inequality constraints, and $\mathbf{h}$, a vector of M equality constraints and the fitness of any candidate solution $\mathbf{x}$ can be evaluated directly from the functions $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$.

For problems with potentially competing and possibly non-commensurate

objectives (i.e. $d > 1$), an objective vector $\mathbf{f}_a$ is said to *dominate* another objective vector $\mathbf{f}_b$, denoted by $\mathbf{f}_a \prec \mathbf{f}_b$ (component-wise), if and only if:

$$\mathbf{f}_{a,i} \leq \mathbf{f}_{b,i} \; \forall \, i \in \{1, \ldots, d\} \; \wedge \; \mathbf{f}_{a,j} < \mathbf{f}_{b,j} \; \exists \, j \in \{1, \ldots, d\} \qquad (4.7)$$

In words, an objective vector is called *non-dominated* if there are no other objective vectors that can increase the value of any one of the d objective functions without decreasing the value of another of the d objective functions (Tan et al., 2005). The set of all *non-dominated* solutions is called the *Pareto set*, *Pareto front* or *efficient frontier* (Eiben and Smith, 2003).

For this examination, $\mathbf{f}(\mathbf{x})$ is a vector of length two (i.e. $d = 2$) containing the entropy, which we want to maximize, and the cost to obtain the sample, which we want to minimize. The decision variable $\mathbf{x}$, is a $2M$ partitioned vector that includes M $0/1$ entries representing which of the M sample locations are included in the sample, a $\sum_{i=1}^{M} x_i = n$ partition representing the tour (i.e. visit sequence) of the selected sample locations, and a $M - n$ partition containing the unsampled location labels from the first M entries from $\mathbf{x}$ and the order is unimportant. Two inequality constraints, $\mathbf{g}(\mathbf{x}) \equiv \{g_1(\mathbf{x}), g_2(\mathbf{x})\}$ were included to bound the cost between \$200 and \$2000 (i.e. budget restrictions). No cardinality constraints were included, since our interest in the maximum entropy for a specific cost, and not for a specific sample size.

The complete formulation for this specific problem is:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{S}}{\operatorname{argmin}} \left\{ \mathbf{f}(\mathbf{x}) \in \mathbb{R}^d \middle| \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0} \right\} \qquad (4.8)$$

where the bi-objective vector, $\mathbf{f}(\mathbf{x})$ is:

$$\mathbf{f}(\mathbf{x}) \equiv \left\{ f_1(\mathbf{x}), f_2(\mathbf{x}) \right\} \equiv \left\{ ln(det(\mathbf{C}[\mathbf{x}, \mathbf{x}])), \sum_{\ell \in M} \mathcal{P}_\ell^c + \sum_{i \in M} \sum_{j \in M} c_{i,j} x_{ij} \right\} \qquad (4.9)$$

subject to the inequality constraints vector, $\mathbf{g}(\mathbf{x})$, defined as:

$$\mathbf{g}(\mathbf{x}) \equiv \left\{ g_1(\mathbf{x}), g_2(\mathbf{x}) \right\} \equiv \left\{ f_2(\mathbf{x}) - 200.0, 2000.0 - f_2(\mathbf{x}) \right\} \qquad (4.10)$$

where $\mathcal{P}_\ell^c$ is the cost to measure the diameters and heights, and thus the volume for the sample location, (P_ℓ^v) for $p_j(\,\cdot\,)$, $c_{i,j}$ is the cost to travel between plot $\mathcal{P}_i$ and $\mathcal{P}_j$, and $ln(det(\mathbf{C}[\mathbf{x},\mathbf{x}]))$ is the logarithm of the determinant of the covariance matrix for the M sample locations, described in Section 4.3.1. Here, we do not examine the within-plot travel costs for stems on the same plot, thus assuming travel time among stems within the plot is negligible.

4.4   HEURISTIC DESCRIPTION

Exact solution methods (i.e. mixed-integer programming) perform satisfactorily for small single-objective problems ($M \leq 75$), and for larger single-objective problems, Lagrangian relaxation and branch-and-bound methods work well (Lawler et al., 1985). For larger problems, or for problems with multiple objectives that cannot be solved exactly, heuristic methods are commonly used to obtain satisfactory solutions, which includes an explosion of work using Evolutionary Algorithms (EA) (Michalewicz and Fogel, 2004).

For small populations of potential sampling locations, exact solution methods such as linear/integer programming, can be used to obtain the set of samples, or tour distances, that minimizes costs for a general unequal probability sample design (Knottnerus, 2003; Tillé, 2006). Once the resulting set of potential samples is identified, enumeration can be used to obtain the Pareto frontier between the unbiased minimum variance estimates for the population totals and the sample cost (Gentle et al., 1997). For larger populations

of potential sample locations ($M \geq 75$ Lee (1998)), the Pareto frontier for a general multi-objective sampling scheme can be effectively determined using heuristic methods (de Gruijter et al., 2006). In forest planning, determination of the Pareto frontier is approximated using heuristic methods such as Tabu Search (TS) (Hansen, 1997), Simulated Annealing (SA) (Deusen, 1999; Nam and Park, 2000), Genetic Algorithms (GA), (Ducheyne et al., 2004), and Evolutionary Algorithms (Stewart et al., 2004; Ducheyne et al., 2006).

We use a class of heuristics known as Evolutionary Algorithms (EA), inspired from Darwinian evolution, which selects, mutates and promotes candidate solutions based on competition, fitness, and reproductive success (Eiben and Smith, 2003; Falcão and Borges, 2001). While similar to GA, The major difference between GA and EA results from the use of the mutation operator. In GA, the crossover operator is used to promote both diversification and intensification, whereas EAs rely on the mutation operator alone to change mutation parameters depending on a variety of metrics of convergence (Igel et al., 2007), stopping (Laumanns et al., 2002), diversification (Farhang-Mehr and Azarm, 2002), and optimality (Deb et al., 2007) criteria.

Several comprehensive reviews of multi-objective evolutionary algorithms are available from Coello Coello (2002), Tan et al. (2005), and Deb (2001) and all suggest that evolutionary algorithms have many advantages for high dimensional vector valued problems with potentially highly disconnected, non-uniformly distributed, and concave Pareto frontiers. For the TSP, evolutionary algorithms have been widely examined (Michalewicz and Fogel, 2004), but in forest operations, examination of EA for both single objective (Falcão and Borges, 2001) and multi-objective (Ducheyne et al., 2004) formulations are rare. To examine the applicability, behavior, and performance of a generalized multi-objective evolutionary algorithm (MOEA) for solving the

> **Data**: $\mathbf{x}, \mathbf{f(x)}, \mathbf{g(x)}, \mathbf{h(x)}$
> **Data**: plot attributes, distance matrix and costs, constraint formulae, software
> **Result**: $(\mu + \lambda)$-PAES Multi-Objective Evolutionary Algorithm (Knowles and Corne, 2000)
> $g \leftarrow 1;$
> $\mu_g = \emptyset;$
> $\lambda_g \leftarrow GenerateRandom();$
> $Evaluate(\lambda_g);$
> $\mu_g \leftarrow UpdateParetoArchive(\lambda_g);$
> **for** $g \leftarrow 2$ **to** G **do**
> $\quad \mid \quad \lambda_g \leftarrow SelectCandidatesFromArchive(\mu_{g-1});$
> $\quad \mid \quad \lambda_g \leftarrow Mutate(\lambda_g);$
> $\quad \mid \quad Evaluate(\lambda_g);$
> $\quad \mid \quad \mu_g \leftarrow UpdateParetoArchive(\lambda_g);$
> **end**
> $ExportArchive(\mu_G);$

Algorithm 2: Pareto archiving multi-objective evolutionary algorithm.

formulation in Section 4.3.3, we used the evolutionary algorithm presented by Knowles et al. (2000).

The $(\mu + \lambda)$-Pareto Archive Evolution Strategy $((\mu + \lambda)$-PAES), requires no assumptions about the decision maker or their preferences in the form of weights or scaling factors to reduce the solution space to a scalar function (Steuer, 1985). The method, presented in Algorithm 2, is an elitest strategy, which guarantees convergence (Rudolph and Agapie, 2000), is simple to program, generates λ mutations at each generation, maintains a set of μ non-dominated solutions throughout the search, and has been shown to work well for a multitude of Pareto front types (Tan et al., 2005).

The initial population of λ candidate solutions are generated randomly and evaluated using the formulations presented in Section 4.3.3. The set of initial λ candidate solutions are then added to the Pareto archive (i.e.

$\mu = \lambda$). A scalar rank is then assigned to each candidate solution, where the rank is defined as the number of $\mu + \lambda$ individuals that dominate the candidate solution. When there are no solutions that dominate the candidate solution, (i.e. non-dominated) a rank of 0 is assigned to the candidate solution. All candidate solutions with $rank > 0$ are then removed from the archive (i.e. dominated). The remaining non-dominated (μ) candidate solutions, which define the current Pareto frontier, are then selected for promotion and further mutation.

Selection for promotion is performed using tournament selection where a set of h candidate solutions enter a tournament and a single winner is selected with probability p (Dumitrescu et al., 2000). We used deterministic binary tournaments ($h = 2, p = 1$), where two solutions are selected with uniform probability from the Pareto archive (i.e. uniform, $P(Sel = 1) = 1/|\mu|$), and one of two candidate solutions is selected, with a probability of one, to be copied into the population buffer (λ) for promotion and mutation. The number of individual objectives that dominated the tournament for each candidate solution determined the winner with probability of one. In the event of a tie (i.e. weakly non-dominated), the first contestant was deemed the winner. The process of selection, promotion, and mutation is performed until some stopping criteria is met. To simplify our examination of the heuristic, we used a deterministic stopping criteria defined by the number of generations (Deb, 2001).

A candidate solution (i.e. candidate plan, policy, or decision vector), can be represented by the the partitioned vector, described in Section 4.3.3, and a variable representing a mutation strategy, is expressed as:

$$\mathbf{x}_c = \langle x_1, \ldots, x_\ell, \ldots, x_M | x_{M+1}, \ldots, x_{M+n}, | x_{M+n+1}, \ldots, x_{2M}, | \sigma \rangle \quad (4.11)$$

Parameter	Specification
Representation	Real/Integer-valued vector
Population Size (Initial,Subsequent)	$10, 7\mu$ Candidate solutions
Generations	100
Recombination	None
Mutation Strategy Parameter (σ)	95%
Parent Selection	Uniform random
Tournament Type	Binary deterministic (k=2,p=1)
Survivor Selection	$(\mu + \lambda)$-PAES
Specialty	None
Stopping Criteria	100 generations

Table. 4.1: Evolution strategy parameters.

where $x_1, \ldots, x_\ell, \ldots, x_M$ represents the binary decision to include $\mathcal{P}_\ell$ from the population of potential M sampling locations, $x_{M+1}, \ldots, x_{M+n}$ represents the tour sequence for the included sample locations that minimizes the Hamiltonian cycle length of $\mathcal{G}$, and $x_{M+1}, \ldots, x_{M+n}$ contains the elements of $\mathcal{P}$ not included in the sample.

The last entry in the candidate solution, σ represents the probability of either swapping two nodes in an existing solution (i.e. attempting to improve the TSP problem) or exchanging a sampled/non-sampled location into or, out of, the solution, effectively forcing a new TSP solution as a result of the new CMESP candidate solution.

4.5 DATA AND METHODS

To examine the Pareto frontier between the maximum average sample information content and the sample cost, a bounded 8.24 ha polygon, denoted $\mathcal{A}$, was located on the Oregon State University McDonald-Dunn Research Forest (Figure 4.1). The boundary for $\mathcal{A}$ and the stem positions, $\mathcal{T}^{\mathbf{u}} \forall k \in \mathcal{U}$,
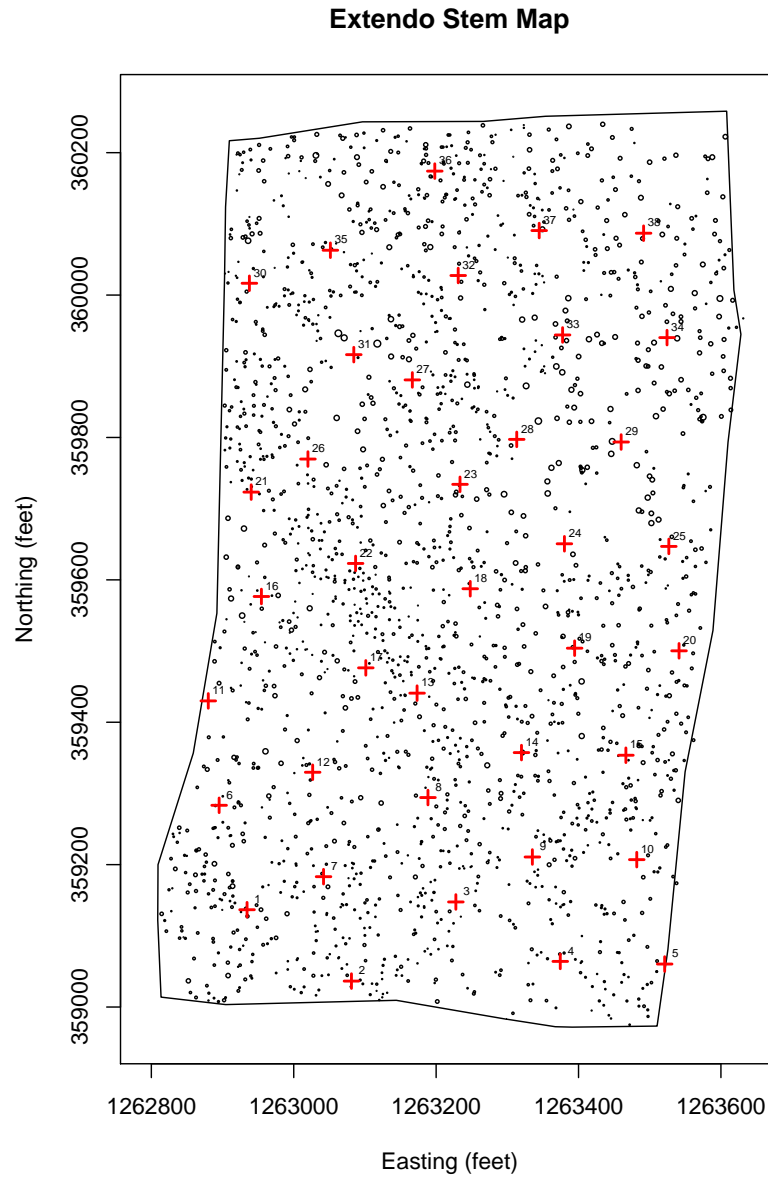
Figure. 4.1: Stem map of "Extendo" harvest unit with the diameter of the stems represented with circles of proportional size. The + symbols represent the locations for the population of M potential sample locations $\mathcal{P}$

for all stems over 15 cm , were recorded using standard survey methods (Davis et al., 1981). Attributes collected for each $\mathcal{T}_k$ included the species, $\mathcal{T}_k^s$, the diameter at breast height $\mathcal{T}_k^d$, the total height, $\mathcal{T}_k^h$, the height to live crown $\mathcal{T}_k^c$, the number of pieces in the main stem, $\mathcal{T}_k^n$, and surface characteristics for each piece $\mathcal{T}_k^{\mathbf{s}} \forall \mathbf{s} \in \mathcal{T}_k^n$.

To determine the volume for each stem, $\mathcal{T}_k^v$ all stems were merchandised into log lengths of four m, to a two cm top, using the taper equation presented by Kozak et al. (1968). If the stem could not be cut into a round number of standard log lengths, any remaining stem length was bucked into a short log no less than two m. Stems were merchandised with a stump height of 0.3 mand each log included 0.2 mof trim. For each log in the stem, the starting and ending height of the log, the nominal length and the actual length (nominal length plus trim), the small and large end diameters, and the Smalian volume was recorded. For this examination, the Smalian volume from each of the logs were then tallied and assigned to $\mathcal{T}_k^v$.

For this study, only Douglas-fir (*psudeotsuga mensezii* (Mirb. Franco)) stems were used resulting in a final count of 2053 stems with a spatial mean density $\overline{V}_\mathcal{A}$ of 249.17 stems ha^{-1} . The total volume for the study area $V_\mathcal{A}$, was 7560.07 m^3 yielding an average stem volume of 3.68 m^3 and an mean volume density of 917.57 m^3 ha^{-1}.

For each possible sample design, the N vector of first-order inclusion probabilities, π, were computed using the "tree-concentric method" (Schreuder et al., 1993) for each $p_j(\cdot) \forall j \in J$. Common definitions of $p(\cdot)$ include "split-plot designs" (i.e. $p(D, \theta_1, \theta_2)$ where $\mathcal{T}$ is stratified into two populations based on some diameter D, so that when $T_k^d \leq D$, stems are included on a fixed-area plot, and when $T_k^d > D$, stems are included on a variable-radius or "prism" plot, where θ_1 is the fixed-area plot radius, and θ_2 is an angle

gauge constant, or basal area factor, that defines the radius of the inclusion probability polygon associated with stem $\mathcal{T}_k$, when $T_k^d \geq D$. To correct for edge-bias, the polygon defining the probability of detection for $\mathcal{T}_k$, $\mathcal{T}_k^p$, was intersected with the edge of $\mathcal{A}$, as described by Beers (1966).

To generate a potential network of "transmitters of a particular message" for *all* information (i.e. possible messages) within $\mathcal{A}$ (Caselton and Hussain, 1980), a non-aligned grid of M potential sample locations was generated using the `spsample` function in the `sp` package (Pebesma and Bivand, 2005) with the `nonaligned` option. At these sample locations, each of the 36 sample designs were installed and the resulting plot-level summary statistics were obtained to determine the volume density ($\mathcal{P}_\ell^v$), the stem density, $\mathcal{P}_\ell^v$, and the number of stems measured, $P_\ell^n \; \forall \; \ell \in \mathcal{W}$. The covariance matrix $\mathbf{C}$ was obtained by calculating the covariance of the results for $\widehat{V}_\mathcal{A}$ at each potential sample location P_ℓ across all 36 sample designs.

The cost of a sample $\mathbf{s} \in \mathcal{S}$ was determined assuming the measurement time required for each stem in the sample was considered fixed at 2 min per stem regardless of stem size (Iles and Bell, 2004). The time to obtain a tree count on each plot was assumed to 2 min per plot and the time required to travel among the stems on a given plot was assumed to be one minute for each stem. The ground velocity to travel among the plots was assumed to be 3.2 km $^{-1}$h. Sampling costs were assumed to be \$60 h^{-1} (Iles and Bell, 2004) regardless of the activity.

4.6   RESULTS AND DISCUSSION

Since examination of sampling is traditionally not examined using this method, we present and discuss our results by examination of the shapes of

the Pareto frontiers, the performance of the estimates for the set of samples, and basic sampling metrics, familiar to practitioners (e.g. mean tree count the measurement cost/travel cost ratio), and finally the performance of the heuristic itself.

4.6.1 Pareto Solutions

The shape of the Pareto frontier determines the potential gains available between samples for a given cost. In Figure 4.2, the curves represent the best possible average information to be gained from a single sample within $\mathcal{A}$, where the primary goal is to obtain the best guess of $V_{\mathcal{A}}$, regardless of the inference type (i.e. design-based estimates or model-based predictions). Since our interest is primarily with determining which curve, or curves, yield the maximum average information content for a given cost, we seek the curve with the steepest ascent.

Upon examination of the Pareto frontiers, the smallest fixed-area plots combined with the largest BAF (i.e. largest variable radius plots and lower tree counts), sample design $p(\,\cdot\,, 5.64, 16))$, increased the information most quickly for all diameter limits. However, examination of Figure 4.2 does not reveal the fact that since the cost to travel for each sample is about half the measurement costs (see Figure 4.5), the number of plots is much larger for that sample design, than for the more "inclusive" sample designs (i.e. larger plot sizes), which span the entire range of plot counts included in the samples as is evident in Figure 4.4.

The panels reveal that for sampling using the largest fixed-area plots (i.e. $p(\,\cdot\,, 17.8, \,\cdot\,)$), the information content for all samples increased more slowly that did samples with smaller fixed radius plots. This is most likely due to

Optimal Information–Sample Cost Pareto Frontiers
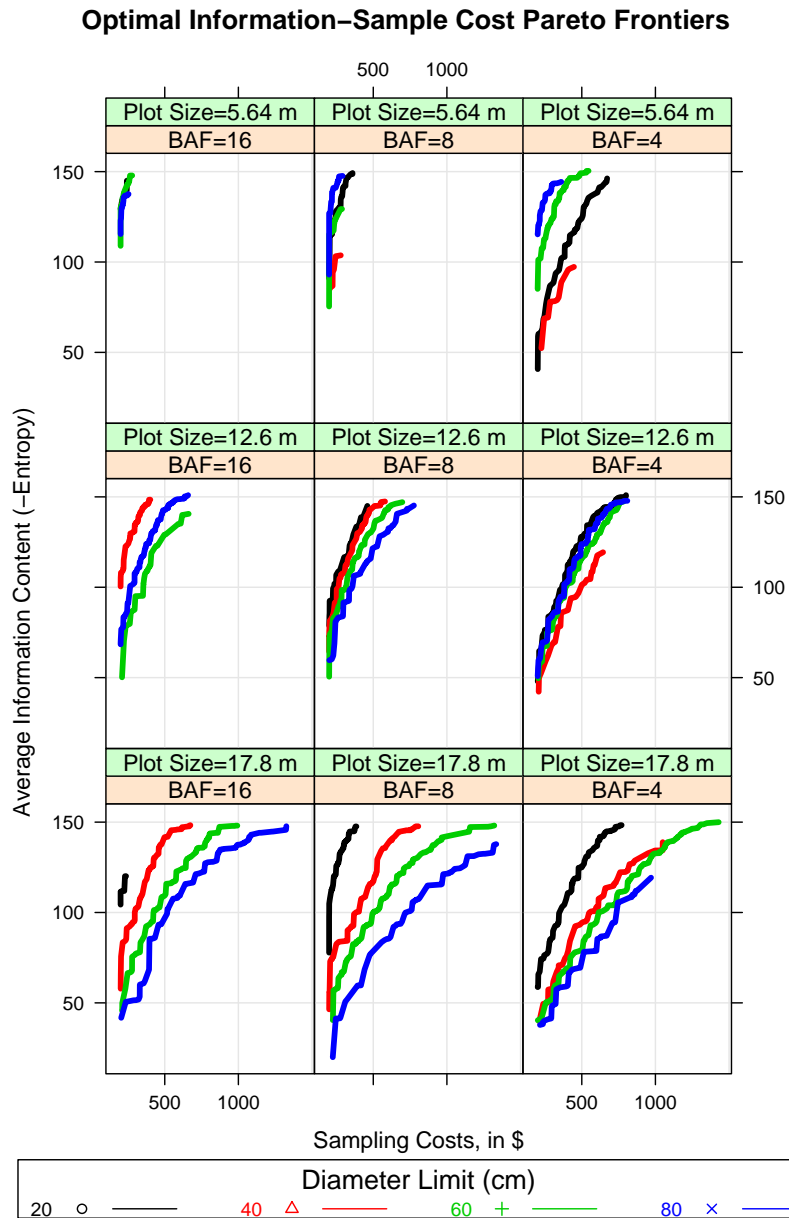
Figure. 4.2: Pareto frontier between the entropy of a sample and the sampling costs to obtain a guess of $V_{\mathcal{A}}$. The four lines in each panel represent the various D values (i.e. 20,40,60,80 cm) that stratify the popuation of $\mathcal{T}$ into two strata (i.e. fixed-area plots or various-radius plots).

the fact that for those plots (i.e. the largest fixed-area plots), the gains to be had from placing plots more closely does not increase the information as quickly as adding more smaller plots spaced farther apart, nor does it add information, per dollar spent obtaining that information, as quickly as plots with slightly a smaller fixed-radius (i.e. $p(\,\cdot\,, 12.6, \,\cdot\,)$).

The informational value of the increasing the size of variable-radius plots (i.e. prism plots) was roughly consistent across angle gauges for each fixed-area plot size. The shape and position of the curves across each fixed-area plot size are similar revealing that the influence of prism-size, for a given fixed radius plot size, has little influence.

The most influential factor for the samples is revealed in the shape and position of the lines relative to each other within a sample design panel. For example, The bottom panels in Figure 4.2, show that for a given fixed-area plot size (17.8 m), the slope of the trade-off of average information per dollar spent to obtain the sample, increases as the size of the angle-gauge increases, thus increasing the size of the variable radius plots, in turn, increasing the number of "prism" trees included in the sample. Conversely, as the size of the fixed-radius plot increases, the number of stems included in the sample likewise increases, but at a different rate than does increasing the variable-radius plot size.

4.6.2 Estimation performance

In this section, we discuss various properties of the estimates of $V_{\mathcal{A}}$, and associated details about the metrics commonly used to compare samples, such as the bias, the mean tree count per plot, and the ratio of the cost to measure the stems versus the cost to travel among the plots (Zeide, 1980).

Since we obtained the samples that contain the optimal average information, we only report the results for the design-based estimation of $V_{\mathcal{A}}$ since this value is the most commonly presented in the literature. It should also be noted here that this does not preclude the same discussion of any results presented for model-based inference, but here, we only discuss design-based results.

4.6.2.1   Bias

Since the explicit goal was to obtain our best guess of $V_{\mathcal{A}}$, and we know $V_{\mathcal{A}}$ as a result of our census, we compared the bias of the resulting samples against $V_{\mathcal{A}}$ (i.e. $V_{\mathcal{A}} - \widehat{V}_{\mathcal{A}}$), for each of the 36 definitions of $p(\,\cdot\,)$. It should be noted here that our value for $V_{\mathcal{A}}$ is the actual value, obtained from all stems in the population of sample locations, and *not* the true stem population, as the value is conditioned on only those stems included in the population of sample plots. The results are presented graphically in Figure 4.3.

The bias, regardless of the definition of $p(\,\cdot\,)$ is almost always positive, except for the smallest fixed area and variable radius plots. This suggests that when large variable radius plots (i.e. small BAF) are used for sampling $V_{\mathcal{A}}$, the results are consistently bias which would occur as a result of biasing the sample towards a particular strata (i.e. larger stems), rather than balancing the samples among the strata or the fact that the population of sample locations is limited to a very small subset of the infinitely large possible set of sample location positions within $\mathcal{A}$.
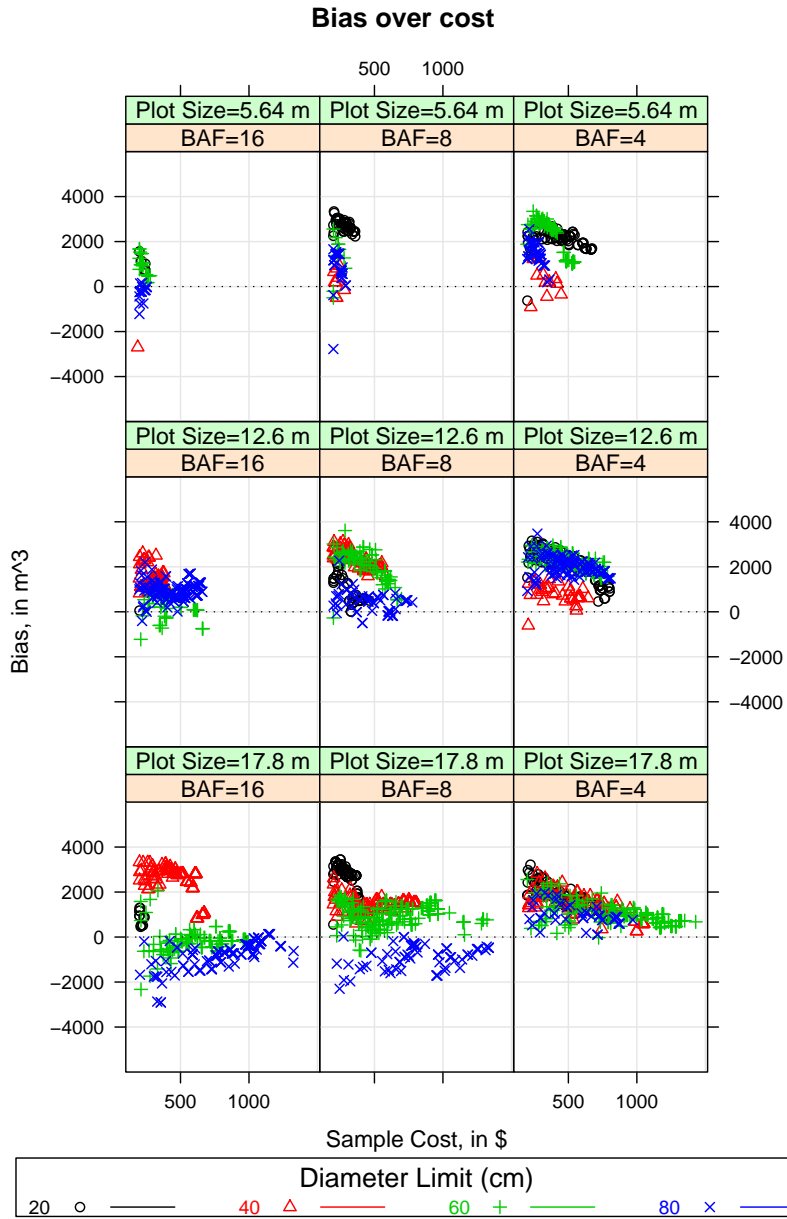
Figure. 4.3: Plot of the bias over the cost of the Pareto optimal solutions.

4.6.2.2   Mean Tree Count per Plot

Another metric often examined in operational sampling is the average number of stems measured on a plot. To examine this metric within the formulation we present, we plotted the mean number of stems measured per sample location, for each of the Pareto optimal sets.

For these solutions, the mean number of measurements on each plot can be excessive as is presented in Figure 4.4 where, for the larger fixed-area plots, the mean number of stems on each plot is about 40 stems. This high number is not commensurate with current practices of targeting between four and ten stems per point as is suggested in variable radius sampling (Iles and Bell, 2004). For the smaller fixed-area plots, the mean tree counts are more in agreement with current practices.

4.6.2.3   Measurement Cost/Travel Cost Ratio

Finally, we examine the ratio of the cost to measure the stems versus the cost to travel among the sample locations (see Figure 4.5). Zeide (1980) optimizes "plot size" and travel time "simultaneously" for a finite area so that the total time will be minimized by expressing both the plot density and the measure time for a square lattice. In contrast, we allowed the travel distance to vary as a function of the next "most valuable" plot within $\mathcal{A}$. This provides the most additional information, resulting in the simultaneous optimization between the cost to obtain a sample and the size of the plot, as defined by $p(\cdot)$ along the Pareto frontier. While, our method is not equivalent to the method used by (Zeide, 1980), our method makes no assumptions about the
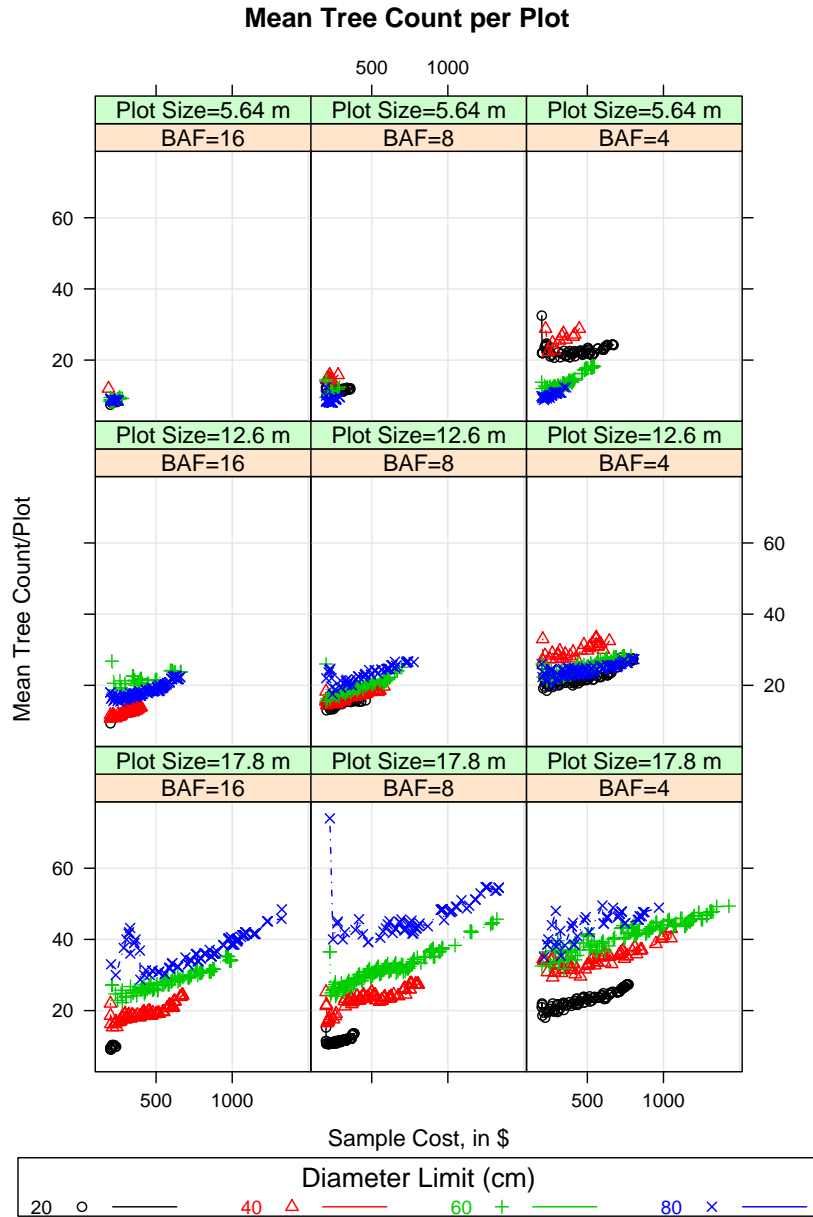
Figure. 4.4: Plot of the mean tree count per plot for the Pareto optimal solutions, presented in Figure 4.2.
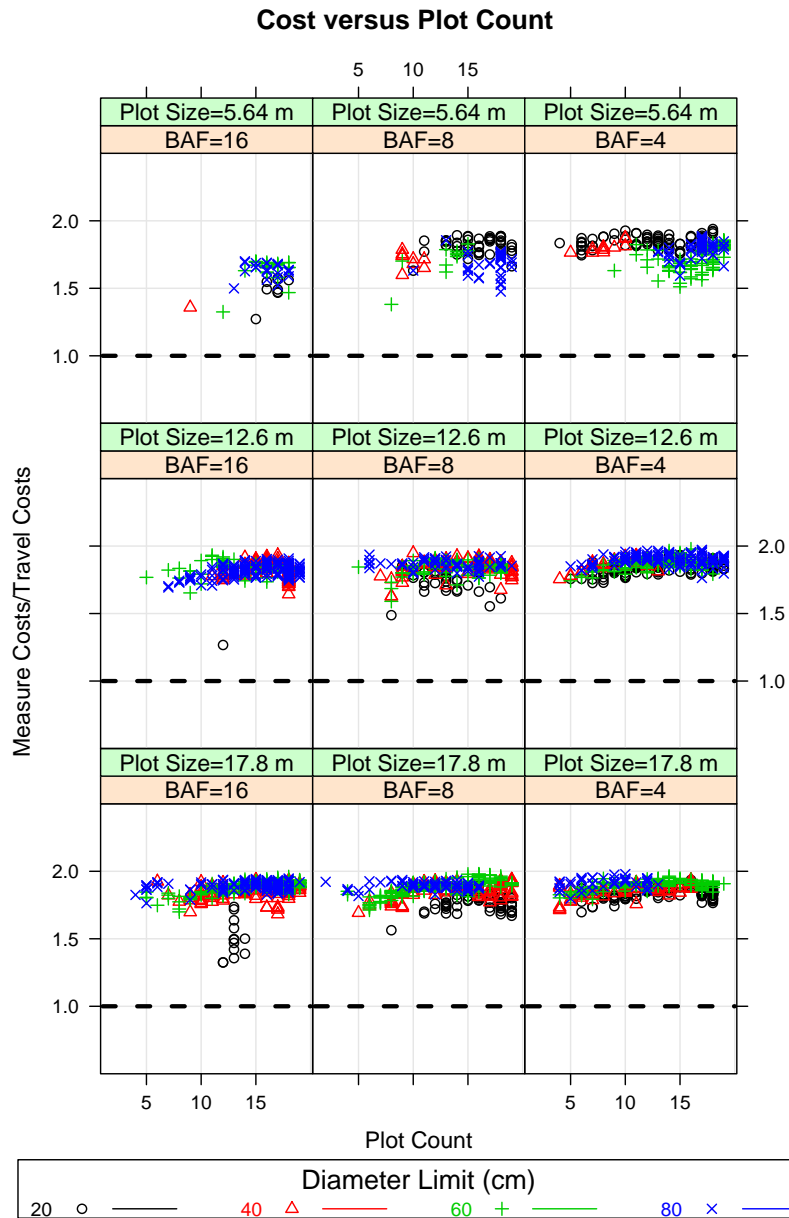
Figure. 4.5: Ratio of the measurement over tour cost for the Pareto optimal solutions. Note that the heavy dashed line where the ratio between the cost to measure all stems and the cost to navigate the tour are equal as concluded by Zeide (1980), is below the Pareto optimal solutions.

configuration of the spatial network of sample locations as being on a regular lattice, nor does out method fix the rate at which information entering the "communication channel" is a function of the ratio between the plot size (i.e. $p(\,\cdot\,)$) and the distance among the plots, by expressing the ratio as a function of $p(\,\cdot\,)$.

The plot in Figure 4.5 contains the ratio, plotted over the number of sample locations, for the cost to measure all stems in a sample over the cost to move among the plots, with time included for determine which stems are to be included and the time required to move among the stems at the selected sample location.

From our plots, we conclude that the optimal, as defined by the Pareto frontier, show that the cost to measure stems is between about 1.4 and 2 times the cost required to move among the plots for these particular solutions. This is in contrast to the conclusions presented by (Zeide, 1980), who says, "the greater the distance between plots the larger they must be," and that the plot size is optimal when the time to travel between the plots is equal to the time to measure the plot. While we find these general conclusions to be true in our research as well, we differ in our conclusions, from our examination of the Pareto frontier, that the time to measure stems on a plot, which is a function of $p(\,\cdot\,)$, should not dictate the spacing of the sample locations, especially when we are interested in more than a single target variable, as suggested by de Gruijter et al. (2006), or when we want to obtain samples for multiple purposes (e.g. simulation, prediction, variogram parameter estimation).

4.6.3 Sample Tours

Next, we examined the proposed samples, with their associated tour paths, for all Pareto optimal solutions. For brevity, we show only two solutions where we selected the solution that contains the highest information content to cost ratio and the solution lowest information to cost ratio, thus representing the best and worst possible choices for sampling $\mathcal{A}$. As expected, the major differences between the two samples is evident by the number and size of plots, the length of the tour, and the distribution of the plots with $\mathcal{A}$.

The sample with the highest information-cost ratio was $p(60, 5.64, 16)$ representing a split plot design where all stems less than 60 cm are selected into a plot with a fixed-radius with a plot radius of 5.64 m and stems over 60 cm are selected into a sample using the "tree-centric" variable radius plot where the plot radius is $\mathcal{T}^d/(2 \times \sqrt{16})$ m. While the information content for this sample was very large, indeed, the tour for this sample is sub-optimal, as the tour is not the least cost tour regarding the travel distances. This is most likely due to the poorly tuned exploration and exploitation strategies used for this problem.

The sample with the lowest information-cost ratio was $p(80, 17.8, 8)$ representing a split plot design where all stems less than 80 cm are selected into a plot with a fixed-radius with a plot radius of 17.8 m and stems over 80 cm are selected into a sample using the "tree-centric" variable radius plot where the plot radius is $\mathcal{T}^d/(2 \times \sqrt{8})$ m. This resulted in only two, albeit large plots, being taken within $\mathcal{A}$.

These results, while not in complete agreement with Zeide (1980), do reinforce the notion that many small plots are far more costly than a few larger plots, which when the goal is to obtain the most information for a given

cost, may be more advantageous as more than a single parameter is of interest and the practitioner is concerned with not only total volume estimation, but variogram parameter estimation, spatial prediction, and capital optimization.

4.6.4 Heuristic Performance

Since we ran the heuristic for each $p(\,\cdot\,)$ only once, we limit the discussion to the graphical results, presented in Figure 4.7, since we cannot make inference about the convergence properties without multiple runs or a completely enumerated solution search/solution space to compare against.

We visually examined the number of Pareto solutions discovered over number of generations to ascertain if the values of $p(\,\cdot\,)$ influenced the rate at which new solutions entered the Pareto archive (see Figure 4.7).

The values for $p(\,\cdot\,)$ appear to influence the rate at which Pareto solutions were discovered over the run of 250 generations (Figure 4.7). For the smaller fixed area plots, the number of Pareto optimal solutions increased more slowly, thus suggesting that for smaller plots, the search for Pareto optimal solutions is more difficult than for larger plot sizes. It should also be noted here that since we ran the heuristic for only 250 generations and did not enumerate the solution space, the exact number of solutions in the global set, for each definition of $p(\,\cdot\,)$ is unknown. In fact, at the end of the run, the number of Pareto optimal solutions, for most definition of $p(\,\cdot\,)$ continues to increase, again suggesting that the heuristic should have been run for more generations.

4.7   CONCLUSIONS

In this manuscript, we determined the Pareto optimal set for a population of sample locations and sample designs that are commonly used for operational planning in the primary forest products supply chain. Since the interest was not only in obtaining arbitrarily small confidence intervals (i.e. standard errors), we used a general information theoretical metric, known as entropy, combined with a traveling salesman problem to determine the optimal trade-offs between the information content of a sample and its cost when our interest is in future realizations of a discrete random field at unsampled locations.

To accomplish these objectives, we examined 36 possible sample designs installed at 38 possible sample locations to determine their ability to yield the highest information for a given cost as a means of determining if there was a single optimal sample design that could be used for any question asked of the data. We found, however, that while we obtained the equivalent D-optimal experimental design sample, our primary interest lay in obtaining precise values of the total volume in $\mathcal{A}$ and data sufficient for production planning and optimization.
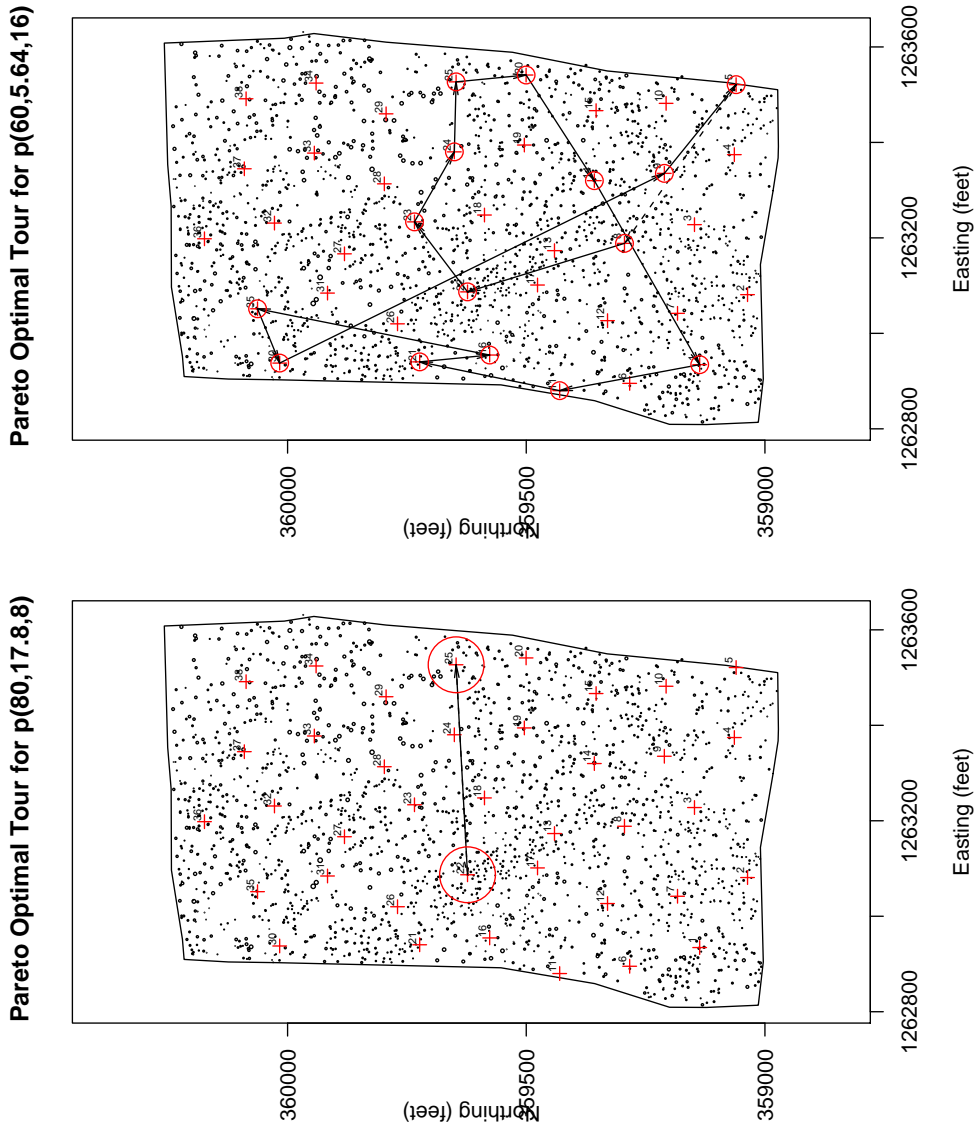
Figure. 4.6: Path and included plots for the minimum costs/maximum standard error and the maximum costs/minimum standard error samples for $\widehat{V}_{\mathcal{A}}$.
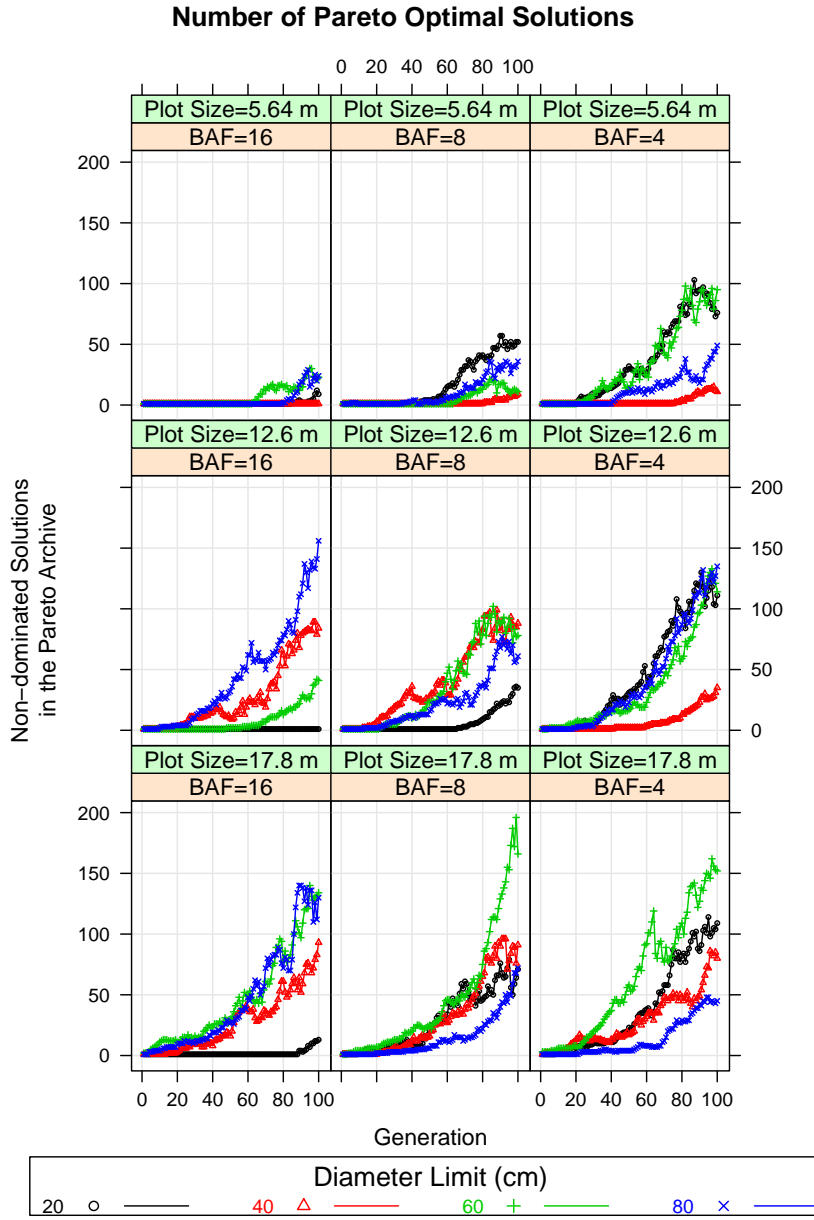
Figure. 4.7: Number of Pareto optimal solutions discovered for each of the CMESP/TSP mutation ratios.

BIBLIOGRAPHY

Abbas, A. E. and Holmes, S. P. (2004). Bioinformatics and management science: some common tools and techniques. *Operations Research*, 52(2):165–190.

Beers, T. W. (1966). The direct correction for boundary-line slopover in horizontal point sampling. Research Progress Report 224, Purdue University, Agricultural Experiment Station, Lafayette, Indiana.

Bettinger, P., Boston, K., and Sessions, J. (1999). Combinatoral optimization of elk habitat effectiveness and timber harvest volume. *Environmental Monitoring and Assessment*.

Blackwell, D. A. (1951). Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102. University of California Press, Berkeley.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Bueso, M., Angulo, J., and Alonso, F. (1998). A state-space model approach to optimal spatial sampling design based on entropy. *Environmental and Ecological Statistics*, 5:29–44.

Bueso, M., Angulo, J., Cruz-Sanjulian, J., and Garcia-Arostegui, J. (1999). Optimal spatial sampling design in a multivariate framework. *Mathematical Geology*, 31:507–525(19).

Caselton, W. F. and Hussain, T. (1980). Hydrological Networks: Information Transmission. *Journal of The Water Resources Planning and Management Division*, 106(2):503–520.

Chapman, R. C. (1982). Plot spacing in systematic sampling. *Journal of Forestry*, 80:409–419.

Chen, X., Dempster, A. P., and Liu, J. S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika*, 81(3):457–469.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons, New York. 428 p.

Coello Coello, C. A. (2002). Evolutionary multi-objective optimization: A critical review. In Sarker, R., Mohammadian, M., and Yao, X., editors, *Evolutionary Optimization*, chapter 5. Kluwer Academic Publishers, Norwell, MA, USA.

Dattorro, J. (2005). *Convex Optimization. mathcalMϵβoo* Publishing,.

Davis, L. S., Johnson, K. N., Bettinger, P. S., and Howard, T. E. (2001). *Forest Management*, chapter 13, page 804. McGraw-Hill series in forest management. McGraw-Hill, New York, New York, 4th edition.

Davis, R. E., Foote, F. S., Anderson, J. M., and Mikhail, E. M. (1981). *Surveying, theory and practice*. McGraw-Hill, Inc.

de Gruijter, J. J., Brus, D. J., Bierkins, M. F. P., and Knotters, M. (2006). *Sampling for Natural Resource Monitoring*. Springer-Verlag.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK. ISBN 0-471-87339-X.

Deb, K., Tewari, R., Dixit, M., and Dutta, J. (2007). Finding Trade-off Solutions Close to KKT Points Using Evolutionary Multi-Objective Optimization. KanGAL Report 2007006, Indian Institute of Technology Kanpur, Kanpur, PIN 208016, India.

Deusen, P. C. V. (1999). Multiple solution harvest scheduling. *Silva Fennica*, 33(3):207–216.

Diestel, R. (2005). *Graph Theory (Graduate Texts in Mathematics)*. Springer.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2004). Even flow scheduling problems in forest management. In Coello, C. A. C. and Lamont, G. B., editors, *Applications of Multi-Objective Evolutionary Algorithms*, pages 701–726. World Scientific, Singapore.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2006). A spatial approach to forest-management optimization: linking gis and multiple objective genetic algorithms. *International Journal of Geographical Information Science*, 20(8):917–928.

Dumitrescu, D., Lazzerini, B., Jain, L., and Dumitrescu, A. (2000). *Evolutionary Computation*. CRC Press.

Ehrgott, M. (2000). Approximation algorithms for combinatorial multicriteria optimization problems. *International Transactions in Operational Research*, 7:5–31.

Eiben, A. E. and Smith, J. E. (2003). *Introduction to Evolutionary Computing.* SpringerVerlag.

Falcão, A. O. and Borges, J. G. (2001). Designing an Evolution Program for Solving Integer Forest Management Scheduling Models: An Application in Portugal. *Forest Science*, 47(2):158–168.

Farhang-Mehr, A. and Azarm, S. (2002). Diversity assessment of pareto optimal solution sets: An entropy approach. In *Congress on Evolutionary Computation (CEC'2002)*, volume 1, pages 723–728, Piscataway, New Jersey. IEEE Service Center.

Gambill, C. W., Harry V. Wiant, J., and Yandle, D. O. (1985). Optimum plot size and baf. *Forest Science*, 31(3):587–594.

Gentle, J. E., Narula, S. C., and Valliant, R. L. (1997). Multicriteria optimization in sampling design. In Ghosh, S., Schucany, W. R., and Smith, W. B., editors, *Statistics of Quality*, volume 153 of *Statistics, textbooks, and monographs*, pages 411–425. Marcel Dekker, Inc.

Gill, C. A. and Joanes, D. N. (1979). Bayesian estimation of shannon's index of diversity. *Biometrika*, 66(1):81–85.

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28:1429–1447.

Hamann, J. D. and Boston, K. (2007). Using spatial statistics to improve the primary forestry supply chain. In Reynolds, K. M., Thompson, A., Shannon, M., Köhl, M., Ray, D., and Rennolls, K., editors, *Sustainable Forestry: from Monitoring and Modelling to Knowledge Management & Policy Science*, chapter 18, pages 285–297. CABI International.

Hansen, M. P. (1997). Tabu Search in Multiobjective Optimisation : MOTS. In *Proceedings of the 13th International Conference on Multiple Criteria Decision Making (MCDM'97)*, Cape Town, South Africa.

Igel, C., Hansen, N., and Roth, S. (2007). Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15(1):1–28.

Iles, D. D. M. K. and Bell, J. F. (2004). Using a large-angle gauge to select trees for measurement in variable plot sampling. *Canadian Journal of Forest Research*, 34(4):840–845.

Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. Master's thesis, Department of Mathematics, University of Chicago, Chicago, Ill.

Knottnerus, P. (2003). *Sample Survey Theory: some Pythagorean perspectives.* Springer-Verlag.

Knowles, J., Oates, M., and Corne, D. (2000). Advanced muli-objecive evolutionary algorithms applied to two problems in telecommunications. *British Telecom Technology Journal*, 18(4):51–65.

Knowles, J. D. and Corne, D. W. (1999). The Pareto Archived Evolution Strategy: A New Baseline Algorithm for Multiobjective Optimisation. In *1999 Congress on Evolutionary Computation*, pages 98–105, Washington, D.C. IEEE Service Center.

Knowles, J. D. and Corne, D. W. (2000). Approximating the nondominated front using the pareto archived evolution strategy. *Evolutionary Computation*, 8(2):149–172.

Ko, C.-W., Lee, J., and Queyranne, M. (1995). An Exact Algorithm for Maximium Entropy Sampling. *Operations Research*, 43(4):684–691.

Kononenko, I. and Kukar, M. (2007). *Machine Learning and Data Mining: Introduction to Principles and Algorithms.* Horwood Publishing Limited, West Sussex, UK.

Kozak, A., Munro, D. D., and Smith, J. (1968). Taper functions and thier application in forest inventory. *Forestry Chronical*, 45:278–283.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In Neyman, J., editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1950*, pages 481–492, Berkeley, California. University of California Press. Reprinted in: *Readings in Mathematical Economics, Vol 1, Value Theory*, (P. Newman, ed.), The Johns Hopkins University Press, Baltimore, 1968, pp. 3–14.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002). Combining convergence and diversity in evolutionary multiobjective optimization. *Evol. Comput.*, 10(3):263–282.

Lawler, E. L., Lenstra, J. K., Kan, A. H. G. R., and Shmoys, D. B., editors (1985). *The Traveling Salesman Problem.* John Wiley, New York.

Le, N. D. and Zidek, J. V. (2006). *Statistical Analysis of Environmental Space-Time Processes*, chapter 8, page 341. Springer Series in Statistics. Springer, New York, New York.

Le, P. G. N., Sampson, P., and Zidek, J. V. (1993). Using entropy in the redesign of an environmental monitoring network. In Patil, G. and Rao, C., editors, *Multivariate Environmental Statistics*, pages 175–202. North Holland.

Lee, J. (1998). Constrained maximum-entropy sampling. *Operations Research*, 46(5):655–664.

Mandallaz, D. and Ye, R. (1999). Forest inventory with optimal two-phase, two-stage sampling schemes based on the anticipated variance. *Candian Journal of Forest Research*, 29:1691–1708.

Mansour, Y. and Schain, M. (1997). Learning with maximum-entropy distributions. In *Computational Learing Theory*, pages 201–210.

Marti, K. (2005). *Stochastic Optimization Methods.* Springer.

Michalewicz, Z. and Fogel, D. B. (2004). *How to Solve It: Modern Heruistics*, chapter 5, page 554. Springer Verlag, 2 edition.

Montes, F., Hernández, M. J., and Canellas, I. (2005). A geostastical approach to cork production sampling estimation in *Quercus suber* forests. *Canadian Journal of Forest Research*, 35:2787–2796.

Nam, D. and Park, C. (2000). Multiobjective simulated annealing: A comparative study to evolutionary algorithms. *International Journal of Fuzzy Systems*, 2(2):87–97.

Oderwald, R. G. and Jones, E. (1992). Sample sizes for point, double sampling. *Canadian Journal of Forest Research*, 22:980–983.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams.* John Wiley and Sons, second edition.

Osyczka, A. (2002). *Evolutionary Algorithms for Single and Multicriteria Design Optimization.* Physica Verlag, Germany. ISBN 3-7908-1418-0.

Paquete, L., Chiarandini, M., and St utzle, T. (2004). Pareto local optimum sets in the biobjective traveling salesman problem: An experimental study. In Gandibleux, X., Sevaux, M., SÃűrensen, K., and T'kindt, V., editors, *Metaheuristics for Multiobjective Optimisation*, volume 535 of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, Berlin, Germany.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.

Rudolph, G. and Agapie, A. (2000). Convergence Properties of Some Multi-Objective Evolutionary Algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, volume 2, pages 1010–1016, Piscataway, New Jersey. IEEE Press.

Schreuder, H. T., Gregoire, T. G., and Wood, G. B. (1993). *Sampling methods for multiresource forest inventory.* John Wiley & Sons, Inc. 446 p.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423.

Shewry, M. and Wynn, H. (1987). Maximum entropy sampling. *Journal of Applied Statistics*, 46:165–170.

Stage, A. R. (1973). Prognosis model for Stand Development. General Technical Report INT-137, USDA Forest Service Research Note, Ogden, Utah.

Steuer, R. (1985). *Multiple Criteria Optimization: Theory, Computation and Application.* John Wiley & Sons, New York, NY.

Stevens, Jr., D. L. (1997). Variable density grid-based sampling designs for continuous spatial populations. *Evironmetrics*, 8:167–195.

Stewart, T. J., Janssen, R., and van Herwijnen, M. (2004). A genetic algorithm approach to multiobjective land use planning. *Comput. Oper. Res.*, 31(14):2293–2313.

Tan, K., Khor, E., and Lee, T. (2005). *Multiobjective Evolutionary Algorithms and Applications.* Springer.

Theil, H. (1971). *Principles of Econometrics.* John Wiley and Sons, Inc.

Tillé, Y. (2006). *Sampling Algorithms.* Springer.

Trujillo-Ventura, A. and Ellis, J. H. (1991). Multiobjective air pollution monitoring network design. *Atmospheric Environment*, 25(2):469–479.

Weintraub, A. and Navon, D. (1976). A forest management planning model integrating silvicultural and transportation activities. *Management Science*, 22(12):1299–1309.

Wu, S. and Zidek, J. V. (1992). Entropy-based analysis of data from selected nadp/ntn network sites for 1983-1986. *Atmospheric Environment*, 26A(11):2089–2103.

Yan, Z., Zhang, L., Kang, L., and Lin, G. (2003). A New MOEA for Multiobjective TSP and Its Convergence Property Analysis. In Fonseca, C. M., Fleming, P. J., Zitzler, E., Deb, K., and Thiele, L., editors, *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pages 342–354, Faro, Portugal. Springer. Lecture Notes in Computer Science. Volume 2632.

Zeide, B. (1980). Plot size optimization. *Forest Science*, 26(2):251–257.

5. MULTI-OBJECTIVE SUPPLY CHAIN OPTIMIZATION FOR FOREST OPERATIONS

5.1 ABSTRACT

Given an operational boundary, a single pre-harvest survey and a daily non-stochastic demand vector, we present a method to examine the Pareto frontier between the maximum expected net revenue and maximum unharvested area (i.e. biological reserves) for a forest operation which we express as a non-linear mixed-integer constrained d-objective ($d \geq 1$) minimization problem assuming the volume density can be modeled using a simple spatial correlation model. The expected net revenue is presented as a continuous review inventory model (i.e. daily), with stochastic supply, non-stochastic demand, product storage costs for undelivered volume (i.e. log yard), salvage costs for perishable products (e.g. blue-stain), and unfulfilled demand is lost (i.e. no backlogging) (Hillier and Lieberman, 1995).

We used block-kriging to refine predictions of the volume density, for smaller partitions (i.e. blocks) of the operational area (i.e. harvest blocks), which yielded unbiased minimum prediction error variance predictions of daily production (i.e. $\text{m}^3 \text{ ha}^{-1}$) and the log-yard inventory distribution (i.e. m^3). Using those values, we generate the Pareto frontier for a set of optimal spatially explicit daily harvest schedules. To approximate the Pareto frontier, we used the $(\mu + \lambda)$-Pareto Archive Evolution Strategy ($(\mu + \lambda)$-PAES), presented by Knowles and Corne (2000).

To examine the effectiveness of the method, we compared the extremes of the Pareto frontier to a "traditional" sinuous harvest pattern as is common in current operations. When the decision maker(s) selected the plan that

maximized net revenue, net revenue increased by 58.65, percent and the unharvested area increased by 23.54 percent as compared to a traditional operation. When the plan that maximized reserves was selected, unharvested area increased by 1473.41 percent, resulting in a decrease in net revenue of -2.13 percent. These results demonstrate that refining the decision making process can be beneficial to decision makers when there are competing and possibly non-commensurate multiple objectives.

5.2   INTRODUCTION

Forest inventories, and how to manage them, often reveal sources of conflict among different ecosystem services customers and managers as each manager has different customer service roles. The logging production manager, faced with minimizing setup costs, production scheduling, often desires long-term stable production runs with adequate log yard inventories to meet demands. The log yard manager, interested in the investment in inventory, minimizing product degradation, ensuring adequate stocking levels, prefers frequent small orders to reduce salvage costs. The marketing manager, emphasizing customer service, wants no stock-outs, backlogs, with minimal lead times, cycle times, and ample stocks to meet demands. Conversely, the financial manager, interested in liquidity, cash flows, inventory turnover, is often willing to accept a reasonable level of stock-out, investing in a minimal inventory to maximize the financial performance of the firm. Finally, the environmental affairs manager, emphasizing environmental aspects of the operation, wants to maintain standing inventories commensurate with certification criteria, marketing, and environmental policies.

We examine a production planning problem for a primary forest prod-

ucts supply chain, by considering the problem as a multi-objective non-linear mixed-integer spatially explicit "inventory" optimization system in which a single product (i.e. log volume) may be either produced (i.e. harvested) or purchased to fulfill non-stochastic service levels (i.e. demands) over a finite planning horizon of T periods (i.e. days).

The objectives are to 1) determine the trade-offs among various optimal harvest plans (i.e. machine routes), so that the production manager may, 2) select an optimal plan that maximizes net revenue for the financial manager, given a set of fixed service levels (i.e. non-stochastic book orders) dictated by the marketing manager, that 4) maximizes the un-harvested area as required by the environmental affairs manager, such that 5) the predicted production levels (i.e. replenishment) are the unbiased with minimum prediction error, as requested by the log-yard manager and finally, 6) to present outputs so that each manager can objectively assess their performance against the other potentially competing and possibly non-commensurate objectives (e.g. $ versus unharvested area) without directly associating or imposing value on the unharvested area.

Constraints include 1) a deterministic feasibility constraint (i.e. penalty function), 2) a set of deterministic equality constraints that produce the unbiased minimum prediction variance replenishment volumes in each period, 3) a deterministic constraint for the unharvested area, using an area restriction model (Weintraub and Murray, 2006, review), 4) a set of constraints to limit the total number of cliques harvested in each period to a single clique, 5) a set of constraints that requires the harvested clique in period t, be adjacent to the set that defines the boundary of the current opening, and finally, 6) a constraint to ensure the first clique harvested in $\mathcal{A}$ is on the boundary of $\mathcal{A}$.

To present our problem, we use definitions that are consistent with multi-

objective optimization (Steuer, 1986), classical design-based sampling texts (Cochran, 1977; Knottnerus, 2003; Tillé, 2006), model-based geostatistics texts (Cressie, 1993; Wackernagel, 1998), combinatorial optimization (Lawler et al., 1985), graph theory (Diestel, 2005), inventory theory (Hillier and Lieberman, 1995; Porteus, 2002), and evolutionary algorithms (Eiben and Smith, 2003). The format of the paper is to first introduce the nomenclature, followed by a presentation of the problem formulation in Section 5.3. In Section 5.4, we present a test dataset, which includes a bounded polygon (i.e. finite area) that contains a finite population of stems (i.e. stem map) and a set of service levels (i.e. log volume demands). Section 5.5 presents the heuristic, developed by (Knowles and Corne, 2000), used to determine the optimal trade-offs or *efficient frontier* for a set of machine paths. Our results of the Pareto optimal production plans, heuristic performance, and a discussion are presented in Section 5.6. Finally, we present our conclusions in Section 5.7.

Nomenclature

Harvest Unit and Block Attributes

$\mathcal{A}$	A harvest unit, stand, or general operational polygon
$\lambda(\mathcal{A})$	A function that returns the surface area of a polygon $\mathcal{A}$
$\mathcal{A}_b$	A partition of $\mathcal{A}$
$\lambda(\mathcal{A}_b)$	A function that returns the surface area of a polygon $\mathcal{A}_b$
$\mathcal{A}^v$	The total volume in $\mathcal{A}$
$\mathcal{A}^v$	Intrinsic regionalized variable over $\mathcal{A}$ with point support
$\mathcal{A}_b^{\mathcal{T}}$	The stems of T in $\mathcal{A}_b$ ($\mathcal{T}_k \cap \mathcal{A}_b$)
$\mathcal{A}_b^v$	The sum of T^v in $\mathcal{A}_b$ ($\mathcal{T}_k^v \cap \mathcal{A}_b$)
$\mathcal{A}_b^{\bar{v}}$	The mean spatial density of $\mathcal{A}_b^v$
$\mathcal{A}_b^d$	The number of adjacent cutting blocks (i.e. degree of the node)
$\mathcal{C}_t^a$	Harvest clique a in period t ($\mathcal{C}_t^a \in B$)

$\mathcal{C}_t^{|a|}$	The number of cutting blocks in harvest clique in period t
$\mathcal{C}_t^v$	The volume of the harvest clique in period t (i.e. $\sum_{i=1}^{\mathcal{C}_t^{|a|}} \mathcal{A}_b^v$)

Tree Attributes

N	The number of stems in $\mathcal{A}$
$\mathcal{T}$	Set of N stems in $\mathcal{A}$ that define the population and universe of inference
$\mathcal{T}_k$	Stem k in a population of N stems
$\mathcal{T}_k^{\mathbf{u}}$	Position of $\mathcal{T}_k$ $\mathbf{u} \in \mathbb{R}^2$
$\mathcal{T}_k^d$	Outside bark stem diameter at 1.3 m for $\mathcal{T}_k$ (i.e. DBH), in mm
$\mathcal{T}_k^h$	Total height for $\mathcal{T}_k$, in m
$\mathcal{T}_k^v$	Total volume for $\mathcal{T}_k$ ($\mathcal{T}_k^v = f(\mathcal{T}_k)$), in m^3
$\mathcal{T}_k^p$	The detection polygon for $\mathcal{T}_k$
$\mathcal{T}_k^{\pi_j}$	The inclusion probability for $\mathcal{T}_k$ as a function of $p_j(\cdot)$
$\mathcal{U}$	The set of unique labels for each $\mathcal{T}$

Inventory Yard Attributes

T	The number of periods in the planning horizon
J	The number of product classes in $\mathcal{I}$
$\mathcal{D}$	A T vector of service levels (i.e. demands)
$\mathcal{I}$	A $T \times J$ matrix representing the log yard inventory
$\mathcal{I}_{t,j}$	Product class $j \in (1, \ldots, J)$, in time period t
$\mathcal{I}_{t,\cdot}$	Total volume available, less salvaged volume $\mathcal{I}_{t,j} \in (1, \ldots, J-1)$, in time period t
$\mathcal{I}_{t,1}$	The replenishment volume resulting from harvesting clique $\mathcal{C}_t$ in period t
$\mathcal{I}_t^+$	The total volume stored (i.e. unsold) in period t
$\mathcal{I}_t^-$	Volume shortage in period t
$\mathcal{I}_t^{s\pm}$	Volume salvaged in period t
$\mathcal{I}_t^s$	Volume delivered or sold in period t

Plot Attributes

M	The number of sampling locations in $\mathcal{A}$
$\mathcal{P}$	Set of M sampling locations in $\mathcal{A}$
$\mathcal{P}_\ell$	Plot ℓ in a sample of M locations
$\mathcal{P}^{\mathbf{u}}$	Position of $\mathcal{P}_\ell$ $\mathbf{u} \in \mathbf{R}^2$
$\mathcal{P}_\ell^v$	A realization of a regionalized random process V, for the total volume of plot $\mathcal{P}_\ell$
$\mathcal{W}$	The set of unique labels for each $\mathcal{P}$

Estimator Formulation

$\mathcal{A}_b^{\hat{\bar{v}}}$ — The predicted spatial average density for $\mathcal{A}_b$

$\mathcal{A}_b^{\hat{v}}$ — The predicted value of $\mathcal{A}_b^v$

$\mathcal{A}_b^{\hat{\sigma}_v^2}$ — The prediction error variance for the predicted value of $\mathcal{A}_b^v$

$\mathcal{C}_t^{\hat{v}}$ — The predicted volume of the harvest clique in period t (i.e. $\sum_{i=1}^{\mathcal{C}_t^{|a|}} \mathcal{A}_b^v$)

$\mathcal{C}_t^{\hat{\sigma}_v^2}$ — The prediction error variance for the predicted value of $\mathcal{C}_b^v$

$\mathbf{I}_{k,\ell}(D,\theta_1,\theta_2)$ — An $N \times M$ indicator vector or matrix for sample inclusion of $\mathcal{T}_k$ at sample location $\mathcal{P}_\ell$

$p(\,\cdot\,)$ — A sample design function that assigns a probability of inclusion, π, to each element of $\mathcal{T}$

D — Minimum $\mathcal{T}^d$ to be included in fixed area plot or variable radius plot

θ_1 — Fixed area plot radius when $\mathcal{T}^d \leq D$

θ_2 — Angle gauge constant for variable radius plots when $\mathcal{T}^d > D$

$d_{k,\ell}$ — $N \times M$ distance matrix from $\mathcal{T}_k$ to $\mathcal{P}_\ell$

$\mathbf{h}$ — Lag or distance between $\mathcal{P}_i^{\mathbf{u}}$ and $\mathcal{P}_j^{\mathbf{u}}$

Operators and Miscellaneous Notation

$\mathbf{0}$ — A vector of zeros

$\mathbb{N}, \mathbb{R}, \mathbb{C}$ — The set of natural, real, and complex numbers

$\forall, \exists, \nexists, \in, \notin \subset$ — for all, exists, not exists, in, not in, and subset

$\vee, \wedge, \neg$ — logical OR, AND, and NEGATION operators

$\mathbf{f}_a \prec \mathbf{f}_b$ — Component-wise less-than vector inequality

$\mathbb{E}[\,\cdot\,], \mathbb{V}[\,\cdot\,], \mathbb{C}[\,\cdot\,]$ — Expected value, variance, and covariance operator

$\mathbb{G}_k[\,\cdot\,]$ — Heaviside operator (Osyczka, 2002)

$\mathbf{C}', \mathbf{c}'$ — transpose of matrix $\mathbf{C}$ or vector $\mathbf{c}$

$\mathbf{C}^{-1}$ — Inverse of $\mathbf{C}$

α — Coordinates for a location $(\alpha_1, \alpha_2) \in \mathcal{A}$

$\mathcal{F}$ — The feasible region

$\mathcal{S}$ — The search space $\subset \mathbb{R}^{r+\nu}$

5.3 PROBLEM FORMULATION

Let $\mathcal{A} \subset \mathbb{R}^2$, represent an operational harvest polygon with a finite boundary, where the surface area, measured in ha, is $\lambda(\mathcal{A})$. The polygon $\mathcal{A}$, is partitioned into A polygons (i.e. cutting blocks). Each cutting block $\mathcal{A}_b$ is defined by a unique label and $\mathcal{B} = \{1, \ldots, b, \ldots, A\}$ is the set of labels that is used to uniquely identify each cutting block. The total area of each cutting block, measured in hectares, is denoted as $\lambda(\mathcal{A}_b)$, where $\bigcup_{\mathcal{A}_b} = \mathcal{A}$ $(\mathcal{A}_b \subset \mathbb{R}^2)$, and $\bigcap_{\mathcal{A}_b} = \emptyset$ $\forall b \in (1, \ldots, B)$.

Let $\mathcal{D} \subset \mathbb{R}^T$, represent a non-stochastic vector of target service levels for a single product (i.e. log volume) for T periods. Service levels can represent customer demands for a single product, a vector of products, including categorical satisfaction variables (i.e. binary "satisfied=1 or not satisfied=0"), probabilities of operational success, and *any* combination thereof. These are typically defined by the ecosystem services manager (e.g. loads per day, pieces per shift, \$ per week, percentage of success in meeting production goals, residual stems per stand, etc.).

The polygon $\mathcal{A}$ contains a finite population of N stems denoted by the set $\mathcal{T}$. Each stem $\mathcal{T}_k$ is defined by a unique label and $\mathcal{U} = \{1, \ldots, k, \ldots, N\}$ is the set of labels that is used to uniquely identify each stem. The attributes for each stem includes a position $\mathcal{T}_k^{\mathbf{u}} \in \mathbb{R}^2$, a diameter $\mathcal{T}_k^d$, and volume $\mathcal{T}_k^v$ and are assumed to be known or measured without error. The elements of $\mathcal{T}$ within each block (i.e. actuals) are denoted $\mathcal{A}_b^{\mathcal{T}} = \mathcal{T}_k \cap \mathcal{A}_b$.

Let $\mathcal{P}$ represent a set of M potential measurement locations placed within $\mathcal{A}$, where each measurement location (i.e. point/plot) $\mathcal{P}_\ell$ is defined by a unique label and $\mathcal{W} = \{1, \ldots, \ell, \ldots, M\}$ is the set of labels that are used to uniquely identify each sampling location. For this examination, each sam-

pling location contains a fixed position vector $\mathcal{P}_\ell^{\mathbf{u}} \in \mathbb{R}^2$ and a volume density observation $\mathcal{P}_\ell^v$ (i.e. m^3 ha^{-1}) that is measured without error. Again, to reduce the notation, we shall write $\ell \in \mathcal{W}$ to represent all sampling locations in $\mathcal{A}$.

Let $\mathcal{I}$ be a $T \times J$ matrix $I_{t,j} \in \mathbb{R}$, where the T rows represent a vector set of J age classes that contain the quantity of available product, and its associated variance, in each of the j product age classes over T periods. The log-yard inventory represents the volume that has been harvested (i.e. $J = 1$) in period t, plus the volume left unsold (i.e. $J = 2, \ldots, J-1$), and the oldest unsold (i.e. salvage volume $\mathcal{I}_{t,J}$) is the sum of the unsold volume in each of the $J - 1$ class in period $t - 1$. The value of the residual volume in $\mathcal{A}$ as a result of any retention tree policy, remains unrealized. All salvage volume in realized at the end of period t where the salvage price represents the least valuable product (i.e. chips).

To obtain $\mathcal{A}_b^{\hat{v}}$ from a single pre-harvest survey, we assume $\mathcal{A}^{v(\alpha)}$ is a *regionalized variable or function* defined by a random process on the domain $\mathcal{A}$ for point sample support where $\mathcal{A}^{v(\alpha)}$ produces a density (i.e. m^3 ha^{-1}) at some point location $\alpha \in \mathcal{A}$ (Armstrong, 1998; Wackernagel, 1998; de Gruijter et al., 2006). This assumption accounts for both local irregularities (e.g. randomness) and a structured aspect (e.g. large scale tendencies) and that, for any distance $|\mathbf{h}| = |\mathbf{u}_i - \mathbf{u}_j|$, the distribution of the random variables $V(\mathbf{u}_1), V(\mathbf{u}_2), \ldots, V(\mathbf{u}_k)$ is assumed to be the same as $V(\mathbf{u}_1 + \mathbf{h}), V(\mathbf{u}_2 + \mathbf{h}), \ldots, V(\mathbf{u}_k + \mathbf{h})$ for the first two moments (i.e. constant mean and covariance) (Reed and Burkhart, 1985). These two assumptions, known as second-order stationarity, are critical in determining the optimal weights so that the prediction at some unsampled location $\mathbf{u}_0$, over a region $\mathcal{A}_b$, is unbiased ($\mathbb{E}[\mathcal{A}^{\hat{v}} - \mathcal{A}^v] = 0$) and the error or *prediction* variance ($\mathbb{V}[\mathcal{A}^{\hat{v}} - \mathcal{A}^v]$)

is minimum (Armstrong, 1998).

In forest operations, it is common to conduct a pre-harvest survey where the sampling unit is the individual stem $\mathcal{T}_k$ and all sample units that are observed, taken collectively, is referred to as a sample (Tillé, 2006). A general function $p(\,\cdot\,)$, commonly referred to as a sample design or plot design in forest sampling, assigns an inclusion probability to each stem $\mathcal{T}_k$ for each unique combination of $\mathcal{T}$ and $p(\,\cdot\,)$ (de Gruijter et al., 2006; Tillé, 2006). The weight used to obtain estimates or predictions of target parameters is then the inverse of the inclusion probability (Husch et al., 1982; Shiver and Borders, 1996). Common definitions of $p(\,\cdot\,)$ include "split-plot designs" (i.e. $p(\,\cdot\,) = f(D, \theta_1, \theta_2)$) where $\mathcal{T}$ is stratified into two populations based on some diameter D, so that when $T_k^d \leq D$, stems are included on a fixed-area plot, and when $T_k^d > D$, stems are included on a variable-radius or "prism" plot, where θ_1 is the fixed-area plot radius, and θ_2 is an angle gauge constant, or basal area factor, that defines the radius of the inclusion probability polygon associated with stem $\mathcal{T}_k$, when $T_k^d \geq D$.

Finally, given only the harvest unit boundary, $\mathcal{A}$, a vector of required service levels (i.e. demands), $\mathcal{D}$ and a single pre-harvest sample, $\mathcal{P}$, from $p(\,\cdot\,)$ and $\mathcal{T}$, a supply chain optimization problem can be expressed as a general non-linear constrained deterministic d-objective problem ($d \geq 1$) formulation (Karush, 1939; Kuhn and Tucker, 1951; Osyczka, 2002; Coello Coello, 2002):

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathcal{S}}{\operatorname{argmin}} \left\{ \mathbf{f}(\mathbf{x}) \in \mathbb{R}^d \middle| \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0} \right\} \tag{5.1}$$

where the goal is to obtain the arguments $\mathbf{x}^*$ that yield the minimum $\mathbf{f}$, a vector of d-objectives, subject to $\mathbf{g}$, a vector of K inequality constraints, and $\mathbf{h}$, a vector of M equality constraints and the fitness of any candidate solution $\mathbf{x}$ can be evaluated directly from the functions $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$.

The feasible region $\mathcal{F}$ is a subset of the entire search space $\mathcal{S} \subset \mathbb{R}^r$ and $\mathcal{F} \subseteq \mathcal{S}$. For an inequality constraint that satisfies $g_k(\mathbf{x}) = 0$, then the inequality k is active at $\mathbf{x}$. All equality constraints $h_m(\mathbf{x})$ are considered active at all values of $\mathcal{F}$, regardless of the value of $\mathbf{x}$ (Bishop, 2006).

For potentially competing and possibly non-commensurate objectives ($d \geq 2$), an objective vector $\mathbf{f}_a$ is said to *dominate* another objective vector $\mathbf{f}_b$, denoted by $\mathbf{f}_a \prec \mathbf{f}_b$ (component-wise), if and only if:

$$\mathbf{f}_{a,i} \leq \mathbf{f}_{b,i} \ \forall \ i \in \{1, \ldots, d\} \ \wedge \ \mathbf{f}_{a,j} < \mathbf{f}_{b,j} \ \exists \ j \in \{1, \ldots, d\} \tag{5.2}$$

In words, an objective vector is called *non-dominated* if there are no other objective vectors that can increase the value of any one of the d objective functions without decreasing the value of another of the d objective functions. The set of all *non-dominated* solutions is called the *Pareto set*, *Pareto front* or *efficient frontier* (Eiben and Smith, 2003).

When all of the inputs and outputs are non-stochastic, Equations (5.1) and (5.2) can be used to determine the Pareto optimal policies which satisfy the goals of the decision maker (Deb, 2001a). For this problem however, the inputs contain a combination of r fixed and known values (e.g. service levels, sample locations, and tree measurements) and ν estimates or predictions for fixed, but unknown population parameters (e.g. $\mathcal{A}^v$, $\mathcal{A}_b^v$, and variogram parameters), which can be expressed by a $r + \nu$ partitioned vector:

$$(\mathbf{x}, \xi) = (x_1, x_2, \ldots, x_r, \xi_{r+1}, \xi_{r+2}, \ldots, \xi_{r+\nu}) \tag{5.3}$$

where the input values $x_1, x_2, \ldots, x_r$ are the fixed and known datum and $\xi_{r+1}, \xi_{r+2}, \ldots, \xi_{r+\nu}$ are random variables, used to represent a guess for the last ν fixed, but unknown inputs, given the first r fixed and known values of the data. When an optimization problem contains fixed, but unknown

values, represented by a random vector for some subset of the inputs, the resulting formulation is a stochastic optimization problem (Gen and Cheng, 1997).

A stochastic optimization problem can then be expressed by letting $\xi = (\xi_{r+1}, \xi_{r+2}, \ldots, \xi_{r+\nu})$ be the random vector of fixed, but unknown values, with the joint probability mass function ϕ, defined by $\mathbf{x}$:

$$\phi_{\mathbf{x}}(\xi) = \phi_{x_1,\ldots,x_r}(\xi_{r+1}, \ldots, \xi_{r+\nu}) \tag{5.4}$$

Since the maximization or minimization of a random vector, or any resulting function, is meaningless, Equation (5.1), now a non-linear constrained stochastic d-objective ($d \geq 1$) minimization problem, expressed as:

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^{r+\nu}} \{\mathbb{E}[\mathbf{f}(\mathbf{x}, \xi)] \in \mathbb{R}^d | \mathbb{E}[\mathbf{g}(\mathbf{x}, \xi)] \geq \mathbf{0}, \mathbb{E}[\mathbf{h}(\mathbf{x}, \xi)] = \mathbf{0}\} \tag{5.5}$$

which can be solved using a deterministic-substitution formulation where the goal is to obtain risk-neutral (e.g. expected value), or risk-averse (e.g. minimum variance) optimal criterion (Marti, 2005), and $\mathbb{E}[\mathbf{f}(\mathbf{x}, \xi)]$, $\mathbb{E}[\mathbf{g}(\mathbf{x}, \xi)]$, and $\mathbb{E}[\mathbf{h}(\mathbf{x}, \xi)]$ are estimated from $\mathbf{x}$ and $\phi_{\mathbf{x}}(\xi)$ using *a priori*, sample, and structural information since $\mathbf{x}$ contains *all known* information available to the decision maker. The available information includes inputs such as stem measurements, the boundary of the harvest area, product yield predictions, general forest conditions, spatial relationships, current management policies, desired objectives, constraints, and *any* functions thereof.

For this problem, we are interested in risk-neutral solutions and assume that the expected values can be estimated without bias (Marti, 2005), which we describe below, and so, to conserve notation, we will express the deter-

ministic substitution problem using the general formulation:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^{r+\nu}}{\operatorname{argmin}} \left\{ \mathbf{f}(\mathbf{x}, \hat{\xi}) \in \mathbb{R}^d | \mathbf{g}(\mathbf{x}, \hat{\xi}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}, \hat{\xi}) = \mathbf{0} \right\} \qquad (5.6)$$

where $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$ are now defined as statistical or econometric models of the data (White, 1999; Konishi and Kitagawa, 2008).

Here, the elements of the objective vector, $\mathbf{f}$, are:

$$\mathbf{f}(\mathbf{x}, \hat{\xi}) = (f_1(\mathbf{x}, \hat{\xi}), f_2(\mathbf{x}, \hat{\xi}), f_3(\mathbf{x}, \hat{\xi})) \qquad (5.7)$$

where the first objective, f_1, is to minimize the constraint violations so that an objective vector, for any given candidate solution, is as close as possible to, or is within the feasible region (Osyczka, 2002). Specifically, the first objective function is expressed as:

$$f_1(\mathbf{x}, \hat{\xi}) \equiv \sum_{m=1}^{M} [h_m(\mathbf{x}, \hat{\xi})]^2 + \sum_{k=1}^{K} \mathbb{G}_k [g_k(\mathbf{x}, \hat{\xi})]^2 \qquad (5.8)$$

where $\mathbb{G}_k$ is defined as:

$$\mathbb{G}_k = \begin{cases} 0 & \text{for } g_k(\mathbf{x}, \hat{\xi}) \geq 0 \\ 1 & \text{for } g_k(\mathbf{x}, \hat{\xi}) < 0 \end{cases} \qquad (5.9)$$

so that when the predicted value of the first objective equals zero (i.e. $f_1(\mathbf{x}, \hat{\xi}) = 0$, the candidate solution yields an objective vector $\mathbf{f}(\mathbf{x}, \hat{\xi})$ within, or on the boundary of the feasible region $\mathcal{F}$.

The second objective function, f_2, is to maximize the unbiased, minimum prediction error variance net revenue for the operation, which we model as a continuous review "first-in, first-out" inventory, with stochastic supply (i.e. predicted harvest or replenishment volumes), given independent non-stochastic demands, with product storage costs for undelivered volume, and

perishable products must be realized upon expiration (i.e. salvage value), with no backlogging (i.e. shortages incur additional costs). For this problem, we assume the cycle length is a single day, over the finite planning horizon of T periods (Tersine, 1988; Hillier and Lieberman, 1995). The second objective function is expressed as:

$$
\begin{aligned}
f_2(\mathbf{x}, \hat{\xi}) &\equiv \sum_{t=1}^{T} \beta_1(t) \max\left(D_t, \sum_{j=1}^{J-1} \mathcal{I}_{t,j}\right) + \sum_{t=1}^{T} \beta_2(t) \max\left(0, \mathcal{I}_{t,J}\right) - \\
&\sum_{t=1}^{T} \beta_3(t) \max\left(0, \sum_{j=1}^{J-1} \mathcal{I}_{t,j} - D_t\right) - \sum_{t=1}^{T} \beta_4(t) \max\left(0, D_t - \sum_{j=1}^{J-1} \mathcal{I}_{t,j}\right) - \sum_{t=1}^{T} \beta_5(t) \mathcal{I}_{t,1}
\end{aligned}
$$

$$(5.10)$$

where $\beta_1(t)$ and $\beta_2(t)$ are the sales and salvage prices (i.e. $ m^{-3}), $\beta_3(t)$, $\beta_4(t)$, and $\beta_5(t)$ are the storage (i.e. $ m^{-3}d^{-1}), shortage costs (i.e. $ m^{-3}), and production costs (i.e. $ m^{-3}), D_t is the demand level in period t (i.e. m^{3}d^{-1}), $\sum_{j=1}^{J-1} \mathcal{I}_{t,j}$ is the volume available to fill the demand (i.e. m^3) from the $J-1$ product classes, $\mathcal{I}_{t,J}$ is the salvage product class volume (i.e. m^3), and $\mathcal{I}_{t,1}$ is the volume harvested in period t (i.e. m^3), resulting from the harvest of clique t. The values β are presented in Table 5.2.

The first and second terms in Equation (5.10) represent the volume sold and volume salvaged in period t, respectively. The first term represents the larger of the demand or if short, the volume currently within the available inventory that can be shipped, in period t. The second term represents the volume in age class J, period t that must be realized in period t as a consequence of product deterioration.

The last three terms represent the storage, shortage, and production costs. The third term in Equation (5.10) represents the volume that remains after meeting $\mathcal{D}_t$ (i.e. storage). If all of the demand is met, and there remains

volume in the inventory, the remaining volume must be stored for one more period (i.e. day). The fourth term represents the unmet demand in period t, and since our model does not permit back-logging (i.e. demand is lost), we must incur the cost of not meeting demand, in the form of potentially lost future sales, or demand is either met by another vendor (i.e. competitor), priority shipment costs, or general loss in revenue. The last term represents the production, or logging, costs which include the cost to fall, buck, yard, load and transport the volume from the stump in $\mathcal{A}_b$, to the mill (i.e. delivery point).

The third objective is to maximize the unharvested area in $\mathcal{A}$, expressed as:

$$f_3(\mathbf{x}, \hat{\tilde{\xi}}) \equiv \lambda(\mathcal{C}_T) \tag{5.11}$$

where $\lambda(\mathcal{C}_T)$ denotes the surface area (i.e. ha) for the cliques assigned to period T, the last period, which in this problem, remains unharvested.

To obtain the unbiased minimum prediction variance product volume in each potential cutting block, (i.e. $\mathcal{A}_b^{\hat{v}}$), $\mathcal{A}$ can be partitioned into $\mathcal{B}$ smaller potential cutting blocks. Then for each potential cutting block, a set of block-kriging equations for $\mathcal{A}_b^v$ can be expressed as:

$$\begin{bmatrix} \widehat{C}_{1,1} & \cdots & \widehat{C}_{1,M} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \widehat{C}_{M,1} & \cdots & \widehat{C}_{M,M} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{A}_b^{\omega_1} \\ \vdots \\ \mathcal{A}_b^{\omega_M} \\ \mathcal{A}_b^{\lambda} \end{bmatrix} - \begin{bmatrix} \widehat{C}_{1,M} \\ \vdots \\ \widehat{C}_{n,M} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \ \forall \ \mathcal{A}_b \in \mathcal{A} \tag{5.12}$$

or more compactly as:

$$\mathbf{C}\mathcal{A}_b^{\omega} - \mathbf{D} = \mathbf{0} \ \forall \ \mathcal{A}_b \in \mathcal{A} \tag{5.13}$$

where $\mathcal{A}_b^\omega$ is the vector of block- kriging weights for production block $\mathcal{A}_b$, λ_b is the sum of the kriging weights, which when $A_b^{\bar{v}}$ is unknown, must sum to one, $\mathbf{C}$ is the covariance matrix of the volume density realizations (i.e. observations) and $\mathbf{D}$ is the vector of the covariances at the sample points themselves, that is, $\mathbf{C}(\mathbf{h}) = 0$ where $\mathbf{h}$ is the distance between sample points, and $\mathbf{0}$ is a $n + 1$ vector of zeros (Isaaks and Srivastava, 1989; Armstrong, 1998; Armstrong and Champigny, 1989).

For each block in $\mathcal{A}$, a set of these constraints generates a system of $n+1$ equations that can be easily solved for $\mathbf{C}^{-1}$ to obtain the weights $\mathcal{A}_b^\omega$:

$$\mathcal{A}_b^\omega = \mathbf{C}^{-1}\mathbf{D} \; \forall \; \mathcal{A}_b \in \mathcal{A} \tag{5.14}$$

provided $\mathbf{C}$ is a positive definite function ($\mathbf{x}'\mathbf{A}\mathbf{x} > 0; \mathbf{A}, \mathbf{x} \in \mathbb{C}^n$), where the resulting values for $\mathcal{A}_b^\omega$ produce the unbiased predictions with the minimum prediction variance for $\mathcal{A}_b^{\hat{v}}$ (Cressie, 1993).

Finally, the predicted volume for a set of harvested cutting blocks (i.e. harvest clique) can then be expressed as:

$$\mathcal{C}_t^{\hat{v}} = \sum_{i=1}^{|\mathcal{C}_t|} A_i^{\hat{v}} \tag{5.15}$$

where $\mathcal{C}_t^{\hat{v}}$ is the total volume from the set of harvest cliques in period t, $|\mathcal{C}_t|$ represents the number of potential cutting blocks included in harvest clique t, which need not be contiguous, and $A_i^{\hat{v}}$ represents the unbiased, minimum prediction error predicted total volume in $\mathcal{A}_b$.

Using these predicted cutting block volumes, grouped together by a set of harvest cliques, a set of identity equations (i.e. equality constraints) can generated to represent the inventory replenishment, salvage, storage, and shortage, as:

$$\mathbf{h}_1(\mathbf{x}, \hat{\xi}) \equiv \qquad \mathcal{I}_{t,\cdot} - \sum_{j=1}^{J-1} \mathcal{I}_{t,j} \qquad = 0 \tag{5.16}$$

$$\mathbf{h}_2(\mathbf{x}, \hat{\xi}) \equiv \qquad \mathcal{I}_t^s - \max\left(D_t, \sum_{j=1}^{J-1} \mathcal{I}_{t,j}\right) \qquad = 0 \quad \forall t \in (1, \dots, T) \tag{5.17}$$

$$\mathbf{h}_3(\mathbf{x}, \hat{\xi}) \equiv \qquad \mathcal{I}_t^{s\pm} - \max\left(0, \mathcal{I}_{t,J}\right) \qquad = 0 \quad \forall t \in (1, \dots, T) \tag{5.18}$$

$$\mathbf{h}_4(\mathbf{x}, \hat{\xi}) \equiv \qquad \mathcal{I}_t^+ - \max\left(0, \sum_{j=1}^{J-1} \mathcal{I}_{t,j} - D_t\right) \qquad = 0 \quad \forall t \in (1, \dots, T) \tag{5.19}$$

$$\mathbf{h}_5(\mathbf{x}, \hat{\xi}) \equiv \qquad \mathcal{I}_t^- - \max\left(0, D_t - \sum_{j=1}^{J-1} \mathcal{I}_{t,j}\right) \qquad = 0 \quad \forall t \in (1, \dots, T) \tag{5.20}$$

$$\mathbf{h}_6(\mathbf{x}, \hat{\xi}) \equiv \quad \mathcal{I}_{t+j,j} = \begin{cases} \sum_T^{t=1} \sum_{J-1}^{j=1} \mathcal{I}_{t,1} & \text{if } i+j < T \\ 0 & \text{elsewhere} \end{cases} = 0 \quad \forall t \in (1, \dots, T) \tag{5.21}$$

where Equation (5.16) defines the log-yard inventory available for shipping to the customer. Equations (5.17), (5.18), (5.19), and (5.20) are described above. Equation (5.21) describes the inventory level after the volume available has been shipped.

To meet the criteria that the initial harvest clique contains a potential cutting block on the boundary of $\mathcal{A}$, and that subsequent harvest cliques are adjacent to the opening, defined by $\cup_{i=1}^t \mathcal{C}_i$, an adjacency list for each potential cutting block is required. Here, we define adjacency as strong adjacency, which requires two adjacent potential cutting blocks share more than a single common point. The equation to ensure the harvest cliques meet the constraints can be expressed as:

$$\mathbf{g}_1(x, \hat{\xi}) \equiv \quad 3|\mathcal{C}_t| > \sum_{j=1}^{|\mathcal{C}_t|} \mathcal{A}_j^d \quad = 0 \tag{5.22}$$

where $|\mathcal{C}_t|$ defines the harvest clique in period t, and the coefficient 3 represents a maximum number of adjacent cutting blocks, developed from a triangulation (described below).

Finally, an adjacency constraint to restrict the cliques that border the current opening is required so that all previous harvested cliques are combined into a single "opening" clique. For each period, the "opening" clique was then created as the sum of the cutting blocks from the previously harvested cliques.

5.4   DATA AND METHODS

Daily Demands
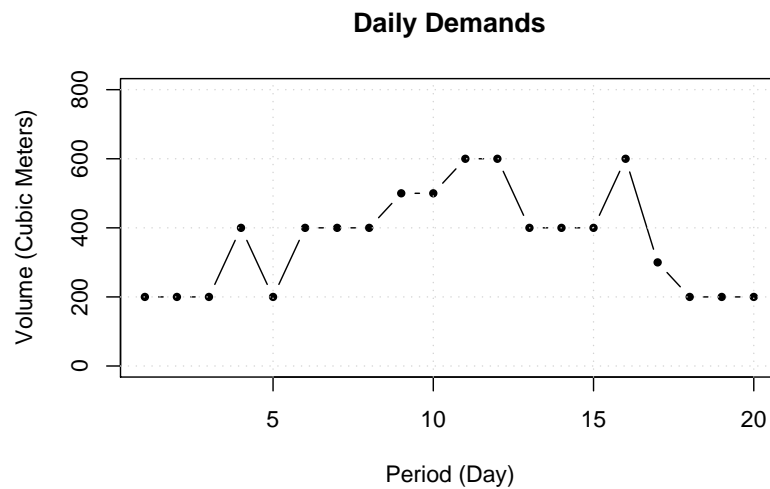
Figure. 5.1: Stem map of the "Extendo" harvest unit (i.e. $\mathcal{T} \cap \mathcal{A}$) with the diameter of the stems (i.e. $\mathcal{T}_k^d$) represented with circles of proportional size.

To examine the method for obtaining a set of optimal harvest plans, and thus approximate the Pareto frontier between the net revenue and the area left unharvested, a bounded 8.24 ha polygon, denoted $\mathcal{A}$, was located on the Oregon State University McDonald-Dunn Research Forest (Figure 5.1). The
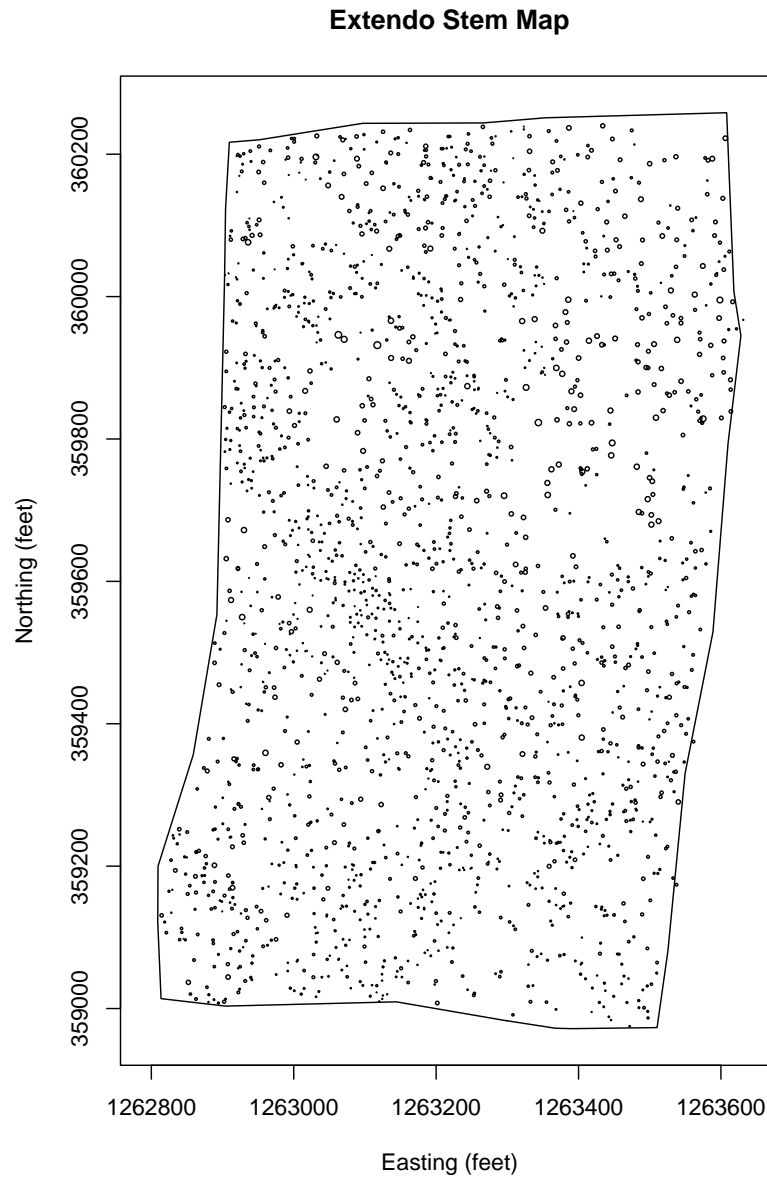
Extendo Stem Map

Figure. 5.2: The "Extendo" stem map displaying the production blocks, the locations of the pre-harvest sample. Boundary cutting blocks are light grey, with the cutting blocks that are adjacent to only one block, shaded in dark grey. Non-shaded (i.e. interior blocks) cannot be harvested first, but can be included in a harvest clique that also includes the boundary blocks.

boundary for $\mathcal{A}$ and the stem positions, $\mathcal{T}^{\mathbf{u}} \,\forall\, k \in \mathcal{U}$, for all stems over 15 cm, were recorded using standard survey methods (Davis et al., 1981). Attributes collected for each $\mathcal{T}_k$ included the species, $\mathcal{T}_k^s$, the diameter at breast height $\mathcal{T}_k^d$, the total height, $\mathcal{T}_k^h$, the height to live crown $\mathcal{T}_k^c$.

To determine the volume for each stem, $\mathcal{T}_k^v$, all stems were "merchandised" into log lengths of 4 m, to a 2 cm top, using the taper equation presented by Kozak et al. (1968). If the stem could not be cut into a round number of standard log lengths, any remaining stem length was bucked into a short log no less than 2 m in length. Stems were merchandised with a stump height of 0.3 m and each log included 0.2 m of trim. For each log in the stem, the starting and ending height of the log, the nominal length and the actual length (nominal length plus trim), the small and large end diameters, and the Smalian volume were recorded. For this examination, the Smalian volume from each of the logs were then tallied and assigned to $\mathcal{T}_k^v$.

For this study, only Douglas-fir (*psudeotsuga mensezii* (Mirb. Franco)) stems were used (i.e. $\mathcal{T}_k^s \in \{DF\}$; $k \in \mathcal{U}$) resulting in a final count of 2053 stems with a spatial mean density (i.e. $\mathcal{A}^{\mathcal{T}} \lambda(\mathcal{A})^{-1}$ of 249.17 stems ha^{-1} . The total volume for the study area $\mathcal{A}^v$, was 7560.07 m^3 yielding an average stem volume of 3.68 m^3 and a spatial mean volume density of 917.57 m^3 ha^{-1}, denoted $\mathcal{A}^{\bar{v}}$.

We simulated the installation of a single pre-harvest sample from $\mathcal{A}$ where the goal was to obtain the maximum constrained entropy sample (i.e. *D*-optimal experimental design sample (see Lee, 1998)), where the sample locations were selected from a set of potential sample locations. The set of potential sample locations were obtained using a non-aligned grid pattern (Figure 5.2). We used the `spsample` function in the `sp` package (Pebesma and Bivand, 2005), to determine the sample locations (i.e. $P_\ell^{\mathbf{u}} \,\forall\, \ell \in \mathcal{W}$).

The sample design for this examination, $p(\,\cdot\,) \equiv p(D = 20, \theta_1 = 5.64, \theta_2 = 4)$, meant that all detectable stems under 20 cm were tallied on a circular fixed-area plot with a radius of 5.64 m, and all detectable stems over 20 cm DBH where tallied using a BAF of 4 m^{3}ha^{-1}.

A total of 38 plots were installed with the predicted total volume of 7350.35 m^3, with a minimized prediction error variance of 7350.35 resulting in a bias of 209.72 m^3. The mean volume density, denoted $\mathcal{A}_b^{\hat{\bar{v}}}$, corrected for "edge-bias" using the method described by (Beers, 1966), was 858.76 m^3 ha^{-1}.

The boundaries for the set of potential cutting blocks were developed using Delaunay triangulation (Shewchuk, 1996; Okabe et al., 2000) which generated 67 potential cutting blocks to be grouped into the harvest cliques. We then used irregular block-kriging (Armstrong, 1998) on the resulting set of potential cutting blocks, using a 12 m $\times$ 20 m grid of prediction locations in $\mathcal{A}$ where the prediction grid was intersected with boundary of each $\mathcal{A}_b$, in order to compute $\mathcal{A}_b^{\hat{\bar{v}}}$, the predicted spatial average density for each potential cutting block (Montes et al., 2005). For $\mathbf{C}$, we assumed an isotropic two-parameter exponential theoretical variogram model:

$$\gamma(\mathbf{h}) = c\Big(1 - \exp\Big(\frac{-|\mathbf{h}|}{a}\Big)\Big) \tag{5.23}$$

where $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$, $\mathbf{C}$ is the covariance between two sample locations P_i^v and P_j^v (Goovaerts, 1997), and a and c represent unknown, but fixed population parameters (Cressie, 1993) to be estimated from the single pre-harvest sample $\mathcal{P}$. We used a maximum range of 1800 m and a distance interval width of 100 m. The variogram parameter estimates and fit statistics are presented in Table 5.2. Once the value of $\mathcal{A}_b^{\hat{\bar{v}}}$ was obtained, $\mathcal{A}_b^{\hat{v}}$ was computed as the product of $\mathcal{A}_b^{\bar{v}} \cdot \lambda(\mathcal{A}_b)$. The demand vector, $\mathcal{D}$, presented graphically in Figure 5.3, required a total of 7350 m^3.

All calculations to obtain the unknown inputs (i.e. $\hat{\xi}$) were performed prior to solving the operational plans, using R (Ihaka and Gentleman, 1996). The sample locations obtained using the `spsample` function in the `sp` package (Pebesma and Bivand, 2005), the grid generation function `gridcentres` from the `spatstat` package was used and for the variogram estimation and kriging prediction functions, we used the `gstat` package (Pebesma, 2004).

5.5 HEURISTIC DESCRIPTION

While exact solution methods for spatially explicit planning formulations have been presented for problems with two objectives (Tóth et al., 2006), examples of exact solution methods for complex problems are rare. For multiple objective formulations, heuristic methods such as Tabu Search (TS) (Hansen, 1997), Simulated Annealing (SA) (Deusen, 1999; Nam and Park, 2000), Genetic Algorithms (GA), (Ducheyne et al., 2004), and Evolutionary Algorithms (Stewart et al., 2004; Ducheyne et al., 2006) where $d \geq 2$, while rare in forest planning, show promise.

We used a class of heuristics known as Evolutionary Algorithms (EA), inspired from Darwinian evolution, which selects, mutates and promotes candidate solutions based on competition, fitness, and reproductive success (Eiben and Smith, 2003; Falcão and Borges, 2001). The major difference between GA and EA results from the use of the mutation operator. In GA, the crossover operator is used to promote both diversification and intensification, whereas EAs rely on the mutation operator alone to change mutation parameters depending on a variety of metrics such as convergence (Igel et al., 2007), stopping (Laumanns et al., 2002), diversification (Farhang-Mehr and Azarm, 2002), and optimality (Deb et al., 2007) criteria.
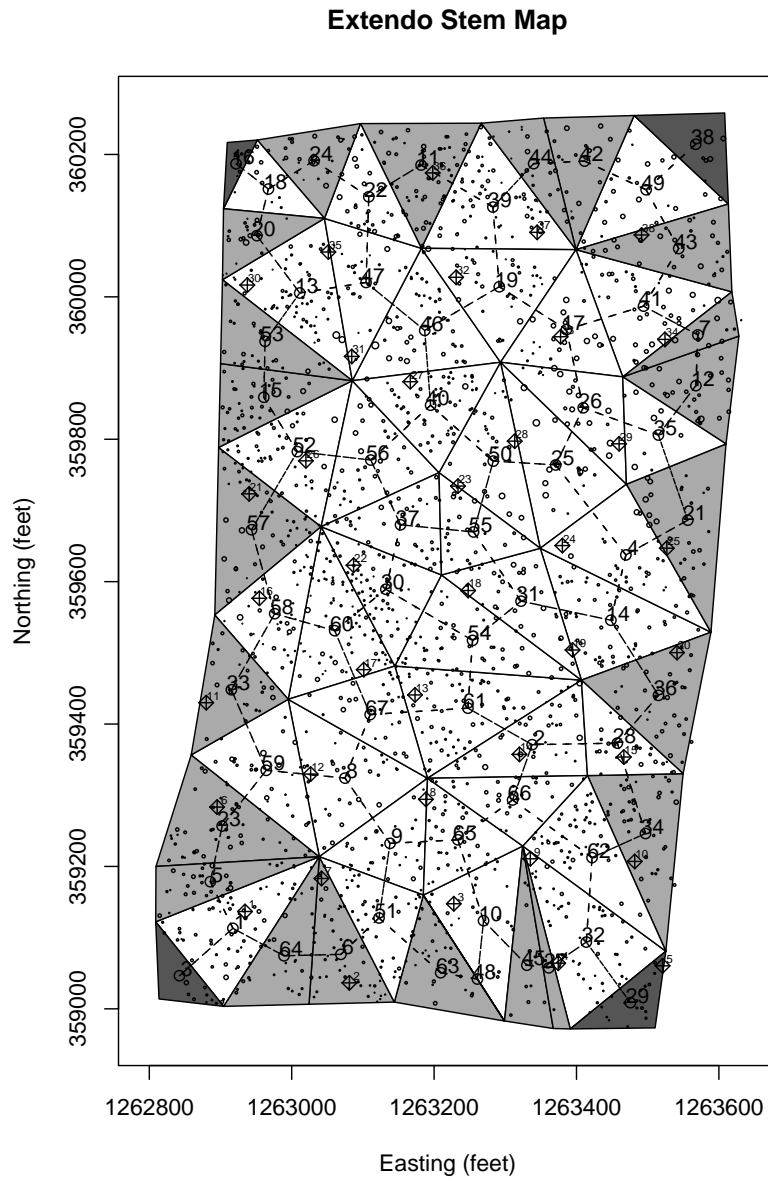
Extendo Stem Map

Figure. 5.3: The vector of demands $\mathcal{D}$.

Several comprehensive reviews of multi-objective evolutionary algorithms are available from Coello Coello (2002), Tan et al. (2005), and Deb (2001a) and all suggest that evolutionary algorithms have many advantages for high dimensional vector valued problems with potentially highly disconnected, non-uniformly distributed, and concave Pareto frontiers. In forest planning however, examination of EA for single objective (Falcão and Borges, 2001) and multi-objective (Ducheyne et al., 2004) formulations, while rare, show promise. To examine the applicability, behavior, and performance of a generalized multi-objective evolutionary algorithm (MOEA), the evolutionary algorithm presented by (Knowles and Corne, 2000) was used to solve the formulation in Section 5.3.

The $(\mu + \lambda)$-Pareto Archive Evolution Strategy $((\mu + \lambda)$-PAES), requires no assumptions about the decision maker or their preferences in the form of weights or scaling factors to reduce the solution space to a scalar function (Steuer, 1985). The method, presented in Algorithm 3, is an elitest strategy, which guarantees convergence (Rudolph and Agapie, 2000), is relatively simple to program, generates λ new mutations at each generation, maintains a set of μ non-dominated solutions throughout the search, and has been shown to work well for multitude of Pareto front types (Tan et al., 2005).

The initial population of λ candidate solutions are generated randomly, potentially repaired, and evaluated using the formulations presented in Section 5.3. The set of initial λ candidate solutions are then added to the Pareto archive (i.e. $\mu = \lambda$). A scalar rank is then assigned to each candidate solution, where the rank is defined as the number of $\mu + \lambda$ individuals that dominate the candidate solution. When there are no solutions that dominate the candidate solution, (i.e. non-dominated) a rank of 0 is assigned to the candidate solution. All candidate solutions with $rank > 0$ are then

Data: $\mathbf{x}, \hat{\xi}, \mathbf{f}(\mathbf{x}, \hat{\xi}), \mathbf{g}(\mathbf{x}, \hat{\xi}), \mathbf{h}(\mathbf{x}, \hat{\xi})$
Data: yield streams, adjacency list(matrix), constraint formulae,
      software
Result: $(\mu+\lambda)$-PAES Multi-Objective Evolutionary Algorithm
      (Knowles and Corne, 2000)
$g \leftarrow 1;$
$\mu_g = \emptyset;$
$\lambda_g \leftarrow GenerateRandom();$
$Evaluate(\lambda_g);$
$\mu_g \leftarrow UpdateParetoArchive(\lambda_g);$
for $g \leftarrow 2$ **to** G **do**
    | $\lambda_g \leftarrow SelectCandidatesFromArchive(\mu_{g-1});$
    | $\lambda_g \leftarrow Mutate(\lambda_g);$
    | $Evaluate(\lambda_g);$
    | $\mu_g \leftarrow UpdateParetoArchive(\lambda_g);$
end
$ExportArchive(\mu_G);$

Algorithm 3: Pareto archiving multi-objective evolutionary algorithm.

removed from the archive (i.e. dominated). The remaining non-dominated (μ) candidate solutions, which define the current Pareto frontier, are then selected for promotion and further mutation.

Selection for promotion is performed using tournament selection where a set of h candidate solutions enter a tournament and a single winner is selected with probability p (Dumitrescu et al., 2000). We used deterministic binary tournaments ($h = 2, p = 1$), where two solutions are selected with uniform probability from the Pareto archive (i.e. uniform, $P(Sel = 1) = 1/|\mu|$), and one of two candidate solutions is selected, with a probability of one, to be copied into the population buffer (λ) for promotion and mutation. The number of individual objectives that dominated the tournament for each candidate solution determined the winner with probability of one. In the event of a tie (i.e. weakly non-dominated), the first solution is selected with

a probability of $\frac{1}{2}$. The process of selection, promotion, and mutation is performed until some stopping criteria is met. For this manuscript, we used a deterministic stopping criteria defined by the number of generations (Deb, 2001a).

Since a portion of the *input data* for this problem is a constant (fixed either known or unknown) regardless of the values of the decision variables needed to evaluate a specific candidate solution, the decision vector included only variables for the harvest period for $\mathcal{A}_b$:

$$\mathbf{x}_c = \langle d_1, \ldots, d_i, \ldots, d_B, \sigma \rangle \tag{5.24}$$

where $d_1, \ldots, d_i, \ldots, d_B$ represents the harvest/reserve day in which to assign cutting block i and cutting blocks assigned to the reserve period (i.e. last day) $d_i = T$. The last entry in the candidate solution, σ represents a mutation strategy.

We implemented two mutation strategies (i.e. $\sigma \in$ (RANDOM_CLUSTERING, RANDOM_RESTART)). The first mutation strategy was to generate a completely random solution, beginning with a new clustering of the potential cutting blocks and to reassign the harvest cliques beginning with period one. The second mutation strategy was to re-assign harvest cliques $t+1, \ldots, T$, given a random starting period t where the harvested cliques were fixed up to period t. We used Baldwinian repair (Ishibuchi et al., 2005), so the candidate solution was feasible regarding the adjacency constraint, described in Equation (5.22). Finally, to examine the performance of the search, we ran the heuristic for 5000 generations, 184 times using the strategy parameters defined in Table 5.1.

Parameter	Specification
Representation	Real/Integer-valued vector
Population size	7μ Candidate solutions
Generations	5000
Recombination	None
Mutation	Random clustering and random starts
Parent Selection	Uniform random
Tournament type	Binary deterministic (k=2,p=1)
Survivor Selection	$(\mu + \lambda)$-PAES
Specialty	Baldwinian repair
Stopping Criteria	5000 generations

Table. 5.1: Evolution strategy parameters.

Input or Parameter	Value	Units
T	21	days
J	5	days
$\mathcal{D}$	$(\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_{T-1})$	m^3
$\mathcal{D}_T$	$\lambda(\mathcal{C}_T)$	ha
D	20	cm
θ_1	5.64	m
θ_2	4	m^{-2}ha^{-1}
a	344149.7	m^2
c	800	m
Sales Price	60.0	\$ m^{-3}
Salvage Price	5.0	\$ m^{-3}
Production Costs	20.0	\$ m^{-3}
Storage Costs	3.0	\$ m^{-3}d^{-1}
Shortage Costs	40.0	\$ m^{-3}

Table. 5.2: Inputs.

5.6   RESULTS AND DISCUSSION

5.6.1   Heuristic Performance

To examine the behavior and performance of the heuristic, we performed 100 replications, and examined the box-and-whisker plots of the objective function distributions as as suggested by Deb (2001b). While others have presented much more complete heuristic performance metrics for Pareto set convergence (Igel et al., 2007), stopping (Laumanns et al., 2002), diversification (Farhang-Mehr and Azarm, 2002), and optimality (Deb et al., 2007) criteria, we limit our discussion to visual examination of the box-and-whisker plots of the maximum objective function values, presented in Figure 5.4. Here, we only present the first 100 generations for brevity.

During the early stages of the search, both the median net revenue and the maximum unharvested area increased rapidly, and then slowed to a steady increase over the 5000 generations. The variation for both the maximum net revenue and maximum unharvested area were relatively equal throughout the entire search for each of the respective objectives. The variance of the net revenue, however, was larger than the variance for the maximum unharvested area as can be seen by the limits of the first and third quartiles in Figure 5.4. The maximum value of the net revenue (i.e. the top whisker), in each generation, increased much more rapidly than did the mean, whereas the maximum for the maximum unharvested area increased more proportional to the mean of the runs.

While the mutation rate and population size parameters are typically the variables of interest in evolutionary computation research (Bartz-Beielstein, 2006), we fixed the mutation rate, mutation types, and population sizes (i.e. 7μ) each run to: 1) minimize potentially confounding factors that might arise

Maximum Net Revenue Objective
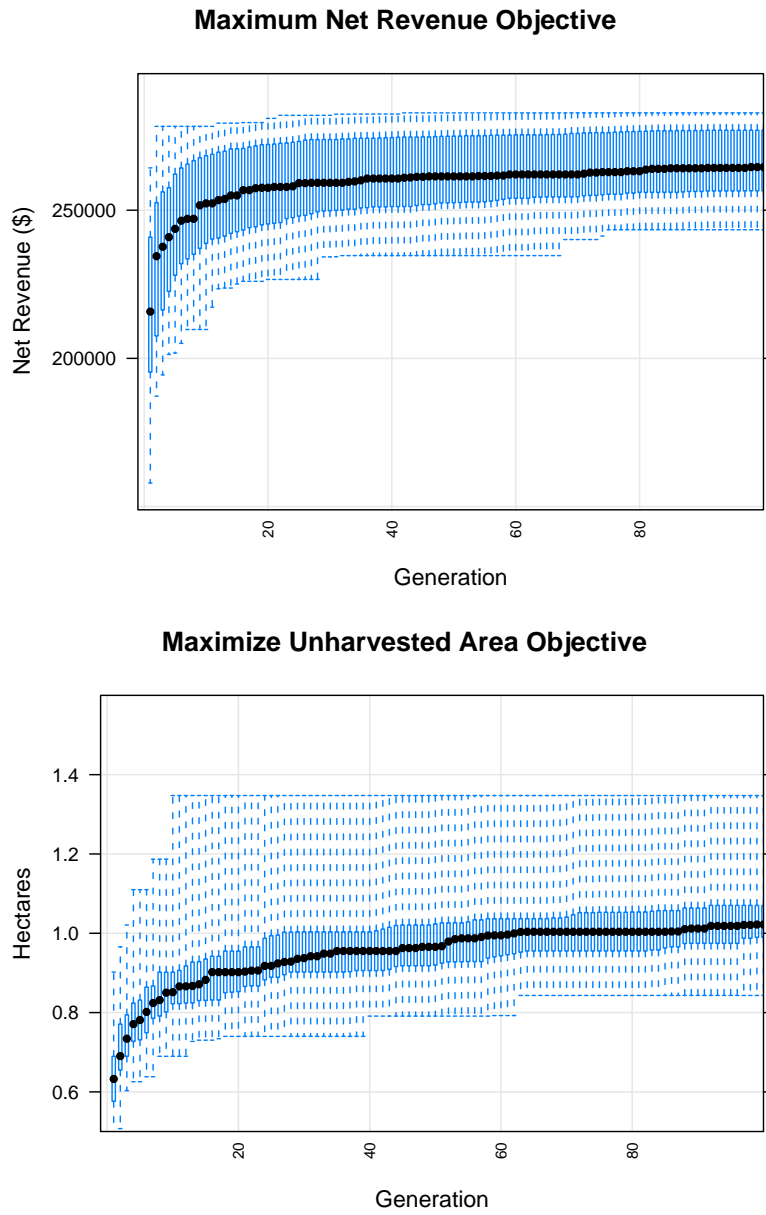

Maximize Unharvested Area Objective

Figure. 5.4: Heuristic performance plot.

from dynamically modifying the mutation rate and population sizes, and 2) because, for a given population size, there are endless methods in which the mutation rate could have been modified.

5.6.2   Operational Plans

To compare the tradeoffs between maximizing net revenue and maximizing the unharvested area, which represents the two extremes of the Pareto set, we present a detailed discussion of the Pareto solution that maximizes net revenue, the Pareto solution that maximized the unharvested area, and for comparison, we include a compromise between the two, representing a harvest pattern similar to current practices.

Since our goal was to determine if increasing the resolution of the decision making process increased our ability to make better decisions by detect salient properties of the solution space (i.e. increase net revenue, the unharvested area, or both), we examined specific differences among solutions for a single Pareto set by selecting the replication with the highest number of discovered Pareto solutions. Figure 5.5 presents the resulting Pareto frontiers from the 100 replications, and the selected Pareto frontier with the associated spatially explicit harvest plans for the extremes of the frontier. Figure 5.6 graphically presents the log yard inventory dynamics (i.e. cumulative net revenue, shipped volume, stored volume, and cumulative shortages) over the planning horizon.

The traditional harvest pattern, based on the spatial mean (i.e. $\mathcal{A}^{\bar{v}}\lambda(\mathcal{A})$), without the incorporation of spatial correlation in the predicted volumes, is presented in the lower left map in Figure 5.5. The plan, developed assuming the volume density is uniform throughout $\mathcal{A}$, begins in the lower left corner,
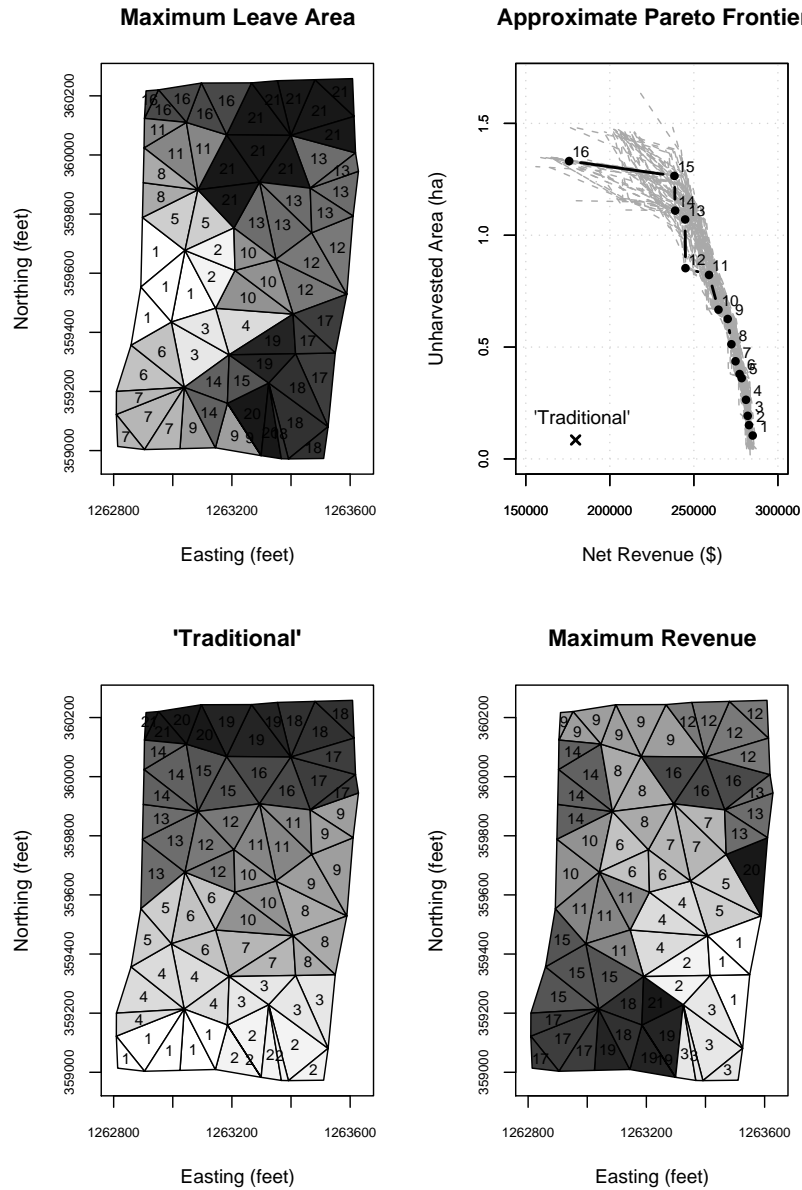
Figure. 5.5: Pareto optimal precision production plan that emphasizes maximum net revenue. The numbers in each $\mathcal{A}_b$ represent the harvest day and a value of 21 (darkest) represents areas to remain unharvested (i.e. reserves).
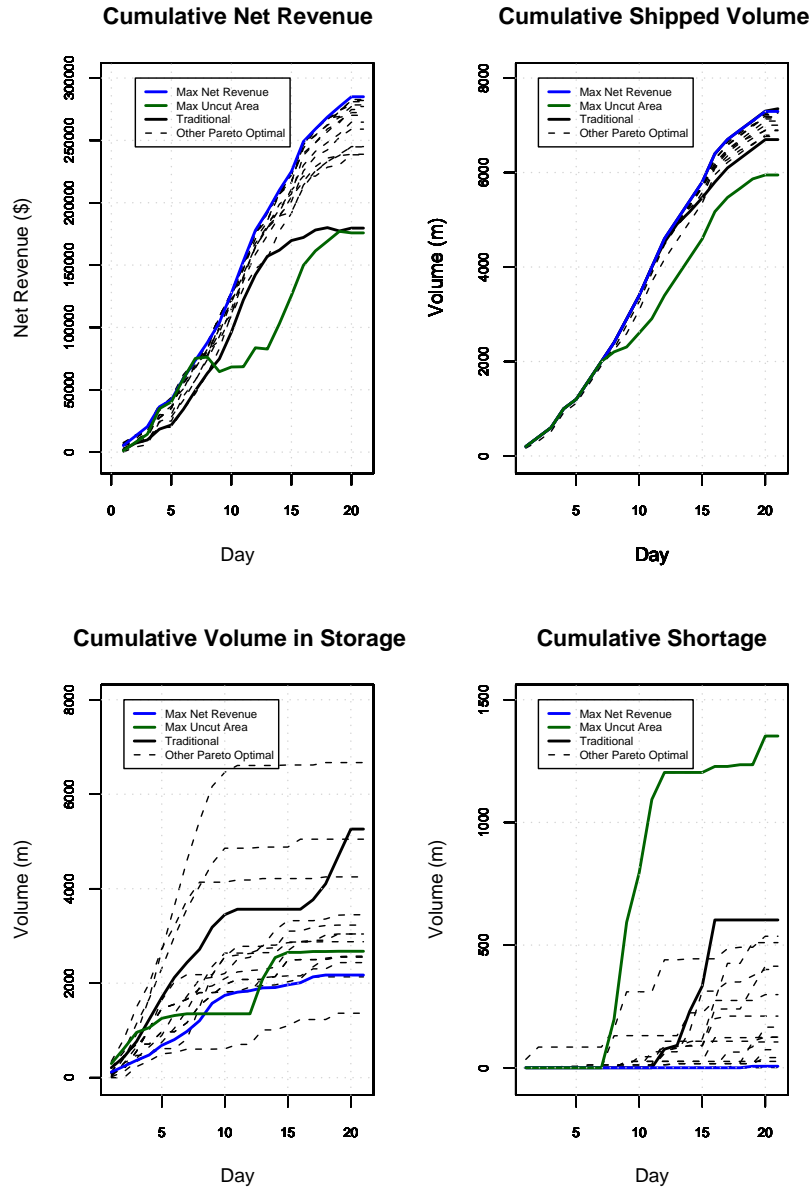
Figure. 5.6: Cumulative net revenue, delivered volume, log-yard inventory, and shortage for the selected Pareto set.

moving from left to right, reversing direction and repeating the process in the opposite direction until. The plan, typical of current practices, leaves the unharvested area in the northeast corner of $\mathcal{A}$. The traditional harvest pattern results in a net revenue of \$ 179,634.56 with an unharvested area of 0.08 ha.

In Figure 5.6, the levels of inventory for the traditional plan are represented as the thick solid black line. The cumulative net revenue for the traditional plan follows the same pattern as the maximum net revenue plan until the mid-point of the planning horizon. The traditional plan then diverges from the maximum revenue plan, as storage costs (lower left Figure 5.6) increases over the planning horizon. The cumulative volume shipped also closely matches the maximum net revenue plan until the mid point of the planning horizon, then falls short as volume is salvaged as a consequence of building a large log-yard inventory which cannot be sold after product degradation.

The harvest plan that maximized net revenue (Plan 1) resulted in a net revenue of \$ 284,992.91 with an unharvested area of 0.1 ha, yielding an increase in net revenue over the traditional plan of 58.65 percent and an increase in reserve area of 23.54 percent. The plan begins in the lower right of the harvest unit, and follows a pattern than harvests the center diagonal of the unit, leaving the south west and north east corners for the second half of the operation. The cumulative net revenue remained the highest throughout the planning horizon as did the cumulative shipped volume. As would be desirable to the log-yard manager, marketing, and financial managers the plan that maximized the net revenue maintained the lowest log-yard inventory and the lowest cumulative volume shortage as well.

The harvest plan that maximized the unharvested area (i.e. Plan 16), as

might desired by the environmental affairs manager attempting the focus on societal goals, is presented in the upper left corner of Figure 5.5, and by the thick light grey solid line in Figure 5.6. Plan 16) resulted in an unharvested area of 1.33 ha yielding an increase in unharvested area over the traditional plan of 14.73 times. Interestingly, the net revenue for Plan 16) was 175,800.71 resulting in decrease in net revenue of -2.13 percent.

The log-yard inventory for the Pareto operations (Figure 5.6), shows that both the maximum unharvested area and the traditional harvest plan both had larger log-yard inventories (i.e. cumulative volume in storage), higher cumulative shortages, and less cumulative shipped volumes than did the maximum net revenue plan, suggesting that when there are competing objectives, the maximization of net revenue can simultaneously increase net revenue, while maintaining the current unharvested area as well as the objectives of the other managers. The maximum unharvested area also maintains the same net revenue curve as the maximum net revenue, until the log-yard manager, unable to change "fate", is forced to then incur shortages, fall behind in shipments, and must maintain the largest log-yard inventory, making this maximization of that objective questionable if the organization wants to remain competitive against a traditional operating plan.

Finally, when some intermediate option is selected for implementation (i.e. more unharvested area and less net revenue), increases in shortages and log-yard inventories, as well as decreases in net revenue and cumulative shipped volume are possible as is evident by the dashed lines in Figure 5.6. These alternatives, when examined simultaneously against each of the other plans, yields a more informed decision regarding the competing objectives than does current multi-criteria decision making practices.

5.7   CONCLUSIONS

We presented a precision planning method to examine the trade-offs among potentially competing and non-commensurate objectives in forest operations, using a mathematical model of the resource, a heuristic search technique, and visual examination of the resulting plans from which we can select. By making certain assumptions regarding spatial correlation within the population of sample plots installed as part of the pre-harvest survey, we were able to compare the resulting set of Pareto optimal operational plans against a traditional operational plan and found that by increasing the resolution of the input data, assuming a simple spatial model of the volume within the harvest unit, and efficiently exploring the solution space, we were able to increase both objective function values over the values resulting from a traditional plan as well as provide an objective assessment of the various management options.

The method identified a set of plans that optimized the production schedule to provide stable production (e.g. logging production manager objectives), maintain a small inventory, thus minimizing product degradation, ensuring adequate stocking levels (e.g. log-yard manager objectives). The pareto optimal plans also allow the production manager to meet the cumulative customer service levels, minimize, or eliminate, shortages while maintaining ample stocks to meet service levels (e.g. marketing manager objectives), and maintain a sufficient unharvested area similar to or better than traditional practices (e.g. environmental affairs manager objectives).

While the problem examined here was small, the advantages of this method can be examined for larger problems, problems with multiple harvest areas, and for problems with additional metrics are included in the objective vector

(e.g. financial, ecological, and social objectives). Finally, as our focus was on the problem formulation and not the solution method specifically, we realize the need to examine mutation operators, selection operators, and methods to visualize Pareto frontiers in forest planning problems is large indeed, as there are many complex relationships among input data, decision variables, and policies defined by $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$ and is the subject of future research.

BIBLIOGRAPHY

Armstrong, M. (1998). *Basic Linear Geostatistics.* Springer.

Armstrong, M. and Champigny, N. (1989). A study on kriging on small blocks. *Canadian Institute of Mining Bulletin,* 82(923):128–133.

Bartz-Beielstein, T. (2006). *Experimental Research in Evolutionary Computation: The New Experimentalism (Natural Computing Series).* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Beers, T. W. (1966). The direct correction for boundary-line slopover in horizontal point sampling. Research Progress Report 224, Purdue University, Agricultural Experiment Station, Lafayette, Indiana.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Cochran, W. G. (1977). *Sampling techniques.* John Wiley & Sons, New York. 428 p.

Coello Coello, C. A. (2002). Evolutionary multi-objective optimization: A critical review. In Sarker, R., Mohammadian, M., and Yao, X., editors, *Evolutionary Optimization,* chapter 5. Kluwer Academic Publishers, Norwell, MA, USA.

Cressie, N. A. C. (1993). *Statistics for Spatial Data,* chapter 2. John Wiley and Sons, Inc., New York, New York, revised edition edition. 900.

Davis, R. E., Foote, F. S., Anderson, J. M., and Mikhail, E. M. (1981). *Surveying, theory and practice.* McGraw-Hill, Inc.

de Gruijter, J. J., Brus, D. J., Bierkins, M. F. P., and Knotters, M. (2006). *Sampling for Natural Resource Monitoring.* Springer-Verlag.

Deb, K. (2001a). *Multi-Objective Optimization using Evolutionary Algorithms.* John Wiley & Sons, Chichester, UK. ISBN 0-471-87339-X.

Deb, K. (2001b). *Multi-Objective Optimization using Evolutionary Algorithms.* Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Chichester.

Deb, K., Tewari, R., Dixit, M., and Dutta, J. (2007). Finding Trade-off Solutions Close to KKT Points Using Evolutionary Multi-Objective Optimization. KanGAL Report 2007006, Indian Institute of Technology Kanpur, Kanpur, PIN 208016, India.

Deusen, P. C. V. (1999). Multiple solution harvest scheduling. *Silva Fennica*, 33(3):207–216.

Diestel, R. (2005). *Graph Theory (Graduate Texts in Mathematics)*. Springer.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2004). Even flow scheduling problems in forest management. In Coello, C. A. C. and Lamont, G. B., editors, *Applications of Multi-Objective Evolutionary Algorithms*, pages 701–726. World Scientific, Singapore.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2006). A spatial approach to forest-management optimization: linking gis and multiple objective genetic algorithms. *International Journal of Geographical Information Science*, 20(8):917–928.

Dumitrescu, D., Lazzerini, B., Jain, L., and Dumitrescu, A. (2000). *Evolutionary Computation*. CRC Press.

Eiben, A. E. and Smith, J. E. (2003). *Introduction to Evolutionary Computing*. SpringerVerlag.

Falcão, A. O. and Borges, J. G. (2001). Designing an Evolution Program for Solving Integer Forest Management Scheduling Models: An Application in Portugal. *Forest Science*, 47(2):158–168.

Farhang-Mehr, A. and Azarm, S. (2002). Diversity assessment of pareto optimal solution sets: An entropy approach. In *Congress on Evolutionary Computation (CEC'2002)*, volume 1, pages 723–728, Piscataway, New Jersey. IEEE Service Center.

Gen, M. and Cheng, R. (1997). *Genetic Algorithms and Engineering Design*. John Wiley and Sons, Inc., New York.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics Series. Oxford University Press, New York, New York.

Hansen, M. P. (1997). Tabu Search in Multiobjective Optimisation : MOTS. In *Proceedings of the 13th International Conference on Multiple Criteria Decision Making (MCDM'97)*, Cape Town, South Africa.

Hillier, F. S. and Lieberman, G. J. (1995). *Introduction to Operations Research*. McGraw-Hill, Inc., sixth edition.

Husch, B., Miller, C. I., and Beers, T. W. (1982). *Forest Mensuration.* John Wiley and Sons, Inc., third edition.

Igel, C., Hansen, N., and Roth, S. (2007). Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15(1):1–28.

Ihaka, R. and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics.* Oxford University Press, Inc., New York, New York.

Ishibuchi, H., Kaige, S., and Narukawa, K. (2005). *Evolutionary Multi-Criterion Optimization*, volume 3410 of *Lecture Notes in Computer Science*, chapter Comparison Between Lamarckian and Baldwinian Repair on Multiobjective 0/1 Knapsack Problems, pages 370–385. Springer.

Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. Master's thesis, Department of Mathematics, University of Chicago, Chicago, Ill.

Knottnerus, P. (2003). *Sample Survey Theory: some Pythagorean perspectives.* Springer-Verlag.

Knowles, J. D. and Corne, D. W. (2000). Approximating the nondominated front using the pareto archived evolution strategy. *Evolutionary Computation*, 8(2):149–172.

Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling.* Springer.

Kozak, A., Munro, D. D., and Smith, J. (1968). Taper functions and thier application in forest inventory. *Forestry Chronical*, 45:278–283.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In Neyman, J., editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1950*, pages 481–492, Berkeley, California. University of California Press. Reprinted in: *Readings in Mathematical Economics, Vol 1, Value Theory*, (P. Newman, ed.), The Johns Hopkins University Press, Baltimore, 1968, pp. 3–14.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002). Combining convergence and diversity in evolutionary multiobjective optimization. *Evol. Comput.*, 10(3):263–282.

Lawler, E. L., Lenstra, J. K., Kan, A. H. G. R., and Shmoys, D. B., editors (1985). *The Traveling Salesman Problem.* John Wiley, New York.

Lee, J. (1998). Constrained maximum-entropy sampling. *Operations Research*, 46:655–664.

Marti, K. (2005). *Stochastic Optimization Methods*. Springer.

Montes, F., Hernández, M. J., and Canellas, I. (2005). A geostastical approach to cork production sampling estimation in *Quercus suber* forests. *Canadian Journal of Forest Research*, 35:2787–2796.

Nam, D. and Park, C. (2000). Multiobjective simulated annealing: A comparative study to evolutionary algorithms. *International Journal of Fuzzy Systems*, 2(2):87–97.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, chapter 6, page 671. John Wiley and Sons, second edition.

Osyczka, A. (2002). *Evolutionary Algorithms for Single and Multicriteria Design Optimization*. Physica Verlag, Germany. ISBN 3-7908-1418-0.

Pebesma, E. J. (2004). Multivariable geostatistics in s: the gstat package. *Computers & Geosciences*, 30:683–691.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.

Porteus, E. L. (2002). *Foundations of Stoahstic Inventory Theory*. Stanford University Press.

Reed, D. D. and Burkhart, H. E. (1985). Spatial autocorrelation of individual tree charactersitics in loblolly pine stands. *Forest Science*, 31(3):575–587.

Rudolph, G. and Agapie, A. (2000). Convergence Properties of Some Multi-Objective Evolutionary Algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, volume 2, pages 1010–1016, Piscataway, New Jersey. IEEE Press.

Shewchuk, J. R. (1996). Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In Lin, M. C. and Manocha, D., editors, *Applied Computational Geometry: Towards Geometric Engineering*, volume 1148 of *Lecture Notes in Computer Science*, pages 203–222. Springer-Verlag. From the First ACM Workshop on Applied Computational Geometry.

Shiver, B. D. and Borders, B. E. (1996). *Sampling Techniques for Forest Resource Sampling*. John Wiley and Sons, Inc., New York.

Steuer, R. (1985). *Multiple Criteria Optimization: Theory, Computation and Application.* John Wiley & Sons, New York, NY.

Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation and Application.* John Wiley, New York.

Stewart, T. J., Janssen, R., and van Herwijnen, M. (2004). A genetic algorithm approach to multiobjective land use planning. *Comput. Oper. Res.*, 31(14):2293–2313.

Tan, K., Khor, E., and Lee, T. (2005). *Multiobjective Evolutionary Algorithms and Applications.* Springer.

Tersine, R. (1988). *Principles of inventory and materials management.* Elsevier Science Publishing Co, Inc., third edition.

Tillé, Y. (2006). *Sampling Algorithms.* Springer.

Tóth, S., McDill, M. E., and Rebain, S. (2006). Finding the efficient frontier of a bi-criteria, spatially explicit, harvest scheduling problem. *Forest Science*, 52(1):93–107.

Wackernagel, H. (1998). *Multivariate Geostatistics*, chapter 1. Springer, Berlin, Germany, 2nd edition.

Weintraub, A. and Murray, A. T. (2006). Review of combinatorial problems induced by spatial forest harvesting planning. *Discrete Appl. Math.*, 154(5):867–879.

White, H. (1999). *Asymptotic Theory for Econometricians.* Acandemic Press.

6. CONCLUSIONS

This collection of manuscripts completes the initial work set forth by researchers at Oregon State University, to develop a framework for the examination, formulation, study of, management, and optimization of the primary forest products supply chain. To provide these tools, these manuscripts make original contributions to the following subject areas:

- Multiple-objective Optimization

    - A general framework for non-linear mixed-integer multi-objective NP-hard planning problems with deterministic and stochastic elements;

    - The use of Multi-Objective Evolutionary Algorithms (MOEA), in forest engineering, forest sampling, and forest operations for examination of the trade-offs among potentially competing and possibly non-commensurate objectives;

    - Basic research regarding evolution strategy parameter values and their influence on the performance of the search of the solution space; and

    - A completely enumerated non-trivial (i.e. spatially correlated) example problem which can be used to examine and compare multi-objective solution search techniques.

- Forest Planning

    - Examination of design-based and model-based inference for the estimation or prediction of missing values which are then used for planning, simulation, and optimization, of forest operations;

- The inclusion of simple spatial correlation structures to assess forest product availability in an operational environment;

- The examination of various spatial correlation models, from which more precise predictions of forest products can be obtained;

- The application of entropy in forest sampling where the goal is to obtain data from a finite region for multiple purposes;

- Optimal experimental design sampling, expressed as a sample data acquisition problem for supply chain optimization activities.

- Supply-chain Optimization

  - A framework that can combine potentially disparate forest operations (e.g. production and log yard inventory) into a single problem where the practitioner can examine the consequences of various decisions objectively;

  - A framework that provides ability to combine deterministic and stochastic inputs and outputs (i.e. predictions/estimates of forest inventories, probability maximization problems, single and joint-chance constraint problems), and general non-linear mixed-integer multi-objective planning problems into a single framework;

  - Examination and communication of the Pareto set for high dimensional solution spaces (i.e. many objectives), which are common in forest engineering; and

As these manuscripts represent the combination of concepts from a variety of disciplines, as is often the case in forest engineering, I have made every attempt to unify the nomenclature from the various disciplines, when possible, to address the combination of biological, quantitative, managerial

and social sciences applied to the management and conservation of forest resources where the goal is to produce some desired societal, biological, or economic objectives. It is by no means complete.

While the problems examined within these manuscripts were relatively small, the advantages of these methods are clear when applied to larger problems with multiple financial, ecological, and social objectives. This then presents the need to examine mutation operators, selection operators, stopping criteria, Pareto diversity measures, convergence criteria, and visualization methods for high-dimensional Pareto frontiers, as large forest planning problems contain complex relationships among input data, decision variables, and policies, defined by $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$, and is the subject of future research.

Finally, as the focus was on problem formulation, solution techniques, and presentation methods in the individual manuscripts, the collection of the manuscripts provide tools for practitioners wanting to optimize forest ecosystem service supply chains, with methods to query, formulate, search for, solve, and disseminate information about the quality and quantity of services forest ecosystem managers can and do provide.

BIBLIOGRAPHY

Abbas, A. E. and Holmes, S. P. (2004). Bioinformatics and management science: some common tools and techniques. *Operations Research*, 52(2):165–190.

Armstrong, M. (1998). *Basic Linear Geostatistics*. Springer.

Armstrong, M. and Champigny, N. (1989). A study on kriging on small blocks. *Canadian Institute of Mining Bulletin*, 82(923):128–133.

Avery, T. E. and Burkhart, H. E. (2003). *Forest measurements, 5th Ed.* McGraw-Hill.

Bang-Jensen, J., Chiarandini, M., Goegebeur, Y., and Jørgensen, B. (2007). Mixed models for the analysis of local search components. In Stützle, T., Birattari, M., and Hoos, H. H., editors, *SLS*, volume 4638 of *Lecture Notes in Computer Science*, pages 91–105. Springer.

Bartz-Beielstein, T. (2006). *Experimental Research in Evolutionary Computation: The New Experimentalism (Natural Computing Series)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Beers, T. W. (1966). The direct correction for boundary-line slopover in horizontal point sampling. Research Progress Report 224, Purdue University, Agricultural Experiment Station, Lafayette, Indiana.

Bettinger, P., Boston, K., and Sessions, J. (1999). Combinatoral optimization of elk habitat effectiveness and timber harvest volume. *Environmental Monitoring and Assessment*.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Blackwell, D. A. (1951). Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102. University of California Press, Berkeley.

Block, R., Van Rees, K. C. J., and Pennock, D. J. (2002). Quantifying Harvesting Impacts using Soil Compaction and Disturbance Regimes at a Landscape Scale. *Soil Sci Soc Am J*, 66(5):1669–1676.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Bueso, M., Angulo, J., and Alonso, F. (1998). A state-space model approach to optimal spatial sampling design based on entropy. *Environmental and Ecological Statistics*, 5:29–44.

Bueso, M., Angulo, J., Cruz-Sanjulian, J., and Garcia-Arostegui, J. (1999). Optimal spatial sampling design in a multivariate framework. *Mathematical Geology*, 31:507–525(19).

Caselton, W. F. and Hussain, T. (1980). Hydrological Networks: Information Transmission. *Journal of The Water Resources Planning and Management Division*, 106(2):503–520.

Chapman, R. C. (1982). Plot spacing in systematic sampling. *Journal of Forestry*, 80:409–419.

Chen, X., Dempster, A. P., and Liu, J. S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika*, 81(3):457–469.

Clark, M. M., Meller, R. D., McDonald, T. P., and Ting, C. C. (2000). A new harvest operation cost model to evaluate forest harvest layout alternatives. *Annals of Operations Research*, 95(1):115–129.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons, New York. 428 p.

Coello Coello, C. A. (2002). Evolutionary multi-objective optimization: A critical review. In Sarker, R., Mohammadian, M., and Yao, X., editors, *Evolutionary Optimization*, chapter 5. Kluwer Academic Publishers, Norwell, MA, USA.

Coello Coello, C. A. and Toscano Pulido, G. (2001). A Micro-Genetic Algorithm for Multiobjective Optimization. In Zitzler, E., Deb, K., Thiele, L., Coello, C. A. C., and Corne, D., editors, *First International Conference on Evolutionary Multi-Criterion Optimization*, pages 126–140. Springer-Verlag. Lecture Notes in Computer Science No. 1993.

Connaughton, K. P. and Fight, R. D. (1984). Applying Trade-Off Analaysis to National Forest Planning. *Journal of Forestry*, 82(11):680–683.

Cooper, C. (2006). Sampling and variance estimation on continous domains. *Environmetrics*, 17:539–553.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17(5):563–586.

Cressie, N. A. C. (1993a). *Statistics for Spatial Data.* John Wiley and Sons, Inc., New York, New York, revised edition edition.

Cressie, N. A. C. (1993b). *Statistics for Spatial Data*, chapter 2. John Wiley and Sons, Inc., New York, New York, revised edition. 900.

Dattorro, J. (2005). *Convex Optimization. mathcalM$\epsilon\beta$oo* Publishing,.

Davis, L. S., Johnson, K. N., Bettinger, P. S., and Howard, T. E. (2001). *Forest Management*, chapter 13, page 804. McGraw-Hill series in forest management. McGraw-Hill, New York, New York, 4th edition.

Davis, R. E., Foote, F. S., Anderson, J. M., and Mikhail, E. M. (1981). *Surveying, theory and practice.* McGraw-Hill, Inc.

de Gruijter, J. J., Brus, D. J., Bierkins, M. F. P., and Knotters, M. (2006). *Sampling for Natural Resource Monitoring.* Springer-Verlag.

Deb, K. (2001a). *Multi-Objective Optimization using Evolutionary Algorithms.* John Wiley & Sons, Chichester, UK. ISBN 0-471-87339-X.

Deb, K. (2001b). *Multi-Objective Optimization using Evolutionary Algorithms.* Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Chichester.

Deb, K., Tewari, R., Dixit, M., and Dutta, J. (2007). Finding Trade-off Solutions Close to KKT Points Using Evolutionary Multi-Objective Optimization. KanGAL Report 2007006, Indian Institute of Technology Kanpur, Kanpur, PIN 208016, India.

Deusen, P. C. V. (1999). Multiple solution harvest scheduling. *Silva Fennica*, 33(3):207–216.

Diestel, R. (2005). *Graph Theory (Graduate Texts in Mathematics).* Springer.

Dimov, L. D., Chambers, J. L., and Lockhart, B. R. (2005). Spatial continuity of tree attributes in bottomland hardwood forests in the southeastern united states. *Forest Science*, 51(6):532–540.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2004). Even flow scheduling problems in forest management. In Coello, C. A. C. and Lamont, G. B., editors, *Applications of Multi-Objective Evolutionary Algorithms*, pages 701–726. World Scientific, Singapore.

Ducheyne, E. I., Wulf, R. R. D., and Baets, B. D. (2006). A spatial approach to forest-management optimization: linking GIS and multiple objective genetic algorithms. *International Journal of Geographical Information Science*, 20(8):917–928.

Dumitrescu, D., Lazzerini, B., Jain, L., and Dumitrescu, A. (2000). *Evolutionary Computation*. CRC Press.

Ehrgott, M. (2000). Approximation algorithms for combinatorial multicriteria optimization problems. *International Transactions in Operational Research*, 7:5–31.

Eiben, A. E. and Smith, J. E. (2003). *Introduction to Evolutionary Computing*. SpringerVerlag.

Falcão, A. O. and Borges, J. G. (2001). Designing an Evolution Program for Solving Integer Forest Management Scheduling Models: An Application in Portugal. *Forest Science*, 47(2):158–168.

Farhang-Mehr, A. and Azarm, S. (2002). Diversity assessment of pareto optimal solution sets: An entropy approach. In *Congress on Evolutionary Computation (CEC'2002)*, volume 1, pages 723–728, Piscataway, New Jersey. IEEE Service Center.

Gambill, C. W., Harry V. Wiant, J., and Yandle, D. O. (1985). Optimum plot size and baf. *Forest Science*, 31(3):587–594.

Gen, M. and Cheng, R. (1997). *Genetic Algorithms and Engineering Design*. John Wiley and Sons, Inc., New York.

Gentle, J. E., Narula, S. C., and Valliant, R. L. (1997). Multicriteria optimization in sampling design. In Ghosh, S., Schucany, W. R., and Smith, W. B., editors, *Statistics of Quality*, volume 153 of *Statistics, textbooks, and monographs*, pages 411–425. Marcel Dekker, Inc.

Gill, C. A. and Joanes, D. N. (1979). Bayesian Estimation of Shannon's Index of Diversity. *Biometrika*, 66(1):81–85.

Gjedtjernet, A. M. F. (1995). Forest operations and environmental protection. *Water, Air, & Soil Pollution*, 82(1):35–41.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics Series. Oxford University Press, New York, New York.

Gregoire, T. (2002). Forestry. In El-Shaarawi, A. H. and Piegorsch, W. W., editors, *Encyclopedia of Environmetrics*, volume 2. John Wiley and Sons.

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28:1429–1447.

Hamann, J. D. and Boston, K. (2007). Using spatial statistics to improve the primary forestry supply chain. In Reynolds, K. M., Thompson, A., Shannon, M., Köhl, M., Ray, D., and Rennolls, K., editors, *Sustainable Forestry: from Monitoring and Modelling to Knowledge Management & Policy Science*, chapter 18, pages 285–297. CABI International.

Hansen, M. P. (1997). Tabu Search in Multiobjective Optimisation : MOTS. In *Proceedings of the 13th International Conference on Multiple Criteria Decision Making (MCDM'97)*, Cape Town, South Africa.

Helms, J. A., editor (1998). *The Dictionary of Forestry*. Society of American Foresters, Bethesda, MD.

Hillier, F. S. and Lieberman, G. J. (1995). *Introduction to Operations Research*. McGraw-Hill, Inc., sixth edition.

Hock, B., Payn, T., and Shirley, J. (1993). Using a geographic information system and geostatistics to estimate site index of *Pinus radiata* for kaingaroa forest. *New Zealnad Journal of Forest Science*, 23:264–277.

Hoef, J. M. V. (2002). Sampling and geostatistics for spatial data. *Ecoscience*, 9(2):152–161.

Hof, J. and Bevers, M. (1994). Optimal timber harvest scheduling with spatially designed sediment objectives. *Canadian Journal of Forest Research*, 30(1494-1500).

Hof, J. G. and Bevers, M. (1998). *Spatial Optimization for Managed Ecosystems*. Columbia University Press, New York.

Hof, J. G., Kent, B. M., and Pickens, J. B. (1992). Chance constraints and chance maximization with random yield coefficients in renewable resource optimization. *Forest Science*, 38:305–323(19).

Holmgren, P. and Thuresson, T. (1997). Applying Objectively Estimated and Spatially Continuous Forest Parameters in Tactical PLanning to Obtain Dynamic Treatment Units. *Forest Science*, 43(3):317–326.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statisical Assocaition*, 47.

Hudak, A. T., Lefsky, M. A., Cohen, W. B., and Berterretche, M. (2002). Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, 82(2-3):397–416.

Husch, B., Beers, T., and Kershaw Jr., J. (2003). *Forest Mensuration*. John Wiley & Sons, Inc., Hoboken. 443 p.

Husch, B., Miller, C. I., and Beers, T. W. (1982). *Forest Mensuration*. John Wiley and Sons, Inc., third edition.

Hyoun-Jin, S., Kyung-Joon, C., Jae-Eung, O., and Je-Seon, R. (2005). Kriging with meta-heuristic methods for optimal design to reduce the noise of the engine cooling fan. *JSME international journal. Series C, Mechanical systems, machine elements and manufacturing*, 48(3):363–370.

Igel, C., Hansen, N., and Roth, S. (2007). Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15(1):1–28.

Ihaka, R. and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314.

Iles, D. D. M. K. and Bell, J. F. (2004). Using a large-angle gauge to select trees for measurement in variable plot sampling. *Canadian Journal of Forest Research*, 34(4):840–845.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, Inc., New York, New York.

Ishibuchi, H., Kaige, S., and Narukawa, K. (2005). *Evolutionary Multi-Criterion Optimization*, volume 3410 of *Lecture Notes in Computer Science*, chapter Comparison Between Lamarckian and Baldwinian Repair on Multiobjective 0/1 Knapsack Problems, pages 370–385. Springer.

Journel, A. and Huijbregts, C. J. (1978). *Mining Geostatistics*, chapter 5, page 600. Academic Press, Inc., London.

Karlsson, J., Ronnqvist, M., and Bergstrom, J. (2003). Short-term harvest planning including scheduling of harvest crews. *International Transactions in Operational Research*, 10(19):413–431.

Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. Master's thesis, Department of Mathematics, University of Chicago, Chicago, Ill.

Kim, Y. C. and Baafi, E. Y. (1984). Combining local kriging variances for short-term mine planning. In Verly, G., David, M., Journel, A., and Marechal, A., editors, *Geostistics for natural resources characterization*, volume 122 of *Series C: Mathematical and Physical Sciences*, pages 185–199, P.O. Box 17, 3300 AA Dorrecht, Holland. NATO Advanced Study Institute on Geostistics for natural resources characterization, Reidel Publishing Company.

King, S. L. (2000). Sequential gaussian simulation vs. simulated annealing for locating pockets of high-value commercial trees in Pennsylvania. *Annals of Operations Research*, 95(1-4):117–203.

Knottnerus, P. (2003). *Sample Survey Theory: some Pythagorean perspectives.* Springer-Verlag.

Knowles, J., Oates, M., and Corne, D. (2000). Advanced muli-objecive evolutionary algorithms applied to two problems in telecommunications. *British Telecom Technology Journal*, 18(4):51–65.

Knowles, J. D. and Corne, D. W. (1999). The Pareto Archived Evolution Strategy: A New Baseline Algorithm for Multiobjective Optimisation. In *1999 Congress on Evolutionary Computation*, pages 98–105, Washington, D.C. IEEE Service Center.

Knowles, J. D. and Corne, D. W. (2000). Approximating the nondominated front using the pareto archived evolution strategy. *Evolutionary Computation*, 8(2):149–172.

Ko, C.-W., Lee, J., and Queyranne, M. (1995). An exact algorithm for maximium entropy sampling. *Operations Research*, 43(4):684–691.

Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling.* Springer.

Kononenko, I. and Kukar, M. (2007). *Machine Learning and Data Mining: Introduction to Principles and Algorithms.* Horwood Publishing Limited, West Sussex, UK.

Kozak, A., Munro, D. D., and Smith, J. (1968). Taper functions and their application in forest inventory. *Forestry Chronicle*, 45:278–283.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In Neyman, J., editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1950*, pages 481–492, Berkeley, California. University of California Press. Reprinted in: *Readings in Mathematical Economics, Vol 1, Value Theory*, (P. Newman, ed.), The Johns Hopkins University Press, Baltimore, 1968, pp. 3–14.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002). Combining convergence and diversity in evolutionary multiobjective optimization. *Evol. Comput.*, 10(3):263–282.

Lawler, E. L., Lenstra, J. K., Kan, A. H. G. R., and Shmoys, D. B., editors (1985). *The Traveling Salesman Problem.* John Wiley, New York.

Le, N. D. and Zidek, J. V. (2006). *Statistical Analysis of Environmental Space-Time Processes*, chapter 8, page 341. Springer Series in Statistics. Springer, New York, New York.

Le, P. G. N., Sampson, P., and Zidek, J. V. (1993). Using entropy in the redesign of an environmental monitoring network. In Patil, G. and Rao, C., editors, *Multivariate Environmental Statistics*, pages 175–202. North Holland.

Lee, J. (1998a). Constrained maximum-entropy sampling. *Operations Research*, 46(5):655–664.

Lee, J. (1998b). Constrained maximum-entropy sampling. *Operations Research*, 46:655–664.

Lin, S. and Kernighan, B. W. (1973). An effective heuristic for the traveling salesman problem. *Operations Research*, 21:498–516.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley and Sons, Inc.

Lohr, S. L. (1999). *Sampling: Design and Analysis*, chapter 8. Brooks/Cole Publishing Company.

Mandallaz, D. (2000). Estimation of the spatial covariance in Universal Kriging: Application to forest inventory. *Environmental and Ecological Statistics*, 7:263–284.

Mandallaz, D. and Ye, R. (1999). Forest inventory with optimal two-phase, two-stage sampling schemes based on the anticipated variance. *Candian Journal of Forest Research*, 29:1691–1708.

Mansour, Y. and Schain, M. (1997). Learning with maximum-entropy distributions. In *Computational Learing Theory*, pages 201–210.

Marti, K. (2005). *Stochastic Optimization Methods*. Springer.

Mawson, J. C. and Mack, R. J. (1982). Inventory costs on small forests. *Journal of Forestry*, 80(3):165–167.

Michalewicz, Z. and Fogel, D. B. (2004). *How to Solve It: Modern Heruistics*, chapter 5, page 554. Springer Verlag, 2 edition.

Montes, F., Hernández, M. J., and Canellas, I. (2005). A geostastical approach to cork production sampling estimation in *Quercus suber* forests. *Canadian Journal of Forest Research*, 35:2787–2796.

Nam, D. and Park, C. (2000). Multiobjective simulated annealing: A comparative study to evolutionary algorithms. *International Journal of Fuzzy Systems*, 2(2):87–97.

Oderwald, R. G. and Jones, E. (1992). Sample sizes for point, double sampling. *Canadian Journal of Forest Research*, 22:980–983.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000a). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley and Sons, second edition.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000b). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, chapter 6, page 671. John Wiley and Sons, second edition.

Osyczka, A. (2002). *Evolutionary Algorithms for Single and Multicriteria Design Optimization*. Physica Verlag, Germany. ISBN 3-7908-1418-0.

Paquete, L., Chiarandini, M., and St utzle, T. (2004). Pareto local optimum sets in the biobjective traveling salesman problem: An experimental study. In Gandibleux, X., Sevaux, M., SÃűrensen, K., and T'kindt, V., editors, *Metaheuristics for Multiobjective Optimisation*, volume 535 of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, Berlin, Germany.

Pebesma, E. J. (2004). Multivariable geostatistics in s: the gstat package. *Computers & Geosciences*, 30:683–691.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.

Porteus, E. L. (2002). *Foundations of Stoahstic Inventory Theory*. Stanford University Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK) and New York, 3rd edition.

Reed, D. D. and Burkhart, H. E. (1985). Spatial autocorrelation of individual tree charactersitics in loblolly pine stands. *Forest Science*, 31(3):575–587.

Reeves, L. H. and Haight, R. G. (2000). Timber harvest scheduling with price uncertainty using Markowitz portfolio optimization. *Annals of Operations Research*, 95:229–250.

Riopel, D., Langevin, A., and Campbell, J. F. (2005). The network of logistics decisions. In Langevin, A. and Riopel, D., editors, *Logistical Systems:Design and Optimization*, number 2 in GERAD, pages 1–38. Springer, New York, New York.

Robinson, A. P. and Wykoff, W. R. (2004). Imputing missing height measures using a mixed-effects modeling strategy. *Canadian Journal of Forest Research*, 34:2492–2500.

Rudolph, G. and Agapie, A. (2000). Convergence Properties of Some Multi-Objective Evolutionary Algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, volume 2, pages 1010–1016, Piscataway, New Jersey. IEEE Press.

Samra, J., Gill, H., and Bhatia, V. (1989). Spatial stochastic modelling of growth and forest resource evaluation. *Forest Science*, 35:663–676.

Sarkar, D. (2008). *lattice: Lattice Graphics*. R package version 0.17-4.

Schreuder, H. T., Gregoire, T. G., and Wood, G. B. (1993). *Sampling methods for multiresource forest inventory*. John Wiley & Sons, Inc. 446 p.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423.

Shewchuk, J. R. (1996). Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In Lin, M. C. and Manocha, D., editors, *Applied Computational Geometry: Towards Geometric Engineering*, volume 1148 of *Lecture Notes in Computer Science*, pages 203–222. Springer-Verlag. From the First ACM Workshop on Applied Computational Geometry.

Shewry, M. and Wynn, H. (1987). Maximum entropy sampling. *Journal of Applied Statistics*, 46:165–170.

Shiver, B. D. and Borders, B. E. (1996). *Sampling Techniques for Forest Resource Sampling*. John Wiley and Sons, Inc., New York.

Stage, A. R. (1973). Prognosis model for Stand Development. General Technical Report INT-137, USDA Forest Service Research Note, Ogden, Utah.

Steuer, R. (1985). *Multiple Criteria Optimization: Theory, Computation and Application*. John Wiley & Sons, New York, NY.

Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation and Application*. John Wiley, New York.

Stevens, Jr., D. L. (1997). Variable density grid-based sampling designs for continuous spatial populations. *Evironmetrics*, 8:167–195.

Stewart, T. J., Janssen, R., and van Herwijnen, M. (2004). A genetic algorithm approach to multiobjective land use planning. *Comput. Oper. Res.*, 31(14):2293–2313.

Stringer, C. (2006). Forest certification and changing global commodity chains. *J Econ Geogr*, 6(5):701–722.

Tan, K., Khor, E., and Lee, T. (2005). *Multiobjective Evolutionary Algorithms and Applications*. Springer.

Tersine, R. (1988). *Principles of inventory and materials management*. Elsevier Science Publishing Co, Inc., third edition.

Theil, H. (1971). *Principles of Econometrics*. John Wiley and Sons, Inc.

Tillé, Y. (2006). *Sampling Algorithms*. Springer.

Tóth, S., McDill, M. E., and Rebain, S. (2006). Finding the efficient frontier of a bi-criteria, spatially explicit, harvest scheduling problem. *Forest Science*, 52(1):93–107.

Trujillo-Ventura, A. and Ellis, J. H. (1991). Multiobjective air pollution monitoring network design. *Atmospheric Environment*, 25(2):469–479.

Tuominen, S., Fish, S., and Poso, S. (2003). Combining remote sensing, data from earlier inventories and geostatistical interpolation in multisource forest inventory. *Canadian Journal of Forest Research*, 33:624–634.

Vera, J. R., Weintraub, A., Koenig, M., Bravo, G., Guignard, M., and Barahona, F. (2003). A lagrangian relaxation approach for a machinery location problem in forest harvesting. *Pesquisa Operacional*, 23(1):111–128.

Wackernagel, H. (1998). *Multivariate Geostatistics*, chapter 1. Springer, Berlin, Germany, 2nd edition.

Wallerman, J., Joyce, S., Vencataswamy, C. P., and Olsson, H. (2002). Prediction of forest stem volume using kriging adapted to detected edges. *Canadian Journal of Forest Research*, 32:509–518.

Weintraub, A. and Murray, A. T. (2006). Review of combinatorial problems induced by spatial forest harvesting planning. *Discrete Appl. Math.*, 154(5):867–879.

Weintraub, A. and Navon, D. (1976). A forest management planning model integrating silvicultural and transportation activities. *Management Science*, 22(12):1299–1309.

White, H. (1999). *Asymptotic Theory for Econometricians*. Acandemic Press.

Wu, S. and Zidek, J. V. (1992). Entropy-based analysis of data from selected nadp/ntn network sites for 1983-1986. *Atmospheric Environment*, 26A(11):2089–2103.

Yan, Z., Zhang, L., Kang, L., and Lin, G. (2003). A New MOEA for Multiobjective TSP and Its Convergence Property Analysis. In Fonseca, C. M., Fleming, P. J., Zitzler, E., Deb, K., and Thiele, L., editors, *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pages 342–354, Faro, Portugal. Springer. Lecture Notes in Computer Science. Volume 2632.

Zeide, B. (1980). Plot size optimization. *Forest Science*, 26(2):251–257.

Zhang, B.-T. (2003). *A unified Bayesian framework for evolutionary learning and optimization*, pages 393–412. Springer-Verlag New York, Inc., New York, NY, USA.

Appendix

APPENDIX A

THE SEARCH SPACE ($\mathcal{S}$) AND THE GLOBAL PARETO SET FOR THE EXAMPLE PROBLEM

For the example problem presented in this manuscript, the cardinality of $\mathcal{S}$, $|\mathcal{S}|$, is defined by the number of partitions for each cutting schedule permutation.

Since the eight stems must be assigned to one of four non-empty sets (i.e. harvest on day one, two or three, or assigned as a residual stem), the total number of ways of partitioning a set of n elements into m non-empty subsets is defined by the Stirling number of the second kind, denoted by $S(n, k)$, where $S(n, k)$ is defined as:

$$S(n, k) = \frac{1}{k!} \sum_{i=1}^{k} (-1)^i \binom{k}{i} (k - i)^n \tag{A.1}$$

and $\binom{k}{i}$ is the binomial coefficient.

Each permutation of n stems defines a unique assignment sequence, in which the total number of permutations of the n stems equals $n!$ The product of the two terms results in the total number of possible candidate solutions, from which the global Pareto optimal set can be determined using the methods describe in Section 2.3. For this examination, the cardinality of the input space was $S(n, k) \times n! = S(8, 4) \times 8! = 1701 \times 40320$, or 68,584,320 possible candidate solutions. This result was verified by complete enumeration.

The total number of feasible and infeasible solutions in $\mathcal{S}$ were 2,949,120 and 65,635,200, respectively, constituting 4.30 and 95.70 percent of the solutions in the search space. The total number of global Pareto optimal solutions was 465, which constituted only 0.0007 percent of the solutions in $\mathcal{S}$ and only 0.0158 of feasible solutions in $\mathcal{S}$. The low ratio of feasible solutions could in-

dicate that a search strategy focusing on minimizing infeasibility may be more efficient at finding Pareto optimal solutions than less directed random searches since optimal solutions are found on the feasibility-infeasible boundary of $\mathcal{S}$ and less optimal, yet feasible solutions would be evaluated when searching within the feasible region.