

DISS. ETH NO.

**Integration of Small Area Estimation Procedures of Timber
Volume Resources in Large Scale Forest Inventories**

A dissertation submitted to the
ETH ZURICH

for the degree of
Doctor of Sciences

presented by

ANDREAS CHRISTIAN HILL
MSc. of Sciences

born 15 January 1986
citizen of Germany

accepted on the recommendation of
Prof. Dr. H.R. Heinimann, examiner
PD Dr. D. Mandallaz, supervisor, co-examiner
Prof. Dr. J. Saborowski, co-examiner
Dr. Johannes Breidenbach, co-examiner

Zurich, 2018

Contents

Acknowledgements	1
1 General Introduction	2
1.1 History and state of the art of forest inventory	3
1.2 Thesis objective and structure	4
2 Synthesis	9
2.1 Goals	10
2.2 Main results	11
2.3 Outlook	12
Bibliography	13
Curriculum Vitae	15

List of Figures

List of Tables

Acknowledgements

The thesis reveals the aggregated results of my work at the Section of Forest Growth Modelling and Computer Sciences of the Northwest German Research Institute and the Department of Forest Economics and Forest Management of the University of Göttingen. Along the way, many people made significant contribution to the presented work. Its completion would have been unlikely without the constructive help and the continuous support from my colleagues, family and friends, whom I thank herewith. In particular, I thank

- Prof. Dr. Jürgen Nagel - for the supervision of the thesis, critical revision of my workings and the freedom to explore new ideas.
- Prof. Dr. Bernhard Möhring - for acting as co-referee, the constructive and rational support of economic problems and the opportunity to keep going.
- Prof. Dr. Thomas Kneib - for being oral examiner in the disputation.
- Prof. Dr. Hermann Spellmann - for the decision support to write this thesis and the trustful and cooperative working environment.
- Christoph Fischer - for valuable help in countless tasks.
- Dr. Jan Hansen - for the numerous hours of help to connect the TreeGrOSS packages with R.
- My co-authors Alexander Lange, Elmar Spiegel and Sabine Rumpf - for the successful submissions.
- Andreas Hill - for many hours with constructive and interesting discussions and for proofreading.
- Dr. Matthias Schmidt and Dr. Egbert Schönfelder - for supporting statistical and methodological questions.
- Jan Butschkow, Tim Koddenberg, Michael Hill and Sebastian Ohrmann - for proofreading.
- The partners from the BioEconomy research project - for supporting practical questions regarding forestry, logistics and wood processing.
- My family - for miscellaneous supports.
- Laura - for giving me the courage to tackle the challenges and the strength to carry on.

Chapter 1

General Introduction

“She may not look like much, but she’s got it where it counts.”

— Han Solo

1.1 History and state of the art of forest inventory

1.2 Thesis objective and structure

The objective of this thesis was to contribute methodologies to the recent developments in combining existing forest inventory data from field surveys with auxiliary data, such as those derived from remote sensing data. The particular focus was thereby on investigating the potential increase in value of already existing large scale terrestrial forest inventories to be used on small-scale management levels. The main question of this thesis to be addressed was: what estimation accuracies can be realized using forest inventory data on much smaller spatial levels as their sampling intensities have originally been designed for? This question was investigated in two main studies: The first study (study 1) constitutes the main part of this thesis and concentrated on exploring the capabilities and performances of design-based multi-phase regression estimators in the service of small area estimation. The second study (study 2) investigated a new approach for evaluating the estimation accuracies of forest attribute maps, which are considered to be a special case of small area estimation. The following subsections will give a more detailed introduction to each of the studies.

1.2.1 Study 1: Design-based small area estimation

This study constituted the major work of the thesis and had the objective to develop and evaluate a double-sampling estimation procedure for the German National Forest Inventory (German NFI). The particular objective of the study was to investigate whether the German NFI data can provide acceptable estimation precision on two forest district levels when incorporated in small area estimation procedures. Similar studies have been conducted in Norway (Breidenbach & Astrup, 2012) and Switzerland (Magnussen et al., 2014; Steinmann et al., 2013), but no extensive study had yet been available for Germany. The results from this study were considered to provide valuable evidence whether a double-sampling extension of the German NFI might be a cost-saving alternative to a regional terrestrial forest district inventory (FDI). It was thus a prerequisite to gather information for a sufficiently large number of small area units in order to allow for reliable conclusions. For this reason, we conducted a study in the German federal state Rhineland-Palatinate (RLP) where we extended the German NFI to a double-sampling design and applied three types of small area regression estimators in order to derive point and variance estimates of mean standing timber volume on two forest district levels comprising 45 and 405 units respectively.

In this framework, we decided to particularly explore the performances of design-based regression estimators in the infinite population approach. Methods for this family of estimators have considerably been contributed to by the works of Mandallaz (2008, 2013a,b) and Mandallaz et al. (2013), and applications of the suggested regression estimators for global estimations have been intensively investigated by Massey (2015). Thus, our study was also a continuation in the application of these estimators for the special case of small area estimation. The design-based double-sampling estimators suggested by Mandallaz were also favored for the following reasons: First, the estimators are explicitly formulated for cluster sampling designs such as applied in the German NFI, which has not yet been the case for frequently used model-dependent estimators. Second, the design-based frame considerably relaxes requirements on the regression model which seemed appropriate facing severe quality restrictions in the auxiliary data of the study area. Third, the estimators provide the asymptotically unbiased g-weight variance estimator which a) accounts for the design-dependency of the regression coefficients on the sample under the commonly applied *internal model approach*, and b) is also robust to heteroscedasticity of model residuals.

The conduction of this study was divided into 3 work packages that each addressed major milestones towards the overall study objective. The work and results of these work packages are respectively presented in chapter ??, chapter ?? and chapter ?? and have also been published in or submitted to scientific peer-reviewed journals. In the following, we will give an introduction

and some additional background information to each of these chapters.

Work Package 1: Software implementation

With respect to the study objective, work package 1 addressed the need of a robust and flexible software implementation of the design-based regression estimators that could handle large inventory data sets and process a large number of small area estimations at once. Whereas several of the estimators suggested by Mandallaz had been applied in simulations and real-world case studies (Mandallaz, 2013a; Mandallaz et al., 2013; Mandallaz, 2013b; Massey et al., 2014; Massey & Mandallaz, 2015a,b), there had yet not been a unified and consistent implementation of the estimators in the same software environment. The work on this study was thus taken as an opportunity to implement the full range of these regression estimators as a *package* in the statistical software *R* (R Core Team, 2017). The implementation procedure comprised three steps in general: First, a comprehensive review of all published regression estimators by Mandallaz; second, the completion of yet missing estimators for three-phase small area estimation; and third, the actual implementation of the estimators in *R*. The latter seemed to be the software of choice, as it constitutes one of the most intensively used statistical softwares and also provides interfaces to data base systems in which inventory and geodata are commonly stored. A review of existing software for multi-stage and multi-phase estimation revealed that in comparison to official statistics, applications particularly suited for forest inventories have been rare. Exceptions are the *R* package **JoSAE** by Breidenbach (2015) and the **maSAE** package by Cullmann (2016). However, a more comprehensive software package covering a larger variety of sample designs and estimators - particularly in the design-based infinite population frame - has not yet been available. In order to address this lack between availability and recent interest in such methods, we also made our software package freely available as the *R* package **forestinventory** that can be installed from the CRAN server (<https://CRAN.R-project.org/package=forestinventory>). Chapter ?? describes the implementation and the application of the two-phase and three-phase estimators in *R* and provides a comprehensive review of the design-based regression estimators for global and small area estimation published by Mandallaz (2008, 2013a,b) and Mandallaz et al. (2013). The availability of the software package in combination with its comprehensive documentation also had the objective to support the transparency and the reproducibility of the methods applied in this study.

Work Package 2: Derivation of auxiliary data and modeling

The objective of work package 2 was to find a suitable ordinary least square (OLS) regression model to be used as *internal model* in the small area regression estimators. In order to apply the estimators to all management units in RLP, the regression model had to allow for predicting the standing timber volume of a German NFI sample plot at any location over the federal state forest area. This also imposed the restriction on the auxiliary data to be available at the federal state level. A similar large-scale study with the same model purpose had recently been published by Maack et al. (2016) for the German federal state of Baden-Württemberg. Likewise in this study, we also derived explanatory variables from country-wide airborne laser scanning (ALS) data, which were in our case however characterized by severe quality variations as well as time gaps of up to 10 years between the acquisition and the terrestrial survey date. Our study thus specifically addressed techniques to improve the performance of ordinary least square regression models under such restricting conditions. Another novelty of our study was the use of tree species information derived from a country-wide tree species classification map as additional explanatory data. The integration of tree species information in timber volume prediction models has often been stated as some of the most promising but often missing and thus not well investigated information (Koch, 2010; White et al., 2016). One yet existing gap of knowledge concerned the effect of species misclassifications on the regression model accuracy. We addressed this question by developing and evaluating a calibration

technique for removing a potential bias in the regression coefficients caused by misclassifications in the tree species variable. A third challenge that further increased the complexity of the model selection procedure was to identify optimal extraction areas for the explanatory variables under the variable plot sizes of angle count sampling applied in the German NFI. Besides the modeling-specific aspects, the work on this study basically comprised the integration and storage of both the terrestrial NFI data and the remote sensing data in a PostgreSQL database using a PostGIS extension. The latter allowed for a georelational storage and query of both data sources and provided fast computation of explanatory variables for large data sets.

Work Package 3: Small area estimation

1.2.2 Study 2: Mapping

Forest attribute maps provide an area-wide overview of important information such as development stages, tree species or growing conditions and have thus always been of high interest for forest practitioners. For a long time, such maps were exclusively produced by hand and required expensive field visits and visual inspection of aerial photography. This amount of work also hampered a frequent updating of the maps. However, the production of maps (*mapping*) covering large areas has lately been substantially supported by the availability of exhaustive remote sensing data in combination with modeling techniques (Brosofske et al., 2014). Especially maps of predicted forest attributes in high resolution are considered to support the spatially precise allocation of management operations such as harvesting. An example is the use of rasterized timber volume prediction maps for the optimal allocation of cable roads in the frame of harvesting in steep slope mountainous terrains (Bont & Heinimann, 2012; Bont et al., 2015).

Despite the advantages of providing such high-resolution predictions maps, one should however have in mind that the predictions are often made for considerably small spatial units (map pixels). In most cases, the map pixels match the extent of an inventory sample plot on which the prediction model has been calibrated. One can thus interpret mapping as an extreme case of *small area estimation*. Design-based double-sampling estimators for small area estimation (Study 1) provide closed-form variance formulas that allow for quantifying the estimation precision for every small area unit individually by its estimation error and confidence interval. However, these concepts of quantifying the uncertainty cannot be transferred to mapping approaches in the particular case of a small area unit (i.e. map pixel) corresponding to the size of a sample plot. It is thus necessary that similar efforts than for model building are invested in methods that reliably quantify the resulting map accuracies. A common way to characterize the map accuracy is to use metrics such as the coefficient of determination (R^2) or cross-validated root mean square error (RMSE), which rather address the overall prediction performance than the accuracy of individual predictions. For this reason, the specification of confidence intervals on pixel level has been stated as an important contribution to map accuracy assessment (McRoberts, 2010).

In case of continuous response variables such as standing timber volume, the application of linear regression models allows for providing a confidence region for each prediction (i.e. pixel) based on the *prediction interval* (Fahrmeir et al., 2013, pp.136–139). The objective of the study presented in Chapter ?? was to investigate an alternative approach of deriving pixel-wise confidence intervals by applying well-known concepts of accuracy assessment for categorical classification results (Congalton & Green, 2008). The core of the suggested approach was to define intervals within the range of terrestrial data and their respective model predictions, and subsequently calculate the *user's accuracy* for each of those intervals. The calculated users' accuracies can then be regarded as the confidence levels for the chosen intervals. In this framework, we demonstrate an optimization algorithm (heuristic search method) that - given a pre-defined number of intervals - automatically identifies the interval boundaries with respect to the best possible classification accuracies. The motivation for the development of this method was twofold: First, to provide a map user with the possibility to evaluate the map detail, i.e. number of intervals/classes, in dependence of the realizable prediction accuracies; and second, to allow for identifying intervals of the response variable for which the map produces considerably high or low prediction accuracies.

The suggested methods were applied in a mountainous study site in the canton of Grisons (Switzerland). We used the regional forest district inventory data in combination with data from an airborne laser scanning acquisition to produce a map of the standing timber volume on sample plot level, on which we subsequently applied the developed accuracy assessment. The setup of this study also addressed the overall question of the thesis, i.e. what prediction accuracies can be realized on small spatial scales when using forest inventory data that are only available in comparatively low sampling frequencies.

Chapter 2

Synthesis

2.1 Goals

2.2 Main results

2.3 Outlook

Bibliography

- Bont, L. & Heinimann, H. R. (2012). Optimum geometric layout of a single cable road. *European journal of forest research*, 131(5), 1439–1448.
- Bont, L. G., Heinimann, H. R., & Church, R. L. (2015). Concurrent optimization of harvesting and road network layouts under steep terrain. *Annals of Operations Research*, 232(1), 41–64.
- Breidenbach, J. (2015). **JoSAE**: *Functions for some Unit-Level Small Area Estimators and their Variances*. R package version 0.2.3.
- Breidenbach, J. & Astrup, R. (2012). Small area estimation of forest attributes in the norwegian national forest inventory. *European Journal of Forest Research*, 131(4), 1255–1267.
- Brosofske, K. D., Froese, R. E., Falkowski, M. J., & Banskota, A. (2014). A review of methods for mapping and prediction of inventory attributes for operational forest management. *Forest Science*, 60(4), 733–756.
- Congalton, R. G. & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.
- Cullmann, A. D. (2016). **maSAE**: *Mandallaz’ Model-Assisted Small Area Estimators*. R package version 0.1-5.
- Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: models, methods and applications*. Springer Science & Business Media.
- Koch, B. (2010). Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 581–590.
- Maack, J., Lingenfelder, M., Weinacker, H., & Koch, B. (2016). Modelling the standing timber volume of baden-württemberg—a large-scale approach using a fusion of landsat, airborne lidar and national forest inventory data. *International Journal of Applied Earth Observation and Geoinformation*, 49, 107–116.
- Magnussen, S., Mandallaz, D., Breidenbach, J., Lanz, A., & Ginzler, C. (2014). National forest inventories in the service of small area estimation of stem volume. *Canadian Journal of Forest Research*, 44(9), 1079–1090.
- Mandallaz, D. (2008). *Sampling techniques for forest inventories*. CRC Press.
- Mandallaz, D. (2013a). Design-based properties of some small-area estimators in forest inventory with two-phase sampling. *Canadian Journal of Forest Research*, 43(5), 441–449.
- Mandallaz, D. (2013b). A three-phase sampling extension of the generalized regression estimator with partially exhaustive information. *Canadian Journal of Forest Research*, 44(4), 383–388.

- Mandallaz, D., Breschan, J., & Hill, A. (2013). New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based monte carlo approach with applications to small-area estimation. *Canadian Journal of Forest Research*, 43(11), 1023–1031.
- Massey, A. & Mandallaz, D. (2015a). Comparison of classical, kernel-based, and nearest neighbors regression estimators using the design-based monte carlo approach for two-phase forest inventories. *Canadian Journal of Forest Research*, 45(11), 1480–1488.
- Massey, A. & Mandallaz, D. (2015b). Design-based regression estimation of net change for forest inventories. *Canadian Journal of Forest Research*, 45(12), 1775–1784.
- Massey, A., Mandallaz, D., & Lanz, A. (2014). Integrating remote sensing and past inventory data under the new annual design of the swiss national forest inventory using three-phase design-based regression estimation. *Canadian Journal of Forest Research*, 44(10), 1177–1186.
- Massey, A. F. (2015). *Multiphase estimation procedures for forest inventories under the design-based Monte Carlo approach*. PhD thesis, ETH Zurich.
- McRoberts, R. E. (2010). The effects of rectification and global positioning system errors on satellite image-based estimates of forest area. *Remote Sensing of Environment*, 114(8), 1710–1717.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Steinmann, K., Mandallaz, D., Ginzler, C., & Lanz, A. (2013). Small area estimations of proportion of forest and timber volume combining lidar data and stereo aerial images with terrestrial data. *Scandinavian journal of forest research*, 28(4), 373–385.
- White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., & Tompalski, P. (2016). Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, 42(5), 619–641.

Curriculum Vitae

Personal Details

Name	Andreas Christian Hill
Date of birth	15.01.1986
Place of birth	Trier (Germany)

Education

02/2014–03/2018	Ph.D. Student ETH Zurich, Department of Environmental Systems Science, Chair of Forest Engineering
04/2015–05/2017	Diploma of Advanced Studies ETH ETH Zurich, Applied Statistics
04/2010–08/2013	Master of Science (M.Sc.) University of Göttingen, Forest Sciences and Forest Ecology with study focus on Forest Ecosystem Analysis and Information Processing
10/2006–03/2010	Bachelor of Science (B.Sc.) University of Göttingen, Forest Sciences and Forest Ecology

Professional Experience

Since 10/2013	Researcher Department of Environmental System Science, Chair of Forest Engineering, ETH Zurich
10/2011–04/2012	Research Assistant Department of Environmental System Science, Chair of Forest Engineering, ETH Zurich
09/2007–03/2011	Research Assistant Department of Forest Inventory and Remote Sensing, Faculty of Forest Sciences, University of Göttingen

Publications in Conference Proceedings and Workshops

- [1] **Hill, A.**, Mandallaz, D., Buddenbaum, H., Stoffels, J., Langshausen, J. (2017): Implementation of design-based small area estimations on forest district level in Rhineland-Palatinate by combining remote sensing data with data of the Third German National Forest Inventory. In *3rd International Workshop on Forest Inventory Statistics*. Freiburg, Baden-Württemberg Germany.
- [2] **Hill, A.**, Stoffels, J., Langshausen, J. (2016): . In *CARISMA-workshop on large-scale mapping and estimation of forest resources*. Ås, Norway
- [3] **Hill, A.**, Breschan, J. (2014): Automatic Design of Efficient Harvesting Units using Remote Sensing and Field Data. In *24th IUFRO World Congress*. Salt Lake City, Utah, USA
- [4] Breschan, J., **Hill, A.** (2014): Validation of timber volume maps derived from remote sensing data. In *24th IUFRO World Congress*. Salt Lake City, Utah, USA

Publications in Scientific Journals

- [1] Lamprecht, S., **Hill, A.**, Stoffels, J., Udelhoven, T.(2017): A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data. *Remote Sensing*, 9 (5). doi: 10.3390/rs9050505
- [2] **Hill, A.**, Breschan, J., Mandallaz, D. (2014): Accuracy Assessment of Timber Volume Maps using Forest Inventory Data and LiDAR Canopy Height Models. *Forests*, 5 (9). 2253-2275. doi: 10.3390/f5092253
- [3] Mandallaz, D., Breschan, J., **Hill, A.** (2013): New Regression Estimators in Forest Inventories with Two-Phase Sampling and Partially Exhaustive Information: a Design-Based Monte Carlo Approach with Applications to Small-Area Estimation. *Canadian Journal of Forest Research*, 43 (11). 1023-1031. doi: 10.1139/cjfr-2013-0181

Other Publications

- [1] Mandallaz, D., **Hill, A.**, Massey, A. (2016): Design-based properties of some small-area estimators in forest inventory with two-phase sampling - revised version. *Technical Report*, Department of Environmental Systems Science, ETH Zurich. doi: 10.3929/ethz-a-010579388
- [2] **Hill, A.**, Massey, A., Mandallaz D. (2016): forestinventory: Design-Based Global and Small-Area Estimations for Multiphase Forest Inventories. R package version 0.1.0 *CRAN Repository* url: <https://CRAN.R-project.org/package=forestinventory>

