

A NEW MODEL-BASED METHOD FOR ESTIMATING THE ABUNDANCE
OF STANDING DEAD TREES

BY

HONG SU AN

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Forestry

2011

ABSTRACT

A NEW MODEL-BASED METHOD FOR ESTIMATING THE ABUNDANCE OF STANDING DEAD TREES

BY

Hong Su An

Standing dead trees (SDT) are an important component of forest ecosystems. However, it can be a challenge to develop reliable estimates of population parameters because dead trees are generally lower in abundance and have more complex spatial distributions (e.g., are more clustered) than live trees. In addition, most forest inventories are designed for sampling live trees. Previous studies (e.g., Bull et al. 1990) have recommend using a relatively higher sampling intensity or larger plot sizes for dead versus live trees, but this is more time consuming and costly. Adding new plots, increasing plot sizes or otherwise modifying plot designs can be especially costly in the case of large scale (e.g., national forest inventories) and other permanent plot network. This thesis sought to explore approaches to improving estimation of standing dead tree abundance, other than adding more plots or modifying plot designs, in the context of the US National Forest Inventory and Analysis Program (USDA Forest Service 2008) and US Forest Health Monitoring (FHM) Program (now merged with the FIA plot design) and other similar permanent plot networks.

One major consequence of using sampling plots that are either too small or too few for sampling standing dead trees is that it is likely that there will be a large proportion of zero observations in data, typically referred to as “zero-inflated” data. Excess zero observations increases variation of estimates of standing dead tree parameters. To reduce this variability

caused by zero-inflated data, a new model, Expected-Zero Hurdle (EZ-Hurdle) method, is proposed. The EZ-Hurdle method replaced the observed zero proportion in data with an expected zero probability obtained from auxiliary information describing the distance from a random point (plot center) to the nearest standing dead tree.

The EZ-Hurdle method greatly improved the precision and showed less average bias than fixed-area sampling, with or without adjustment using the standard Hurdle model when tested with both simulation and field studies. The EZ-Hurdle method improved the precision without adding fixed-area plot but it required additional information to explain the uncertainty caused by zero observations in data. Especially, EZ-Hurdle methods improved the precision when only additional information was applied without adding points. Therefore, it can be applied to improve the precision of estimates without changing plot design such as FIA and FHM program. The EZ-Hurdle method performs best when the density of standing dead trees is low or a small fixed-area plot size is used to collect the data because the expected zero probability which is modeled from auxiliary information showed less variation than observed zero proportion in data.

Although EZ-Hurdle method showed better precision, it is less cost and sampling efficiency than fixed-area sampling method due to time to search the nearest standing dead tree. Therefore, distance-limited EZ-Hurdle method which restricts the search radius to find the nearest standing dead tree was proposed to reduce time to collect auxiliary information. Distance-limited EZ-Hurdle method showed better precision than fixed-area sampling for all circumstances such as densities and spatial patterns. It also has better time efficiency than the original EZ-Hurdle method. Therefore, the EZ-Hurdle method with a distance-limited method can be an alternative method to improve the precision for estimating the density of standing dead trees without changes of plot design using reasonable cost and time to collect the data.

Copyright by

Hong Su An

2011

ACKNOWLEDGMENTS

It was fortunate that I have had the opportunity to interact with many people who have inspired and have been inspirational to my life and research at Michigan State University. Especially, I would like to gratefully and sincerely appreciate to Dr. David W. MacFarlane for his guidance and patience. Especially, I really appreciate his friendship and understanding during my graduate study. His advice gave me great challenge and led me to complete Ph.D program. I would also like to give many thanks to my advisory committee, Dr. Richard K. Kobe, Dr. Daniel Hayes, and Dr. Andrew O. Finley. They gave me broad view of science and a lot of comments and advice for my research.

I am grateful to the Forest Modeling Lab group: Zhonglei Wang, Neil Ver Planck, and Lisa Parker. It was an exciting and enjoyable time with them. Especially, it was exciting experience for data collection not only my research but also other projects. I also appreciate Dr. Kyung-Hwan Han, Dr. Sang Choon Jeon, Dr. Sang Yeob Lee, and Dr. Hwan Jung. They encouraged me when I fell into a slump and prayed for me. In addition, I would like to thank my friends in New Hope Baptist church.

Especially, I am grateful to Dr. Man Yong Shin. He is my mentor not only science but also life. He encourages me to pursue the Ph.D program and gives me tremendous amounts of advice for my life and research. He is my role model for my life. I also would like to thank Dr. Alla Sikorskii. She gave me the opportunity to expand my knowledge from forestry to nursing research. It was an unforgettable experience working as a member of her research group.

Lastly, I would like to thank my family for all their love and encouragement. Their endless commitment and efforts to make me a scholar.

TABLE OF CONTENTS

List of Tables.....	viii
List of Figures.....	x
1. Introduction	1
1.1. The Importance of Dead Trees in Forest Ecosystems.....	1
1.2. Challenges for Estimating the Abundance of Standing Dead Trees	2
1.3. Goals and Objectives of this Dissertation	4
2. Review of General Sampling Methods and Models for Dealing with Zero-Inflated Data	7
2.1. Overview of Sampling Inference	7
2.1.1. Design-Based Estimation.....	7
2.1.2. Model-Based Estimation.....	10
2.2. Statistical Models for Count Data and the Problem of Zero-Inflation	13
2.2.1. Overview.....	13
2.2.2. Poisson Model.....	13
2.2.3. Negative Binomial Model.....	14
2.2.4. Hurdle Model	15
3. Development of the EZ-Hurdle Method	17
3.1. Model Development	17
4. A Simulation Study to Understand the Properties of the EZ-Hurdle Method and Compare it to Related Approaches	20
4.1. Overview	20
4.2. Data	21
4.3. Methods	22
4.3.1. Estimating the Expected Zero Probability	22
4.3.2. Comparison of EZ-Hurdle to Existing Methods and Models	25
4.3.3. Sensitivity Analysis	25
4.4. Results	27
4.4.1. Inclusion probability	27
4.4.2. Comparison of Estimated Density of SDT/ha by Methods	32
4.5. Discussion of Simulation Study	43
5. An Application of the EZ-Hurdle Method for the Estimation of Standing Dead Tree	

Abundance in a Real Forest Setting	46
5.1. Introduction	46
5.2. Methods	47
5.2.1. Data Collection	47
5.2.2. Statistical Analysis	51
5.3. Results	53
5.4. Discussion for Applied Study	57
6. Cost and Sampling Efficiency of EZ-Hurdle Method	59
6.1. Overview	59
6.2. Data	60
6.3. Analytical Methods	62
Estimation of time costs from field data for bootstrapping and simulation	63
6.4. Results	71
6.4.1. Comparison of Inclusion Probabilities Between Different Forest Type Conditions .	71
6.4.2. Results of Time Requirement Studies	74
6.4.3. Comparison of the Coefficient of Variation by Method in Simulation	80
6.4.4. Comparison Sampling Efficiency by Method	82
6.5. Discussion for Cost and Sampling Efficiency	84
7. Developing a Distance-Limited EZ-Hurdle Method	87
7.1. Overview	87
7.2. Methods	87
7.3. Results	88
7.4. Discussion for Distance-Limited EZ-Hurdle Method	99
8. Conclusion	101
9. References	104

LIST OF TABLES

Table 4.1. Parameters to create clustering patterns from a Matern Cluster Point Process.	20
Table 4.2. Number of samples by sampling intensity (%) and plot size (radius, m).	22
Table 4.3. Models to estimate the inclusion probability of a standing dead tree.	23
Table 4.4. Mean RMSE and AIC by model to estimate the inclusion probability of a standing dead tree.	27
Table 5.1. Guide to define the decay class by conditions (USDA Forest Service 2005).	50
Table 5.2. The number of fixed-area plots and random points by sampling intensity and method.	52
Table 5.3. Summary of statistics from fixed-area sampling method by infestation status.	53
Table 5.4. The results of variance-to-mean ratio and index of dispersion.	55
Table 5.5. The estimated parameters and standard errors () of reduced Gompertz function.	55
Table 6.1. The classification of basal area (BA) class.	62
Table 6.2. The estimated parameters and R-square for regression model to estimate the time requirement for a 0.08 ha fixed area plot.	65
Table 6.3. The estimated parameters and R-square for regression model to estimate the search time to measure the nearest SDT.	69
Table 6.4. Summary of results for BA for live trees, density of standing dead trees (No.SDT/ha), average distance (Dist.) from point to the nearest SDT, and average time (Time) to search for and measure the nearest standing dead tree with standard deviation in () after the value.	71
Table 6.5. Estimated coefficients and standard errors () for reduced Gompertz function fitting by forest cover type and basal area class.	72
Table 6.6. Average of estimated time requirements (minutes) by survey type from calibrated simulations under different spatial patterns and standing dead tree densities.	74
Table 6.7. Estimated average time requirement (min.) per plot (& point) by method and plot type when time models are applied to the BBDMS data.	75
Table 6.8. Estimated time requirement (hr) for FAS with 7.32 m radius subplots and the additional time requirement for EZP by PPR for a specific number of fixed-area	

plots, with two-person crews.....	76
Table 6.9. Relative sampling efficiency of EZ-Hurdle method to compare FAS by density and spatial pattern.	82
Table 6.10. Relative sampling efficiency of EZ-Hurdle method to compare FAS by infected status.....	83
Table 7.1. Maximum search radii to find the nearest standing dead tree.....	88
Table 7.2. The CV of estimates (N/ha) and expected zero probability (EZP) by PPR and search radius (m) for EZ-Hurdle method when the spatial pattern of standing dead trees is random and density is 12/ha.....	89
Table 7.3. The CV of estimates (N/ha) and expected zero probability (EZP) by PPR and search radius (m) for EZ-Hurdle method when the spatial pattern of standing dead trees is cluster and density is 12/ha.	91

LIST OF FIGURES

Figure 4.1. The distribution of simulated standing dead trees by spatial pattern when the density is 12/ha. The area depicted for each pattern is about 18 ha.....	21
Figure 4.2. The change of inclusion probability of a standing dead tree estimated by the reduced Gompertz function under different conditions.	24
Figure 4.3. The mean and 95% CI of inclusion probabilities by spatial patterns and densities of standing dead trees when the search radius is 7.32m.....	29
Figure 4.4. The mean and 95% CI of inclusion probabilities by spatial patterns and densities of standing dead trees when the search radius is 17.95m.....	30
Figure 4.5. Observed and expected zero proportion in 3,000 simulated data sets by spatial patterns, when the plot radius is 7.32m and density is 24/ha.	31
Figure 4.6. The distribution of estimated density of standing dead trees by sampling intensities (1%, 3%, 5%, 7%) and methods when the plot radius is 7.32m and spatial pattern is random. True density is 12 trees per ha. SRS is simple random sampling method with fixed-area circular plot. HDP is Hurdle model with Poisson distribution and EZP is EZ-Hurdle model with Poisson distribution.....	33
Figure 4.7. The distribution of estimated density of standing dead trees by sampling intensities (1%, 3%, 5%, 7%) and methods when the plot radius is 7.32m and spatial pattern is Cluster I. True density is 49 trees per ha. SRS is simple random sampling method with fixed-area circular plot. HDP is Hurdle model with Poisson distribution and EZP is EZ-Hurdle method with Poisson distribution.....	34
Figure 4.8. RMSE by spatial patterns, sampling intensities and plot sizes. SRS is simple random sampling method, EZP is Expected-zero Hurdle model with Poisson distribution, Ran. is random pattern, C-I is clustered pattern I, and C-II is clustered pattern II. .	36
Figure 4. 9. The errors by sampling intensity when density is 12/ha and plot radius is 7.32 m...	38
Figure 4. 10. The errors by sampling intensity when density is 12/ha and plot radius is 17.95 m.	39
Figure 4. 11. The errors by sampling intensity when density is 49/ha and plot radius is 7.32 m.	40
Figure 4. 12. The errors by sampling intensity when density is 49/ha and plot radius is 17.95 m.	41
Figure 4.13. The changes of errors (%) by the changes of inclusion probability when sampling intensity is 5% and plot radius is 7.32m.	42
Figure 5.1. The selected locations among beech bark monitoring plot (BBDMS) and Pigeon	

River Country Forest (PRC).....	47
Figure 5.2. Plot design with 30 sampling points in the transect layout.	48
Figure 5.3. A standing dead tree is defined as a dead tree greater than 12.7 cm DBH, which is taller than 1.37 m and leans less than 45 degrees from vertical (USDA Forest Service 2005).....	49
Figure 5.4. The number of standing dead trees (No. SDT/ha) in five decay classes in two forest types. AB is American beech.....	54
Figure 5.5. Estimated inclusion probability by infected status. Dashed line is 7.32 m.	56
Figure 5.6. The change of coefficient of variations by the number of fixed-area plots. FAS is fixed-area sampling method, PPR is point to plot ratio of data for EZ-Hurdle method. Solid line is infected forests and dashed line is non-infected forests.....	57
Figure 6.1. The location of Pigeon River Country Forest (PRC) in MI.	61
Figure 6.2. Distribution of the number of SDT per plot and time requirement to measure the SDT within plot (0.08 ha). Dashed lines are medians.	65
Figure 6.3. The scatter plot between time (second) and number of SDT within plot and residuals of regression model to estimate time requirement for fixed area (0.08 ha).	66
Figure 6.4. The distribution of search area (ha) and time (second) to measure the nearest SDT in PRC. Dashed lines are median values.....	68
Figure 6.5. The scattered plot between search time (Sec.) and search radius (m) and residuals of selected regression model to estimate time to search the nearest SDT.	69
Figure 6.6. The inclusion probabilities by forest type and basal area class. Dashed lines are the radii of subplot (7.32 m) and annular plot (17.95 m).....	73
Figure 6.7. The distribution of distance between a random point to the nearest standing dead tree by density and spatial pattern under the simulation study. The dashed line is average distance.....	78
Figure 6.8. Additional time requirement for EZ-Hurdle method above the time cost for FAS by PPR from field study. Solid line is infected forests and dashed line is non-infected forests.	79
Figure 6.9. The change of coefficient of variation of estimated density by PPR and the number of fixed-area plots.....	81
Figure 7.1. The coefficients of variation (CV) of estimated density of standing dead trees by spatial pattern and search radius. FAS is the fixed-area sampling method, PPR is the point to plot ratio, and Inf. means unlimited search radius.	93

Figure 7.2. The coefficients of variation (CV) of estimated expected zero probability (EZP) by spatial pattern and search radius. FAS is the fixed-area sampling method, PPR is the point to plot ratio, and Inf. means an unlimited search radius was used.....	94
Figure 7.3. The change of inclusion probability of a standing dead tree by increasing the search radius when density of standing dead trees is 24/ha and spatial pattern is clustered.	96
Figure 7.4. Addition time requirement for EZ-Hurdle method by maximum search radius. PPR is point to plot ratio, FAS is fixed-area sampling method, and Inf. is unlimited distance.	98

1. Introduction

1.1. The Importance of Dead Trees in Forest Ecosystems

Dead trees influence many aspects of forest ecosystems such as soil fertility, hydrology, and wildlife habitat (Kimmins 1992). Tree death changes the resource availability for other organisms in the ecosystem (Franklin et al. 1987). Resources such as light, nutrients, and water are increased by tree death. For example, dead trees release nutrients and energy as they decompose (Harmon et al. 1986), so dead trees influence nutrient cycling (Maser and Trappe 1984). Consequently, dead wood is an important component to the global carbon (C) cycle (Oswalt et al. 2008). Following international agreements to control C emissions, quantifying the C source and sink by region is an important issue (Rothstein et al. 2004) and dead wood holds a substantial amount of C, which is slowly released to the atmosphere and soil during decomposition (Harmon et al. 1990; Keenan et al. 1993; Krankina and Harmon 1995). Therefore, the death of trees causes increases in C levels in the atmosphere (Clark et al. 2004). In some cases, dead components of the forest such as dead wood and litter may have more C than live components (Delaney et al. 1997).

Dead trees also affect the species diversity in forest ecosystems because they offer habitat for a variety of organisms (Franklin et al. 1987; Green and Peterken 1997; McCarthy and Bailey 1994). For example, standing dead trees and fallen dead trees provide habitat for birds (McClelland 1977), amphibians (Jaeger 1980), small mammals (Dueser and Shugart Jr 1978; McComb et al. 1993), and reptiles (Harmon et al. 1986). Brown (1985) reported that about 100 species of vertebrates use standing dead trees and 150 species use logs. Maser and Trappe (1984) reported that insects use the dead trees for their habitat, and fungi and mosses are living in

dead trees.

Dead tree populations are also critical to assess and monitor the health of forest ecosystems (Gray 2003; Greif and Archibold 2000). The abundance or size of dead trees has been used to assess the stage of forest development. For example, the density of dead trees suggests one of the criteria to define the old-growth Douglas-fir and mixed-conifer forest in the Pacific Northwest and California (Franklin et al. 1986). Dead wood is often used as part of criteria and indicators of sustainable forest management by some international initiatives, such as Montreal Process or International Tropical Timber Organization (ITTO) (Bütler and Schlaepfer 2004). Dead wood and dead trees have been suggested as an indicator for biodiversity in forest by the Fourth Ministerial Conference on the Protection of Forests in Europe in 2003. Several forest health indicators can be obtained from inventory data such as stand density, species composition, mortality/growth data, and growth-to-removals ratio (O'Langhlin and Cook 2003). Dead trees and dead wood are closely connected with mortality/growth rates. Therefore, reliable estimation of dead tree population attributes are important to assess and monitor forest ecosystem process (e.g., carbon cycling) and values (e.g., biodiversity).

1.2. Challenges for Estimating the Abundance of Standing Dead Trees

It can be a challenge to develop reliable estimates of population parameters because dead trees are generally lower in abundance and have more complex spatial distributions (e.g., are more clustered) than live trees. The abundance and spatial pattern of dead trees are different by geographical location, stand age, forest type and management regime (Cline et al. 1980, Guby and Dobbartin 1996, Green and Peterken 1997, Fridman and Walheim 2000). Fridman and Walhem (2000) reported that the volume of dead wood (standing dead trees and logs) is different

by vegetation type and geographical location in Sweden. Stephens (2004) reported that standing dead tree abundance is increased with increasing stand age in pine-mixed conifer forests. However, the number of standing dead trees may not correlate with the age of stand (Vasiliauskas et al. 2004). The abundance of dead trees are different by management regimes in hemlock-hardwood forest (Tyrrell and Crow 1994), temperate deciduous forest (Green and Peterken 1997), pine forest (Montes and Canellas 2006; Reid et al. 1996), and Norway spruce (Ranius et al. 2003). The spatial pattern of mortality is also different by the agent that caused tree death (Franklin et al. 1987). For example, forest fires or wind throw causes clustering of dead trees. This wide degree of variation in standing dead tree abundance between forest ecosystems increases sampling error. However, it suggests that auxiliary information regarding forest condition might allow for better estimation of standing dead tree abundance.

Another issue for sampling dead trees is that most forest inventories are designed for sampling live trees, with dead tree data often collected under the live-tree design(e.g., standing dead trees on the US National Forest Inventory and Analysis Program (FIA) plots (Bechtold and Patterson 2005)). Since the abundance of standing dead trees is generally lower and they have a more variable spatial pattern than live trees, larger plot sizes or higher sampling intensities should be needed to get the same relative accuracy for dead trees as achieved for live ones. For example, 0.4 or 1.0 ha plots have been used to estimate the abundance of standing dead trees (Ganey 1999; Spiering and Knight 2005; Stephens 2004). Bull et al., (1990) recommended a factor-5 prism (using horizontal point sampling), 1 ha fixed-area plot size, or an overall sampling intensity of about 5%, when the mean density of standing dead trees is between 0.5 to 5 standing dead trees per hectare in order to estimate standing dead tree density within 24% of the actual density, and a complete count survey has been recommended when the mean density of standing

dead tree is less than 0.5 per hectare.

While intensification of sampling efforts or changing plot designs for estimating attributes of standing dead trees is a straightforward solution to the problem, ultimately the cost of intensification of sampling must be considered and weighted against the value of increased accuracy (Curtis and Marshall 2005; Gregoire and Valentine 2008). Adding new plots, increasing plot sizes or otherwise modifying plot designs can be especially costly in the case of large scale (e.g., national) forest inventories and other permanent plot networks and may be infeasible in some cases. Thus, a solution based utilizing relatively cheap auxiliary data or models to enhance estimates derived from standard forest inventory data might provide an efficient alternative to dramatically increasing sampling intensity or modifying plot designs.

One major consequence of surveying too-small an area (e.g., using plot sizes that are too small) is that there may be a large number of zero observations of standing dead trees. Excess zero observations (a.k.a. zero-inflated data) will increase variation in estimates of standing dead tree parameters (Eskelson et al. 2009; Potts and Elith 2006). Because of the tendency for standing dead trees to be aggregated in space and a generally lower abundance of standing dead trees relative to live ones, the problem of zero-inflated data is likely large. For example, about 44% of National Forestry Inventory (FIA) plots observed no dead trees on them (Woodall et al. 2011). Therefore, an estimator, which uses relatively cheap auxiliary information to clarify uncertainty created by zero observation of standing dead trees under standard forest inventory designs, could provide cost-efficient estimation of standing dead tree abundance.

1.3. Goals and Objectives of this Dissertation

This thesis explores a new approach to improving estimation of stand dead tree

abundance, other than adding more plots or modifying plot designs, in the context of the US National Forest Inventory and Analysis Program (Bechtold and Patterson 2005) and US Forest Health Monitoring (FHM) Program (now merged with the FIA plot design) and other similar permanent plot networks. The main goal of this research was to develop a new estimator to predict the abundance of standing dead trees more precisely using auxiliary data and a model to reduce estimation uncertainty associated with excessive zero observations in data. The underlying hypothesis of this thesis was that auxiliary information regarding the distance to the nearest standing dead tree could reduce the variation of estimates caused by large zero observations in data under a given plot design.

In order to meet the main research goal several objectives were undertaken and met:

1. Review the general problem of zero-inflated or large zero observations in data and review sampling methods and models and examine this problem in the specific context of estimating standing dead tree abundance.
2. Develop a new estimation method for standing dead trees which expands on existing zero-inflated modeling methods and which is compatible with typical estimation methods used by forestry practitioners.
3. Use simulation studies and field data to explore the problem of zero-inflated or large zero observations in data for estimating the density of standing dead trees from fixed-radius plot sampling under different simulated and real forest conditions.
4. Examine the properties of the proposed new estimator, under simulated and real forest

conditions.

5. Compare the new estimator with existing methods such as simple random sampling method and simple random sampling combined with existing zero-inflated models, in terms of estimation error and costs of data collection.

6. Finally, this research suggests a sampling strategy for estimating standing dead tree abundance with reference to the FIA/FHM plot design.

2. Review of General Sampling Methods and Models for Dealing with Zero-Inflated Data

2.1. Overview of Sampling Inference

Most forest inventories are focused on estimating population parameters, such as mean per unit area (or plot), and its associated variance, from sample data, which are then extrapolated to the larger population of interest (the reference population). Sampling inference can be classified into design-based and model-based inference. Model-based inference is usually applied at the estimation stage. Gregoire (1998) argued that "sample selection cannot be inherently model based". Sample selection can be described as probabilistic, sequential, or purposive (Gregoire 1998). On the other hand, models can be applied to estimate the population parameters during the estimation stage. Ratio and regression estimators are well-known model-based estimators. In this research, population parameters estimated by both design-based and model-based estimators were compared.

2.1.1. Design-Based Estimation

Suppose the population is consisted of N sample units. Each unit is identified by a label i , which is $i = \{1, \dots, N\}$. Let y_i be the observed characteristics or measurement such as number of standing dead tree at the i th unit. The population is represented by N sample units and observed characteristics y , $Y = \{y_i, \dots, y_N\}$ with unknown parameters. Let y be the sample data, $y = \{y_1, \dots, y_n\}$ where n is number of selected sample units that are selected by specific sampling design. In survey sampling, the interest is making inference of population parameters from the sample data (y).

In the design-based view, the y_i of populations are considered as a fixed set of unknown constants. Therefore, the estimated population parameters are also considered as a fixed constants (Dorazio 1999; Gregoire 1998). Generally, the population mean is defined as $\mu = 1/N \cdot \sum_{i=1}^N y_i$. An estimator which is used to calculate the estimates of population parameters arises directly from the sampling design. The choice of sampling design is important to having precise population parameter estimates because the sampling design defines the selection method for an individual sample unit which represents the population. Sampling design is a system of sampling methods selecting the sample unit from the sample space under the same probability of being sampled. Therefore, the design-based approach can be defined as probability-based sampling which is called randomness. Therefore, no assumptions about the population Y are needed in design-based approach (Dorazio 1999). Based on selected sampling design, an estimator is a rule to calculate the population parameters from the sample data. The term "sampling strategy" is used to define the combination of sampling design and estimator.

An estimator of the population parameter is commonly derived to ensure that the expected value of the estimator equals the value of population parameter. Such estimators are said to be a design-unbiased estimators. For example, the estimator of population mean (μ) under simple random sampling is $\mu = \sum_{i=1}^n y_i / n$. Since $E(\mu)$, which is the expected value of the estimate (μ) estimated by all possible sample y under simple random sampling, is equal to the population mean (μ), the estimator is design-unbiased for μ . The estimators are unbiased regardless of the nature of population. It is an important consequence of linkage between sampling design and estimation.

Most current forest inventory methods for standing dead trees are based on design-based

inference using sampling strategies for standing live trees (Kenning et al. 2005). These methods include simple random sampling with fixed-area plot, variable radius sampling (point sampling), and strip cruising. Fixed-area sampling and point sampling have been recommended to estimate the density of standing dead trees (Bull et al. 1990). For fixed-area sampling, arguably the most common plot design, the shape and size of plot and sampling intensity should be decided at the sampling design stage. Any shape of plots can be used, but circular or rectangular shapes are commonly used in practice. Ganey (1999), for example, used square 1 ha plots to collect the information of standing dead trees, and Bull *et al.* (1990), Stephens (2004), Kenning *et al.* (2005), and the FIA program use a circular plot to collect the standing dead tree information. The size of plot and sampling intensity are usually decided by the variable of interest, allowable costs, and desired precision (Avery and Burkhart 1983). After plot size and sampling intensity have been decided, one needs to decide the sampling design to be applied to establish the plot in the research area. In most situations, sampling locations are point locations. For example, a sampling location is the plot center of circular plot or a corner point of rectangular plot. In this research, simple random sampling (SRS) and systematic sampling (SS) designs were applied to define the baseline method for estimating the abundance of standing dead trees.

Generally, the population parameters of interest are mean per sampling unit or mean per unit area and the population total. When we know at least one population parameter, we can extrapolate the other population parameters. The estimator of sample mean and the variance estimator of sample mean are as follows:

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i$$

$$s_y^2 = S_y^2 / n$$

Where y_i is measured characteristic of interest on sampling unit i such as number of trees, n is number of sampling units (plots) in the sample, and S_y^2 is the variance of y . The mean per unit area is estimated by

$$\bar{y}_{unit} = \bar{y} \cdot E$$

$$E = 1/A$$

where E is an expansion factor and A is the area of plot.

2.1.2. Model-Based Estimation

In model-based view, a model is a theoretical construction which defines the differential probability of objects being sampled because the y_i of population is considered as a realization of one or more stochastic processes. Therefore, the population of Y is modeled as a random variable whose joint distribution $f_y(y|\theta)$ is defined by one or more unknown, fixed parameter(s) θ , where y is observed data set, $y = \{y_1, \dots, y_n\}$. The model is built based on the information of interest such as density, age, or distribution pattern. The information can be obtained from prior researches and/or assumptions suggested by the structure of population. After the model is defined, the common objective is to estimate the unknown parameter θ from sample y (Dorazio 1999). For example, we assume that the underlying distribution of standing dead tree population (Y) is the Poisson distribution, $Y \sim Pois(\lambda)$. We assume the observed data, $y_i = \{y_1, \dots, y_n\}$, is also Poisson distribution. Given this assumption, the number of standing dead trees for each y_i is modeled as a random outcome of a Poisson process, which has the density function $f(y; \lambda) = \lambda^y \cdot e^{-\lambda} / y!$. The unknown parameter λ of Poisson distribution

corresponds to the probability that number of standing dead trees tallied within plot. Under the Poisson model, $E(\mu) = \lambda$ and inference about λ and μ are equivalent. In other words, the model which is fitted with estimated parameter λ is used to make inference about the standing dead tree population. The parameter (θ) and population parameters are commonly estimated by the method of maximum likelihood (ML), which defines the value of θ that is most likely for the given data and model. ML is estimated by maximizing the likelihood function $L(\theta|y)$:

$$L(\theta|y) = f(y|\theta),$$

where y is observed or selected data set. Because the likelihood function is a function of θ with fixed y , we have different parameter θ for the different sample set. The ML method provides the most likely value for θ . In addition, the sampling design is not relevant to inference because the estimation is based on the likelihood function. Gregoire (1998) stated that in model-based inference, an estimator of population mean ($\hat{\mu}$) is to be model unbiased when

$$E(\hat{\mu} - \mu) = 0.$$

There are many advantages for model-based estimators and inferential procedures. Since population Y is a single realization of one or more stochastic processes, there is considerable flexibility in identifying and selecting classes of models for approximating the true value which is underlying processes believed to have generated Y . The prior and observed information can be used to determine the model. For example, according to the several studies and prior information, the number of standing dead trees is observed in each plot and y values are often dominated by counts of 0 or 1 (e.g., Eskelson *et al.* 2009) and y values are nonnegative integers. Based on the prior information, we can assume that the underlying distribution can be Poisson or negative binomial distribution. Based on the assumption, the model is fitted to sample data.

However, it is not always easy to find the best model. If the model of data is not correct, the estimated parameters may be biased (Dorazio 1999). Therefore, we need to be careful when we select the model to make inference from the sampling data.

Count regression models have been applied to estimate the abundance of species, dead trees, and mortality. Count regression models have been widely used in fields such as econometrics, epidemiology, and nursing, but have only recently been introduced into forest ecosystem sciences (e.g., Affleck, 2006). In a comparison of five count regression models for estimating the abundance of a vulnerable plant species, Potts and Elith (2006) demonstrated that the Hurdle model, which assumes that the observed data set follows mixture of two statistical distributions, showed the best fit to model the observed data set. Negative binomial regression model and nearest neighbor imputation method were compared by Eskelson et al. (2009) to predict the abundance of standing dead trees and cavity trees (Eskelson et al. 2009). Affleck (2006) also compared different count regression models to predict the stand mortality, finding the Hurdle and negative binomial models to provide good fits to data. These previous studies aimed to find the best fit model for observed data set to explain the relationship between the abundance data or mortality and some environmental variables. These indicate that the model performance is different by underlying assumption for population parameters. Hence, the assumption for the underlying statistical distribution of the model is important factor to predict the abundance of standing dead tree population. Several count regression models used to model standing dead tree data are introduced next chapter.

2.2. Statistical Models for Count Data and the Problem of Zero-Inflation

2.2.1. Overview

As the statistical model, count data, $y = (y_1, \dots, y_n)$, is a random variable. The count data y is modeled by the probability mass function ($f(y|\theta)$) such as Poisson or negative binomial (NB) distribution which is characterized by one or more unknown and fixed parameter θ . Statistical models for count data are a kind of discrete response models that aim to explain the number of occurrences or count of events (Hilbe 2007). Since count data has only non-negative integer values, such as the number of standing dead trees per plot, the probability mass which uses non-negative integer values is applied to model the count data, such as Poisson and negative binomial distribution. Here, the counts of interest are counts of standing dead trees in fixed-area plots, which form the base data for making inference about standing dead tree density (standing dead trees per unit area).

2.2.2. Poisson Model

The Poisson model is a basic count regression model derived from the Poisson distribution. For example, let's say that Y_i is the number of standing dead trees at site i . Under simple random sampling (SRS), we typically assume that Y_i is Poisson-distributed with a mean μ_i and an associated variance. The number of occurrences in count data (i.e., counts in the fixed-area plots) is estimated by Poisson distribution. The probability function of Poisson distribution can be characterized as

$$P\{Y = y\} = \frac{e^{-\mu} \mu^y}{y!}$$

where $y = \{0, 1, 2, 3, \dots\}$, the random variable Y is the count response, and the parameter μ is the expected value and variance. An important property of the Poisson probability is that the mean ($\mu = E[Y]$) and variance ($Var[Y]$) are equal.

One of the central problems associated with both spatial aggregation and low abundance of standing dead trees is the excess zero observations in sample data (i.e., plots where no dead trees were observed). When there is a larger proportion of zero count observations in the data than assumed under the statistical distribution (i.e., Poisson distribution), this data is called ‘zero inflated data’ (Tu 2002). Typical statistical methods are unsuitable or difficult to apply when analyzing zero inflated data. For example, Ganey (1999) used medians and interquantile ranges instead of means and variances for inference, because the frequency of standing dead trees in his sample data was highly skewed. Zero-inflated data usually do not hold to normality or homoscedasticity assumptions. In addition, insufficient nonzero observations increase the uncertainty of estimation. Therefore, statistical models, which apply other statistical distributions, have been developed to model large proportions of zero observations in data.

2.2.3. Negative Binomial Model

The negative binomial (NB) distribution can be applied when there is over-dispersion in a Poisson regression model. The NB distribution is as a combination of two distributions, giving a combined Poisson-gamma distribution. It assumes that count responses (y) are Poisson distributed with μ , and μ is assumed to follow a gamma distribution. The NB distribution is characterized as

$$P\{Y = y\} = \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)} \cdot \left(\frac{k}{\mu+k}\right)^k \cdot \left(1 - \frac{k}{\mu+k}\right)^y$$

where $y = \{0, 1, 2, 3, \dots\}$, the random variable Y is the count response. The expected value is μ and the variance is $\mu + \mu^2/k$ where k is a dispersion parameter. When k is large, the term μ^2/k is approximately 0. Although the NB model is more flexible than the Poisson model, it may also problems where zero probability in data is greater than or less than the zero probability estimated by a regular count distribution.

2.2.4. Hurdle Model

The Hurdle model, proposed by Mullahy (1986), was developed to account for excess zero observations in count data; it consists of two component models representing two processes defining the proportion of zero and non-zero counts and the “hurdle” is a partition between them. For the case of standing dead trees, one process is causing the absence of standing dead trees at a location and another process is influencing the number of standing dead trees where they occur. In the Hurdle model, the binomial distribution can be used to model the absence and presence of standing dead trees and a zero-truncated count distribution, such as Poisson and NB distribution, can be used to model the portion of the count data where at least one standing dead tree was found. The combined probability function for the Hurdle model is defined as follows:

$$\Pr\{Y = y\} = \begin{cases} \pi & y = 0 \\ (1 - \pi) \times \frac{f(Y = y | \mu)}{1 - f(Y = 0 | \mu)} & y > 0 \end{cases}$$

where π is observed zero probability in count data according to binomial distribution, $f(Y = y | \mu) / [1 - f(Y = 0 | \mu)]$ is the zero-truncated form of Poisson or NB distribution, and $f(Y = 0 | \mu)$ is the probability of zero estimated by Poisson or NB distribution. The expected value and variance of Y in Poisson Hurdle model are estimated by:

$$E_{hdp}[Y] = (1 - \pi) \cdot \frac{\mu}{1 - e^{-\mu}}$$

$$Var_{hdp}[Y] = (1 - \pi) \cdot \frac{\mu + \mu^2}{1 - e^{-\mu}} - \left[(1 - \pi) \frac{\mu}{1 - e^{-\mu}} \right]^2$$

where π is the observed zero probability estimated by binomial process and μ is the expected value of the Poisson distribution. In case of NB Hurdle model, the expected value and variance of Y in NB Hurdle model are estimated by:

$$E_{hdNB}[Y] = (1 - \pi) \cdot \frac{\mu}{1 - P_0} \quad \text{where } P_0 = \left(\frac{k}{\mu_i + k} \right)^k$$

$$Var_{hdNB}[Y] = \frac{(1 - \pi)}{1 - P_0} \cdot \left(\mu^2 + \mu + \frac{\mu^2}{k} \right) - \left[(1 - \pi) \frac{\mu}{1 - P_0} \right]^2$$

where π is observed zero probability estimated by binomial process, μ is the expected value of NB distribution, k is dispersion parameter which is estimated by NB distribution. The zero probabilities (π) for both Poisson Hurdle and NB Hurdle models are estimated from binomial process using observed absence and presence data.

3. Development of the EZ-Hurdle Method

3.1. Model Development

Since the Hurdle model estimates the zero-probability in the count data based on the observed proportion of 0's in the sample data, the zero-probability is a population parameter also subject to sampling error due to, e.g., a too-small plot size and/or variation in standing dead tree spatial pattern and abundance. To reduce this variability, we proposed the Expected-Zero (EZ) Hurdle model, which replaces the observed proportion of 0's from the sample data with an expected zero-probability obtained from auxiliary information. This expected zero-probability provides additional information which, theoretically, should reduce random variation associated with observed zero-proportion in the count data.

To use the additional information regarding the zero probability, the EZ-Hurdle method adds an additional stage to the estimation process. The first step is estimating the expected zero-probability observations (P_e) given the sampling intensity. Here, the expected zero-probability is estimated by the predicted inclusion probability of a standing dead tree for a given search radius, but it could be estimated otherwise (e.g., from another data source or model). The relationship between the expected zero-probability and the inclusion probability of a standing dead tree is:

$$P_{eid} = 1 - P_{Inc}(id)$$

where P_{eid} is the expected zero probability for the given search radius d with restricted condition i such as spatial pattern and density of standing dead trees and $P_{Inc}(id)$ is the inclusion probability of a standing dead tree at the search radius d . In this application, a model is applied to find the inclusion probability of a standing dead tree for a given search radius at

each sample point, using the observed distance from a sample point to the nearest standing dead tree as the source of auxiliary information. As the fixed-area plot radius increases, the inclusion probability of a standing dead tree approaches one and the number of non-observations, i.e., the number of 0's, approaches zero. Thus, for any given search radius employed during sampling, an expected number of 0's can be estimated from a data set consisting of point-to-dead tree distances, or from a function derived from such data.

The second step is estimating non-zero counts using zero-truncated count distributions such as Poisson or NB distribution by adjusting the zero-probability estimated during the first step. An expected zero-probability is obtained as a function of a search radius d , and the probability (P_{eid}) is plugged into the following probability function:

$$P\{Y = y\} = \begin{cases} P_{eid} & y = 0 \\ (1 - P_{eid}) \cdot \frac{f(Y = y | \mu_{id})}{1 - f(Y = 0 | \mu_{id})} & y > 0 \end{cases}$$

where P_{eid} is the modeled expected zero-probability for the given search radius d and μ_{id} is estimated mean by truncated count distributions such as Poisson and NB distribution for the given search radius d and i is the restricted condition such as spatial pattern and density of standing dead trees. The log-likelihood function of the EZ-Hurdle model is as follows:

$$L(y_{i1}, \dots, y_{in}) = \sum_{j=1}^n \left\{ \ln f(y_{ij} | \mu_i) - \ln [1 - f(0 | \mu_i)] \right\}$$

If one can estimate the expected zero-probability for the given search radius (d) precisely, one can obtain a more precise estimate of the number of standing dead trees. For the EZ-Hurdle with Poisson distribution (EZP) model, the expected value and variance for the count distribution are:

$$E_{EZP}[Y_i] = (1 - P_{eid}) \cdot \frac{\mu_{id}}{1 - e^{-\mu_{id}}}$$

$$Var_{EZP}[Y_i] = (1 - P_{eid}) \cdot \frac{\mu_{id} + \mu_{id}^2}{1 - e^{-\mu_{id}}} - \left[(1 - P_{eid}) \frac{\mu_{id}}{1 - e^{-\mu_{id}}} \right]^2$$

and for the EZ-Hurdle model with NB distribution (EZNB):

$$E_{EZNB}[Y_{id}] = (1 - P_{eid}) \cdot \frac{\mu_{id}}{1 - P_0} \quad \text{where } P_0 = \left(\frac{k_{id}}{\mu_{id} + k_{id}} \right)^{k_{id}}$$

$$Var_{EZNB}[Y_i] = \frac{(1 - P_{eid})}{1 - P_0} \left(\mu_i^2 + \mu_i + \frac{\mu_i^2}{k} \right) - \left[(1 - P_{eid}) \frac{\mu_i}{1 - P_0} \right]^2$$

where P_{eid} is the expected zero-probability estimated by model for the given search radius d ,

μ_{id} is the expected value of truncated Poisson or NB distribution for restriction i , such as spatial

pattern and density, and k_{id} is dispersion parameter which is estimated by NB distribution.

4. A Simulation Study to Understand the Properties of the EZ-Hurdle Method and Compare it to Related Approaches

4.1. Overview

A simulation study was chosen to illustrate the properties of the EZ-Hurdle method and to compare it to some existing approaches. A simulation study was advantageous because the true density of standing dead trees was known and because different spatial patterns and abundances of dead trees could be devised. A simulated area of 324ha (1,800×1,800m) was chosen to represent a large forest stand and in order to avoid the problem caused by small data set which is less than 30 observations when we applied large plot size. Three different spatial patterns random, clustered and highly clustered were applied to generate the standing dead trees at three different abundance levels ranging from 12/ha to 49/ha (Table 4.1).

Table 4.1. Parameters to create clustering patterns from a Matern Cluster Point Process.

Spatial Pattern	No. of clusters/ha	Radius of cluster(m)	Density/ha
Cluster I	3	30	12, 24, 49
Cluster II	2	30	12, 24, 49

With typical tree abundances in forests ranging from hundreds to thousands of trees per hectare, this represents a range of dead tree relative abundances ranging from about 1 to 25% of all standing trees. Random patterns were generated by a Poisson process, and clustered patterns were generated by the Matern-cluster point process. A Matern-clustered point process was used, where the number of clusters, cluster size, and mean density are fixed for each cluster (Matern 1986). The Spatstat package in R (Baddeley and Turner 2005) was used to create simulated

populations of standing dead trees. Figure 4.1 shows an example of standing dead trees generated for the simulation. Cluster pattern II has more aggregation than cluster pattern I when the density is 12.36/ha.

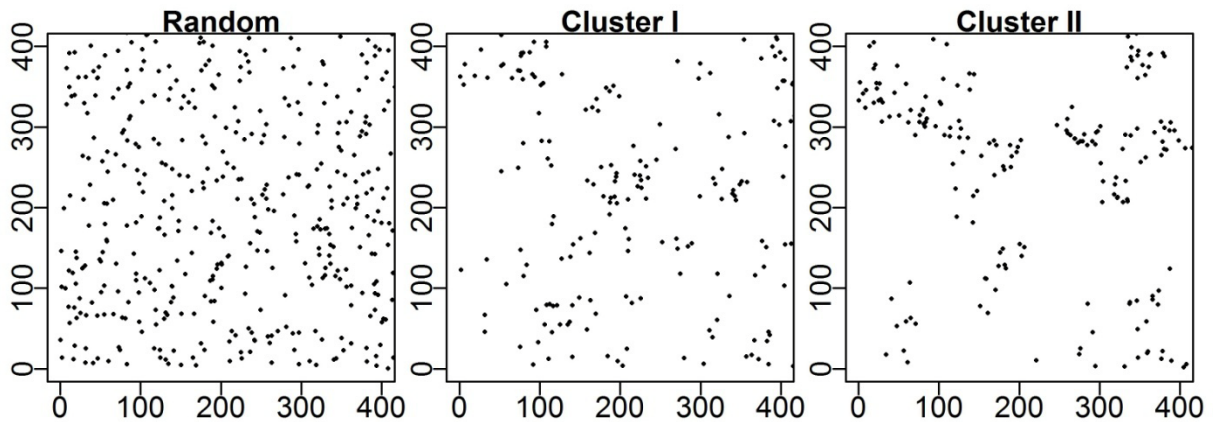


Figure 4.1. The distribution of simulated standing dead trees by spatial pattern when the density is 12/ha. The area depicted for each pattern is about 18 ha.

4.2. Data

In fixed radius plot sampling method, sampling locations are point locations, which are commonly selected randomly or systematically from the continuous area frame. A sampling location is the center of circular plot or a corner point of a rectangular plot. In this simulation study, two different fixed-radius sample plots were used. One is a 7.32 m radius circular plot which is same as the ‘subplot’ in the FIA design. Another is a 17.95 m radius circular plot which is same as the ‘annular plot’ in the FIA design. In order to reduce the edge effect, a 50 m buffer area was applied. Therefore, plot locations (plot centers) were randomly selected within a core area which was 289 ha. Ten different sampling intensities were applied representing 1% to 10% of the total forest area covered by the sum of the areas of all the plots, by a 1% interval for the two specified plot sizes (Table 4.2). The number of standing dead trees was counted and the

distance from the plot center to the nearest standing dead tree was measured. Three thousand data sets were generated for each sampling intensity and plot size by spatial pattern using a custom algorithm written in the R statistical computing language (R Development Core Team 2011). The three thousand data sets were used to estimate the population parameters under each sampling method.

Table 4.2. Number of samples by sampling intensity (%) and plot size (radius, m).

Intensity (%) \ radius(m)	7.32	17.95
1	193	32
2	385	64
3	578	96
4	771	128
5	964	160
6	1,156	192
7	1,349	224
8	1,542	256
9	1,735	288
10	1,927	320

4.3. Methods

4.3.1. Estimating the Expected Zero Probability

The inclusion probability of a standing dead tree was modeled based on the distance from random points to the nearest standing dead tree. Models fit to the simulation data were used to calculate the inclusion probability of at least one standing dead tree ($P_{Inc}(d)$) and the corresponding expected zero probability ($1 - P_{Inc}(d)$) for each search radius. We considered an array of potential models that are sigmoid-type functions (Table 4.3), which indicate an initially

low change in the inclusion probability over relatively short distances and then exponentially increasing inclusion probability as the search radius continues to increase, ultimately saturating as the inclusion probability approaches one; this pattern was observed in the data. Two models were selected and the best of the two was determined based on root mean squared error (RMSE) and Aikake's Information Criterion (AIC).

Table 4.3. Models to estimate the inclusion probability of a standing dead tree.

Name	Model
Model I (Reduced Logistic function)	$\begin{cases} P_{Inc}(id) = 0, & d = 0 \\ P_{Inc}(id) = \frac{1}{1 + b \cdot \exp^{-c \times d}}, & d > 0 \end{cases}$
Model II (Reduced Gompertz function)	$\begin{cases} P_{Inc}(id) = 0, & d = 0 \\ P_{Inc}(id) = \exp(-b \times c^d), & d > 0 \end{cases}$

$P_{Inc}(id)$ is the inclusion probability of a standing dead tree in forest i and d is search radius (m).

Figure 4.2 shows the change of inclusion probability of a standing dead tree estimated by the reduced Gompertz function. Both models are sigmoid curves. The change of inclusion probability has strong relationship with the inflection point. For example, a inflection point moves to the left (close to 0) when the density of standing dead trees is getting high and the spatial pattern of standing dead trees is random because there are high inclusion probabilities.

In order to find the best model, 964 points were randomly selected by spatial pattern and density. There is one reason that the maximum number of random points is 964 because according to a previous study, 5% sampling intensity is recommended to estimate the abundance of standing dead trees when fixed-area sampling method is applied (Bull et al. 1990). Hence, it assumed that 964 random points should be enough to test the model performance to estimate the

inclusion probability of a standing dead tree. Three hundred iterations had been applied by spatial pattern and density of standing dead trees. For each iteration, RMSE and AIC were calculated. Finally, the best model was selected which has better RMSE and AIC.

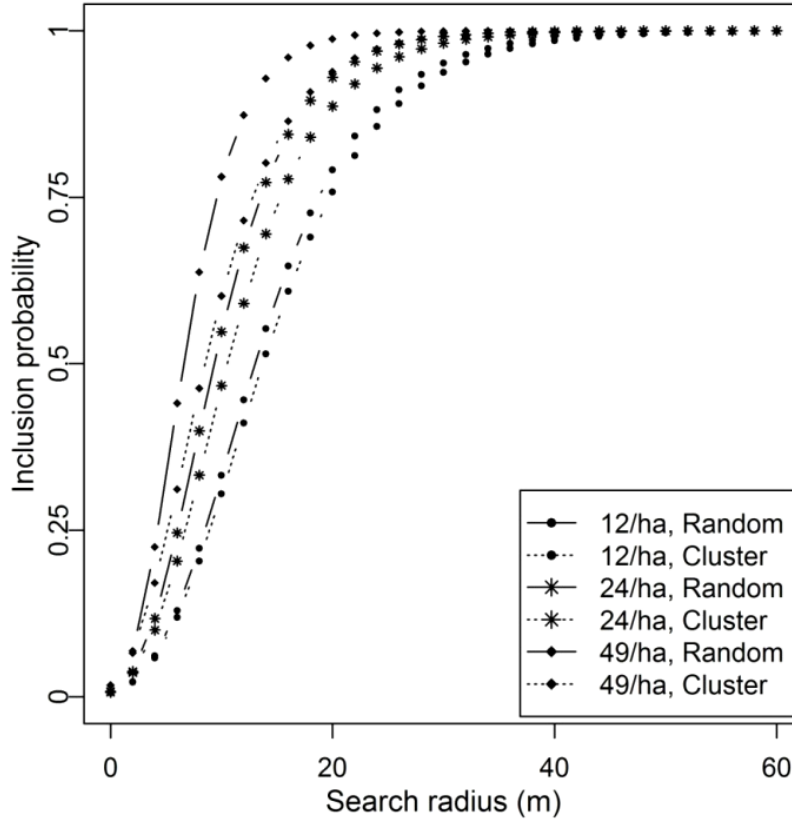


Figure 4.2. The change of inclusion probability of a standing dead tree estimated by the reduced Gompertz function under different conditions.

It was expected that the precision of estimated inclusion probability should be increased with increasing the number of random points. In order to find a reliable number of random points to estimate the inclusion probability, the change of variation for the estimated inclusion probability was examined by spatial pattern and density using the different number of random points from 30 to 1,050. The inclusion probability for given search radii 7.32 and 17.95 m were

estimated from 3,000 data sets by spatial pattern and density using selected model.

4.3.2. Comparison of EZ-Hurdle to Existing Methods and Models

Using 3,000 simulated data sets, the abundance of standing dead trees per ha was estimated by spatial pattern and density of standing dead trees and for each method including SRS, Poisson, Poisson-Hurdle, NB-Hurdle, Poisson EZ-Hurdle, and NB EZ-Hurdle method. In the simulation, we knew the true abundance of standing dead trees per ha, so, root mean square error (RMSE) and error (Error) were used to compare the methods. The RMSE and average bias were computed as follows for each simulation environment:

Error = $(E_{ijkmp} - T_{jkmp})$, and

$$RMSE_{jkmp} = \sqrt{\frac{\sum_{i=1}^n (E_{ijkmp} - T_{jkmp})^2}{n}}$$

where E is estimated density of standing dead trees per ha by method, T is the true density of standing dead trees, i is the number of iterations (3,000), j is the density of standing dead trees, k is a spatial pattern, m is a sampling intensity, and p is a plot size.

4.3.3. Sensitivity Analysis

Due to the EZ-Hurdle method being a model-based estimator, any estimate is sensitive to the modeled inclusion probability. Therefore, a sensitivity analysis was conducted to examine the sensitivity of the EZ-Hurdle method to errors in specifying the corrected inclusion probability, i.e., when the expected probability of zero is different from the “true” zero probability under a fixed sample size and plot size. This was quantified by the estimated % change in the estimation

error of standing dead tree density estimated by EZ-Hurdle method as a function of changing the inclusion probability of a SDT under holding all other factors constant such as spatial pattern, density, and plot size. The percent error (%) in standing dead tree density was calculated as:

$$\text{percent error} = \frac{|T_D - E_D|}{T_D} \times 100,$$

where T_D is the true density and E_D is the estimated density by EZ-Hurdle method.

4.4. Results

4.4.1. Inclusion probability

The reduced Gompertz function (Model II) was selected to model the inclusion probability. Model II tended to have the smallest RMSE and AIC for overall spatial pattern and density (Table 4.4). Both models show that RMSE increases with increasing clustering of dead trees relative to a random pattern (Table 4.4), since there is a lower inclusion probability of a standing dead tree under a clustered pattern and a correspondingly greater zero inflation for any given density.

Table 4.4. Mean RMSE and AIC by model to estimate the inclusion probability of a standing dead tree.

Spatial Pattern	Density/ha	Model I		Model II	
		RMSE	AIC	RMSE	AIC
Random	12.4	0.159	-2932	0.061	-3479
	24.7	0.042	-678	0.042	-2454
	42.0	0.092	-1486	0.031	-1805
	49.4	0.082	-1544	0.032	-1822
Clustered I	12.4	0.651	-3513	0.093	-5119
	24.7	0.980	-2719	0.276	-3642
	42.0	1.353	-3130	0.574	-3875
	49.4	1.423	-2488	0.641	-3074
Clustered II	12.4	0.842	-3321	0.412	-4835
	24.7	0.652	-2343	0.326	-3538
	42.0	0.843	-2874	0.756	-3427
	49.4	1.321	-2784	0.847	-3173

The variation of the estimated inclusion probabilities decreased with increasing number of random points for all spatial pattern and densities for both search radii (Figs. 4.3 and 4.4). With an increasing number of random points, the mean of inclusion probability approaches the true value. For example, the inclusion probability approaches to 0.176 with increasing the

random points when spatial pattern is random, density is 12/ha, and search radius is 7.32m (Fig. 4.3). The results clearly show that the 95% CI are decreased with increasing the random points and that the confidence intervals are overlapped and do not dramatically decrease after 500 random points. However, in absolute terms, the 95% CI were less than 1% over all sampling intensities, spatial patterns, and densities (Figs. 4.3 and 4.4), so a much smaller number of points than 500 can yield good estimates of the inclusion probabilities. Therefore, the 500 random points were used to estimate the inclusion probability of a standing dead tree during the simulation.

The inclusion probabilities were different by the spatial patterns and densities of a standing dead tree for both radii (Figs. 4.3 and 4.4). Inclusion probabilities were higher under a random distribution of trees than that found for the two clustered patterns for all densities and search radii. In addition, clustered pattern I showed relatively higher inclusion probabilities than clustered pattern II because clustered pattern II has more aggregation than clustered pattern I. These results indicate that the inclusion probabilities are highest when the density of standing dead trees is high and randomly distributed in area and lowest with a few trees in tight clusters. However, the increase in inclusion probabilities with increasing dead tree density was much smaller when trees were clustered than randomly distributed (compare the top and bottom rows of sub-figures in Figs. 4.3 and 4.4), because in the former case, the inclusion probability was largely dependent on a sample point landing in or near a cluster, rather than how many dead trees were in a cluster.

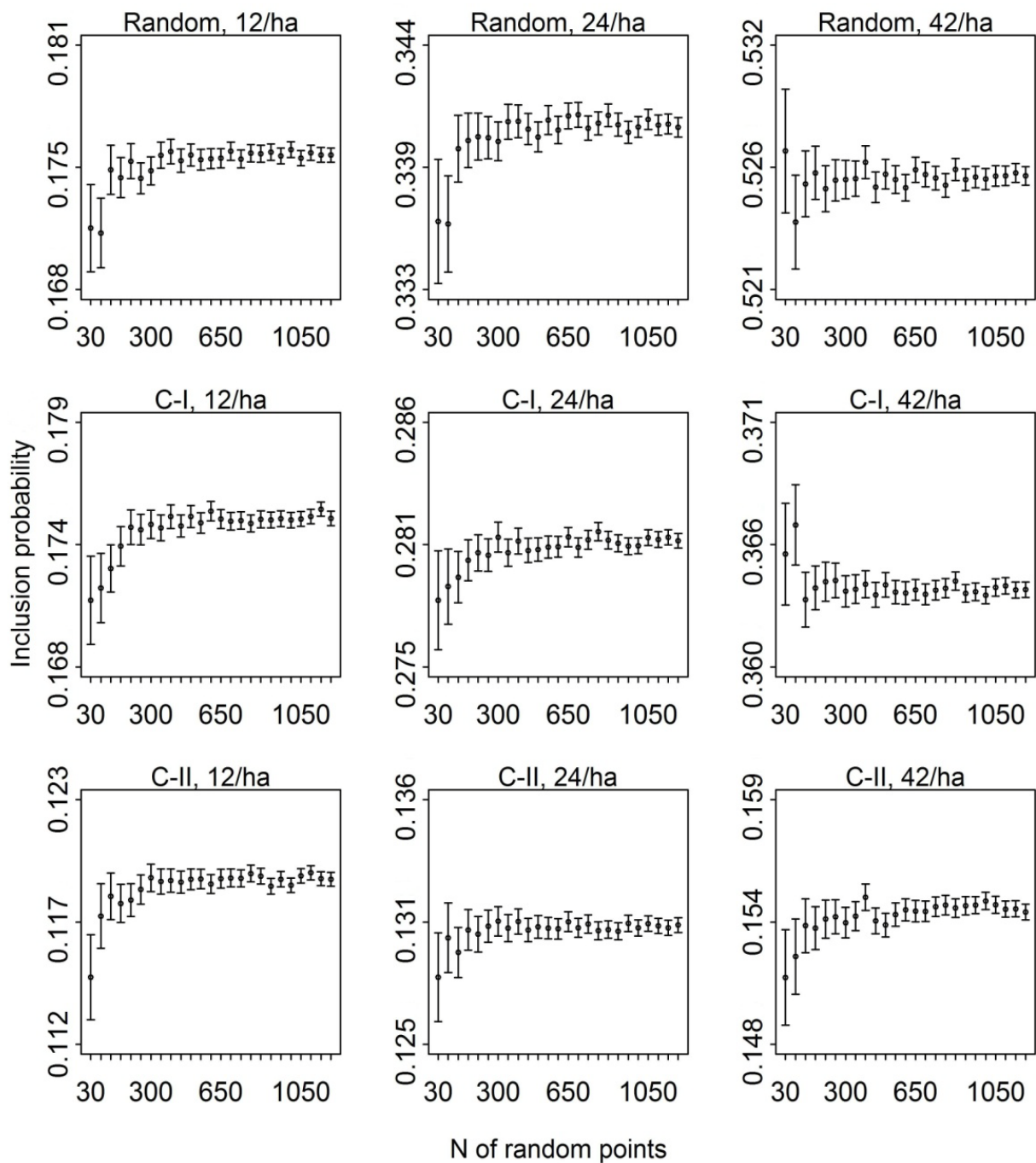


Figure 4.3. The mean and 95% CI of inclusion probabilities by spatial patterns and densities of standing dead trees when the search radius is 7.32m.

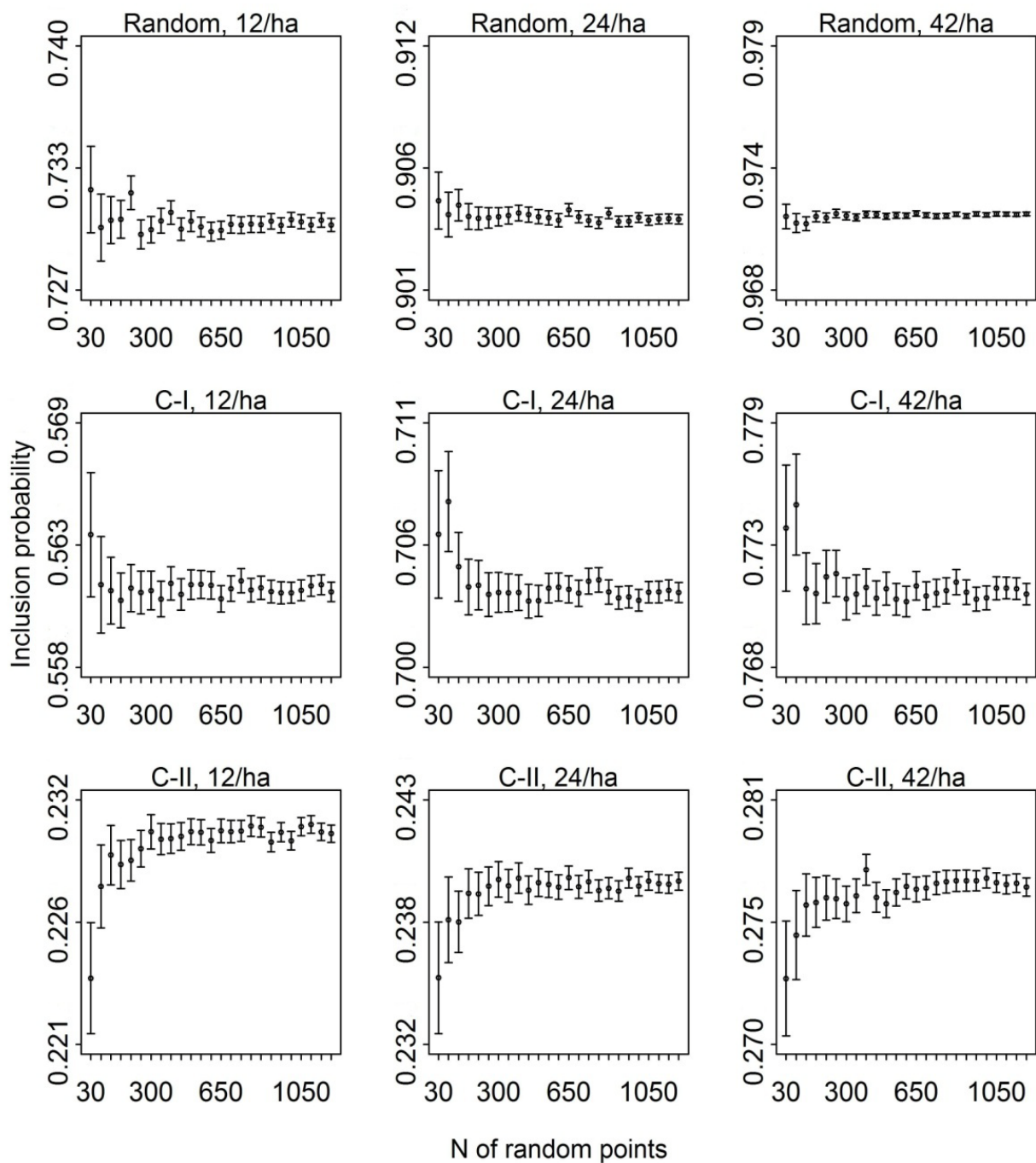


Figure 4.4. The mean and 95% CI of inclusion probabilities by spatial patterns and densities of standing dead trees when the search radius is 17.95m.

One way to understand how the EZ-Hurdle method works is to see how the observed zero proportion varies over many samples and compared that to the variation in the expected zero probability predicted from the model. Figure 4.5 shows the variation of observed and expected zero proportion (i.e., the expected zero probability) by spatial patterns in 3,000 simulated data sets.

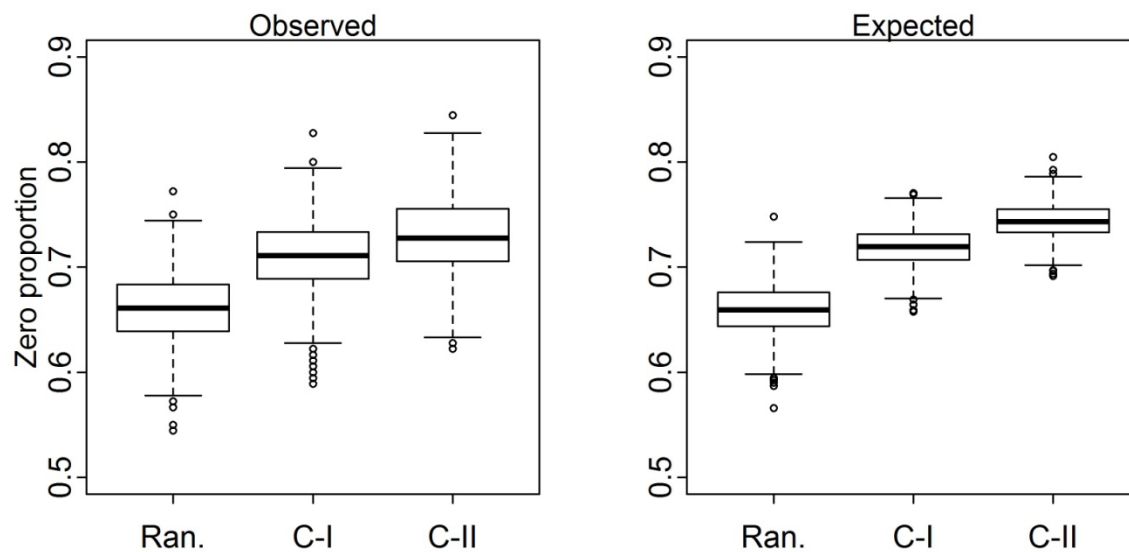


Figure 4.5. Observed and expected zero proportion in 3,000 simulated data sets by spatial patterns, when the plot radius is 7.32m and density is 24/ha.

In this example, there are 578 random plots with a sampling intensity of 3% and a plot radius is 7.32 m. The mean of observed zero proportion is increased with decreasing density of standing dead trees. The mean of observed zero proportion is approximately 0.67 or 67% when the spatial pattern is random. It means that approximately 67% of plots among 3,000 plots have no dead trees. The mean of modeled expected zero probability when 500 random points are used is approximately 0.66 at the same spatial pattern, so only slightly different than the observed. However, and the expected zero probability from the model has an order of magnitude smaller

variation than observed zero proportion. Especially, there is more improvement of zero proportion when the standing dead trees are more clustered. Therefore, EZ-Hurdle method can have better estimates than SRS or other methods which used observed zero proportion because there is less variation in the estimate of the observed zero proportion between samples.

4.4.2. Comparison of Estimated Density of SDT/ha by Methods

The EZ-Hurdle method showed better precision in estimating standing dead tree density than SRS and various forms of the Hurdle model over all scenarios examined. Figure 4.6 and 4.7 show the distribution of estimated densities of standing dead trees per ha using different estimators, based on 3,000 samples generated at different sampling intensities, with lower (Fig. 4.6) and higher dead tree densities (Fig. 4.7), respectively. The data clearly show a narrower distribution of estimates with a greater proportion of estimates concentrated around the mean when the EZ-Hurdle method is employed (Figs. 4.6 and 4.7). The EZ-Hurdle with Poisson (EZP) or EZ-Hurdle with NB distribution (EZ-NB) yielded nearly identical distributions of estimates to each other, so only results for EZP are shown. SRS, the Hurdle model with Poisson (HDP) or the Hurdle model with negative binomial distribution (NB-Hurdle) methods yielded nearly identical distributions of estimates to each other. The lack of difference of Hurdle-based approaches from SRS is not surprising, since they all use the observed zero proportion from the plot data, while the EZ-Hurdle method uses nearest standing dead tree distances to reduce the variation associated with zero observations in the data.

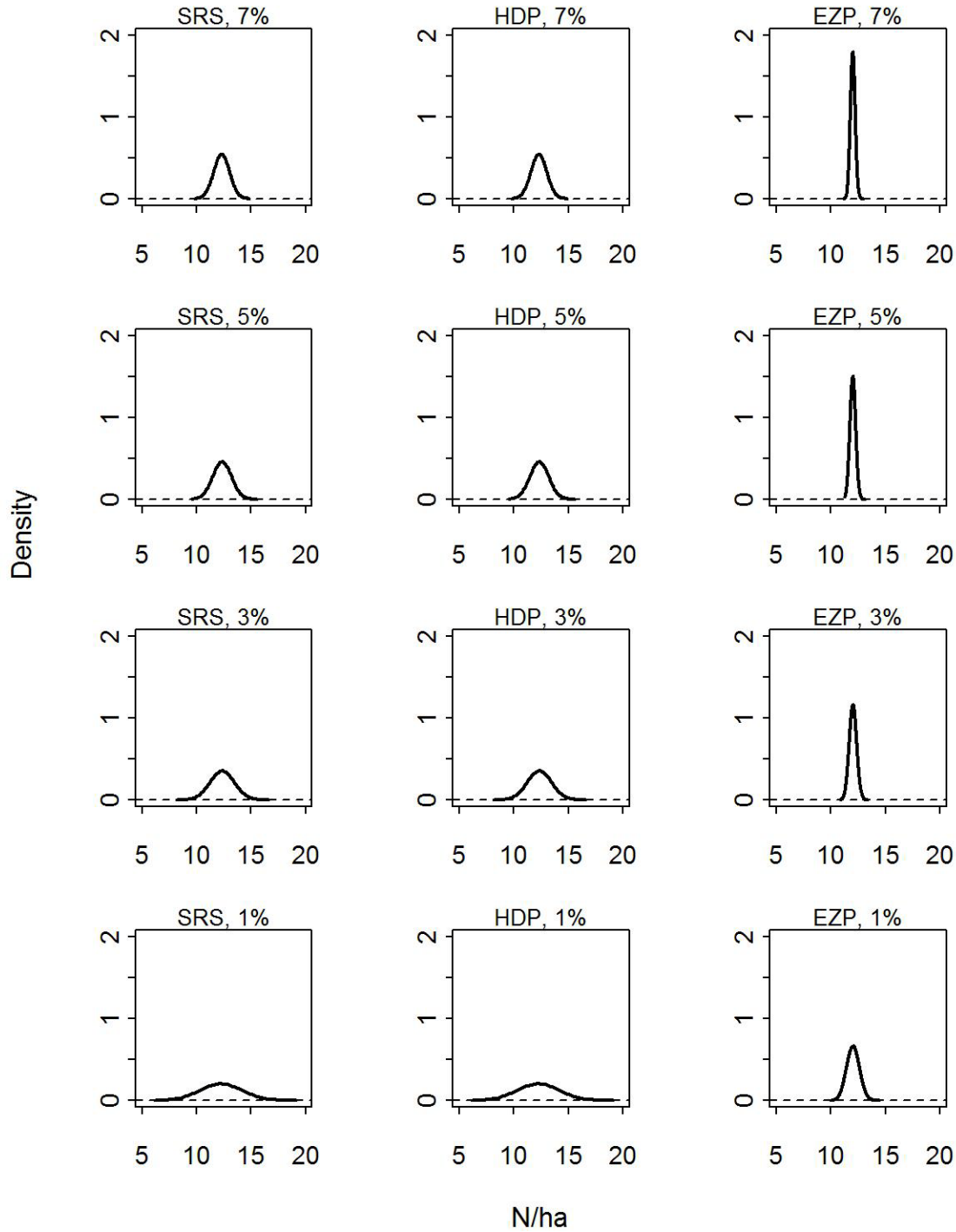


Figure 4.6. The distribution of estimated density of standing dead trees by sampling intensities (1%, 3%, 5%, 7%) and methods when the plot radius is 7.32m and spatial pattern is random. True density is 12 trees per ha. SRS is simple random sampling method with fixed-area circular plot. HDP is Hurdle model with Poisson distribution and EZP is EZ-Hurdle model with Poisson distribution.

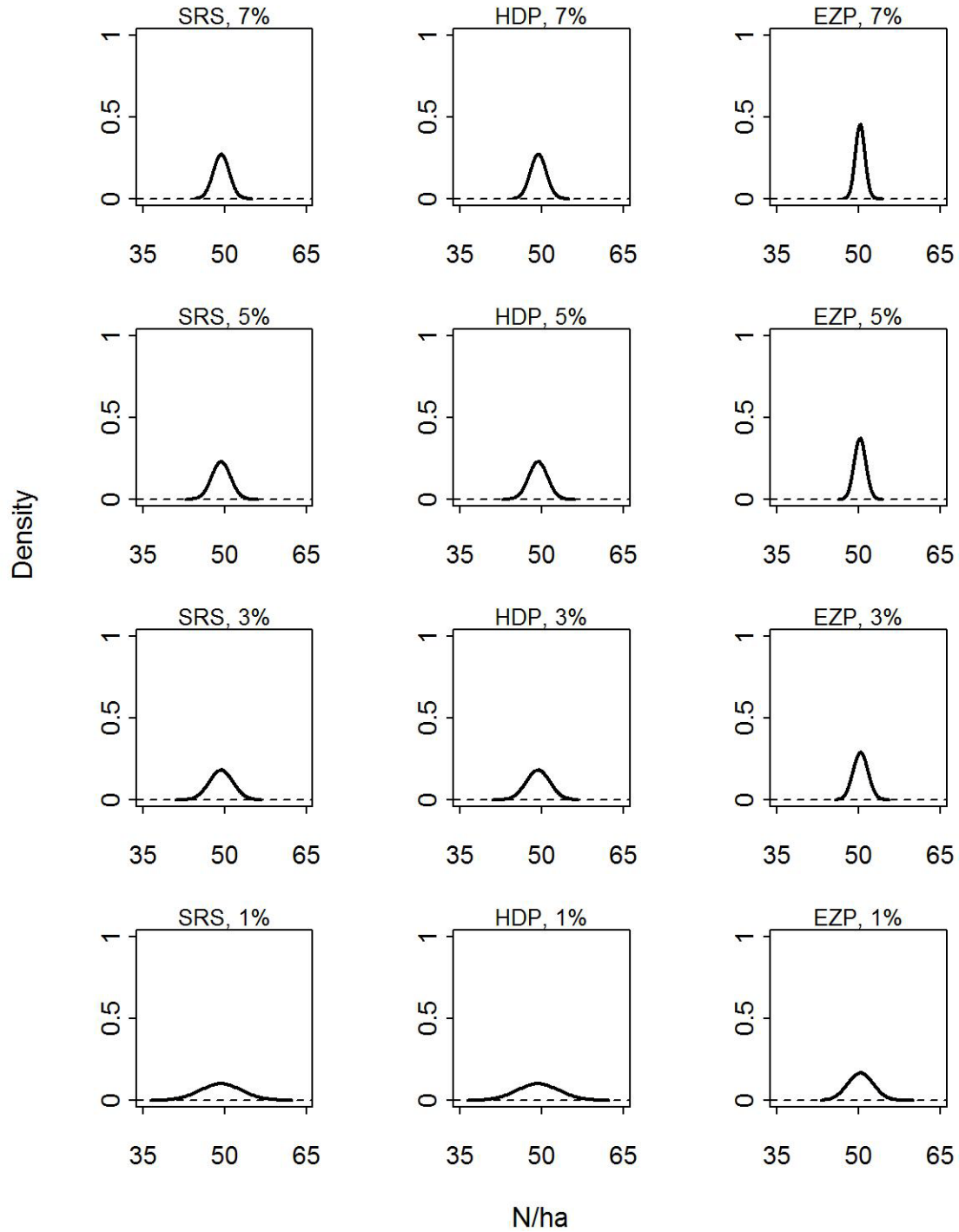


Figure 4.7. The distribution of estimated density of standing dead trees by sampling intensities (1%, 3%, 5%, 7%) and methods when the plot radius is 7.32m and spatial pattern is Cluster I. True density is 49 trees per ha. SRS is simple random sampling method with fixed-area circular plot. HDP is Hurdle model with Poisson distribution and EZP is EZ-Hurdle method with Poisson distribution.

As expected, the precision of all methods increased with increasing sampling intensity. However, under the same sampling intensity, the distribution of estimates was more different between EZ-Hurdle and other methods when plot size was small and density was low (Figs. 4.6 and 4.7). Conversely, the distribution of estimates between all the methods were much more similar when a larger plot size was applied and a lower sampling intensity was employed, although the EZ-Hurdle model did not converge on the same distribution of estimates given by the other methods for any other the scenarios examined. Theoretically, SRS and the EZ-Hurdle method should converge only when the expected and observed zero proportions are the same.

The RMSE captures the overall difference in estimation error between the methods as compared to the true estimate. The RMSE of EZP was almost always lower than that from SRS under all of the different scenarios examined, except at the highest density and largest plot size when sampling intensity was greater than 5% (Fig. 4.8, the other Hurdle methods are omitted from the figure because they give nearly identical results to SRS). The difference in RMSE of all the methods decreases as the sampling intensity increases under the same conditions (plot size, density and spatial pattern), with EZ-Hurdle always outperforming SRS when sampling intensity is less than 5% and in many cases still outperforming SRS up to the maximum sampling intensity examined (10%). All the methods worked the best when dead trees were distributed randomly in space and the superiority of the EZ-Hurdle method was greater when standing dead trees were clustered in space (for reasons explained previously, see Fig. 4.6). When zero inflation was highest (low density and small plots, upper left sub-figure, Fig. 4.8), EZ Hurdle was superior regardless of spatial pattern.

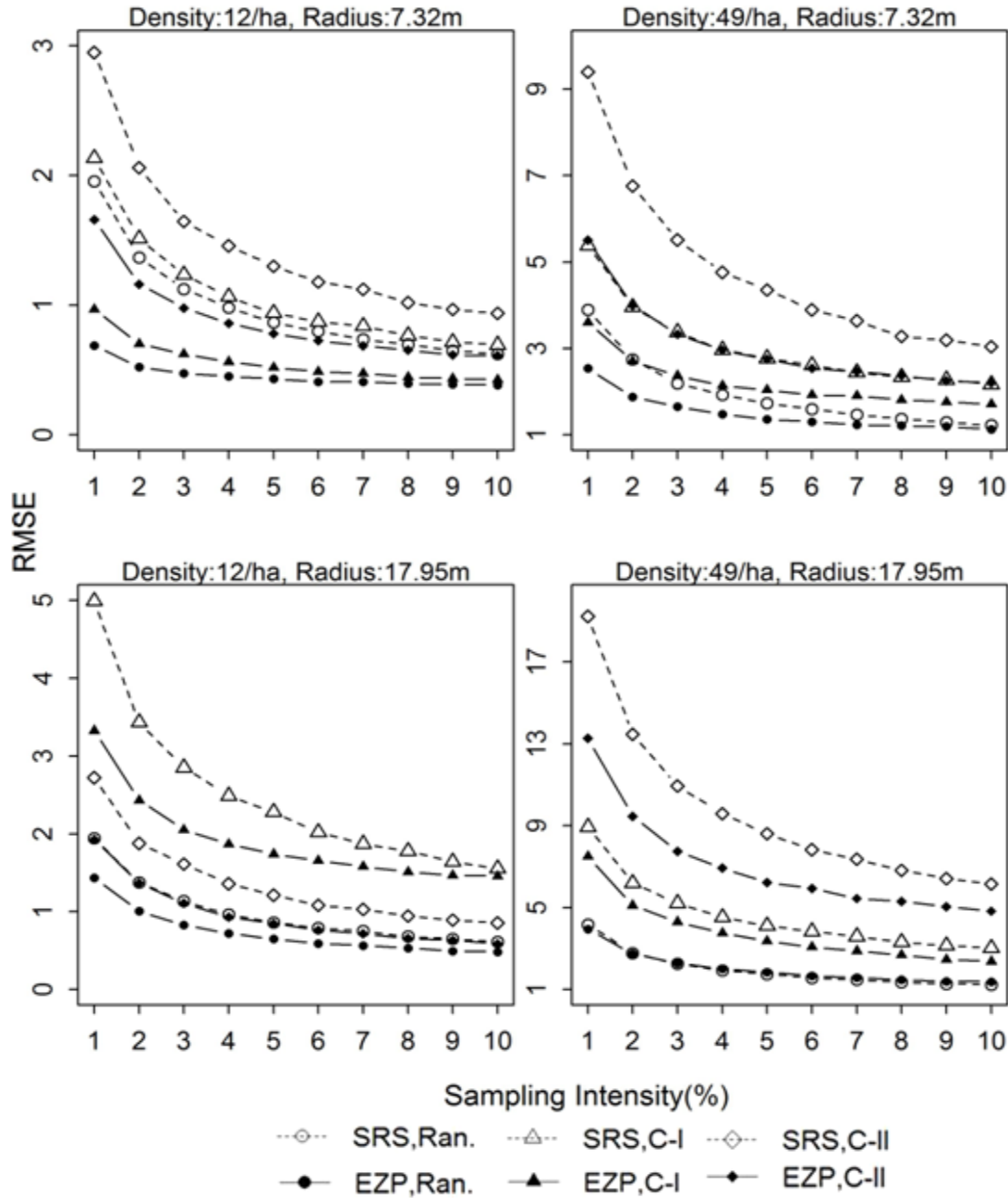


Figure 4.8. RMSE by spatial patterns, sampling intensities and plot sizes. SRS is simple random sampling method, EZP is Expected-zero Hurdle model with Poisson distribution, Ran. is random pattern, C-I is clustered pattern I, and C-II is clustered pattern II.

The EZ-Hurdle method has less variation in estimation error for all spatial patterns, densities, and plot sizes (Figs. 4.9, 4.10, 4.11, and 4.12). Variation in estimation error was decreased with increasing sampling intensity for both SRS and EZ-Hurdle method. Variation in estimation error also increased when standing dead trees were clustered for both methods. Under the same density of standing dead trees, the EZ-Hurdle method showed greater improvement in estimation error over SRS when a smaller plot size (7.32 m radius) was applied. The difference between the methods was also larger when standing dead tree density was lower (compare Figs. 4.9, 4.10 to 4.11 and 4.12). These results were expected because the EZ-Hurdle method should perform better when there are more zero observations in the data. Under either method, there was higher variation of estimation error when the density of standing dead trees was 49 per ha than compared to 12 per ha (Figs. 4.9, 4.10, 4.11, and 4.12), which was expected because larger variation is typically associated with a larger population mean (Avery and Burkhart 1983).

SRS is a design-unbiased estimator, while the EZ-Hurdle method is model-based estimator, which is expected to be biased (Gregoire 1998). The data show that errors for both methods were clustered around zero (Figs. 4.9, 4.10, 4.11, and 4.12), such that the bias in the EZ-Hurdle method is apparently low.

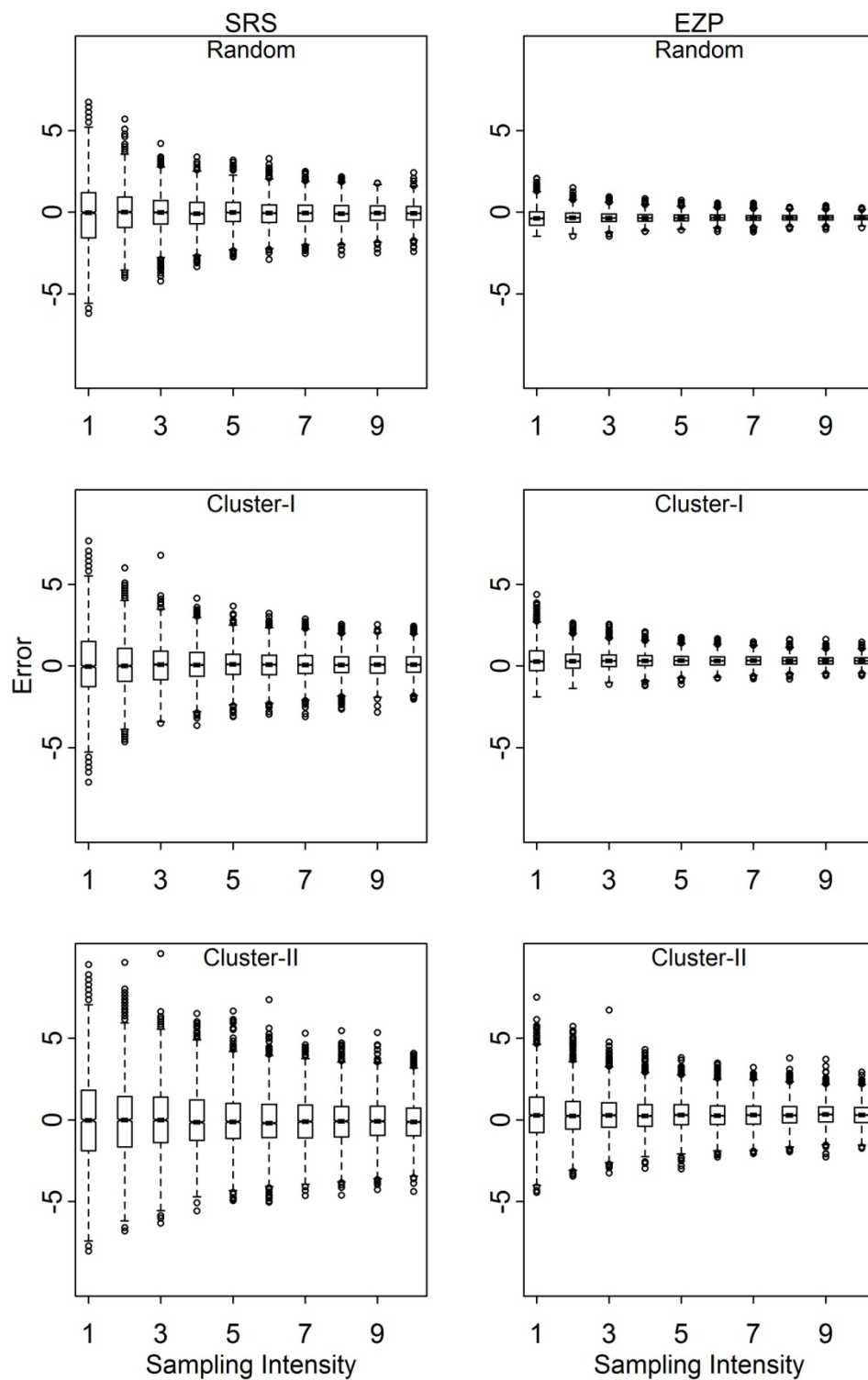


Figure 4. 9. The errors by sampling intensity when density is 12/ha and plot radius is 7.32 m.

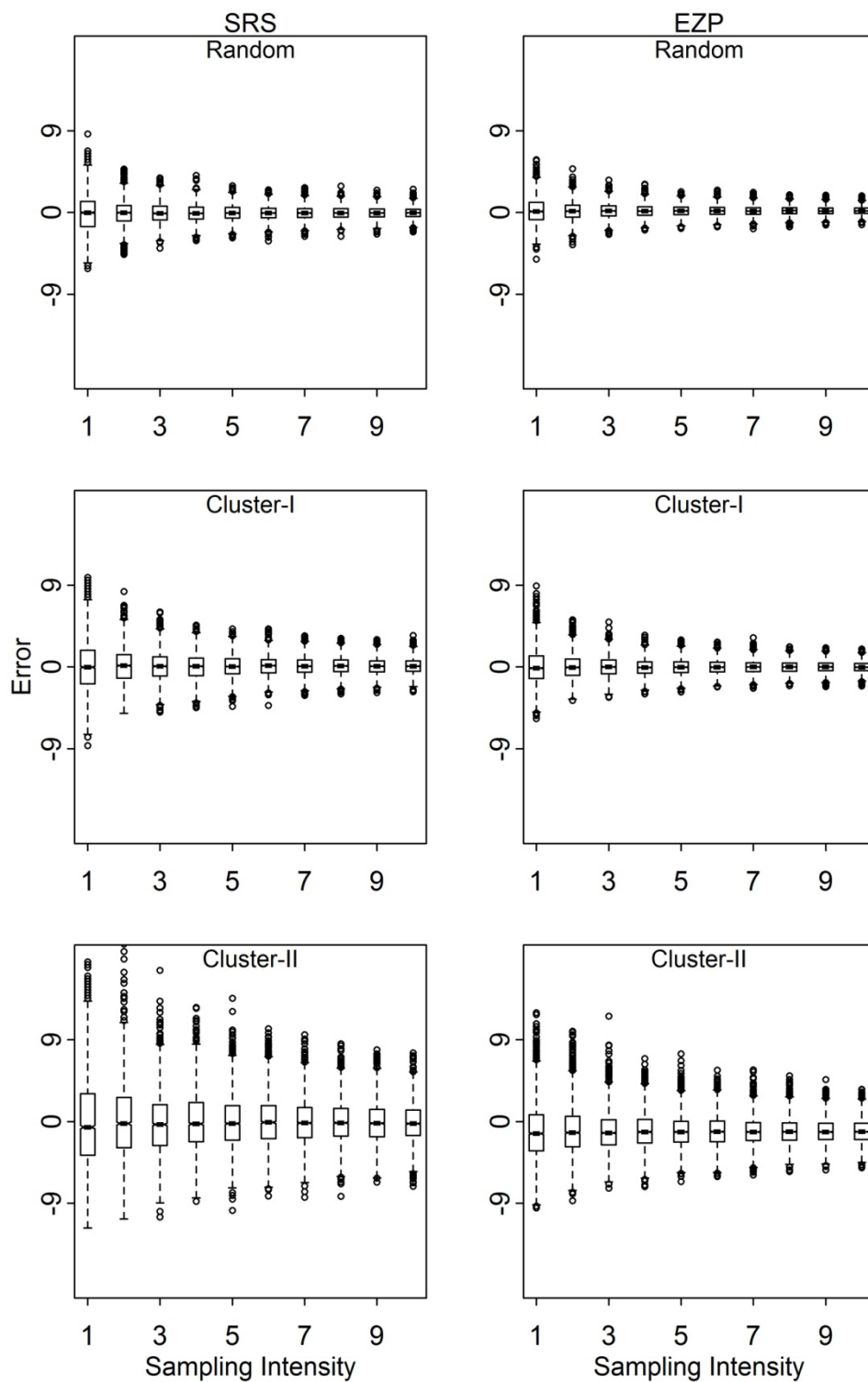


Figure 4. 10. The errors by sampling intensity when density is 12/ha and plot radius is 17.95 m.

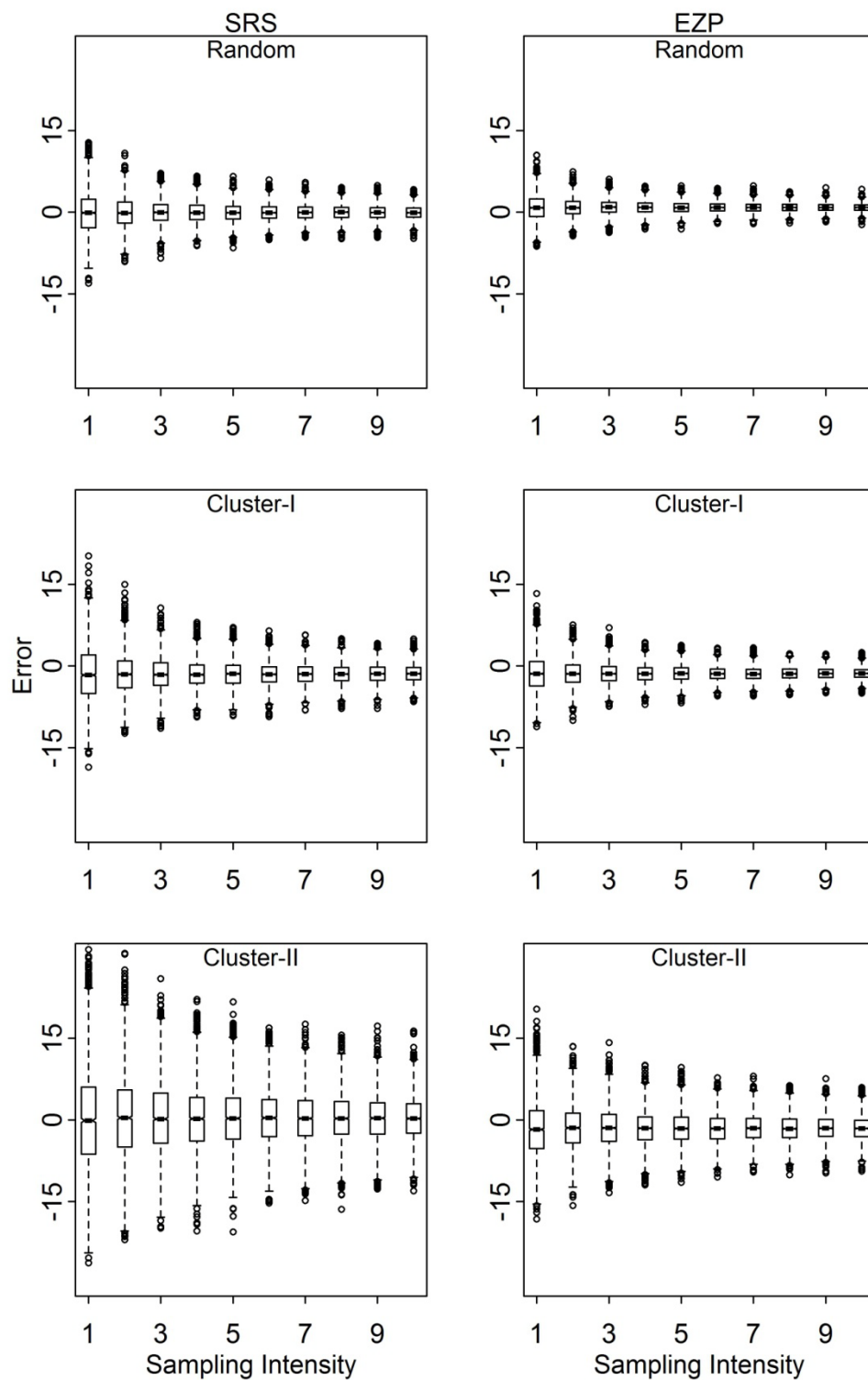


Figure 4. 11. The errors by sampling intensity when density is 49/ha and plot radius is 7.32 m.

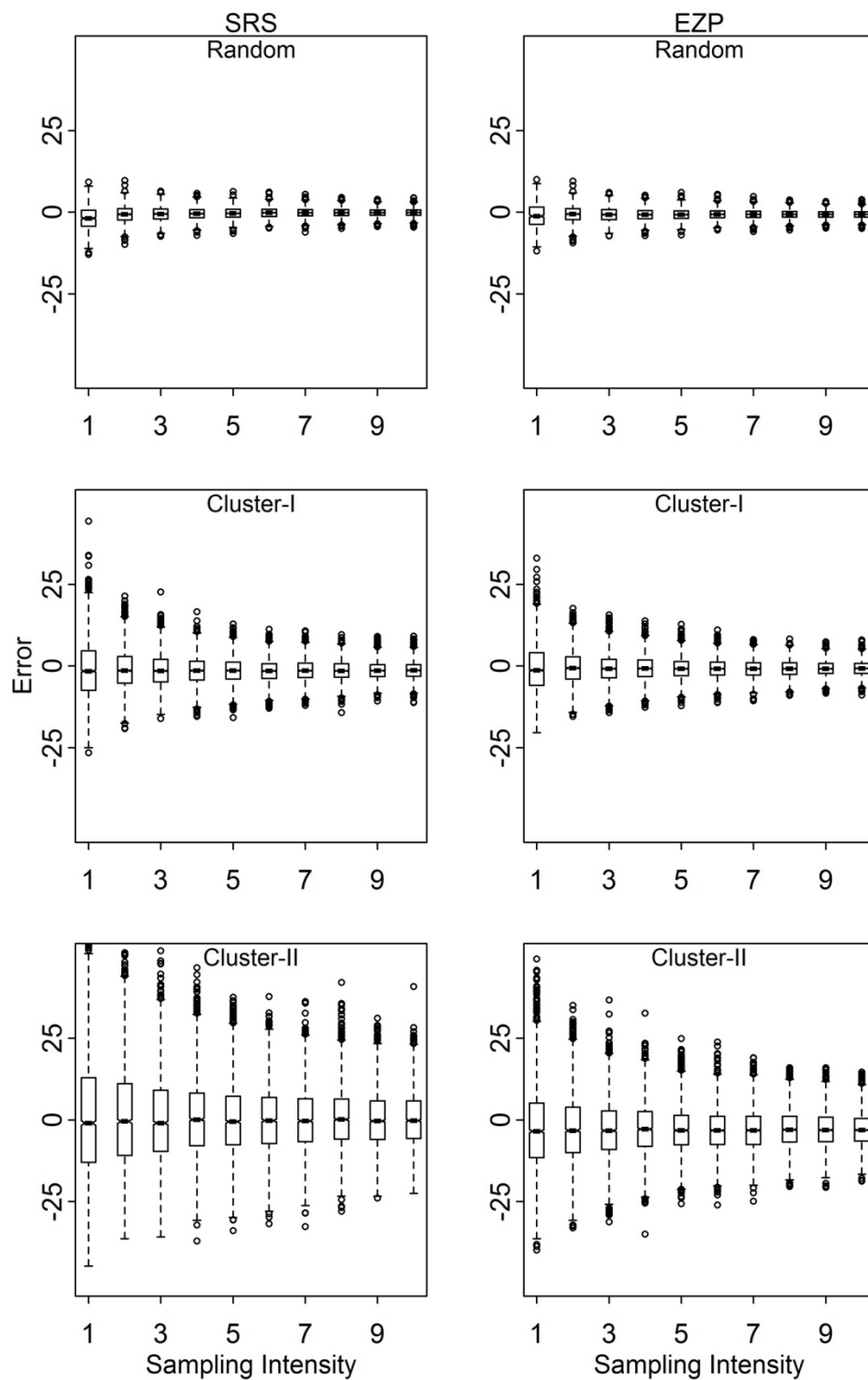


Figure 4. 12. The errors by sampling intensity when density is 49/ha and plot radius is 17.95 m.

Given that a relatively large number ($n = 500$) random sample points were used to estimate the inclusion probability of a standing dead tree under a given search radius, it is important to consider how sensitive the EZ-Hurdle method is to errors in specifying the expected zero probability, i.e., if lower numbers of random point to dead tree distances were employed to calibrate the EZ-Hurdle model. The results of the sensitivity analysis (Fig. 4.13) show that a 1% change of the estimated inclusion probability corresponds to less than a 0.1% change in error in actual standing dead tree density, indicating that the method is robust to relatively small errors in estimating the expected zero-probabilities.

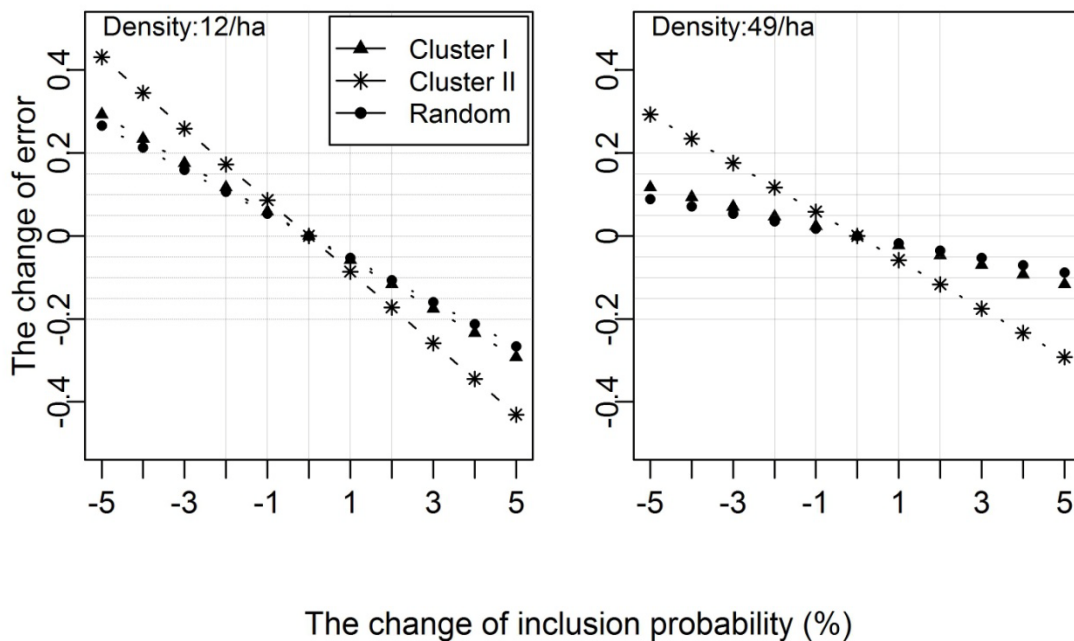


Figure 4.13. The changes of errors (%) by the changes of inclusion probability when sampling intensity is 5% and plot radius is 7.32m.

This is supported by the results in Figure 4.3 and 4.4, which show that even with sample sizes as low as 30 the expected difference between the estimated and true inclusion probability is only a few percentage points. The results (Fig. 4.10) also show that the percent error is more sensitive

to the change of inclusion probability when the density of standing dead trees is lower and when the standing dead trees are more clustered in space.

4.5. Discussion of Simulation Study

The EZ-Hurdle method employs the EZ-Hurdle model to clarify the variability associated with estimating the true proportion of zeros in count data, in this case counts of standing dead trees in fixed radius plots. The results clearly show that additional information regarding the true proportion of zeros reduces estimation error both in terms of increased precision and reduced average bias, i.e., improved accuracy, except under conditions where the data has a small number of zero observations, either because dead tree density is high or because plot sizes are large and trees are distributed randomly in space. On the other hand, the standard Hurdle model shows similar results to compare SRS for all combinations examined because it uses the observed zero proportion from the sample data when estimating the population parameter. This latter result was expected because previous studies estimating the abundance of snags (Eskelson et al. 2009), species abundance (Potts and Elith 2006), and modeling stand mortality functions (Affleck 2006), reported that the Hurdle model performs well for fitting the distribution of observed data, i.e., Hurdle models parameterized with the observed proportion of zeros in the data should have similar expected value as estimated by SRS.

The EZ-Hurdle method gave superior estimates in almost all cases examined when the sampling intensity was less than 5%, which is consistent with the recommendation by Bull *et al.*, (1990) that a 5% sampling intensity is necessary for reliable estimates of standing dead trees using standard (e.g., simple random) sampling methods. Recognizing that the EZ-Hurdle method relies on an underlying sampling design, it can allow for greater flexibility in choosing a sampling design; for example deciding the plot size to estimate the abundance of standing dead

trees. The EZ-Hurdle method can improve the precision when using smaller plot sizes than typically recommended for standing dead trees (e.g., 0.1 to 1 ha plots,. (Ganey 1999; Stephens 2004)). In effect, the EZ-Hurdle method helps to mitigate increased variation and possible bias associated with too many non-observations of standing dead tree within plot, which is associated with decreasing plot size (Kenning et al. 2005). Hence, the EZ-Hurdle method can allow for reductions in plot size or plot numbers when the auxiliary information, such distances to nearest standing dead trees, is available from previous studies or additional data collection.

The improved results shown here for estimating standing dead tree abundance are likely general, in that we believe that the EZ-Hurdle method should work well whenever zero observations in data are a major source of uncertainty in the data (i.e., zero-inflated data). For example, it may allow for increased accuracy in estimating the abundance of rare species by enhancing the Hurdle model-based methods of Potts and Elith, (2006). However, it is important to emphasize that the EZ-Hurdle method requires that the expected zero-probability be estimable through some auxiliary data source, which in this case was as simple as measuring the distance to the nearest standing dead tree. For other situations, estimating the expected zero probability may not be so simple. According to a study by Kenning et al. (2005), the time requirement to measure the nearest dead tree was increased with a decreasing density of dead trees and the latter is the best application of the EZ-Hurdle model. Therefore, further work is needed beyond this simulation to examine the cost efficiency of applying the EZ-Hurdle method, when the auxiliary information used to estimate the expected zero-probability is calculated from locally collected distances to the nearest dead tree (this is dealt with in later chapters of this dissertation).

It might be possible to estimate the inclusion probability of a standing dead tree from stand level attributes or models calibrated elsewhere, e.g., in similar stands (see e.g., Eskelson et

al. 2009). There were some clear trends in the data that the inclusion probability of a standing dead tree under a given plot design was different by both the density of standing dead trees and their spatial pattern. In fact, many studies suggest that the abundance and spatial pattern of standing dead trees varies by geographical location, stand age, forest type and management regime (Cline et al. 1980, Guby and Dobbertin 1996, Green and Peterken 1997, Fridman and Walheim 2000). Thus, it might be more cost-effective to predict the expected zero probability under a given sampling design and set of forest conditions from simple auxiliary variables, rather than collecting additional data such as nearest dead tree distances on site. The biggest concern would likely be that estimated population parameters from the EZ-Hurdle method would be more likely to be biased with the additional auxiliary data than with the observed data, and thus not representative of local conditions (Dorazio 1999). Further studies along these lines are needed. Nonetheless, studies of the relationship between expected zero-probabilities for dead trees and forest attributes have an intrinsic value, in that they might serve to establish expected baseline mortality rates for comparing observed rates, which could provide benchmarks for assessing and monitoring the health of forest ecosystems.

5. An Application of the EZ-Hurdle Method for the Estimation of Standing Dead Tree Abundance in a Real Forest Setting

5.1. Introduction

Beech bark disease is a disease of American beech (*Fagus grandifolia* Ehrhart) which is caused by an interaction of the exotic sap-feeding beech scale (*Cryptococcus fagi* Baer) with at least three species of *Nectria* fungi (*Nectria galligena*, *N. coccinea* var. *faginata*, and *N. ochroleuca*) (McCullough et al. 2001). Beech bark disease has several effects on beech trees: reducing leaf size, discoloring foliage, causing stem and crown dieback, reducing tree growth, reducing masting, and eventually causing tree mortality (McCullough et al. 2001). In 2000, beech bark disease was reported in Michigan. More than 7.5 million beech trees could die after the introduction of beech bark disease (Petrillo et al. 2004). Therefore, the beech bark disease monitoring system (BBDMS) was developed to monitor the beech bark disease. From 2001 to 2003, 202 monitoring plots were established in Michigan. This monitoring system has been used to find the temporal and spatial changes, and determine the impacts that the beech disease has on beech trees and northern hardwood forest ecosystem.

In this study, a new estimator, the EZ-Hurdle method was used to estimate the density of standing dead trees in the BBDMS. A conventional simple random sampling method with fixed-area plots and was compared to the EZ-Hurdle method to estimate the abundance of standing dead trees. In addition, the inclusion probabilities of a standing dead tree on fixed radius plots were examined by disease occurrence.

5.2. Methods

5.2.1. Data Collection

Based on the previous data from the BBDMS, 20 forests were selected which have similar stand characteristics, such as basal area and location, in the Lower Peninsula of Michigan (Fig. 5.1); 10 infected forests and 10 non-infected forests were selected.

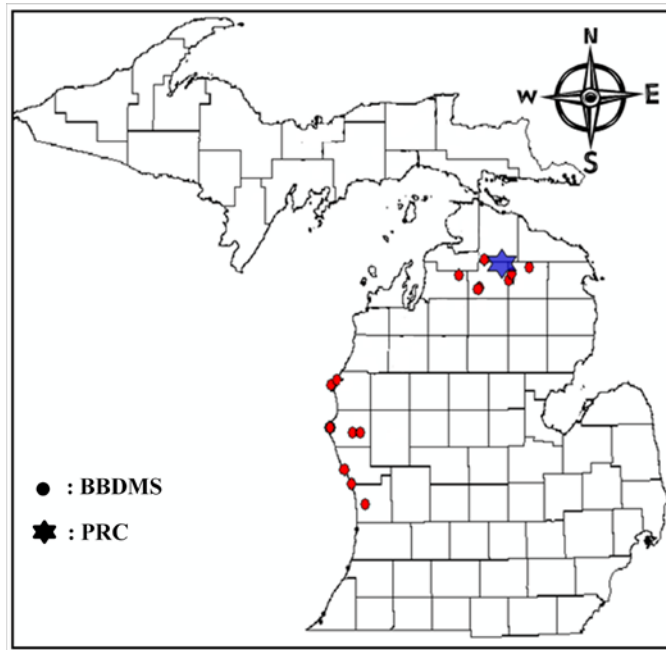


Figure 5.1. The selected locations among beech bark monitoring plot (BBDMS) and Pigeon River Country Forest (PRC).

The BBDMS used two transect matrix layouts 5×6 or 10×3 (Fig. 5.2) employing systematic sampling with fixed-area plots (abbreviated as fixed-area sampling, FAS). Where possible, the matrix was positioned as parallel to the nearest road. In this study, new sample plots were superimposed over the old BBDMS grid, consisting of 30 plot centers spaced 40 m apart. The first plot center was always at least 40 m from a road or another forest type to remove

the edge effects. The areas of forest were chosen to be at least 6 ha in order to insure that all 30 plot centers were located within same forest type. Two circular plots were established for each plot center. FIA-sized subplots (7.32 m radius = 0.017 ha) and annular plots (17.95 m radius = 0.100 ha) were employed to make the study relevant to the FIA program (USDA Forest Service 2005) and Forest Health Monitoring (FHM) programs, which mostly use subplots to collect data on standing dead trees, except in the Pacific Northwest states such as Oregon and California, where the annual plots are sometimes used to collect additional standing dead tree data for dead trees whose DBH is greater than 54 cm (24 in).

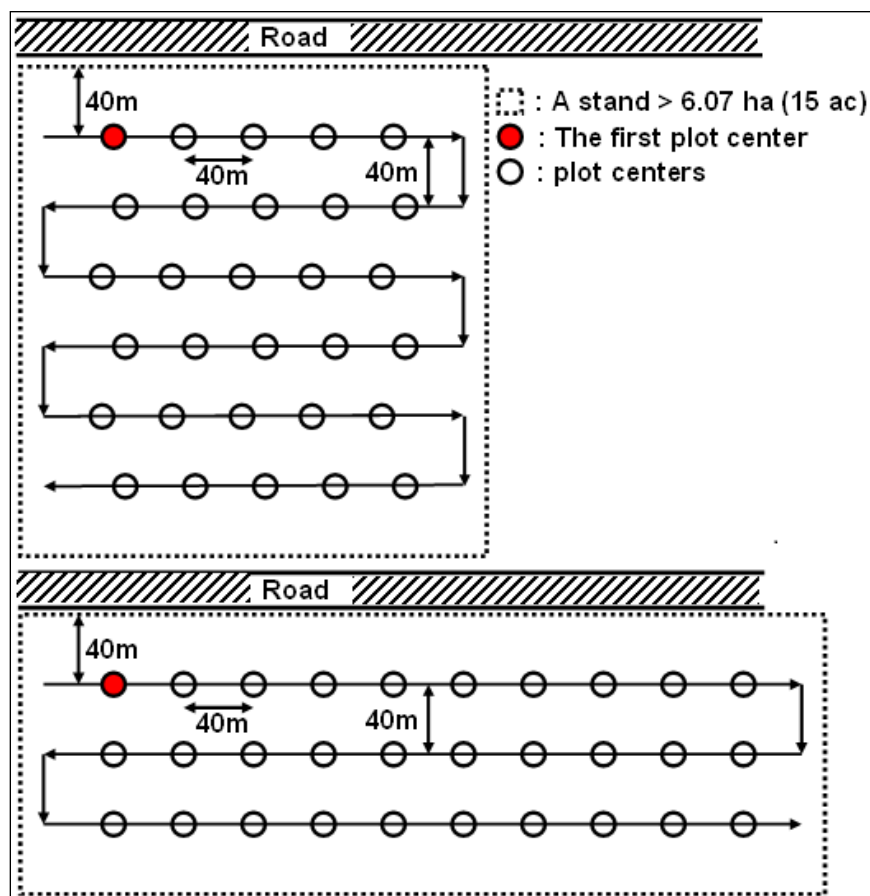


Figure 5.2. Plot design with 30 sampling points in the transect layout.

All data collection was conducted to match some of the basic protocols of the FIA program. All live and standing dead trees with a DBH of ≥ 12.7 cm (5 in) were tallied at each subplot. Since the definition of dead material in forests differs by researcher (Voller and Harrison 1998), the definition of a standing dead tree used by the FIA program was applied to define standing dead trees in the field studies. Thus a standing dead tree had to be taller than 1.3 m, have a DBH of ≥ 12.7 cm and a lean of less than 45 degrees from vertical; otherwise it is classified as coarse woody debris (CWD) (Fig. 5.3, from USDA Forest Service 2005).

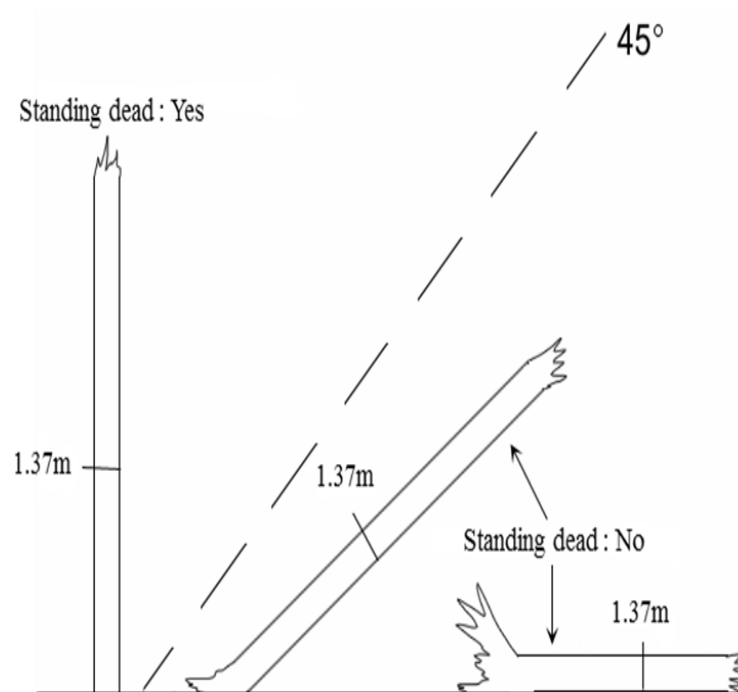


Figure 5.3. A standing dead tree is defined as a dead tree greater than 12.7 cm DBH, which is taller than 1.37 m and leans less than 45 degrees from vertical (USDA Forest Service 2005).

For both live and standing dead trees, characteristics recorded were species, DBH, and distance and azimuth from plot center. In case of standing dead trees, decay class was also recorded

according to the five-class system following the FIA program (Table 5.1, from USDA Forest Service 2005); Distance was measured with an electronic distance measuring device (Vertex III, Haglof, inc.) or a measuring tape, when the electronic distance measurer did not work because the distance was too far to measure. Unlike the FIA protocol, the distance to the nearest standing dead tree was also measured when the standing dead trees were outside of the plot boundaries. The time to completely measure the subplot and annular plot was recorded.

Table 5.1. Guide to define the decay class by conditions (USDA Forest Service 2005).

Decay Class	Limbs and branches	Top	Remaining bark (%)	Sapwood condition	Heartwood condition
1	All present	Pointed	100	Intact; sound, incipient decay, hard, original color	Sound, hard, original color
2	Few limbs, No fine branches	May be broken	Variable	Sloughing; advanced decay, fibrous, firm to soft, light brown	Sound at base, incipient decay in outer edge of upper bole, hard, light to reddish brown
3	Limb stubs only	Broken	Variable	Sloughing; fibrous, soft, light to reddish brown	Incipient decay at base, advanced decay throughout upper bole, fibrous, hard to firm, reddish brown
4	Few or no stubs	Broken	Variable	Sloughing; cubical, soft, reddish to dark brown	Advanced decay at base, sloughing from upper bole, fibrous to cubical, soft, dark reddish brown
5	None	Broken	Less than 20	Gone	Sloughing, cubical, soft, dark brown, OR fibrous, very soft, dark reddish brown, encased in hardened shell

5.2.2. Statistical Analysis

The estimates of the density of standing dead trees per ha was calculated by using an estimator for fixed-area sampling method (FAS) and the EZ-Hurdle method. According to the simulation study, there was no significant difference of parameter estimates between EZ-Hurdle with Poisson vs. negative binomial distributions. Thus, EZ-Hurdle method with a Poisson distribution (EZP) was used to estimate the density of standing dead trees for the field study. The spatial pattern of standing dead trees in the field study was unknown, so several test statistics were explored to test for spatial pattern. To evaluate whether the standing dead trees were aggregated in infected and non-infected forests, a variance-to-mean ratio was calculated using fixed-area plot data as follows:

$$v = \frac{s_x^2}{\bar{x}}$$

where s_x^2 is variance and $\bar{x}$ is mean.

The variance-to-mean ratio under an assumed Poisson distribution has been used to measure spatial pattern, however, it can be a poor indicator of spatial pattern when sampling intensity is too low, or plot sizes are not large enough (Young and Young 1998). An alternative metric, the Index of Dispersion (ID), which directly takes into account the sample size, was calculated as follows:

$$ID = \frac{(n-1) \times s_x^2}{\bar{x}}$$

where n is number of plot. If the variance equals the mean, ID has an approximately χ^2 distribution with $(n-1)$ degrees of freedom (Fisher 1922).

In order to compare the FAS and EZP from the field data, a simple bootstrap re-sampling

procedure was used (Efron and Tibshirani 1993) to estimate the variance of estimated density of standing dead trees for both methods. For the FAS method, plot centers were re-sampled from all collected data using subplot by different sampling intensity. Several different sampling scenarios were tested for the EZP method, consisting of different point-to plot-ratios (PPR), calculated as:

$$PPR = \frac{\text{No. random points}}{\text{No. fixed area plots}}$$

Recall that, in the field study, the distance to the nearest standing dead tree was measured from a point at the center of each fixed area plot; this corresponds to a sampling scenario of $PPR = 1$ (Table 5.2). At $PPR > 1$ additional random points are added outside the plots. 1,000 repetitions were applied to estimate the density of standing dead trees for each scenario (Table 5.2).

Table 5.2. The number of fixed-area plots and random points by sampling intensity and method.

Type SI (%)	Method					
	FAS	EZP				
	No. plot	PPR				
		1.00	1.25	1.50	1.75	2.00
1	36	36/36	45/36	54/36	63/36	72/36
2	72	72/72	90/72	108/72	126/72	144/72
3	108	108/108	135/108	162/108	189/108	216/108
4	144	144/144	180/144	216/144	252/144	288/144
5	180	180/180	225/180	270/180		

SI is sampling intensity, FAS is fixed-area sampling method, EZP is EZ-Hurdle method with Poisson distribution, No. plot is the number of fixed-area plot used FAS, and PPR is point to plot ratio which is the number of random points / the number of fixed-area plots.

Since we did not know the true density of standing dead trees per ha in these forests, the RMSE could not be calculated, so instead the coefficient of variation (CV) of standing dead tree density from FAS vs. EZP was calculated using 1,000 repetitions from bootstrapping with each different sampling scenarios. The CV is expressed as a percentage value and allows for

comparison of the relative variability about mean (Avery and Burkhardt 1983). The CV is calculated by

$$CV = \frac{S_x}{\bar{x}} \times 100$$

where S_x is a standard deviation.

5.3. Results

Table 5.3 summarizes the stand statistics obtained by traditional FAS. Both beech bark disease-infected forests and non-infected forests had similar density of live trees, but the infected forests had more than twice the density of standing dead trees than non-infected areas. Expressed as basal area for standing dead trees, infected forests had a significantly greater basal area (BA, m²/ha); about 3 times greater than non-infected stands. Results were similar whether larger (annular) or smaller (sub) plots were employed (Table 5.3, note: live trees were not measured in the annular plots).

Table 5.3. Summary of statistics from fixed-area sampling method by infestation status.

Plot	Infected	n	No. live/ha	No. SDT/ha	BA (live)	BA(SDT)
Subplot	No	299	452.0±15.3 ^A	19.5±4.4 ^A	26.9±6.5 ^A	1.0±0.6 ^A
	Yes	299	434.7±6.3 ^A	52.3±7.7 ^B	32.4±3.9 ^B	3.0±1.8 ^B
Annular	No	299	-	23.3±2.49 ^A	-	1.2±0.7 ^A
	Yes	299	-	50.8±3.87 ^B	-	3.2±1.6 ^B

Estimates are shown as mean±95% CI, ^A and ^B are significantly different ($\alpha = 0.05$), and Infected is the infection status of beech bark disease.

The decay class was classified by five-class system. The decay class was compared by infestation status and also compared between all standing dead trees and American beech trees.

Figure 5.4 shows the distribution of number of standing dead trees per ha by decay class. More than 50% of standing dead trees was classified as decay class 2 and 3 for both infected and non-infected forests when all species are included. However, more than 50% American beech trees was classified into decay class 1 and 2 in case of infected areas, indicating that a lot of the standing dead American beech trees died recently (between 2003 and 2008) from beech bark disease.

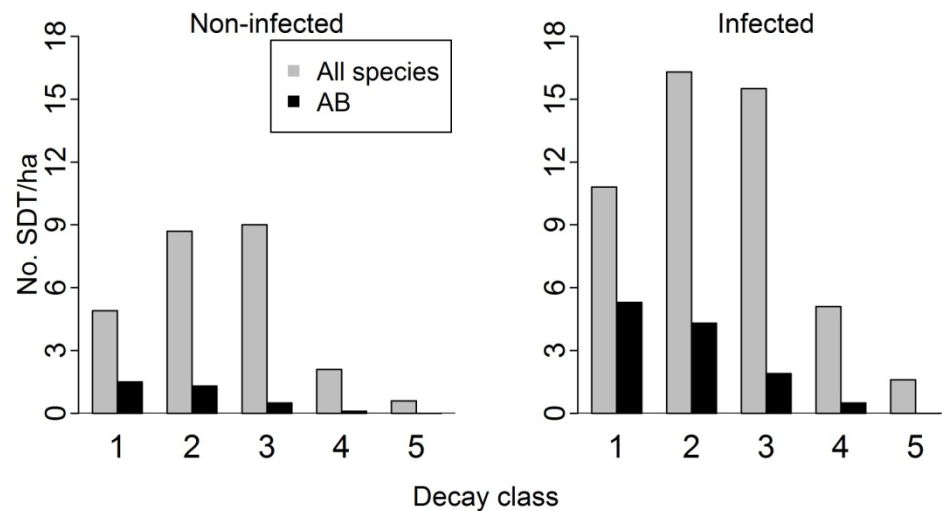


Figure 5.4. The number of standing dead trees (No. SDT/ha) in five decay classes in two forest types. AB is American beech.

Although the density of standing dead trees was significantly different, the standing dead trees for both infected and non-infected forests show clustered pattern (Table 5.4). The variance-to-ratio values are greater than one and index of dispersion is also significantly different from one at 95% probability level for both infected and non-infected forests. In other words, the spatial pattern of the standing dead trees showed aggregation for both infected and non-infected areas.

Table 5.4. The results of variance-to-mean ratio and index of dispersion.

Plot	Infected	n	Variance ratio	ID
Subplot	No	299	77.84	597.60 [*]
	Yes	299	87.96	387.71 [*]
Annular	No	299	20.74	276.22 [*]
	Yes	299	23.01	201.32 [*]

Infected is infected status, ID is index of dispersion and ^{*} is significantly different from one by index of dispersion.

Figure 5.5 shows the inclusion probability a standing dead tree estimated by reduced Gompertz function (Table 5.5) by infected status. The inclusion probability was significantly different by infected status because the 95% confidence interval (CI) of parameters do not overlap (Rubin and MacFarlane 2008). As expected, there is higher inclusion probability in infected forests than non-infected forests because the density of standing dead trees is almost twice as high in infected forests. Note that the plot size is a major determinant of zero-inflation in the data (Fig. 5.5).

Table 5.5. The estimated parameters and standard errors () of reduced Gompertz function.

Infected Status	Parameters	
	b	c
Yes	4.11(0.068)	0.78(0.002)
No	4.85(0.122)	0.85(0.002)

Reduced Gompertz function is $P_{Inc}(id) = \exp(-b \times c^d)$ where b and c are parameters and d is search radius.

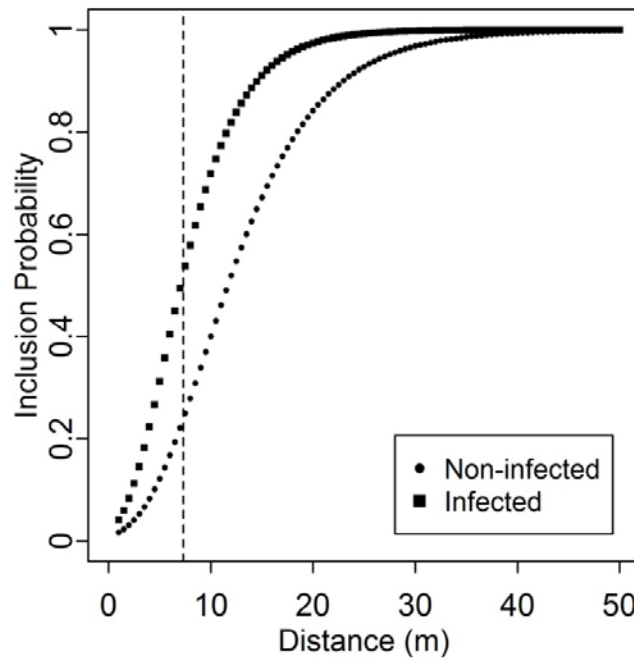


Figure 5.5. Estimated inclusion probability by infected status. Dashed line is 7.32 m.

Figure 5.6 shows the coefficient of variation for estimated density of standing dead trees per ha by method and PPR. The results clearly show that the EZ-Hurdle method improved the precision of estimating standing dead tree abundance over FAS, as indicated by the CV. The CV of standing dead tree abundance decreased with increasing the number of fixed-area plots or increasing the relative number of random points (i.e., PPR) used for the EZP method. The CV was lower under all methods in the infected plots, because standing dead tree density was higher and so zero inflation was lower (Fig. 5.6). One interesting result is that the EZ-Hurdle method showed smaller coefficient of variation when only addition information was applied to estimate the expected zero probability without adding additional random points. The improvement of coefficient of variation was much larger when $PPR > 1$ (compare e.g., difference between $PPR = 1$ and $PPR = 1.25$, Fig. 5.6) because additional sample locations are added to extend the information content of the auxiliary data.

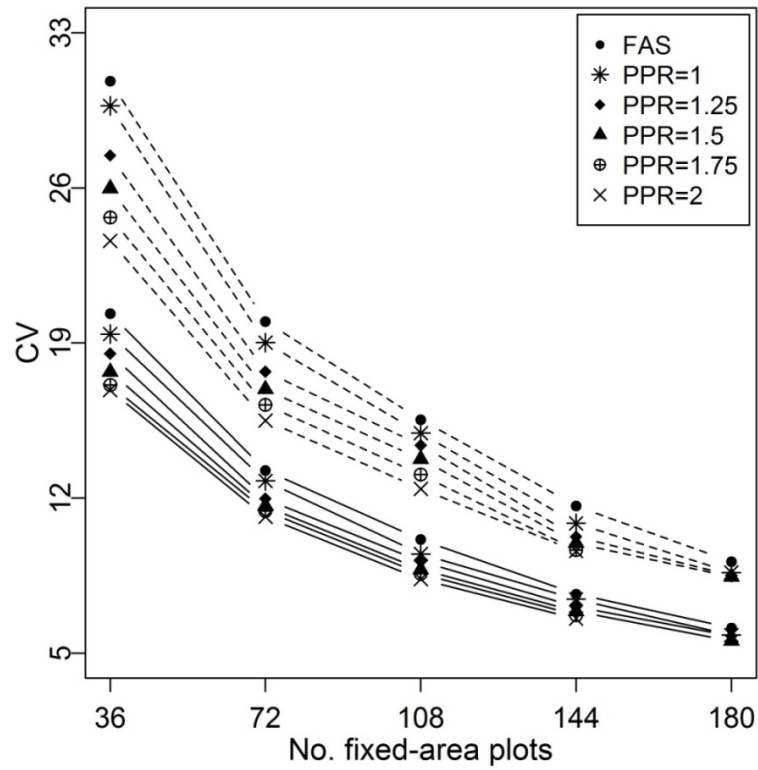


Figure 5.6. The change of coefficient of variations by the number of fixed-area plots. FAS is fixed-area sampling method, PPR is point to plot ratio of data for EZ-Hurdle method. Solid line is infected forests and dashed line is non-infected forests.

5.4. Discussion for Applied Study

The results of the field study reinforce those of the simulation study, demonstrating that the EZ-Hurdle method (EZP) increased the precision of estimates of the density of standing dead trees relative to the FAS method for both infected and non-infected forests. The auxiliary information, which is distance to the nearest standing dead tree, reduces the variation caused by zero observation, which is largest when dead tree densities are lowest (Fig. 5.5). When additional points to collect the auxiliary information are added, EZP method shows greater improvements of precision. In addition, EZ-Hurdle method has better precision than FAS when

only additional information is added without adding points. This property can be an advantage to improve the precision without changing plot design or adding more plots. Although EZ-Hurdle method gave better estimates than FAS, it also required more data, which imposes additional costs. Thus, further examination was needed to confirm that EZ-Hurdle method can be more cost-efficient than FAS. Cost efficiencies for the EZ-Hurdle method and strategies for applying it to FIA plots are discussed in the chapters which follow.

The spatial pattern and density of dead trees were also found different by species and site characteristics (Ganey 1999) and also different due to the agent of tree death (Franklin et al. 1987). In this study, beech bark disease is the major source of beech tree mortality. Both the variance-to-mean ratio and index of dispersion (ID) showed that standing dead trees were clustered in both infected and non-infected forests. Additionally, the shape of inclusion probability function can be an indicator of spatial pattern and density of standing dead trees because the inclusion probability increases rapidly within short distance because generally there are more points within short distances when points are clustered in area (Bailey and Gatrell 1995).

6. Cost and Sampling Efficiency of EZ-Hurdle Method

6.1. Overview

A well designed sampling strategy usually provides a more efficient alternative to a census, but since it leaves significant portions of the population out, the choice of sampling strategy is critical to obtaining reliable parameters of the population of interest. Efficiency relates the amount of information collected to the resources expended (Gregoire and Valentine 2008) and is a critical factor for selection of the sampling strategy. Sampling efficiency is typically evaluated by two aspects: accuracy of estimated population parameters and cost effort to collect sample data.

According to previous studies, EZ-Hurdle method produces better estimates than FAS when additional information is collected, in this case, measuring the distance to the nearest standing dead tree from fixed-radius plot centers and from additional random points. If one can obtain such auxiliary information for improving the estimate at relatively low cost, the EZ-Hurdle method should be efficient as well as accurate.

Auxiliary information describing the expected zero probability can be obtained by two methods. One is from a previous study or prior information, which would be less costly than collecting new data, but this could cause biased estimates if the density and spatial pattern of standing dead trees at the locality of interest is different than that of the prior study. More straightforward, and likely more reliable, would be to derive the auxiliary data from additional field data collected at the site. The latter should be more costly.

In this study, the EZ-Hurdle method is compared side by side with FAS on both precision of estimation and cost to collect the data. The precision was evaluated by the coefficient of variation of standing dead tree density per hectare and cost was evaluated by the time to collect

the data. A priori, it is understood that the time requirement to collect the data for EZ-Hurdle method is always greater than time requirement for FAS, because EZ-Hurdle method enhances FAS plot data through use of auxiliary data. In this study, the time requirement for the EZ-Hurdle method was measured in the field and the efficiency of the method was examined by combining time cost information from field data with another simulation. In the next chapter, this data analysis is extended to determine the best sampling strategy to estimate the density of standing dead trees.

6.2. Data

The time requirement for fixed-area sampling method was modeled from BBDMS data used in the previous study and the time requirement for searching for the nearest standing dead tree was modeled using new field data which was collected from Pigeon River Country Forest (PRC) in Michigan (Fig. 6.1). The density of standing dead trees was estimated from both actual data (BBDMS) and simulation data.

Data for estimating the time requirement of fixed-area plot

As in the FIA design, the (larger) annular plots were superimposed over the subplots at each sample point, so the time to collect all the tree data on both sub- and annular plots in the BBDMS, with a two-person crew, was recorded sequentially as follows: the start time and end time for a subplot and the additional time to measure trees to the boundary of the annular plot. Therefore, time to finish the annular plot is the combination of the time to finish the subplot plus the additional time requirement to measure standing dead trees which are located in the area between subplot and annular plot boundaries. The time to measure the distance to the nearest dead tree within the plot areas was recorded, as this data is collected under the current FIA

design. However, when there was no standing dead tree within annular plot, the nearest standing dead tree was still measured, but the search time to find the nearest standing dead trees was not recorded.



Figure 6.1. The location of Pigeon River Country Forest (PRC) in MI.

Data for estimating the search time to the nearest standing dead tree

Because the time to search for the nearest standing dead tree was not recorded for each tree in the BBDMS study and because the BBDMS only focused on a limited range of beech-dominated northern hardwood stands, additional data was collected at Pigeon River Country State Forest (PRC) in Michigan (Fig. 6.1) to (1) better understand the time-cost-efficiency of the EZ-Hurdle method and (2) to explore the prospects for predicting the expected zero probability from basic stand data (forest type and density). According to previous studies, the density of standing dead trees differs by geographical location, species, and management regime (Cline et al. 1980; Fridman and Walheim 2000; Wisdom and Bate 2008), but is not correlated with stand age (Lee et al. 1997; McCarthy and Bailey 1994; Sturtevant et al. 1997; Vasiliauskas et al. 2004). Therefore, four different forest cover types were selected: Oak, Pine, Aspen, and Northern

hardwood forest based on species composition and each forest type was further subdivided into four stand density types according to basal area (m^2/ha) (Table 6.1) to represent a wide range of stand conditions for sampling.

Table 6.1. The classification of basal area (BA) class.

BA (m^2/ha)	BA class
$\text{BA} \leq 9.2$	1
$9.2 < \text{BA} \leq 16.1$	2
$16.1 < \text{BA} \leq 23.0$	3
$23.0 < \text{BA}$	4

Three replications of each forest cover type – stand density combination were installed to understand within condition variation. The forest stands surveyed were at least 7 ha and a 40 m buffer area was applied to avoid the edge effects, such as roads or other forest types, following protocols similar to that of the BBDMS design. Within each replication, 30 points were randomly selected. At each sample point, species, DBH, decay class, and distance were recorded for the nearest standing dead tree. The characteristics of live trees, DBH and species, were also collected, at five randomly selected points using fixed-area subplots. For each standing dead tree, the combined time to search for the nearest standing dead tree and measure its attributes was recorded with two-person crews.

6.3. Analytical Methods

Boot-strapping and Simulation

Similar to that discussed earlier, a boot strapping approach was used to re-sample the BBDMS data. Five sampling intensities of $n = 1,000$ of a 1% to 5% sampling intensity by 1%

interval (see, Table 5.2) were applied. A 5% sampling intensity was chosen as the upper end based upon the fact that the previous chapters show that the benefit of EZ-Hurdle is low when sampling intensity with FAS is greater than this. Five PPR strategies were applied to collect the distance information (see Table 5.2). For each scenario, 1,000 repetitions were applied.

Another simulation study was run with experimental treatments including two spatial patterns, random and cluster pattern I, three different densities, such as 12, 24, and 49/ha, applied for each spatial pattern in order to compare field study. The same number of plots and points used for bootstrapping study (Table 5.2) were applied.

In both the bootstrapping and simulation study precision was compared using the coefficient of variation for the estimated density of standing dead tree per ha.

Estimation of time costs from field data for bootstrapping and simulation

Time costs to complete both fixed-area sampling and the EZ-Hurdle method varied from place to place depending on several factors. The measured BBDMS and PRC time data were used to develop models which could represent time variability in the simulation / re-sampling environments.

In fixed-area sampling method, time requirement is strongly influenced by the number of tallied standing dead trees (Kenning et al. 2005) and plot size (area). Hence, in this study, the time requirement for FAS is proportional to the number of tallied trees times the plot size:

$$T_{FAS} \propto n_t \cdot plot$$

where T_{FAS} is time requirement for FAS and n_t is the number of tallied trees and $plot$ is the area of the plot. To estimate time requirement for fixed-area plot with a certain plot area i was

modeled as:

$$t_i = \beta_0 + \beta_1 \cdot n_t + \varepsilon$$

where t_i is the time required at a fixed-area plot of a specified size, n_t is the number of standing dead trees tallied, ε is an error term, and β_0 and β_1 are fitted coefficients. Because of the way the time data were collected, the time to count and measure all of the dead trees in a subplot was not directly available (live trees were also measured, see previous chapter for more details); only the time to count and measure the dead trees in the area between the boundaries of the sub- and annular plots was known. Therefore, in the regression model, the time requirement to complete FAS in a plot size of interest, was weighted by w calculated as follows:

$$w = \frac{\text{Area Plot}_i}{\text{Area between subplot and annular plot}}$$

where w is weight.

Finally, the time requirement for a subplot is

$$T_{FAS} = w \cdot t_i$$

Figure 6.2 shows the density distribution for different counts of dead trees within a 0.08 ha plot (the size of the area between annular plot and subplot). Approximately 50% of such fixed-area plots has less than 2 standing dead trees within a 0.08 ha area and the median time requirement to complete the plot was approximately 7.30 minutes (Fig. 6.2).

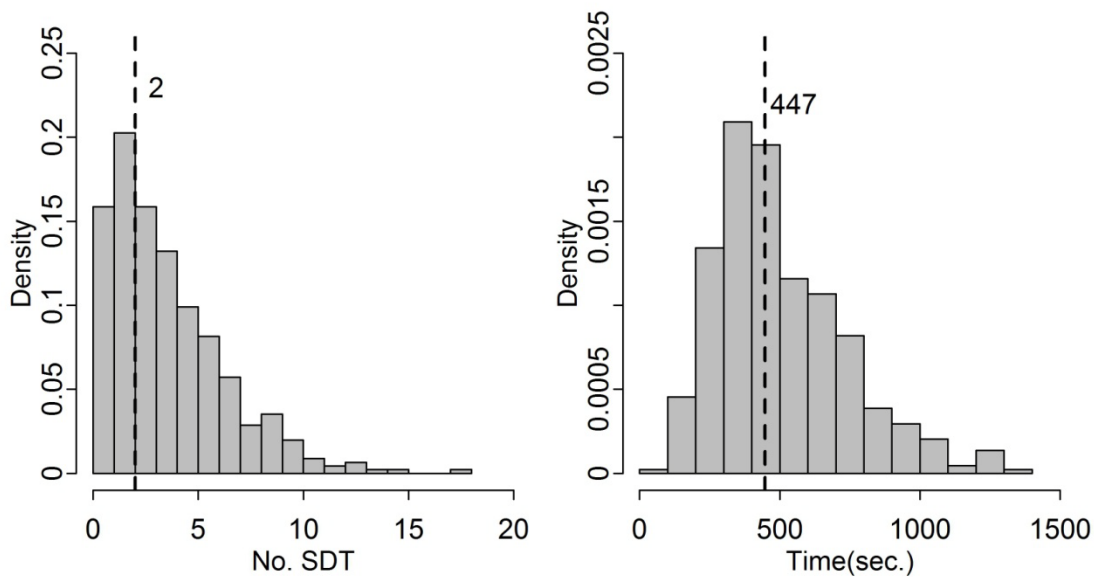


Figure 6.2. Distribution of the number of SDT per plot and time requirement to measure the SDT within plot (0.08 ha). Dashed lines are medians.

Table 6.2 shows the estimated parameters of the regression model fit to the BBDMS time data. The intercept is 303 seconds. Therefore, it was assumed that 303 seconds are required when there is no standing dead tree within 0.08 ha. This parameter determined the minimum search time for FAS for a plot of this size and was extrapolated to FIA plots using the weight (w) described above.

Table 6.2. The estimated parameters and R-square for regression model to estimate the time requirement for a 0.08 ha fixed area plot.

Response	Intercept(β_0)	Parameter (β_1)	R-square
Time (second)	303.4***	63.8***	0.56

The numebr of crews is two. Statistical significant level (***, $p < 0.001$).

Figure 6.3 shows the relationship between the time to complete a 0.08 ha FAS plot and

number of standing dead trees tallied, along with the associated residual errors. The results show a relatively unbiased model but with a considerable amount of scatter in the relationship, since only about 56% of time requirement was explained by the number of standing dead trees tallied (Table 6.3). The spread of residuals was roughly the same across the range of fitted values. Other factors affecting this relationship included the relative difficulty of navigating the terrain the density of the vegetation and the time spent identifying and classifying dead trees. Based on the lack of bias in the model, it was assumed that time requirement for plots of other sizes is roughly proportional to this relationship, i.e., the proportion of the number of standing dead trees times plot size ($T_{FAS} \propto n_t \cdot plot$) was proportional to the time requirement for other (sub- and annular) plot sizes. Thus, for the bootstrapping and simulation procedure the model (Table 6.2) and procedure above was used to estimate the time to complete FAS for every plot in every iteration.

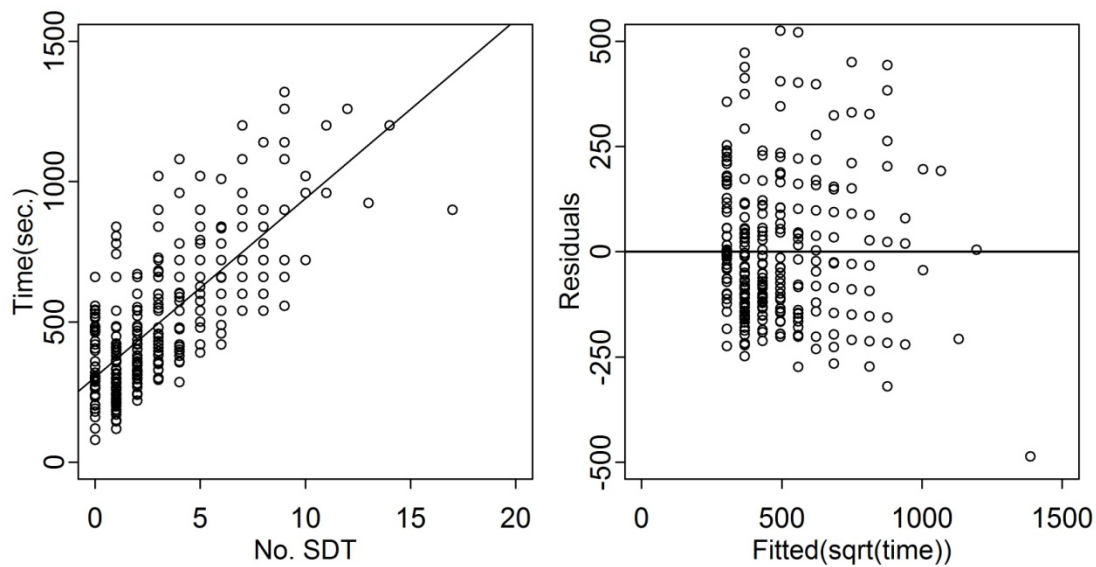


Figure 6.3. The scatter plot between time (second) and number of SDT within plot and residuals of regression model to estimate time requirement for fixed area (0.08 ha).

Estimation of search time for finding and measuring the nearest standing dead tree

For the bootstrapping and simulation procedure, it was also necessary to estimate the time to find and measure the nearest standing dead tree for every plot in every iteration. From the PRC data, the distance to the nearest standing dead trees has a strong influence on time requirements. Therefore, the time requirement to find and measure the nearest standing dead tree was modeled by

$$T_{neari} = \beta_0 + \beta_1 \cdot \pi(d_i)^2 + \varepsilon$$

where T_{neari} is the time required at plot or point i , d_i is the distance from the plot center (point) to the nearest standing dead tree, ε is an error term describing other factors effecting the time requirement, and β_0 and β_1 are coefficients which were obtained from fitting the model to the PRC data. Because DBH, species, and decay class were also recorded for the nearest standing dead tree, the time requirement estimated (T_{neari}) is an overestimate of the time to measure the distance to the nearest standing dead tree. However, since tree classification and measurement times were also included in the time estimates for the FAS method, both T_{neari} and T_{FAS} can be considered as standardized for the purposes of comparing the time cost of FAS to EZ-Hurdle method.

From the raw PRC data, we can see that the average search time to find and measure the nearest standing dead tree was dramatically increased due to the fact that some standing dead trees were located very far away from the random points (Fig. 6.4). Because the search area and time to measure the nearest standing dead tree were right skewed, several data transformations were applied to find the best fit model.

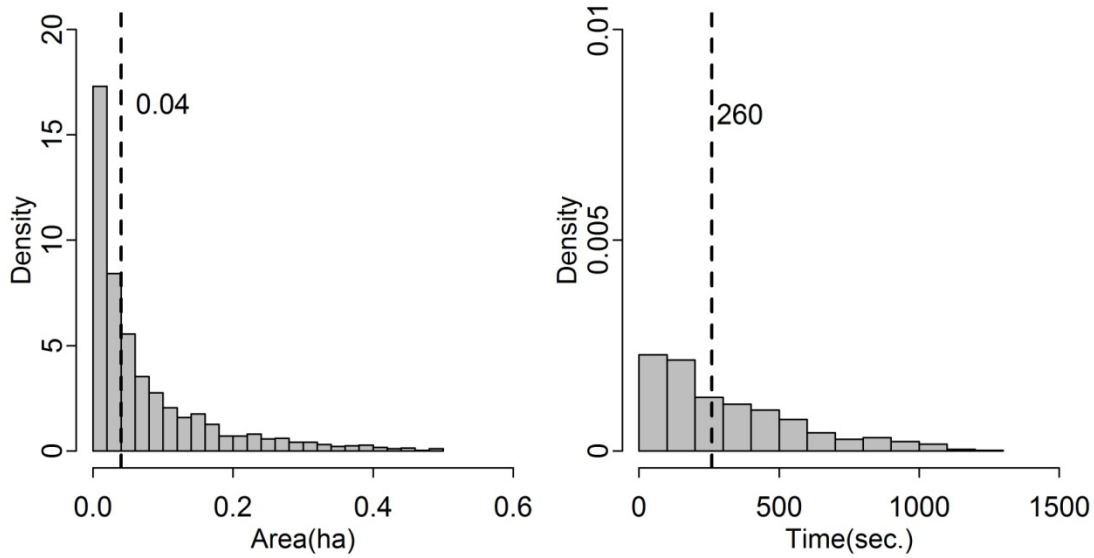


Figure 6.4. The distribution of search area (ha) and time (second) to measure the nearest SDT in PRC. Dashed lines are median values.

Based on the residual of sum of squares (RSS) and AIC, the best fit regression model (Fig. 6.5) was:

$$\sqrt{\text{Time}} = \beta_0 + \beta_1 \cdot \sqrt{\text{radius}} + \varepsilon$$

where time is in seconds, radius, is the radius of search area in meters, ε is an error term, and β_0 and β_1 are coefficients of the model fitted to the PRC data. There was strong linear relationship between search time and search radius when both variables were square root transformed. About 96% of time variation was explained by search radius alone and there was no major pattern in residual plot (Fig. 6.5) to indicate significant bias in the model, except a very small distortion for very small search areas / very short search times (the intercept of the model was -5 seconds, Table 6.3). The estimated parameters (Table 6.3) were used to estimate the search time for both the simulation and the boot-strapping study.

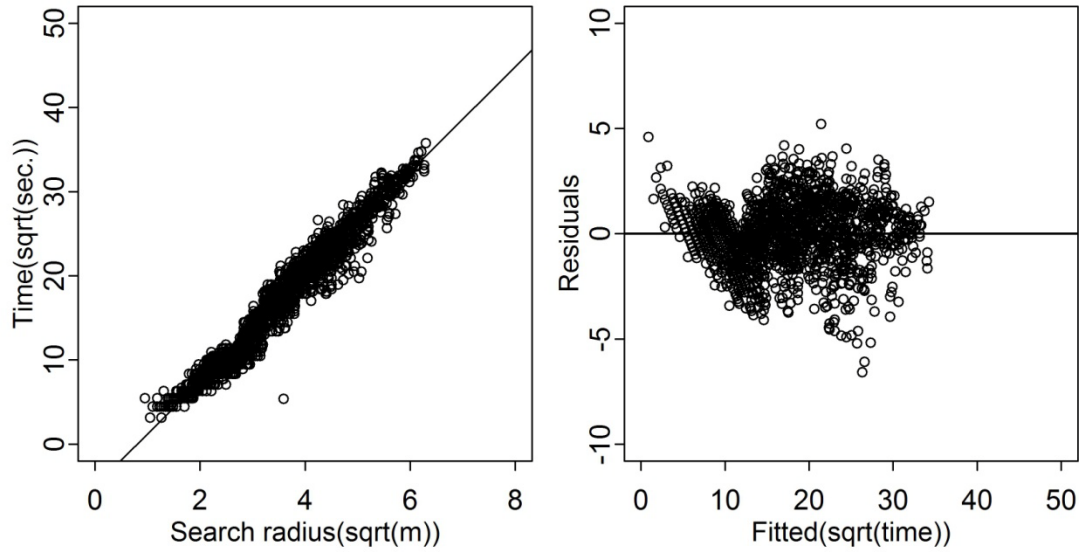


Figure 6.5. The scattered plot between search time (Sec.) and search radius (m) and residuals of selected regression model to estimate time to search the nearest SDT.

Table 6.3. The estimated parameters and R-square for regression model to estimate the search time to measure the nearest SDT.

Response	Intercept(β_0)	Parameter (β_1)	R-square
$\sqrt{\text{Time(sec.)}}$	-5.04***	6.24***	0.96

Statistical significant level (***, $p < 0.001$).

Estimation of time requirement for EZ-Hurdle method

Recall that the EZ-Hurdle method is applied to FAS data, where auxiliary data is collected or otherwise available. The time cost for the EZ-Hurdle method was computed as the sum of three time components:

$$T_{EZ} = T_{FAS} + (n - n_0) \cdot (T_{near} - T_{FAS,0}) + n^* \cdot T_{near}$$

where T_{EZ} is time requirement for EZ-Hurdle model, n is the number of fixed-area plots, n_0 is the number of fixed-area plots with a standing dead tree, T_{near} is time requirement to measure the nearest standing dead tree, $T_{FAS,0}$ is time requirement for FAS when there is no standing dead trees in plot, and n^* is the number of additional points to collect the distance information. T_{EZ} is equal to T_{FAS} when n^* and $(n-n_0)$ are 0.

Estimation of sampling efficiency

Sampling efficiency was evaluated by both the precision of parameters and the time cost to collect data (Lessard et al. 1994). The relative efficiency to compare FAS was calculated by:

$$E_r = \frac{CV_{EZ}^2}{CV_{FAS}^2} \times \frac{\bar{T}_{ez}}{\bar{T}_{FAS}}, \text{ where,}$$

$$\bar{T}_{EZ} = \sum_{i=1}^n T_{EZi} / n, \text{ and}$$

$$\bar{T}_{FAS} = \sum_{i=1}^n T_{FASi} / n$$

and where CV_{EZ} is the coefficient of variation of the parameter estimated by EZ-Hurdle method, CV_{FAS} is the coefficient of variation of the parameter estimated by FAS, i is the number of repetitions (1,000), T_{EZi} is the required time to collect data for repetition i for EZ-Hurdle, and T_{FASi} the required time to collect data for repetition i for FAS. When $E_r < 1$, the EZ-Hurdle method is more efficient than FAS.

6.4. Results

6.4.1. Comparison of Inclusion Probabilities Between Different Forest Type Conditions

Table 6.4 summarizes the results of the random point to nearest standing dead tree data collected from the PRC.

Table 6.4. Summary of results for BA for live trees, density of standing dead trees (No.SDT/ha), average distance (Dist.) from point to the nearest SDT, and average time (Time) to search for and measure the nearest standing dead tree with standard deviation in () after the value.

Type	BA Class	BA (m ² /ha)	No.SDT/ha	Dist. (m)	Time ^a (sec.)
Aspen	1	8.83 (0.34)	6.4 (2.9)	18.0 (8.0)	398 (242)
	2	10.52 (0.71)	13.7 (10.2)	12.4 (8.0)	250 (228)
	3	19.57 (2.05)	33.5 (11.9)	8.1 (5.4)	139 (126)
	4	27.73 (4.23)	34.0 (28.3)	8.0 (5.3)	143 (139)
NHW	1	9.69 (0.37)	8.4 (10.9)	14.8 (8.7)	384 (288)
	2	13.92 (1.63)	16.4 (8.5)	12.0 (7.0)	315 (243)
	3	20.42 (1.77)	18.2 (12.2)	11.5 (6.4)	304 (233)
	4	27.27 (2.88)	15.1 (10.9)	12.5 (6.4)	307 (224)
Oak	1	8.50 (0.83)	5.8 (1.9)	18.6 (8.5)	474 (280)
	2	14.75 (0.51)	6.0 (5.7)	18.4 (9.1)	480 (299)
	3	21.83 (0.39)	12.5 (16.9)	12.1 (8.0)	288 (253)
	4	28.29 (1.94)	44.5 (13.0)	7.2 (4.4)	144 (123)
Pine	1	8.96 (1.18)	6.3 (3.1)	17.7 (9.6)	454 (326)
	2	14.96 (0.42)	13.1 (11.6)	13.4 (7.9)	310 (256)
	3	20.26 (1.78)	7.4 (1.7)	16.2 (8.9)	402 (297)
	4	29.09 (2.67)	23.3 (21.8)	9.9 (6.1)	204 (187)

^a time per two-person crew.

Densities of standing dead trees generally increased with increasing basal area with some variation in the trend for NHW and pine forests, indicating an accumulation of dead trees as the

stands developed, consistent with the even-aged management that has been applied to the stands (even the NHWs). Some of the anomalies in this trend (Table 6.5) may be explained by the fact that most forests in the PRC are managed forest for timber production, sometimes treated with pre-commercial thinning, which often include removals of low vigor and dying trees.

The average distance between random points to the nearest standing dead tree tended to be decreased with increasing the density of standing dead trees (Table 6.4). The decline in mean time to search for and measure the nearest standing dead tree was similar for all forest types, except NHW forest which had very similar time costs, densities of dead trees and distances to dead trees in the higher three basal area classes, which (anecdotally) appeared to be because of relatively thicker understory vegetation in dense forests of this type that may have increased the search time. These data (Table 6.4) indicate that, in general, the inclusion probability of a standing dead tree may be estimable from stand level data.

Table 6.5. Estimated coefficients and standard errors () for reduced Gompertz function fitting by forest cover type and basal area class.

Forest type	BA class	Parameters		Forest type	BA class	Parameters	
		b(se)	c(se)			b(se)	c(se)
Aspen	1	6.88 (0.286)	0.88 (0.002)	Oak	1	7.06(0.195)	0.88(0.001)
	2	3.60 (0.088)	0.86 (0.002)		2	5.26(0.120)	0.89(0.001)
	3	3.61 (0.092)	0.78 (0.003)		3	4.26(0.129)	0.83(0.003)
	4	4.34 (0.083)	0.76 (0.002)		4	4.94(0.074)	0.73(0.002)
NHW	1	3.35 (0.104)	0.89 (0.002)	Pine	1	4.09(0.083)	0.90(0.001)
	2	4.58 (0.070)	0.84 (0.001)		2	4.67(0.110)	0.85(0.002)
	3	4.71 (0.102)	0.83 (0.002)		3	4.49(0.074)	0.88(0.001)
	4	4.32 (0.095)	0.84 (0.002)		4	4.52(0.082)	0.80(0.002)

BA is basal area, NHW is northern hardwood forest, and se is standard error.

When the reduced Gompertz function: $P_{Inc}(id) = \exp(-b \times c^d)$ was fitted to the PRC data, the inclusion probability was significantly different by forest type and basal area class (the 95%

confidence interval of estimated coefficients did not overlap, Table 6.5).

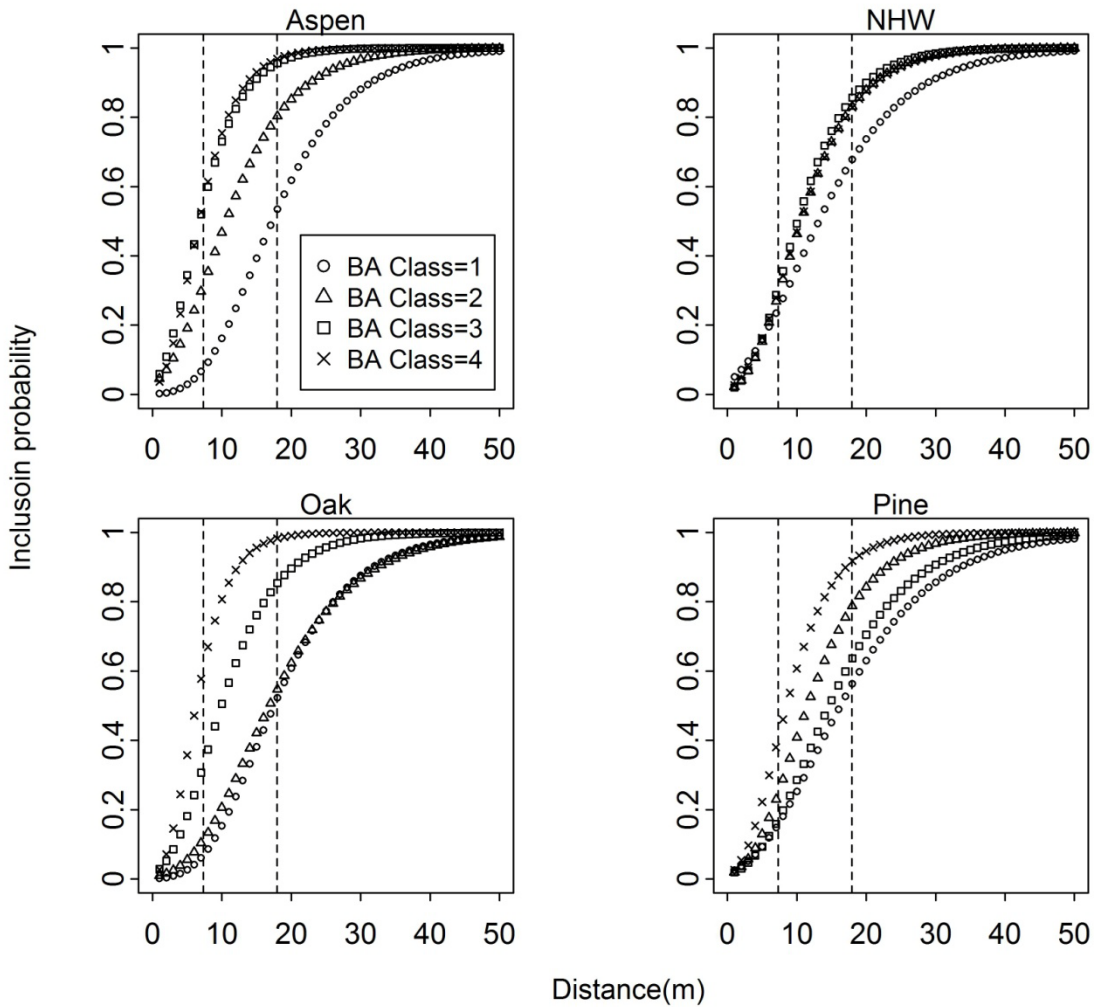


Figure 6.6. The inclusion probabilities by forest type and basal area class. Dashed lines are the radii of subplot (7.32 m) and annular plot (17.95 m).

Figure 6.6 shows the inclusion probability of a standing dead tree by forest type and basal area class predicted from the models in Table 6.5. As expected from previous chapters, the inclusion probability was increased with increasing the density of standing dead trees. For all forest types, stands in the lowest and highest basal area class were clearly different, but there was not always a noticeable difference in the intermediate classes (Fig. 6.6), especially in NHW forests, where

BA class 1 was different, but there was little difference between the other three BA classes. The most unusual case was that of the pine forests, where there was a higher inclusion probability (higher mortality, Table 6.4) in BA class 2 than 3 (Fig. 6.7). The latter indicates that under some circumstances the abundance of standing dead trees is not proportional to stand basal area, and other density-independent mortality factors are at work.

6.4.2. Results of Time Requirement Studies

Table 6.6 summarizes the simulated average required time to count and measure standing dead trees using a 7.32 m (subplot) and 17.95 m radius fixed-area sampling plot and the time to search for and measure the nearest standing dead tree (300 random points were used to generate the average times), in computer-generated forests of varying densities and spatial patterns.

Table 6.6. Average of estimated time requirements (minutes) by survey type from calibrated simulations under different spatial patterns and standing dead tree densities.

Spatial	Density	Subplot	Annular	Nearest
Cluster	12	1.1	9.2	5.8
	24	1.1	10.7	6.1
	49	1.2	13.7	5.3
Random	12	1.1	9.1	5.8
	24	1.1	10.7	3.8
	49	1.2	13.8	2.3

Nearest is time to search for and measure the nearest standing dead tree.

The calibrated simulations (Table 6.6) show only relatively small differences of time requirement in the FAS between the different spatial patterns and densities when subplot in FIA is used. However, there was clear pattern that the time requirement for annular plots increases

with increasing density of standing dead trees, but with no significant difference between the spatial pattern types. In the case of the time to find and measure the nearest standing dead tree, the time requirement tended to decrease with increasing density of standing dead trees in a random pattern, but there was no difference between densities when the dead trees were in a clustered pattern, because most of the time was spent finding the nearest standing dead tree, regardless of the density, since trees within the cluster were relatively close together. In terms of the differences between methods, the simulations showed that finding and measuring the nearest standing dead tree took about 2 to 5.5 times longer than counting and measuring dead trees in a subplot, but only about 1/6 to about 6/10 of the time to complete an annular plot, depending on the conditions of the stand (Table 6.7). The annular plots require from about 8 to about 11.5 times longer to complete than the subplots.

Table 6.7. Estimated average time requirement (min.) per plot (& point) by method and plot type when time models are applied to the BBDMS data.

Infected status	FAS		EZ-Hurdle (w/subplot)
	Subplot	Annular	
Yes	1.9	11.5	2.9
No	1.4	8.5	5.4

Time requirement was measured with two-person crews. FAS is fixed-area sampling method, subplot is 7.32 m radius, annular plot is 17.95 m radius, and EZ-Hurdle is sum of the time to find and measure the nearest standing dead tree and to measure standing dead trees within a subplot.

The time models were also applied to the BBDMS and the average time requirement per plot or point by survey method was estimated by infected status (Table 6.7). For FAS, in both the infected (about 52 SDT / ha) and non-infected plots (about 21 SDT / ha), annular plots required about 6 times longer than the subplots to complete. The estimated time for the EZ-Hurdle method, which includes both the time to complete a subplot and the time to find and measure the

nearest standing dead, was about 1.5 to 3.9 times longer than the time for the subplot alone, for infected vs. non-infected forests, respectively, the latter because with fewer standing dead trees, it took longer to find them. However, the EZ-Hurdle method required about 1.5 to 4 times less time than annular plots for infected and non-infected forests indicating that the EZ-Hurdle method has a much lower time cost than FAS with annular-sized plots.

Table 6.8. Estimated time requirement (hr) for FAS with 7.32 m radius subplots and the additional time requirement for EZP by PPR for a specific number of fixed-area plots, with two-person crews.

N/ha	No. plot	Random						Cluster					
		FAS	EZP					FAS	EZP				
			PPR						PPR				
			1.00	1.25	1.50	1.75	2.00		1.00	1.25	1.50	1.75	2.00
12	36	0.7	3.5	4.4	5.3	6.1	7.0	0.7	6.0	7.5	9.0	10.5	12.0
	72	1.5	7.1	8.8	10.6	12.3	14.1	1.5	12.1	15.2	18.2	21.3	24.3
	108	2.2	10.6	13.2	15.8	18.5	21.1	2.2	18.0	22.5	27.1	31.6	36.1
	144	3.0	14.0	17.5	21.0	24.5	28.0	3.0	24.2	30.2	36.3	42.3	48.3
	180	3.7	17.5	21.9	26.3	30.7	35.0	3.7	30.1	37.6	45.2	52.8	60.3
24	36	0.9	2.1	2.6	3.2	3.7	4.2	0.9	4.4	5.4	6.5	7.6	8.7
	72	1.7	4.2	5.3	6.3	7.4	8.4	1.7	8.7	10.8	13.0	15.2	17.4
	108	2.6	6.3	7.9	9.5	11.1	12.7	2.6	13.1	16.4	19.6	22.9	26.1
	144	3.5	8.4	10.5	12.6	14.7	16.8	3.5	17.3	21.7	26.1	30.4	34.7
	180	4.4	10.5	13.2	15.8	18.4	21.1	4.3	21.8	27.2	32.7	38.1	43.6
49	36	1.1	1.4	1.7	2.0	2.4	2.7	1.1	3.8	4.8	5.7	6.7	7.7
	72	2.3	2.7	3.4	4.1	4.8	5.4	2.2	7.7	9.7	11.6	13.5	15.4
	108	3.4	4.1	5.1	6.1	7.1	8.1	3.4	11.6	14.4	17.3	20.2	23.1
	144	4.5	5.4	6.8	8.2	9.5	10.9	4.5	15.3	19.2	23.0	26.8	30.7
	180	5.7	6.8	8.5	10.2	11.9	13.6	5.6	19.1	23.9	28.7	33.6	38.4

FAS is fixed-area sampling method, No. plot is the number of fixed-area plots, N/ha is density of standing dead trees per ha, EZP is EZ-Hurdle method, and PPR is point to plot ratio.

The time data was further extrapolated via simulation to allow for the time requirement to be explored under different sampling scenarios (Table 6.8). The data show that it can take a

very large amount of additional time to collect the additional point to standing dead tree distances to gain greater precision under the EZ-Hurdle method, and that this is sensitive to both stand conditions (density and spatial pattern) as well as the quantity of additional data collected (Table 6.8). Much of this additional time comes from searching for the occasional dead tree that is very far away from the plot center or random point (when $PPR > 1$). This is particularly problematic when dead trees are in a clustered pattern at low density, because the distribution of distance between random point to the nearest standing dead trees is strongly right skewed (Fig. 6.7). The additional time was much lower when dead trees were randomly distributed (about one-third to one-half of that for clustered trees), but was still quite large compared to the base time for FAS (Table 6.8).

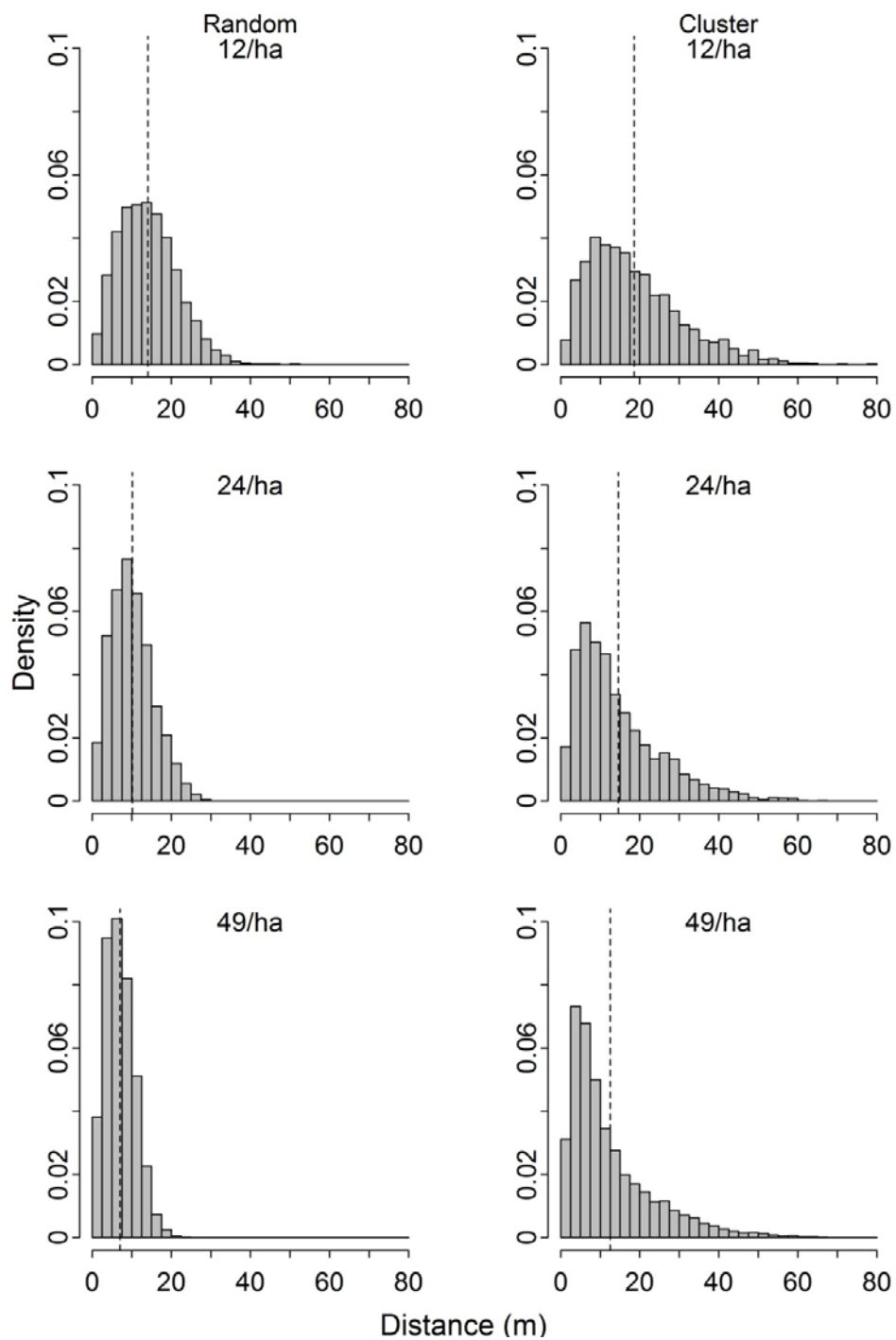


Figure 6.7. The distribution of distance between a random point to the nearest standing dead tree by density and spatial pattern under the simulation study. The dashed line is average distance.

The additional time requirements were also extrapolated from the BBDMS data through the bootstrapping procedure. As in the simulation, times were much lower when the density of standing dead trees was higher (52/ha) in the non-infected forests, for all PPR and numbers of plots, in comparison with non-infected forests where densities were lower. Additional time requirements for the EZ-Hurdle method increased linearly with increasing the number of additional random points outside of the plots (Fig. 6.8).

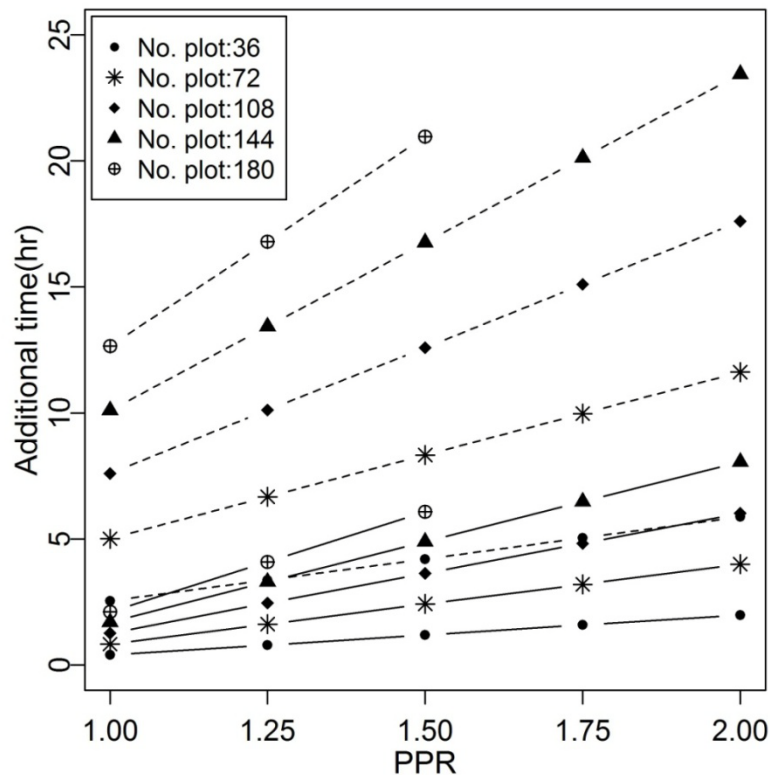


Figure 6.8. Additional time requirement for EZ-Hurdle method above the time cost for FAS by PPR from field study. Solid line is infected forests and dashed line is non-infected forests.

6.4.3. Comparison of the Coefficient of Variation by Method in Simulation

The coefficient of variation for different sampling strategies explored via simulation are shown in Figure 6.9. The coefficients of variation decreased with increasing the number of plots for both methods. The coefficients of variation of EGP method were smaller than FAS for all PPR and the number of fixed-area plot. The improvement of coefficient of variation by EGP method was greater under clustered patterns than random patterns, given the same number of fixed-area plots. As expected from the previous applied study, there were greater or similar improvements of coefficients of variation when PPR is greater than 1, because additional locations are sampled outside of the plot network. However, the data above show that this additional data comes at a very high time cost to collect the auxiliary data.

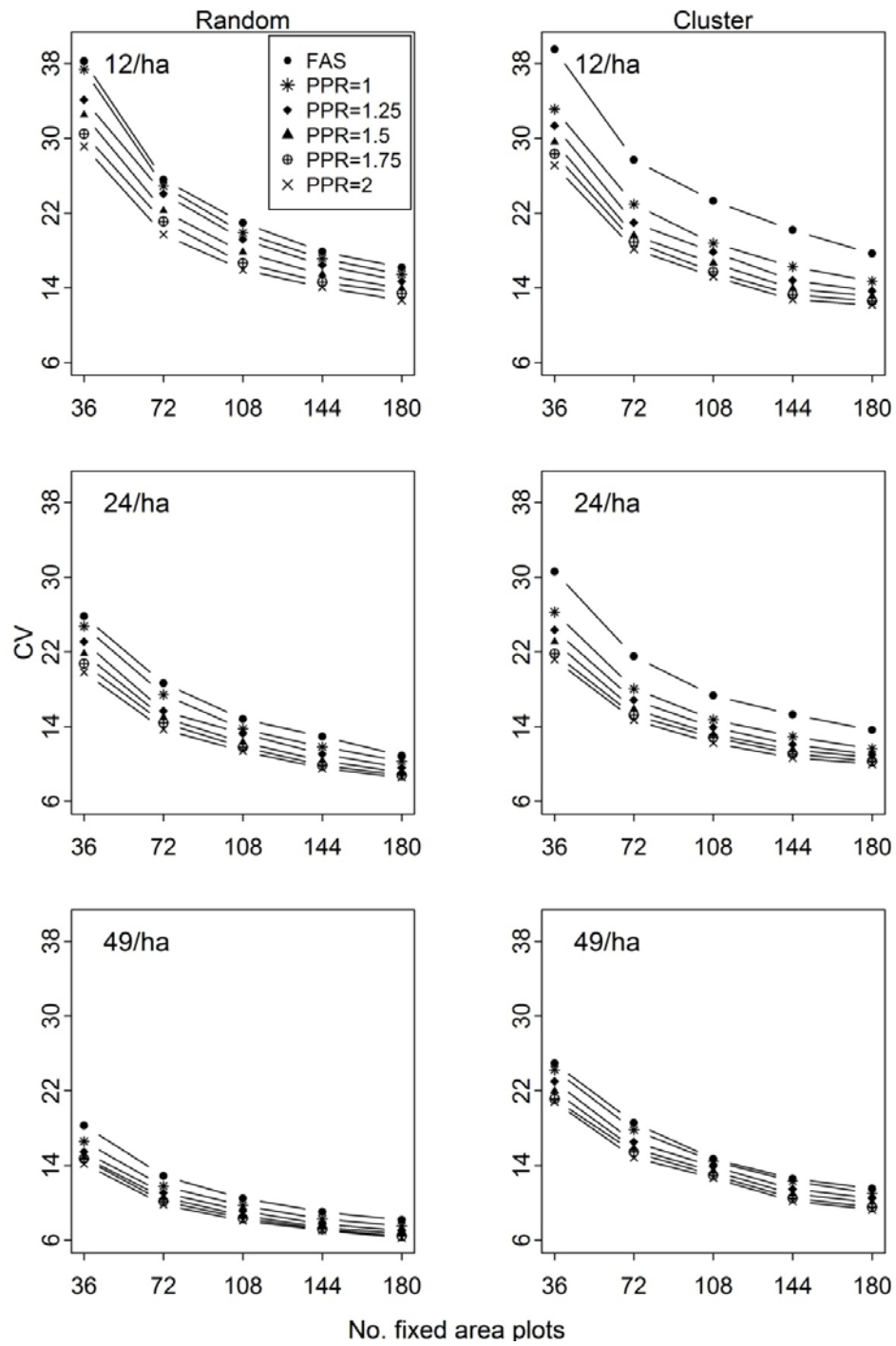


Figure 6.9. The change of coefficient of variation of estimated density by PPR and the number of fixed-area plots.

6.4.4. Comparison Sampling Efficiency by Method

First, relative sampling efficiency was examined via simulation. Relative sampling efficiencies for estimating abundance of standing dead trees are shown in Tables 6.9 when subplot was applied. Because relative efficiency is greater than 1 for all densities and spatial patterns (Table 6.9), the FAS method has a better sampling efficiency than the EZ-Hurdle method because the increase in precision required a relatively much larger increase in the time requirement.

Table 6.9. Relative sampling efficiency of EZ-Hurdle method to compare FAS by density and spatial pattern.

N/ha	No. plot	Random					Cluster				
		PPR					PPR				
		1.00	1.25	1.50	1.75	2.00	1.00	1.25	1.50	1.75	2.00
12	36	5.5	5.5	5.9	5.9	6.1	6.4	7.0	7.4	7.8	8.1
	72	5.5	6.2	6.2	6.4	6.3	5.5	5.6	5.8	6.2	6.4
	108	5.2	5.8	5.9	5.9	6.1	4.9	5.4	5.5	5.7	6.0
	144	5.3	5.9	6.1	6.2	6.5	4.8	4.9	5.1	5.3	5.6
	180	5.2	5.7	6.0	6.4	6.4	5.1	5.4	5.8	6.2	6.5
24	36	3.1	3.2	3.3	3.4	3.5	1.3	1.2	1.2	1.2	1.2
	72	3.0	2.8	3.0	3.1	3.2	1.6	1.6	1.6	1.6	1.7
	108	3.0	3.2	3.2	3.3	3.5	1.9	2.0	2.0	2.2	2.2
	144	2.9	2.9	2.9	3.0	3.2	2.1	2.2	2.3	2.4	2.4
	180	3.0	3.1	3.2	3.4	3.6	2.4	2.5	2.7	2.8	2.9
49	36	1.8	1.8	1.9	2.0	2.1	1.3	1.2	1.2	1.2	1.2
	72	1.8	1.9	1.9	1.9	2.0	1.5	1.4	1.4	1.5	1.5
	108	1.9	1.9	1.9	2.0	2.0	1.8	1.9	2.0	2.0	2.1
	144	1.9	1.8	1.9	2.0	2.1	2.0	2.0	2.0	2.1	2.1
	180	1.9	1.8	1.9	1.9	2.0	2.1	2.2	2.2	2.2	2.3

No. plot is the number of fixed-area plots, N/ha is density of standing dead trees per ha, EZP is EZ-Hurdle method with Poisson distribution, and PPR is point to plot ratio.

In the applied study, the difference in efficiency between EZ-Hurdle and FAS (with small plots) was smaller. However, the FAS method with subplots had better sampling efficiency than EZ-Hurdle method in field study (Table 6.10) and the overall patterns in the results mirrored

those observed in the simulation. The time requirement for the EZ-Hurdle method was more than double in non-infected forests, relative sampling efficiencies were worse when density of standing dead trees is low (non-infected forests) than high density forests (infected forest). In combination, the field study and simulation results, indicate that, holding all else constant, it is more cost-effective to add more smaller (FIA-subplot-sized) plots, where possible, then to collect point-to-dead tree distances to improve precision under the EZ-Hurdle method. However, this is not true for larger plots sizes (e.g., annual-size plots) and is irrelevant under circumstances where additional plots cannot be added, due to other costs or data needs (e.g., permanent sample plots, such as those used by FIA).

Table 6.10. Relative sampling efficiency of EZ-Hurdle method to compare FAS by infected status.

Infected status	No. plot	PPR				
		1.00	1.25	1.50	1.75	2.00
Yes	36	1.55	1.70	1.81	1.93	2.11
	72	1.60	1.69	1.86	2.04	2.18
	108	1.51	1.68	1.81	1.97	2.09
	144	1.63	1.78	1.95	2.11	2.26
	180	1.55	1.82	1.96		
No	36	3.97	3.40	4.49	4.71	4.93
	72	3.82	4.10	4.46	4.76	4.95
	108	3.94	4.56	4.97	5.20	5.36
	144	3.68	4.06	4.57	4.99	5.63
	180	3.83	4.79	5.34		

6.5. Discussion for Cost and Sampling Efficiency

Sampling efficiency is an important factor to decide the sampling strategy for estimating population parameters (Gregoire and Valentine 2008). To estimate the abundance of standing dead trees, inventory methods for estimating live trees have been modified (Kenning et al. 2005). Modifications include increasing the intensity of typical sampling methods such as strip cruising, sampling with fixed-area plots, and horizontal point sampling (prism sampling), which can have poor precision and high variation, despite a considerable time investment, when such methods are applied to areas where individuals in the population are in relatively low abundance or in high variability areas (Bull et al. 1990).

In this study, the EZ-Hurdle method increased the precision of estimates under all conditions by introducing auxiliary data, but at a significantly increased cost under some conditions. Although the EZ-Hurdle model showed better precision than the FAS method when additional information was applied to estimate the expected zero probability, EZ-Hurdle method had worse sampling efficiency than FAS method due to the relatively high search time to find and measure the nearest dead tree, particularly at low standing dead tree density, where zero inflation is most likely, and where EZ-Hurdle performs the best. According to the previous study for N-tree sampling method, one-tree sampling method showed better sampling efficiency than FAS method for estimating the abundance of snags when the density is greater than 70 per ha (Kenning et al. 2005). In this latter case, the FAS method needed more than twice the time requirement than one-tree sampling. In this study, the focus was on stands with lower abundances of standing dead trees which caused the EZ-Hurdle method to be less time efficient when compared with FAS using smaller (FIA-sized subplot) plots, even though EZ-Hurdle produced better estimates. In other words, it's more cost efficient to add more FAS plots, rather

than adding more dead-tree distances for plots of this size, under situations where zero inflation in the data is likely (very low densities of dead trees). On the other hand, the EZ-Hurdle method showed better cost efficiency when compared with FAS using larger (FIA-sized annular) plots, which means, e.g., that EZ-Hurdle method might be beneficial in the Northwestern FIA region where currently annular plots are used in some cases to estimate standing dead tree abundances. Hence, in situations where there is a restriction on the number of plot locations that can be added, the EZ-Hurdle method can be an alternative approach to improve the precision of estimates for the density of standing dead trees because it can improve estimates without the need to establish new plots. But, where there is no such restriction, applying many, smaller FAS plots might be the most effective way to deal with the problem of zero-inflated data sets that arise when dead tree abundances are relatively low. Therefore, EZ-Hurdle method might be most useful to apply to monitoring or inventory programs which use a fixed number of permanent sample plots or cannot change the sampling design, such as the BBDMS in Michigan, the FIA in the USA and the National Inventory System in South Korea (which uses a very similar design to FIA). The EZ-Hurdle method needs less than 5 minutes to find the nearest standing dead tree. In the case of the FIA program, it takes about one day to collect all data to finish one FIA plot, which consists of 4 subplots, so the additional 20 minutes to improve the precision of standing dead tree estimates is a relatively small portion of the total work load for that plot.

Finally, given that the EZ-Hurdle method produces better estimates and its main limitation is the (time) cost of the auxiliary data, it would be beneficial to reduce the cost of the auxiliary data to improve the sampling efficiency of the method. One alternative could be estimating the inclusion probability of a standing dead tree under the sampling design from stand attributes such as forest type, density, spatial pattern, or basal area. In this study, the inclusion

probability was predictably different by density of standing dead trees, species, and spatial pattern, in most cases, which suggests that this is a promising. This will be the subject of further research on the EZ-Hurdle method. Another alternative could be a distance-limited method which restricts the search radius to find the nearest standing dead tree; this is pursued in the next chapter.

7. Developing a Distance-Limited EZ-Hurdle Method

7.1. Overview

Both the field and simulation studies showed that the time requirement to search the nearest standing dead tree was dramatically increased with decreasing the density of standing dead trees because there are points in space where it takes quite a long time to find the nearest dead tree. Therefore, a distance-limited EZ-Hurdle method was explored, where various maximum search radii were applied to find the nearest standing dead tree.

7.2. Methods

Simulation data were used again in this study, following the same basic design as described previously, except that under the original EZ-Hurdle method, there is no distance restriction to search for the nearest standing dead tree. In this new simulation study, a subplot (7.32 m radius) was again used to collect the standing dead tree data, but six different scenarios were used to define the maximum search radius to find the nearest standing dead tree, the first four at +2 meter increments beyond the plot boundary, the fifth out to the boundary of an FIA annular plot and finally to unlimited distance from the point (Table 7.1). Following the previous design (see Table 5.2), five different PPR were applied to collect the data and two different spatial patterns, random and cluster I, were applied with three different densities (12, 24, and 49/ha) to define the underlying populations. For each scenario, 1,000 repetitions were applied to examine the properties of distance-limited EZ-Hurdle method.

Table 7.1. Maximum search radii to find the nearest standing dead tree.

Maximum search radius (m)					
9.32	11.32	13.32	15.32	17.94	Unlimited

The coefficient of variation for estimates and expected zero probability were calculated by search radius. The change of coefficient of variation was compared by density of standing dead trees and spatial patterns. In order to evaluate the time requirement by search radius, time requirement was calculated using regression models developed previously from BBDMS and PRC data (see Tables 6.2 and 6.3).

7.3. Results

Comparison of coefficient of variation by search radii

Tables 7.2 and 7.3 show the coefficient of variation of estimates (N/ha) and expected zero probability (EZP) by the number of fixed-area plot and PPR when the density of standing dead trees is 12 per ha, and when the spatial pattern is random and clustered, respectively. As in previous analyses, the EZ-Hurdle method had a smaller coefficient of variation for all search radii and PPR for both spatial patterns than FAS (Tables 7.2 and 7.3).

There was a strong relationship between the coefficients of variation of the estimates and expected zero probabilities. When the coefficient of variation of expected zero probability decreases, the coefficient of variation of estimate also decreases because the expected zero probability in EZ-Hurdle model reduces the uncertainty caused by zero observations in data. This means that the EZ-Hurdle method still gives a benefit even when one searches only a small distance outside of the FAS plot.

Table 7.2. The CV of estimates (N/ha) and expected zero probability (EZP) by PPR and search radius (m) for EZ-Hurdle method when the spatial pattern of standing dead trees is random and density is 12/ha.

No. fixed	PPR	Value	FAS	Search radius (m)					
				9.32	11.32	13.32	15.32	17.94	∞
36	1.00	N/ha	38.24	34.26	33.39	33.33	33.66	34.34	37.36
		EZP	8.35	8.02	7.57	7.46	7.5	7.62	8.04
	1.25	N/ha	38.24	31.04	30.06	30.09	30.58	31.4	34.12
		EZP	8.35	7.08	6.62	6.57	6.66	6.82	7.21
	1.50	N/ha	38.24	29.46	28.6	28.77	29.28	29.98	32.51
		EZP	8.35	6.51	6.13	6.11	6.22	6.36	6.72
	1.75	N/ha	38.24	27.93	27.18	27.32	27.69	28.19	30.47
		EZP	8.35	6.05	5.72	5.71	5.79	5.89	6.23
	2.00	N/ha	38.24	26.85	26.15	26.24	26.61	27.12	29.14
		EZP	8.35	5.71	5.40	5.38	5.46	5.56	5.85
72	1.00	N/ha	25.58	23.57	22.94	22.85	23.07	23.86	24.91
		EZP	5.55	5.21	4.97	4.93	4.97	5.12	5.43
	1.25	N/ha	25.58	21.83	21.34	21.3	21.43	22.11	24.06
		EZP	5.55	4.75	4.54	4.52	4.54	4.67	4.95
	1.50	N/ha	25.58	20.22	19.72	19.71	19.85	20.38	22.24
		EZP	5.55	4.34	4.13	4.11	4.14	4.25	4.52
	1.75	N/ha	25.58	19.03	18.57	18.58	18.82	19.39	21.1
		EZP	5.55	4.02	3.82	3.81	3.87	3.99	4.25
	2.00	N/ha	25.58	17.91	17.46	17.47	17.68	18.14	19.7
		EZP	5.55	3.73	3.55	3.54	3.59	3.69	3.93
108	1.00	N/ha	20.94	18.85	18.23	18.25	18.54	19.14	19.88
		EZP	4.67	4.19	3.97	3.97	4.04	4.17	4.44
	1.25	N/ha	20.94	17.22	16.68	16.73	17.04	17.58	19.17
		EZP	4.67	3.74	3.57	3.57	3.64	3.76	4.01
	1.50	N/ha	20.94	16.09	15.61	15.64	15.89	16.39	17.82
		EZP	4.67	3.41	3.25	3.26	3.32	3.43	3.66
	1.75	N/ha	20.94	15.12	14.67	14.67	14.9	15.3	16.64
		EZP	4.67	3.15	3.00	3.00	3.05	3.14	3.37
	2.00	N/ha	20.94	14.55	14.11	14.10	14.32	14.71	15.91
		EZP	4.67	2.99	2.85	2.84	2.89	2.98	3.19

No. fixed is the number of fixed-area plots, PPR is point to plot ratio, FAS is fixed-area sampling method, and EZP is expected zero probability.

Table 7.2. Continue.

No. fixed	PPR	Value	FAS	Search radius (m)					
				9.32	11.32	13.32	15.32	17.94	∞
144	1.00	N/ha	17.85	16.31	15.94	15.89	16.16	16.56	17.12
		EZP	3.96	3.59	3.45	3.44	3.50	3.58	3.81
	1.25	N/ha	17.85	14.89	14.52	14.46	14.7	15.06	16.42
		EZP	3.96	3.18	3.05	3.04	3.09	3.17	3.39
	1.50	N/ha	17.85	14.04	13.65	13.58	13.8	14.15	15.42
		EZP	3.96	2.95	2.82	2.80	2.86	2.93	3.14
	1.75	N/ha	17.85	13.44	13.04	13.00	13.22	13.55	14.62
		EZP	3.96	2.79	2.66	2.65	2.70	2.77	2.94
	2.00	N/ha	17.85	12.86	12.48	12.46	12.68	13.00	14.05
		EZP	3.96	2.61	2.49	2.49	2.54	2.60	2.78
180	1.00	N/ha	16.16	14.75	14.26	14.23	14.45	14.9	15.42
		EZP	3.54	3.2	3.03	3.02	3.07	3.17	3.36
	1.25	N/ha	16.16	13.47	12.99	12.99	13.17	13.57	14.71
		EZP	3.54	2.88	2.72	2.72	2.76	2.84	3.02
	1.50	N/ha	16.16	12.75	12.29	12.28	12.45	12.81	13.92
		EZP	3.54	2.68	2.53	2.52	2.56	2.64	2.81
	1.75	N/ha	16.16	12.27	11.86	11.84	12.01	12.37	13.42
		EZP	3.54	2.55	2.41	2.4	2.44	2.52	2.69
	2.00	N/ha	16.16	11.58	11.22	11.19	11.33	11.66	12.66
		EZP	3.54	2.37	2.25	2.24	2.27	2.34	2.51

No. fixed is the number of fixed-area plots, PPR is point to plot ratio, FAS is fixed-area sampling method, and EZP is expected zero probability.

For example, when maximum search radius is 9.32, which is 2 m greater than an FIA subplot, the EZ-Hurdle method had less coefficient of variation than FAS (Tables 7.2 and 7.3). Therefore, distance-limited EZ-Hurdle method can improve the precision without heavy investment of search time to find the nearest standing dead trees. In order to confirm this trend, the coefficient of variation of estimate and expected zero probability were examined with additional two different densities such as 24 and 49 per ha.

Table 7.3. The CV of estimates (N/ha) and expected zero probability (EZP) by PPR and search radius (m) for EZ-Hurdle method when the spatial pattern of standing dead trees is cluster and density is 12/ha.

No. fixed	PPR	Value	FAS	Search radius (m)					
				9.32	11.32	13.32	15.32	17.94	∞
36	1.00	N/ha	39.51	36.54	35.59	34.54	34.02	33.44	33.12
		EZP	7.53	7.32	6.64	6.28	6.16	6.07	6.14
	1.25	N/ha	39.51	34.68	33.61	32.61	32.11	31.67	31.35
		EZP	7.53	6.52	5.93	5.6	5.49	5.43	5.53
	1.50	N/ha	39.51	32.37	31.61	30.94	30.47	29.95	29.59
		EZP	7.53	5.78	5.31	5.07	4.98	4.91	5.02
	1.75	N/ha	39.51	31.07	30.27	29.63	29.14	28.71	28.32
		EZP	7.53	5.45	5	4.78	4.69	4.64	4.73
	2.00	N/ha	39.51	29.73	28.99	28.45	27.99	27.59	27.11
		EZP	7.53	5	4.6	4.41	4.32	4.28	4.34
72	1.00	N/ha	27.69	25.99	25.21	24.77	24.24	23.69	22.93
		EZP	5.22	4.91	4.57	4.44	4.33	4.26	4.31
	1.25	N/ha	27.69	23.82	22.88	22.5	22.07	21.51	20.96
		EZP	5.22	4.26	3.91	3.79	3.71	3.64	3.71
	1.50	N/ha	27.69	22.42	21.55	21.17	20.81	20.33	19.6
		EZP	5.22	3.91	3.59	3.48	3.41	3.35	3.37
	1.75	N/ha	27.69	21.58	20.83	20.38	20.04	19.6	18.91
		EZP	5.22	3.71	3.42	3.29	3.23	3.18	3.19
	2.00	N/ha	27.69	20.62	19.93	19.49	19.16	18.73	18.12
		EZP	5.22	3.50	3.23	3.11	3.05	3.00	3.02
108	1.00	N/ha	23.28	21.46	20.71	20.26	19.86	19.42	18.77
		EZP	4.42	3.98	3.68	3.56	3.50	3.44	3.50
	1.25	N/ha	23.28	20.52	19.79	19.38	19.00	18.58	17.83
		EZP	4.42	3.66	3.38	3.27	3.21	3.15	3.18
	1.50	N/ha	23.28	19.12	18.47	18.07	17.72	17.35	16.63
		EZP	4.42	3.31	3.05	2.95	2.89	2.85	2.87
	1.75	N/ha	23.28	17.89	17.30	16.95	16.63	16.31	15.69
		EZP	4.42	3.06	2.83	2.73	2.67	2.63	2.66
	2.00	N/ha	23.28	17.33	16.79	16.47	16.17	15.85	15.19
		EZP	4.42	2.89	2.68	2.59	2.54	2.50	2.51

No. fixed is the number of fixed-area plots, PPR is point to plot ratio, FAS is fixed-area sampling method, and EZP is expected zero probability.

Table 7.3. Continue.

No. fixed	PPR	Value	FAS	Search radius (m)					
				9.32	11.32	13.32	15.32	17.94	∞
144	1.00	N/ha	20.17	18.67	17.90	17.55	17.22	16.79	16.25
		EZP	3.86	3.49	3.22	3.13	3.08	3.03	3.07
	1.25	N/ha	20.17	17.04	16.28	15.92	15.61	15.24	14.78
		EZP	3.86	3.10	2.84	2.75	2.70	2.65	2.70
	1.50	N/ha	20.17	16.02	15.33	14.99	14.70	14.39	13.92
		EZP	3.86	2.89	2.65	2.57	2.52	2.49	2.53
	1.75	N/ha	20.17	15.36	14.70	14.38	14.11	13.83	13.25
		EZP	3.86	2.72	2.50	2.42	2.37	2.35	2.36
	2.00	N/ha	20.17	14.73	14.11	13.80	13.56	13.30	12.74
		EZP	3.86	2.58	2.37	2.29	2.25	2.23	2.23
180	1.00	N/ha	17.65	16.83	16.29	15.91	15.63	15.32	14.69
		EZP	3.33	3.10	2.90	2.80	2.75	2.72	2.72
	1.25	N/ha	17.65	15.54	15.05	14.73	14.47	14.22	13.66
		EZP	3.33	2.80	2.61	2.53	2.49	2.46	2.47
	1.50	N/ha	17.65	14.82	14.34	14.06	13.83	13.60	13.09
		EZP	3.33	2.63	2.45	2.38	2.34	2.32	2.32
	1.75	N/ha	17.65	14.13	13.73	13.50	13.30	13.08	12.59
		EZP	3.33	2.44	2.28	2.22	2.19	2.17	2.16
	2.00	N/ha	17.65	13.50	13.15	12.95	12.77	12.57	12.14
		EZP	3.33	2.28	2.13	2.07	2.04	2.03	2.03

No. fixed is the number of fixed-area plots, PPR is point to plot ratio, FAS is fixed-area sampling method, and EZP is expected zero probability.

Figures 7.1 and 7.2 show the coefficients of variation of estimated density of standing dead trees (N/ha) by spatial pattern and search radii when density of standing dead trees are 24 per ha and 49 per ha, respectively. The same trends are observed as when the density of standing dead trees was 12 per ha. These results indicate that the coefficient of variation for the estimated density of standing dead trees by EZ-Hurdle method is always less than FAS for both spatial patterns and PPR, whenever one searches beyond the plot radius.

One strange pattern was that the coefficient of variation decreased until the search radius was 13.32 and then increased or was very similar when the search radius was greater than 13.32

up to an unlimited distance for all number of fixed-area plots and PPR under a random pattern (Tables 7.2 and 7.3 and Figs. 7.1 and 7.2).

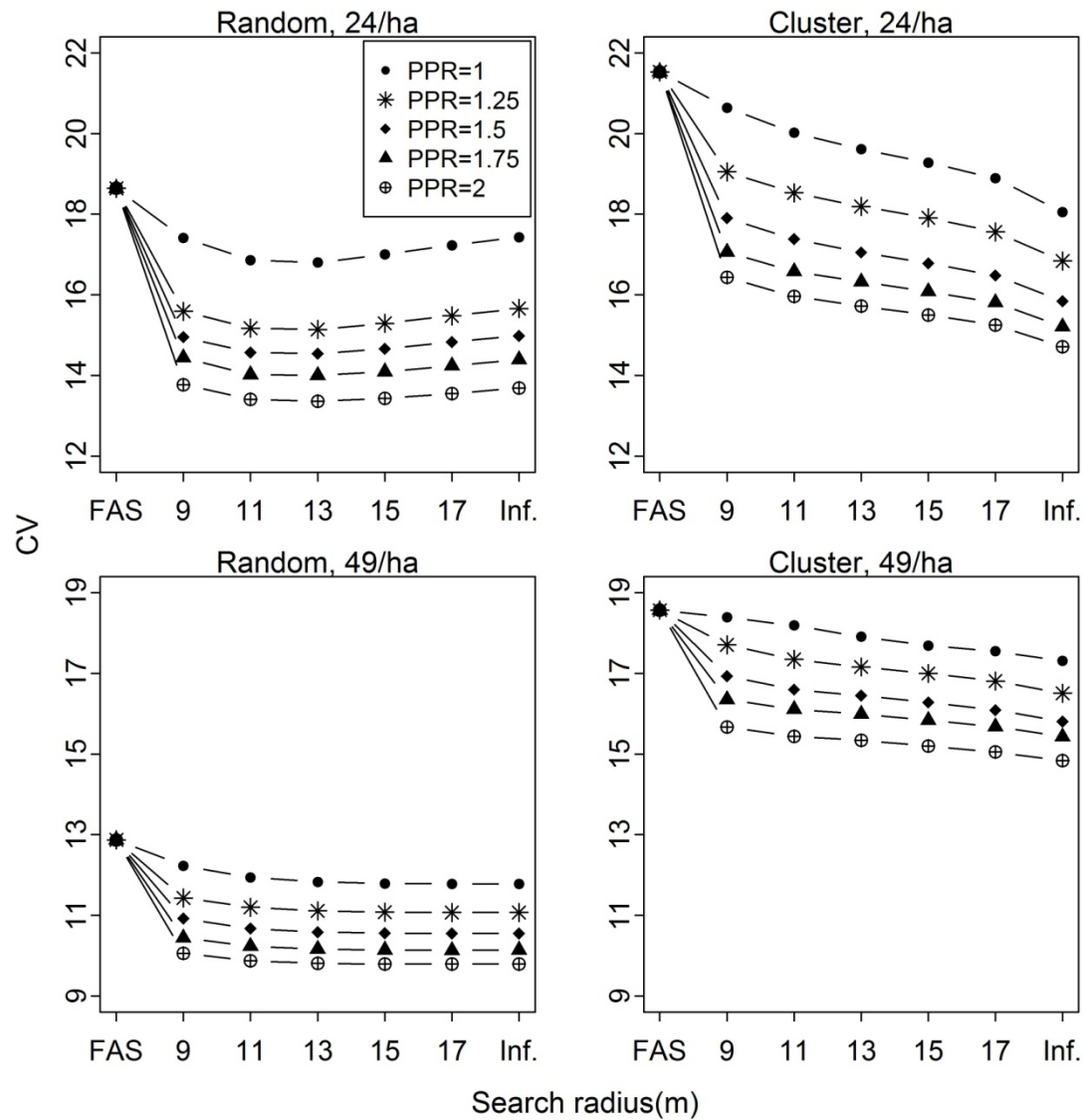


Figure 7.1. The coefficients of variation (CV) of estimated density of standing dead trees by spatial pattern and search radius. FAS is the fixed-area sampling method, PPR is the point to plot ratio, and Inf. means unlimited search radius.

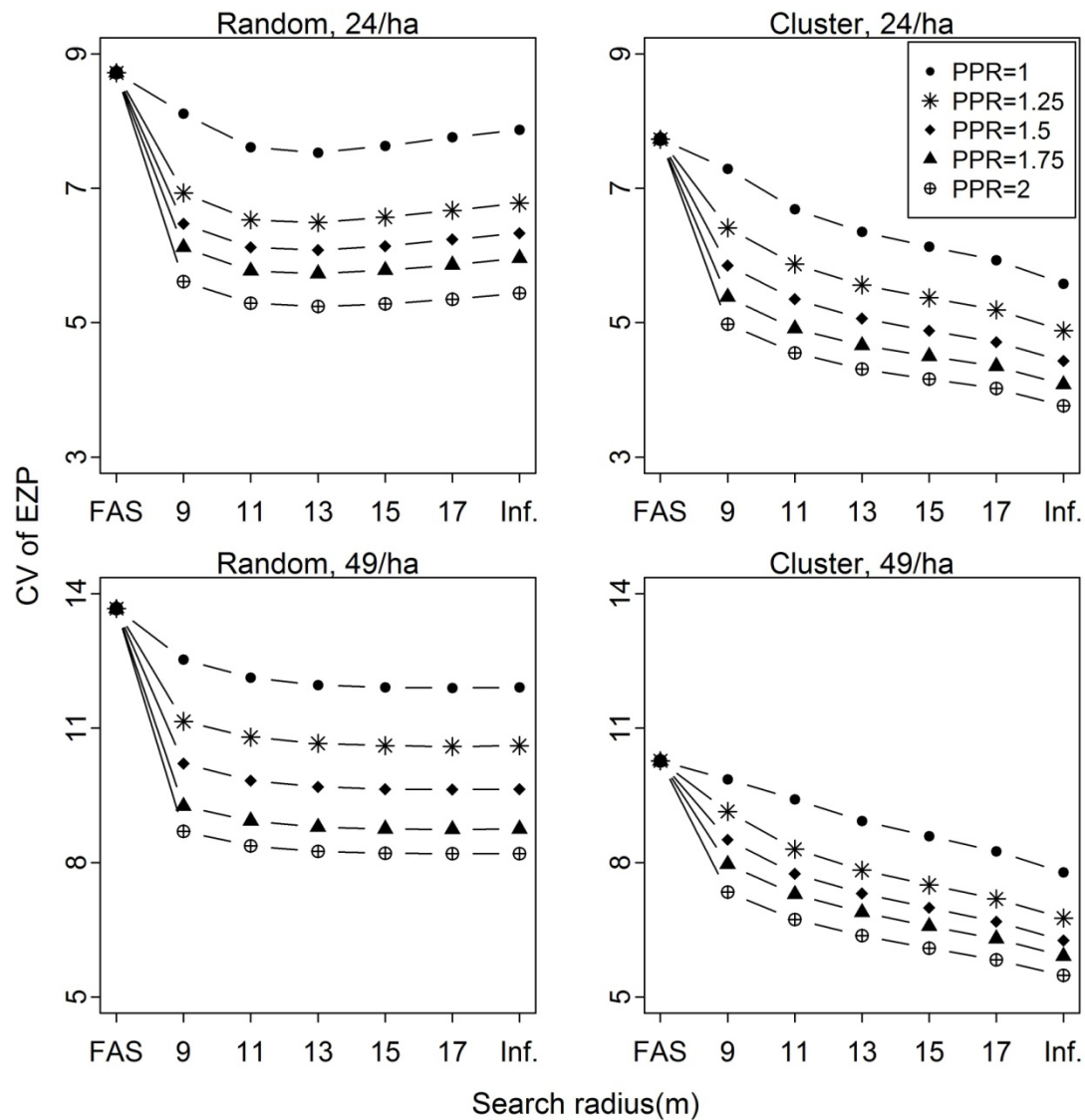


Figure 7.2. The coefficients of variation (CV) of estimated expected zero probability (EZP) by spatial pattern and search radius. FAS is the fixed-area sampling method, PPR is the point to plot ratio, and Inf. means an unlimited search radius was used.

Whereas, the coefficients of variation was decreased continuously when the search radius limitation increased up to an unlimited search under the clustered pattern. It was expected that the expected zero probability for the given search radius (7.32 m radius in this study) should have less variation when there is no distance limit to search the nearest standing dead tree. However, according to Table 7.2 and Figure 7.1, the EZ-hurdle method had the best precision

when the maximum search radius was 11.32 or 13.32 m under a random spatial pattern. This unusual result was checked several times but persisted, suggesting that under some conditions the distance-limited approach can perform better than the distance-unlimited approach in terms of improved precision.

In general, the model selected as the best model for estimating the inclusion probability was the one which has the least square errors for data. Especially, in the case of a parametric method, it is possible that the least square error can be worse when more information or data is added to fit the model when the additional information creates more variability. It is not guaranteed that the inclusion probability in 7.32 m radius is equal to for all search radii (Fig. 7.3). For example, the inclusion probability was 0.34 when the maximum search radius is 9.32 and 11.32 m but it was decreased to 0.33 when the maximum search radius is greater than 11.32 (Fig. 7.3). When events (standing dead trees in this study) are randomly distributed the nearest neighbor distance from random point should be equal to event to event distances. It means that there is no pattern for the inter-event interaction, which is the distance between point to event or event to event in space. It is possible that the nearest neighbor distance from large search areas to find the nearest standing dead tree can bring another inter-event interaction which is different inter-event interaction within small search area because the define of spatial pattern of events (standing dead trees) using inter-event interaction is very sensitive to the spatial scale (Bailey and Gatrell 1995). Therefore, when additional information is collected from the larger search area which is greater than 13.32 m were added to model the inclusion probability, the addition data can introduce the random noise such as another inter-event interaction to the data and can increase the variation to estimate the inclusion probability of give search radius. In other words, it is not guaranteed that one can have better precision to estimate inclusion probability although

add more data.

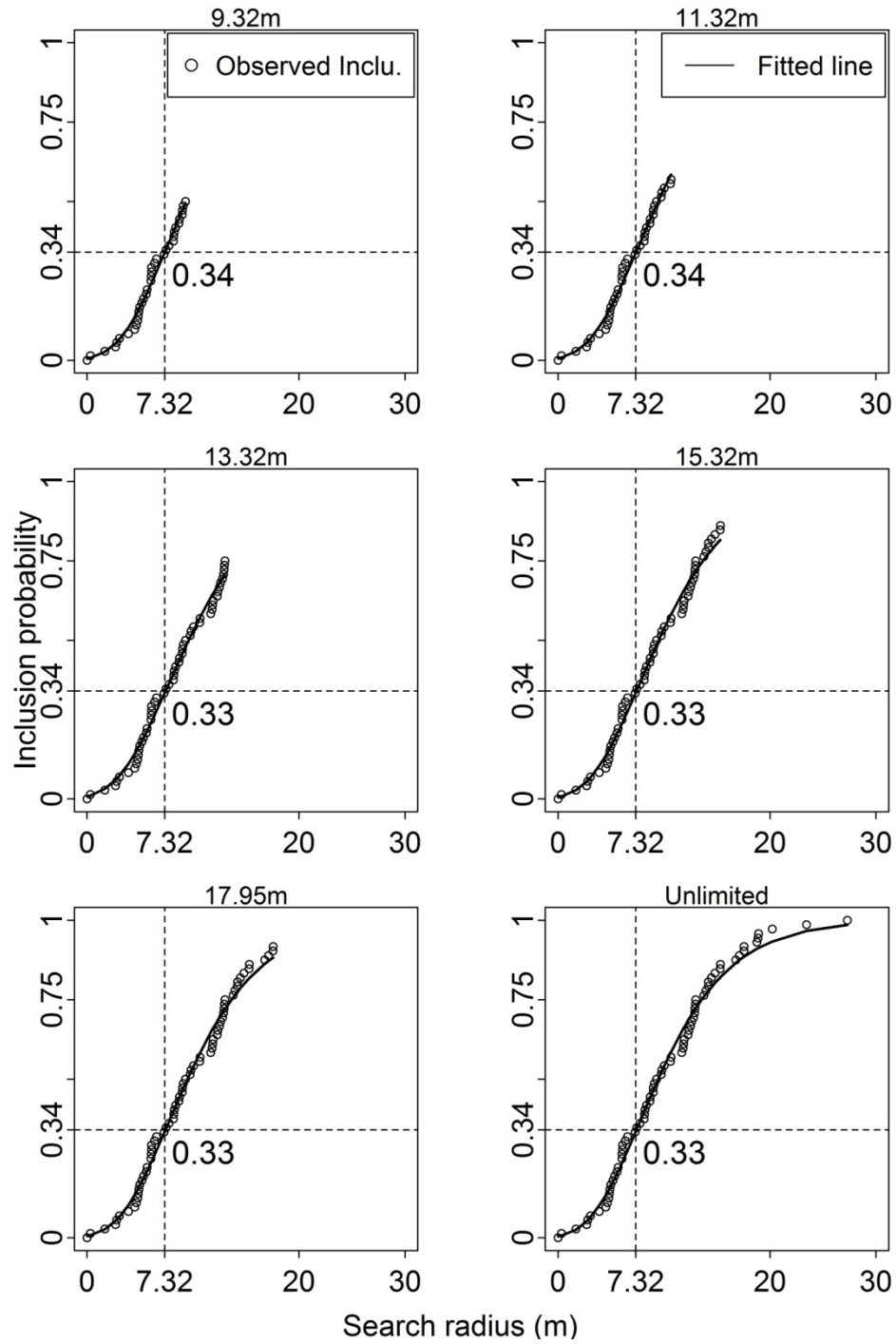


Figure 7.3. The change of inclusion probability of a standing dead tree by increasing the search radius when density of standing dead trees is 24/ha and spatial pattern is clustered.

Comparison of time requirement by search radius

Figure 7.4 shows the additional time requirement for EZ-Hurdle method by maximum search radius. When the search radius is limited, the EZ-Hurdle method takes a lot less time to collect the distance information. For example, when the maximum search radius is 9.32, more than 70% of the time requirement can be saved compared to the distance-unlimited method. If maximum search radius is 17.95 m, more than 50% of the time requirement can be saved.

In general, for both spatial patterns and all densities, additional time requirements were rapidly increased when search radius changes from 17.95 m to unlimited distance (effectively infinite, denote as Inf. in Fig. 7.4). In the case of a random pattern, the additional time requirement was moderately increased until search radius was 17.95 m in comparison to the additional time requirement under a clustered pattern. Especially, when the density of standing dead trees is 49 per ha in random pattern, there was similar time requirement until the search radius was 17.95 m. Thus, much of the additional time cost, which comes from searching for the nearest standing dead tree which is, on occasion, very far away from the point, which may even reduce the quality of the estimate under some conditions (Figs. 7.1 and 7.2), can be saved with distance limiting method.

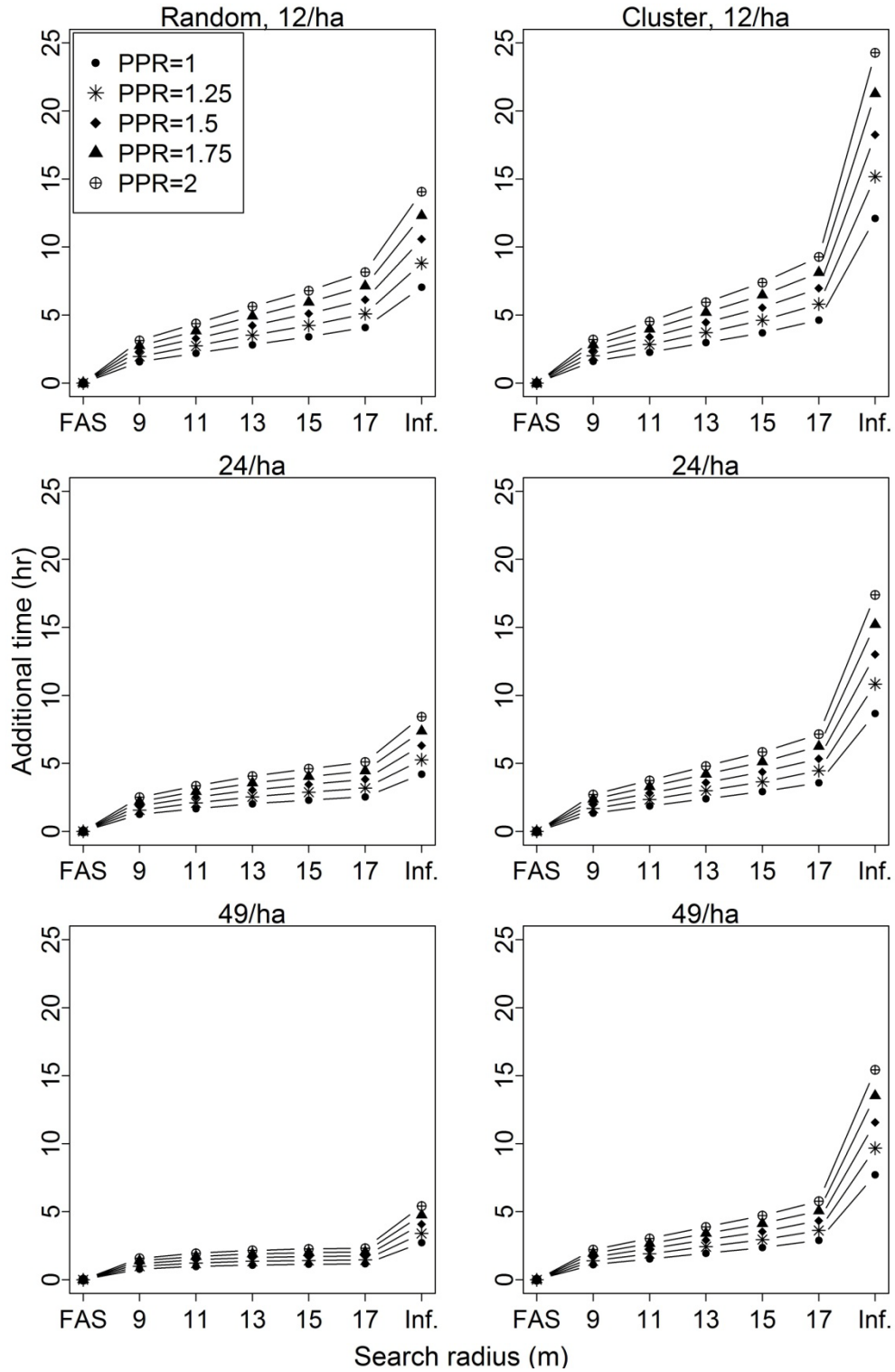


Figure 7.4. Addition time requirement for EZ-Hurdle method by maximum search radius. PPR is point to plot ratio, FAS is fixed-area sampling method, and Inf. is unlimited distance.

7.4. Discussion for Distance-Limited EZ-Hurdle Method

The distance-limited EZ-Hurdle method showed better precision for all restricted search radii, spatial patterns, and densities. Moreover, it still shows better precision even when PPR is 1. Therefore, distance-limited EZ-Hurdle method can be applied to estimate the density of standing dead trees without changes in plot designs such as that used in the FIA and FHM program.

According to the results of time requirement study, the distance-limited EZ-Hurdle method showed great improvement in cost efficiency over the standard EZ-Hurdle method with no distance limit set to find the nearest standing dead tree. The additional time requirement was less than 70% in comparison with the distance-unlimited method when the search radius is 2 m greater than the subplot radius. In addition, we can save more than 50% time requirement when maximum search radius is equal to annular plot (17.95 m) in FIA.

Although the maximum search radius has to be decided based on budget or sampling design, according to the results of this study, 13.32 m should be the best choice when the standing dead trees are randomly distributed. When standing dead trees are clustered, 17.95 m should be the best as the maximum search radius because the precision of estimates is increased with extending the search radius and we can save more than 50% time requirement than standard EZ-Hurdle method. When the maximum search radius is 17.95 m (equal to annular plot), average time investment per plot should be less than 2 minutes based on the BBDMS data. However, we do not know the true spatial pattern of standing dead trees in practice. Therefore, these results suggest that the most efficient application of the method for FIA / FHM program should be the 13.2 m radius distance-limited EZ-Hurdle method, which should prove to be the best all around in the typical case when the spatial pattern of standing dead trees is unknown. It should allow for modest gains in precision in estimates of standing dead tree densities under the

current FIA or FHM plot design, without changing plot design and relatively small investment of time and cost at each plot.

8. Conclusion

Based on both simulation studies and applied studies in real forests, the EZ-Hurdle method can improve the precision of an estimate where zero-inflated data are a source of estimation error. However, the method requires adding additional (auxiliary) data to estimate the expected zero proportion in the data, in this case by collecting additional data describing the distance between a fixed-area plot center or random point and the nearest standing dead tree. This additional data comes at an additional time cost, which can be quite large as shown here and can make the method less cost-efficient than cheaper methods even if the estimates are of lower quality. For the specific case of improving estimates of standing dead tree density from fixed-area plots, the EZ-Hurdle method produces better, but less cost-efficient estimates than FAS method with smaller (e.g., 7 m radius) fixed-area plots, because the time to search for standing dead trees can be quite long compared to establishing a small fixed radius plot which quite often contains few or no dead trees and it proved much more time-cost competitive than using larger (e.g., 17.95 m) fixed area plots. A search distance-limited variant of the EZ-Hurdle method showed promise for reducing the cost inefficiency of the method. The results of this study also suggest that the expected zero probability for including standing dead trees in fixed area plots may also be estimated from simple stand data, such as forest type and basal area, at little or no additional cost, which would make the method even more cost efficient. For the case of standing dead trees, the EZ-Hurdle method is best applied under conditions where zero-inflation in the data is likely and where adding additional samples is unlikely.

In this dissertation, EZ-Hurdle method was been applied to estimate the density of standing dead tree. Because EZ-Hurdle method can perform better when there are large or excess zero observations in data, it might also be used to estimate the density of rare species or

other low-abundance populations. Future work is planned to explore the application of the EZ-Hurdle method to estimate the carbon sequestration of dead trees. Further research is needed to examine the properties of the inclusion probability of a standing dead tree for different forest types, because it may be used to define an “expected mortality” benchmark for forest health monitoring and understanding stand mortality processes.

References

9. References

- Affleck, D.L.R. 2006. Poisson mixture models for regression analysis of stand-level mortality. *Can. J. For. Res.* 36(11):2994-3006.
- Avery, T.E., and H.E. Burkhart. 1983. *Forest measurements*. McGraw-Hill Inc., New York.
- Baddeley, A., and R. Turner. 2005. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 12(6):1-42.
- Bailey, T.C., and A.C. Gatrell. 1995. *Interactive spatial data analysis*. Longman Scientific & Technical Essex.
- Bechtold, W.A., and P.L. Patterson. 2005. The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: USDA For. Serv., Southern Research Station 85.
- Bull, E.L., R.S. Holthausen, and D.B. Marx. 1990. How to determine snag density. *West. J. Appl. For.* 5(2):56-58.
- Bütler, R., and R. Schlaepfer. 2004. Spruce snag quantification by coupling colour infrared aerial photos and a GIS. *For. Ecol. Manag.* 195(3):325-339.
- Clark, D.B., C.S. Castro, L.D.A. Alvarado, and J.M. Read. 2004. Quantifying mortality of tropical rain forest trees using high-spatial-resolution satellite data. *Ecology Letters* 7(1):52-59.
- Cline, S.P., A.B. Berg, and H.M. Wight. 1980. Snag characteristics and dynamics in Douglas-fir forests, western Oregon. *J. Wildl. Manage.* 44(4):773-786.
- Curtis, R.O., and D.D. Marshall. 2005. Permanent-plot procedures for silvicultural and yield research. Gen. Tech. Rep. PNW-GTR-634. USDA For. Serv.
- Delaney, M., S. Brown, A.E. Lugo, A. Torres-Lezama, and N. Bello Quintero. 1997. The distribution of organic carbon in major components of forests located in five life zones of Venezuela. *J. Trop. Ecol.* 13(5):697-708.
- Dorazio, R.M. 1999. Design-based and model-based inference in surveys of freshwater mollusks. *J. N. Am. Benthol. Soc.* 18(1):118-131.
- Dueser, R.D., and H.H. Shugart Jr. 1978. Microhabitats in a Forest-Floor Small Mammal Fauna. *Ecology* 59(1):89-98.
- Efron, B., and R. Tibshirani. 1993. *An introduction to the bootstrap*. Chapman & Hall/CRC, New York.

- Eskelson, B.N.I., H. Temesgen, and T.M. Barrett. 2009. Estimating cavity tree and snag abundance using negative binomial regression models and nearest neighbor imputation methods. *Can. J. For. Res.* 39(9):1749-1765.
- Fisher, R.A. 1922. The Accuracy of the Plating Method of Estimating the Density of Bacterial Populations. *Annals of Applied Biology* 9:325-359.
- Franklin, J.F., F. Hall, W. Laudenslayer, C. Maser, J. Nunan, J. Poppino, C.J. Ralph, and T. Spies. 1986. Interim definitions for old-growth Douglas-fir and mixed-conifer forests in the Pacific Northwest and California. Research note PN-447. USDA For. Serv., Portland, Oregon.
- Franklin, J.F., H.H. Shugart, and M.E. Harmon. 1987. Tree Death as an Ecological Process. *BioScience* 37(8):550-556.
- Fridman, J., and M. Walheim. 2000. Amount, structure, and dynamics of dead wood on managed forestland in Sweden. *For. Ecol. Manag.* 131(1-3):23-36.
- Ganey, J.L. 1999. Snag density and composition of snag populations on two National Forests in northern Arizona. *For. Ecol. Manag.* 117(1-3):169-178.
- Gray, A. 2003. Monitoring stand structure in mature coastal Douglas-fir forests: effect of plot size. *For. Ecol. Manag.* 175(1-3):1-16.
- Green, P., and G.F. Peterken. 1997. Variation in the amount of dead wood in the woodlands of the Lower Wye Valley, UK in relation to the intensity of management. *For. Ecol. Manag.* 98(3):229-238.
- Gregoire, T. 1998. Design-based and model-based inference in survey sampling: Appreciating the difference. *Can. J. For. Res.* 28(10):1429-1447.
- Gregoire, T., G., and H.T. Valentine. 2008. Sampling techniques for natural and environmental resources. Chapman & Hall/CRC.
- Greif, G.E., and O.W. Archibold. 2000. Standing-dead tree component of the boreal forest in central Saskatchewan. *For. Ecol. Manag.* 131(1-3):37-46.
- Harmon, M.E., W.K. Ferrell, and J.F. Franklin. 1990. Effects on Carbon Storage of Conversion of Old-Growth Forests to Young Forests. *Science* 247(4943):699-702.
- Harmon, M.E., J.F. Franklin, F.J. Swanson, P. Sollins, S.V. Gregory, J.D. Lattin, N.H. Anderson, S.P. Cline, and N.G. Aumen. 1986. Ecology of coarse woody debris in temperate ecosystems. *Adv. Ecol. Res.* 15:133-302.
- Hilbe, J. 2007. Negative binomial regression. Cambridge University Press New York.

- Jaeger, R.G. 1980. Microhabitats of a terrestrial forest salamander. *Copeia* 1980(2):265-268.
- Keenan, R.J., C.E. Prescott, and J.P.H. Kimmins. 1993. Mass and nutrient content of woody debris and forest floor in western red cedar and western hemlock forests on northern Vancouver Island. *Can. J. For. Res.* 23(6):1052-1059.
- Kenning, R.S., M.J. Ducey, J.C. Brissette, and J.H. Gove. 2005. Field efficiency and bias of snag inventory methods. *Can. J. For. Res.* 35(12):2900-2910.
- Kimmins, J.P. 1992. *Balancing Act: Environmental Issues in Forestry*. UBC Press, Vancouver.
- Krankina, O.N., and M.E. Harmon. 1995. Dynamics of the dead wood carbon pool in northwestern Russian boreal forests. *Water, Air, & Soil Pollution* 82(1):227-238.
- Lee, P.C., S. Crites, M. Nietfeld, H.V. Nguyen, and J.B. Stelfox. 1997. Characteristics and origins of deadwood material in aspen-dominated boreal forests. *Ecological Applications* 7(2):691-701.
- Lessard, V., D.D. Reed, and N. Monkevich. 1994. Comparing n-tree distance sampling with point and plot sampling in northern Michigan forest types. *North. J. Appl. For.* 11(1):12-16.
- Maser, C., and J.M. Trappe. 1984. The seen and unseen world of the fallen tree. USDA For. Serv. Gen. Tech. Rep. PNW-GTR-164:153p.
- Matern, B. 1986. Spatial variation, volume 36 of *Lecture Notes in Statistics*. New York: Springer-Verlag, second edition.
- McCarthy, B.C., and R.R. Bailey. 1994. Distribution and abundance of coarse woody debris in a managed forest landscape of the central Appalachians. *Can. J. For. Res.* 24(7):1317-1329.
- McClelland, B.R. 1977. Relationships between hole-nesting birds, forest snags, and decay in western larch-douglas-fir forests of the northern Rocky Mountains, Dissertation, University of Montana, Missoula, Montana, USA.
- McComb, W.C., T.A. Spies, and W.H. Emmingham. 1993. Douglas-Fir Forests: Managing for Timber and Mature-Forest Habitat. *J. Forestry* 91(12):31-42.
- McCullough, D.G., R.L. Heyd, and J.G. O'Brien. 2001. Biology and management of beech bark disease. Extension Bull. E-2746, Michigan State University Extension Service.
- Montes, F., and I. Canellas. 2006. Modelling coarse woody debris dynamics in even-aged Scots pine forests. *For. Ecol. Manag.* 221(1-3):220-232.
- O'Langhlin, J., and P.S. Cook. 2003. Inventory-based forest health indicators: Implications for national forest management. *J. Forestry* 101(2):11-17.

- Oswalt, S.N., T.J. Brandeis, and C.W. Woodall. 2008. Contribution of Dead Wood to Biomass and Carbon Stocks in the Caribbean: St. John, US Virgin Islands. *Biotropica* 40(1):20-27.
- Petrillo, H.A., J.A. Witter, and E.M. Thompson. 2004. Michigan beech bark disease monitoring and impact analysis system. Unpublished, University of Michigan.
- Potts, J.M., and J. Elith. 2006. Comparing species abundance models. *Ecological Modelling* 199(2):153-163.
- R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ranius, T., O. Kindvall, N. Kruys, and B.G. Jonsson. 2003. Modelling dead wood in Norway spruce stands subject to different management regimes. *For. Ecol. Manag.* 182(1-3):13-29.
- Reid, C.M., A. Foggo, and M. Speight. 1996. Dead wood in the Caledonian pine forest. *Forestry* 69(3):275.
- Rothstein, D.E., Z. Yermakov, and A.L. Buell. 2004. Loss and recovery of ecosystem carbon pools following stand-replacing wildfire in Michigan jack pine forests. *Can. J. For. Res.* 34(9):1908-1918.
- Rubin, B., and D. MacFarlane. 2008. Using the Space-Time Permutation Scan Statistic to Map Anomalous Diameter Distributions Drawn from Landscape-Scale Forest Inventories. *For. Sci.* 54(5):523-533.
- Spiering, D.J., and R.L. Knight. 2005. Snag density and use by cavity-nesting birds in managed stands of the Black Hills National Forest. *For. Ecol. Manag.* 214(1-3):40-52.
- Stephens, S.L. 2004. Fuel loads, snag abundance, and snag recruitment in an unmanaged Jeffrey pine-mixed conifer forest in Northwestern Mexico. *For. Ecol. Manag.* 199(1):103-113.
- Sturtevant, B.R., J.A. Bissonette, J.N. Long, and D.W. Roberts. 1997. Coarse woody debris as a function of age, stand structure, and disturbance in boreal Newfoundland. *Ecological Applications* 7(2):702-712.
- Tu, W. 2002. Zero-inflated data. In: El-Shaarawi, A.H., Peiegorsch, W.W. (Eds.), *Encyclopedia on Envirometrics*. John Wiley and Sons, Chichester. 4:2387-2391.
- Tyrrell, L.E., and T.R. Crow. 1994. Dynamics of dead wood in old-growth hemlock-hardwood forests of northern Wisconsin and northern Michigan. *Can. J. For. Res.* 24(8):1672-1683.
- USDA Forest Service. 2005. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 3.0. USDA For. Serv.:203p.

- USDA Forest Service. 2008. Forest Inventory and Analysis National Program. USDA Forest Service.
- Vasiliauskas, R., A. Vasiliauskas, J. Stenlid, and A. Matelis. 2004. Dead trees and protected polypores in unmanaged north-temperate forest stands of Lithuania. *For. Ecol. Manag.* 193(3):355-370.
- Voller, J., and S. Harrison. 1998. Conservation Biology Principles for Forested Landscapes. Univ of British Columbia Pr.
- Wisdom, M.J., and L.J. Bate. 2008. Snag density varies with intensity of timber harvest and human access. *For. Ecol. Manag.* 255(7):2085-2093.
- Woodall, C.W., G.M. Domke, D.W. MacFarlane, and C.M. Oswalt. 2011. Comparing field- and model-based standing dead tree carbon stock estimates across forests of the United States. *Forestry*. In Review.
- Young, L.J., and J.H. Young. 1998. Statistical ecology: a population perspective. Kluwer Academic Pub.