



Doctoral Thesis

Optimal sample design for extensive forest inventories

Author(s):

Lanz, Adrian

Publication Date:

2000

Permanent Link:

<https://doi.org/10.3929/ethz-a-004118649> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH No. 13843

Optimal Sample Design for Extensive Forest Inventories

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of
Doctor of Technical Sciences

presented by

Adrian Lanz
dipl. Forst-Ing. ETH
born April 17, 1957
citizen of Roggwil BE

accepted on the recommendation of

Prof. Dr. Peter Bachmann, examiner
PD Dr. Daniel Mandallaz, co-examiner
Prof. Dr. Michael Köhl, co-examiner

2000

Summary

At the planning stage of an extensive forest inventory one has to decide on how to select the sample from the population. In the course of comparing two sample design alternatives, the one resulting in a higher precision of the estimates for the limited budget available, respectively the design resulting in lower cost for given precision requirements of the estimates is considered more efficient. A theoretical framework allowing to optimise sample designs for extensive forest inventories was elaborated in this thesis.

The theory is based on the typical sample selection plan for extensive forest inventories: the selection of sampling points in the investigated forest area and the selection of sample trees in the neighbourhood of these sampling points. It is assumed that the sampling points are randomly distributed over the forest area (random point sampling). It can be shown that the required population parameters are estimated without bias in the framework of probability sampling theory (design-based approach), if the selection of sample trees in the neighbourhood of the sampling points fulfils well-defined rules.

The estimation procedures are elaborated for two-phase, two-stage trakt sampling with varying sampling intensities in the second phase (DSRPS); two-stage sampling means in this context independent and random sub-sampling of sample trees selected in a first stage in the neighbourhood of the sampling points. In principle, the estimator for the mean spatial density of a target variable over the forest area under study can also be used for the ratio of two target variables or the total of a target variable over the forest area. In the latter case, however, the surface area of the investigated forest are has to be known. DSRPS-sampling includes as special cases several sample designs, which are frequently used in practice: one-phase sampling, one-stage sampling, simple trakt sampling (a trakt of size 1), non-stratified sampling.

It is well known from sampling theory that a population model is needed to optimise sample designs, i. e. the anticipated variance of the estimates under a stochastic population model is required. Usually, the target variable is modelled (super-population model). Under the random point sampling approach, however, it is sufficient to assume a Poisson-model for the tree locations leaving the population parameters constant. To get a better adjustment of the model to the real forest under study, one may assume an independent and uniform distribution of the tree locations over arbitrarily chosen population-strata within the forest area (local Poisson forest model). Under two-phase sampling it is assumed that the auxiliary information observed in the first phase can define these population strata (two-phase sampling for stratification).

Under the Poisson-model and DSRPS-sampling, the anticipated variance is a function of relatively few estimable population parameters. Linear cost models are needed to derive analytical solutions for the relative efficiency of sample design alternatives and their design parameters (number and allocation of sampling points and the rules for the selection of sample trees in the neighbourhood of the sampling points).

The findings under the model assumptions confirm several results, which have been suspected, or partially empirically and partially theoretically found in the past. This is the case, for instance, for the selection of sample trees in the neighbourhood of the sampling points. The tree selection probabilities under one-stage sampling are optimal, if they are proportional to the size of the target variable (PPS). Under two-stage sampling a prediction variable is observed on all sample trees (first-stage sample of trees) and the target variable is measured on a sub-set of the first-stage sample of trees (second-stage sample of trees). Ideally the measurement cost per first-stage sample tree should be substantially

smaller than for a second-stage sample tree and that the correlation between the prediction and target variable should be high. Under two-stage sampling, the first-stage tree selection probabilities are optimal if chosen proportionally to the size of the prediction variable (PPP), and if the second-stage tree inclusion probabilities are chosen proportionally to the size of the error (PPE), i. e. the residual value between the target and prediction variable. These proportionality rules are often difficult to implement in practice. Typically, they are approximated by a sampling procedure, which selects the sample trees on two or more concentric circles. In the thesis, an optimisation technique is presented for such sampling procedures, so that the global optimisation of sample designs is available for these sampling procedures, which are of great practical importance.

A case study with data from the first and second Swiss national forest inventory shows that the anticipated variance slightly underestimates the empirical variance for all target variables and for all domains of study. The reason is: the variation of the target variables over the forested landscape in Switzerland is unsatisfactorily explained by the population-strata in the model (stand type, degree of canopy density, degree of mixture), the distribution of the tree locations in Swiss forests is not random, the sampling points of the Swiss national forest inventory are systematically distributed over the forest area.

Optimal sample designs have been calculated for the target variable timber volume. The sample design implemented in the second Swiss national forest inventory is very similar to the optimal sample design calculated under the model assumptions. That a two-phase, two-stage sample design has been chosen for the second Swiss national forest inventory, compared with the one-phase, one-stage sample design applied in the first Swiss national forest inventory, was an important and correct decision.

The Poisson forest model has been proven to be an excellent tool for the optimisation of sample designs for extensive forest inventories. The analytical and empirical results can be understood on intuitive grounds, the model parameters are in most cases easily available, and the calculations are readily done.

Zusammenfassung

Bei der Planung von extensiven Waldinventuren stellt sich die Frage, wie die Stichprobe aus der Population zu entnehmen ist. Im Vergleich zweier Stichprobenverfahren wird dabei dasjenige Verfahren als effizienter betrachtet, das bei vergleichbaren Kosten eine präzisere Schätzung der Populationsparameter liefert, resp. bei vorgegebenen Genauigkeitsansprüchen an die Schätzungen tiefere Kosten verursacht. In der Arbeit werden die theoretischen Grundlagen erarbeitet, die es erlauben, verschiedene Stichprobenpläne für beliebige grossflächige Waldgebiete zu optimieren.

Die Theorie basiert auf der für Waldinventuren typischen Auswahl von Stichprobenpunkten im Waldgebiet, in deren Umgebung dann die Probebäume aufgenommen werden. Dabei wird vorausgesetzt, dass die Lage der Stichprobenpunkte im Waldgebiet dem Zufall überlassen wird (Zufallspunktstichprobe, random point sampling). Es kann gezeigt werden, dass bei geeigneter Auswahl der Probebäume um die Stichprobenzentren, die gesuchten Populationsparameter im Rahmen der Theorie für Zufallsstichproben (design-based approach) unverzerrt geschätzt werden können.

Die Schätzverfahren werden für ein zweiphasiges, zweistufiges Traktauswahlverfahren mit unterschiedlicher Stichprobendichte in der zweiten Phase (DSRPS) erarbeitet; dabei wird unter zweistufig die zufällige und unabhängige Unterauswahl von Probebäumen verstanden, die in einer ersten Stufe in der Umgebung der terrestrischen Stichprobenpunkte ausgewählt wurden. Der Schätzer für die mittlere räumliche Dichte einer Zufallsvariable im Waldgebiet kann grundsätzlich auch für die Schätzung von Quotienten zweier Zufallsvariablen und die Schätzung von Totalen verwendet werden, wobei für die Schätzung von Totalen vorausgesetzt wird, dass die Waldfläche bekannt ist. Das DSRPS-Auswahlverfahren beinhaltet verschiedene Spezialfälle, die in der Praxis häufig angetroffen werden: einphasiges Auswahlverfahren, einstufiges Auswahlverfahren, einfaches Traktauswahlverfahren (Trakt der Grösse 1), nicht-stratifiziertes Auswahlverfahren.

Aus der Stichprobentheorie ist bekannt, dass die Optimierung von Stichprobenplänen nur über Populationsmodelle möglich ist, d.h. über die antizipierte Varianz der Schätzer unter einem stochastischen Populationsmodell. Im Gegensatz zur Modellierung der Zufallsvariable selbst (Superpopulationsmodell), genügt im Fall der Waldinventur ein Poisson-Modell für die Lage der Bäume im Waldgebiet. Die Populationsparameter bleiben unter diesem Modell also unverändert. Die zufällige und unabhängige Verteilung der Lage der Bäume kann auch innerhalb von frei wählbare Populationsstraten innerhalb des Waldgebietes vorausgesetzt werden (local Poisson forest model), womit es besser an die realen Verhältnisse angepasst werden kann. Bei der Optimierung von zweiphasigen Stichprobenplänen wird dann vorausgesetzt, dass die Hilfsinformationen der ersten Phase diese Populationsstraten erfassen (zweiphasige Stichprobe zur Stratifizierung).

Unter dem Poisson-Modell lässt sich die antizipierte Varianz der Schätzer unter dem DSRPS-Auswahlverfahren als Funktion von wenigen, einfach zu schätzenden Populationsparametern ausdrücken. Im Verbund mit linearen Kostmodellen können die relative Effizienz verschiedener Auswahlverfahren und deren optimale Design-Parameter (Anzahl Stichprobenpunkte, Allokation der Stichprobenpunkte auf die Straten, Auswahl der Probebäume) analytisch berechnet werden.

Die unter den Modellannahmen analytisch hergeleiteten Ergebnisse bestätigen bisher vermutete, teilweise theoretisch und teilweise empirisch nachgewiesene Zusammenhänge. Dies trifft etwa auf die Auswahl der Probebäume in der Umgebung der Stichprobenpunkte zu.

Diese ist beim einstufigen Verfahren optimal, falls die Auswahlwahrscheinlichkeiten proportional zur Zielgrösse gewählt werden (PPS: probability proportional to size). Beim zweistufigen Baumauswahlverfahren wird an allen Probebäumen (erste Stufe) eine kostengünstig zu erhebende Hilfsgrösse gemessen, die idealerweise gut mit der Zielgrösse korreliert. Die mit höheren Kosten verbundene Messung der Zielgrösse beschränkt sich auf eine Unterstichprobe (zweite Stufe) der Probebäume. Das optimale Auswahlverfahren wählt die Probebäume der ersten Stufe proportional zur Hilfsgrösse (PPP: probability proportional to prediction) und diejenigen der zweiten Stufe proportional zum Residuum zwischen Ziel- und Hilfsgrösse (PPE: probability proportional to error). In der Praxis können diese Proportionalitätsregeln oftmals nicht genau eingehalten werden. Sie werden typischerweise durch die Auswahl der Probebäume auf konzentrischen Kreisen um den Stichprobenpunkt ersetzt. Ein Verfahren zur Optimierung konzentrischer Kreisproben wird in der Arbeit vorgestellt, so dass die globale Optimierung des Stichproben-Designs auch unter diesem Baumauswahlverfahren ermöglicht wird.

In einer Fallstudie mit Daten aus dem ersten und zweiten Schweizer Landesforstinventar wurde festgestellt, dass die antizipierte Varianz die empirische Varianz für alle untersuchten Zielgrössen in allen Regionen des Landes in etwa dem gleichen Mass leicht unterschätzt. Die Erklärungen dafür sind: die im Modell verwendeten Populationsstraten (Entwicklungsstufen, Deckungsgrad, Mischungsgrad) erfassen die Streuung der Zielgrössen im Schweizer Wald ungenügend, die reale Verteilung der Bäume im Schweizer Wald ist nicht zufällig, die Stichprobenpunkte des Schweizer Landesforstinventars sind systematisch und nicht zufällig über das Waldgebiet verteilt.

Für die Zielgrösse Holzvorrat wurden optimale Stichprobenpläne berechnet. Das im zweiten Landesforstinventar verwendete Stichprobenverfahren kommt dabei dem optimalen Verfahren unter den Modellannahmen erstaunlich nahe. Wichtig war, dass im Vergleich mit dem einphasigen, einstufigen Stichprobenverfahren des ersten Landesforstinventars, im zweiten Schweizer Landesforstinventar ein zweiphasiges, zweistufiges Verfahren gewählt wurde.

Das Populationsmodell erweist sich in den bisherigen Untersuchungen als ausgezeichnetes Werkzeug für die Optimierung von extensiven Waldinventuren. Die analytischen und empirischen Resultate sind gut interpretierbar, die Modellparameter können in der Regel leicht beschafft werden, und die Berechnungen sind wenig aufwendig.