Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2021:6
Faculty of Forest Sciences

Data assimilation of forest variables predicted from remote sensing data

Nils Lindgren

Data assimilation of forest variables predicted from remote sensing data

Nils Lindgren
Faculty of Forest Sciences
Department of Forest Resource Management
Umeå

SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Umeå 2021

Acta Universitatis agriculturae Sueciae
2021:6

ISSN 1652-6880
ISBN (print version) 978-91-7760-690-1
ISBN (electronic version) 978-91-7760-691-8

Umeå
Print: SLU Service/Repro, Uppsala 2021

Data assimilation of forest variables predicted from remote sensing data

Abstract

Forest information for management planning is today gathered through a combination of field inventories and remote sensing, but the available flow of remote sensing data over time is not yet utilized for continuously improving predictions of forest variables. In the thesis, the utility of data assimilation, in particular the Extended Kalman filter, for forest variable prediction is investigated. This is an iterative algorithm, where data are repeatedly merged and forecasted.

The test site was a forest estate in southern Sweden (Lat. 58°N Long. 13°E). Data assimilation of remote sensing predictions of canopy surface models from digital aerial photogrammetry in paper I and predictions based on interferometric synthetic aperture radar in paper II provided a marginally improved accuracy. This gain was, however, far from the theoretical potential of data assimilation. The reason for this was suggested to be correlation of errors of subsequent predictions across time, i.e. residuals from different predictions over a certain forest area had a similar size and sign. In paper III these error correlations were quantified, and an example of the importance of considering them was given. In paper IV, it was shown that classical calibration could be applied to counteract regression toward the mean, and thus reduce the error correlations. In paper V, it was shown that data assimilation applied to a time series of data from various remote sensing sensors could be used to, over time, improve initial predictions based on aerial laser scanning data. It was also shown how the combination of classical calibration and a suggested modified version of the extended Kalman filter, that accounted for error correlations, contributed to these promising results.

Keywords: Forest inventory, remote sensing, growth, data assimilation, prediction, extended Kalman filter.

Author’s address: Nils Lindgren, Swedish University of Agricultural Sciences, Department of Forest Resource Management, Umeå, Sweden

Dataassimilering av skogliga variabler skattade med fjärranalys

Sammanfattning

Information för skogligplanering samlas idag in genom en kombination av fältinventering och fjärranalys, men det finns ett flöde av fjärranalysdata som inte används för att förbättra skattningen av skogstillståndet. I avhandlingen undersöks dataassimilering, särskilt metoden ”Extended Kalman Filter”, för skattning av skogliga variabler. Detta är en iterativ algoritm där data viktas samman och framskrivs upprepade gånger.

Studierna utfördes på en större skogsfastighet i Västergötland (Lat. 58 ° N Long. 13 ° E). Dataassimilering av serier av fjärranalysskattningar från tredimensionella krontaksmodeller från flygburen digitalfotogrammetri i studie I och radar interferometri (InSAR) i studie II gav en något förbättrad noggrannhet. Vinsten var dock långt ifrån den teoretiska potentialen för dataasssimilering. Anledningen antogs vara att felen hos följande skattningar korrelerar, dvs. avvikelser från sanna skogliga variabeln över ett visst skogsområde är en över eller underskattning av liknande storlek som följande skattningar. Felkorrelationerna kvantifierades i studie III och ett exempel gavs på hur viktigt det är att beakta dem. I studie IV visades hur klassisk kalibrering kunde tillämpas för att motverka regressionsanalysens dragning mot medelvärdet, och därmed minska felkorrelationerna. I studie V visades att dataassimilering av en tidsserie med data från olika fjärranalyssensorer kunde användas för att över tid förbättra skattningar där start tidpunkten var en skattning från flygburenlaserskanning. Det visades också hur kombinationen av klassisk kalibrering och en föreslagen modifierad version av metoden ”Extended Kalman Filter” som tar hänsyn till felkorrelationer bidrog till lovande resultat.

Nyckelord: Skogsinventering, fjärranalys, tillväxt, dataassimilering, skattning, extended Kalman filtrering

Författarens adress: Nils Lindgren, Sveriges Lantbruksuniversitet, Institutionen för skoglig resurshushållning, Umeå, Sweden

Dedication

To my beloved family

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Nyström M., Lindgren N., Wallerman J., Grafström A., Muszta A., Nyström K., Bohlin J., Willén E., Fransson J.E.S., Ehlers S., Olsson H., Ståhl G. (2015). Data assimilation in forest inventory: First empirical results. *Forests* 6:12, 4540–4557.

II. Lindgren N., Persson H.J., Nyström M., Nyström K., Grafström A., Muszta A., Willén E., Fransson J.E.S., Ståhl G., Olsson H. (2017). Improved prediction of forest variables using data assimilation of interferometric synthetic aperture radar data. *Canadian Journal of Remote Sensing.* 43:4, 374–383.

III. Ehlers S., Saarela S., Lindgren N., Lindberg E., Nyström M., Persson H.J., Olsson H., Ståhl G. (2018). Assessing error correlations in remote sensing-based estimates of forest attributes for improved composite estimation. *Remote Sensing*. 10:5, 667.

IV. Lindgren N., Nyström K., Olsson H., Ståhl G. Importance of calibration for improving the efficiency of data assimilation for predicting forest characteristics. (manuscript)

V. Lindgren N., Olsson H., Nyström K., Nyström M., Ståhl G. Data assimilation of growing stock volume using a sequence of remote sensing data from different sensors. (manuscript)

Paper II is reproduced with the permission of the publisher. Paper I and III are published as open-source articles.

The contribution of Nils Lindgren to the papers included in this thesis was as follows:

I. Contributed to the planning of the study, the selection of reference data, wrote parts of the code for producing the results and graphics, and wrote parts of the manuscript.

II. Planned the study, made the predictions from InSAR data, handled the sample plot data, and produced the results partly based on existing code, produced the figures and wrote the major part of the manuscript.

III. Contributed to the processing of the remote sensing data and participated in the discussion and preparation of the manuscript.

IV. Contributed to planning the study, wrote the code, performed the simulations, produced the results for the study, and wrote large parts of the manuscript.

V. Did the major part of the planning of the study, the predictions from remotes sensing data, ran the assimilations and adapted the code for a developed method of data assimilation, as well as prepared the results and wrote the major part of the manuscript.

Abbreviations

3D	Three dimensional
ABA	Area-based approach
ALS	Airborne laser scanning
BA	Basal area
DA	Data assimilation
DEM	Digital elevation model
DP	Digital photogrammetry
DSM	Digital surface model
EKF	Extended Kalman Filter
EKFm	Modified Extended Kalman Filter
GPS	Global positioning system
H_L	Lorey's mean height
InSAR	Interferometric Synthetic Aperture Radar
LOOCV	Leave-One-Out Cross-Validation
NFI	National forest inventory
OS	Optical satellite
RMSE	Root mean square error
rRMSE	Relative RMSE
SI	Site index

1. Introduction

Forest owners require knowledge about the state of their forests for management planning, for ascertaining that governmental regulations are followed, and sometimes also for environmental certification purposes, e.g. according to the Forest Stewardship Council (FSC) practices (Svenska FSC 2020). Production of timber and pulpwood is generally the main goal of forestry in the Nordic countries, and thus economic considerations are important for management decisions. A typical management cycle in Swedish forestry starts with plantation of coniferous tree species. Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) sum up to 80% of the growing stock volume. Of the deciduous species birch (*Betula* spp.) is most common. Pre-commercial thinning with brush saws follows upon planting. The established forest stand is often thinned at least once, and, lastly, the stand is finally felled in an operation where almost all trees are removed and the cycle starts again. Thus, even-aged forestry is typically practiced, although uneven-aged forestry is attracting increasing interest.

Wall-to-wall forest mapping has traditionally been carried out through delineation of the forest into compartments, which normally contain one stand each. This step has been guided by visual interpretation of air photos. Forest data for the compartments are then assessed using field surveys, with quick manual measurements at subjectively selected locations. Forest data have also been captured by interpretation and measurements of stereo air photos in photogrammetric work stations (Åge 1985; Ståhl 1992). To support forest owners in their decision making, several computer programs, e.g. the Heureka system developed at the Swedish University of Agricultural Sciences (SLU) (Wikström et al. 2011), are available. These systems require a starting state of the forest stands across the holding. For large holdings, computer aided planning tools are sometimes based on a sample of objectively surveyed stands.

Typical practices have been to make a new stand delineation and assessment of forest characteristics in the stands every 10-30 years. During such intervals, a significant amount of growth and mortality will occur, and some stands will be thinned or final felled before the new inventory cycle starts. At a national level about 1% of the forest area is final felled and about 2% is thinned yearly (Nilsson et al. 2015). Traditionally, forest managers have handled growth and other changes by updating the stand registers using various models. Growth models are used to account for the growth, based on Site Index (SI) and other input variables, as well as any registered management actions during the update period (Wikström et al. 2011). If data would be acquired more frequently better management decisions could be made, but this must be balanced with the cost of acquiring data (Ståhl et al. 1994; Holopainen and Talvitie 2006; Kangas et al. 2018b). In practical forestry, old data have usually been discarded once new data are acquired, and updating relies mainly on growth forecast models. Methods that could utilize new data in combination with the forecasted existing data would contribute to increased accuracy and cost efficiency of forest inventories.

1.1 Remote sensing data commonly available for data assimilation

An increasing amount of remote sensing datasets are currently being supplied, from several sensors that provide data useful for prediction of variables of interest to forestry, such as Lorey's mean height (H_L), basal area (BA) and volume. Airborne Laser Scanning (ALS) has led to a breakthrough in data acquisition for many of the variables needed for forest management planning (Næsset et al. 2004). However, ALS data are typically made available only at long intervals because of the acquisition costs. In Sweden, new ALS data from Lantmäteriet (the Swedish National Land Survey) are currently expected every seven years (Lantmäteriet 2019). Other kinds of remote sensing data are made available more frequently, but they are not as good as ALS data for predicting forest stand characteristics. The remote sensing sensors used in this thesis were ALS, Digital Photogrammetry (DP) from aerial photographs, Interferometric Synthetic Aperture Radar (InSAR) from the TanDEM-X satellite constellation, and Optical satellite data (OS).

ALS provides measurements related to canopy height and stand density (e.g. Nilsson et al. 2017). However, tree species composition information is difficult to obtain through the common Area-Based Approach (ABA) (Næsset 2002) with the ALS-scanners currently in operation. Studies show that predictions of deciduous dominated stands in a mixed coniferous/deciduous forest landscape tend to be biased if scanning is made during seasons with leaf-on conditions (Naesset 2005; Bohlin et al. 2017; Nilsson et al. 2017). However, if the point density of the ALS data is high, crown shapes can give information about tree species (Holmgren and Persson 2004), but 1-2 returns/m^2, used in the current nationwide ALS campaign in Sweden, is too low for single tree detection (Kaartinen et al. 2012).

Acquisition of aerial photos has a long history in many countries, including Sweden, and manual interpretation of orthophotos is still a very important component of forest mapping. Government-sponsored programs make data available at relatively low cost or even freely. New images with 16 x 16 cm pixels are acquired every second year for southern Sweden and the northern coastal parts of Sweden (Lantmäteriet 2019) whereas the remaining parts of Northern Sweden are photographed more seldom and with larger pixels. Digital Photogrammetric processing of images, taken with stereo overlap, can provide three dimensional (3D) point clouds that describe the upper part of the canopy (St-Onge et al. 2008; Bohlin et al. 2012). The ground level must be known to make predictions from DP; such information can be obtained from Digital Elevation Models (DEMs), which are nowadays very accurate through the use of laser scanning. Forest attributes can be predicted from DP point clouds in a similar area-based fashion as from ALS point clouds. For example, H_L is successfully predicted, while density related forest variables such as BA are predicted less accurately, compared to ALS-based predictions (Bohlin et al. 2012, 2017; Vastaranta et al. 2013; Rahlf et al. 2014; Yu et al. 2015; Ali-Sisto and Packalen 2017).

Forest variables can also be predicted from Optical satellite data (OS); examples are the nationwide maps produced in Finland and Sweden that have been trained with National Forest Inventory (NFI) plot data (Reese et al. 2003; Tomppo et al. 2008). Whereas OS data are not as good as ALS- and DP-data for predicting characteristics such as BA and H_L, perhaps the most important remaining roles for OS images in forestry are assessment of tree species information (Reese et al. 2003) and detection of changes (Kennedy et al. 2010). A plethora of OS platforms is at hand; these satellites produces images with pixel sizes ranging from kilometers down to sub-meter.

InSAR is a technique to derive 3D Digital Surface Models (DSM) from the phase shifts in pairs of radar images. A major benefit of radar remote sensing techniques is their independence of cloud cover, which is a substantial obstacle for many of the other sensors. The InSAR data in this thesis were derived from the satellite constellation TanDEM-X, acquiring data at X-band, i.e. 3,7 cm wavelength (Moreira et al. 2004).

Many additional sources of remote sensing data are available, but they did not fit into the scope of this thesis. Drones that can carry sensors (Puliti et al. 2015), and equipment mounted on backpacks or cellphones can measure trees from the ground (Liang et al. 2018). A noteworthy development is the use of harvester machine data to update stand registers after partial harvests such as thinning (Hannrup et al. 2015). Table 1 presents an overview of important sensors available for providing data useful for predicting forest characteristics.

Table 1. A selection of repeated sensor acquisitions readily available over Sweden. Lantmäteriet collects airborne laser scanning (ALS) and digital photogrammetry (DP) data.

Sensor category	Sensor	Frequency	Comment
OS	Sentinel 2	5 days	The revisit time for a given point at
OS	Landsat 8	16 days	northern latitudes is more frequent, but most images are disturbed by clouds.
OS	WorldView-2		Images can be ordered on request
OS	Planet		Images can be received several times per week.
ALS	ALS	7 years	Revisit time for the current nation-wide ALS program in Sweden.
DP	Aerial photographs	2 years	2 years refer to southern Sweden and the coastal part of northern Sweden
InSAR	TanDEM-X	11 days	TanDEM-X is not a long term operational program.

Table 2 presents the typical accuracies of volume assessment reported in some other studies, focusing on the sensors utilized in this thesis. All the studies regards Nordic conditions.

Table 2. Reported accuracies of volume predictions in Nordic conditions, from sensors included in this work. Relative Root Mean Squared Errors (rRMSE) is given in percent of the mean value of the true volume in the validation material.

Sensor type	Reported rRMSEs	Validation	Reference
OS	59–69%	Plots	(Reese et al. 2003)
	33%	Stand	
	50–56%	Small stands	(Hyyppä et al. 2000)
ALS	19%	Plots	(Rahlf et al. 2014)
	12%	Stands	
	16–17%	Large plots	(Yu et al. 2015)
	19-25%	Plots	(Nilsson et al. 2017)
	17-22%	Stands	
	21-25%	Plots	(Persson and Fransson
	14-16%	Stands	2016)*
DP	31%	Plots	(Rahlf et al. 2014)
	18%	Stands	
	29-33%	Plots	(Bohlin et al. 2017)
	22-23	Stands	
InSAR	42%	Plots	(Rahlf et al. 2014)
	18%	Stands	
	22%	Large plots	(Yu et al. 2015)
	26-29%	Plots	(Persson and Fransson
	15-17%	Stands	2016)*

*Aboveground biomass

1.2 Data assimilation

There is a frequent flow of data from various remote sensing sources (Table 1), providing information about forests as often as several times a week. All predictions from remote sensing data contain noise, but if many measurements are combined the noise could, in theory, be reduced. However, since the forest state may change between remote sensing data acquisitions, updates to account for growth and management operations must also be made. Data assimilation (DA) is a group of methods to combine predictions across time in cases where the system evolves over time (Kalman 1960). DA has reached extensive use in several fields, not least in meteorology where the weather forecasts are improved by DA (Ghil and Malanotte-Rizzoli 1991). A schematic figure of the problem that DA tries to solve is given in Figure 1. At each time point, there are noisy data about the state available through measurements, and we want to use the time series of measurements, together with updates, in a fashion so that maximum accuracy is obtained for our state estimates ("results") at each time point. The updating step accounts for changes in the system between the time points where measurements are made, so that the old data can be used in combination with the new data to improve the accuracy of the estimates.

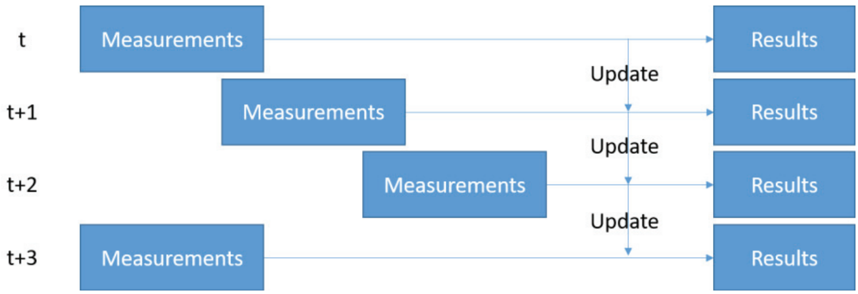

Figure 1. The problem formulation (adapted from Kangas (1991)) using a repeated supply of data and updates to obtain best possible results at each time point.

When applying DA to improve forest stand databases, a forecast model could be used to update existing state predictions until a new remote sensing prediction is available. A new prediction could then be obtained using a combined estimate from both sources of information. The idea is that the merged prediction should be more accurate than each of the individual predictions that are combined. Reducing error by applying DA, caused by forecasting or in the starting forest state prediction, can bring advantages for forest management planning (Saad et al. 2017).

Remote sensing-predictions have their strengths and weaknesses, and combining data from several sensors can bring advantages (Xu et al. 2015). H_L correlates weakly with OS data, and thus the addition of height-related metrics from other sensors than OS is beneficial when predicting variables such as volume and biomass (Magnusson and Fransson 2005). On the other hand, OS metrics contain information that is lacking in other types of remote sensing data. In particular leaf-on ALS data can benefit from the combination with OS metrics (Wallerman and Holmgren 2007; Xu et al. 2015). Leaf-off ALS data suffer less from discrepancies in volume predictions depending on deciduous or coniferous forest (Naesset 2005; Nilsson et al. 2017), but OS data can still improve the accuracy of volume predictions (Kukkonen et al. 2018) and probably of other forest variables as well.

If the properties are similar for a set of remote sensing metrics, concerning sensor and season, a given forest area (e.g. a sample plot) will tend to have errors of the same sign and magnitude. For example, two consecutive ALS predictions over a deciduous sample plot will probably have a positive prediction error in both cases; thus, the prediction errors are likely to be correlated.

A large number of studies have successfully assessed changes in forested landscapes across time (Kennedy et al. 2010; Xu et al. 2015; Deutscher et al. 2017; Grabska et al. 2020). Several satellite sensors, e.g. OS and some InSAR sensors, will provide frequent data at hand for change detection. Site Index and growth have successfully been predicted based on height from remote sensing metrics in combination with age from either long time series of OS data (Lefsky et al. 2002), age from stand registers (Holopainen et al. 2010), or long time series of height data (Véga and St-Onge 2009; Noordermeer 2020). Update through models to account for growth and changes is common practice in Nordic forestry, but updating through models accumulates errors as time passes.

Some earlier studies have advocated DA approaches for forest inventory, such as the Kalman filter for making use of old inventory data to improve precision in new inventories (Dixon and Howitt 1979). Kangas et al. (1991) proposed the use of prior information and growth updates as an estimator for forest inventories, with field-based inventories in mind. Other notable examples are Czaplewski (1990), Czaplewski and Thompson (2008) and Ståhl et al. (1994) who suggested DA in field-based forest inventory contexts.

Ehlers et al. (2013) demonstrated the potential of DA for estimating stand level characteristics in a simulation study, where a time-series of remote sensing-data were utilized together with a growth model. In this study, DA with an Extended Kalman Filter or a Bayesian filter improved the accuracy of predictions compared to discarding old data as new were made available. The simulation study involved new acquisitions every two or five years throughout a 50 years period. Vastaranta (2018) evaluated an approach similar to DA with two predictions of biomass from stereo satellite point clouds: a two years old prediction was updated by a forecast model and combined with a new prediction. This resulted in a gain in accuracy, and a reduction of bias. Series of remote sensing data can also be used for improving the growth part of a DA framework. A method for updating forest variables in a DA framework using best linear unbiased prediction (BLUP) was proposed by Hou et al. (2019). However, this study left out forecasting and worked with data from a single season.

Use of forecasted old information in combination with new data might also be a useful method for improving estimates for large areas, for example in the context of NFIs. In the Finnish multi-source NFI a combination of NFI plot data and optical satellite data have been used for enabling forest statistics to be presented on municipality level. The variation between estimates form different years is, however, quite large. Katila and Heikkinen (2020) showed that more stable estimates could be obtained by modeling trends and combining estimates from three different years. In a simulation study, Kangas et al. (2020) showed the benefit of using Kalman filtering for combining new and old NFI data even if good auxiliary data is available. One of their observations was the necessity to find and compensate for changes like cuttings.

The work in this thesis further developed the application of DA based on remote sensing data, taking a few steps towards practical implementation of DA routines for practical forestry.

2. Aims and objectives

The overall aim of this thesis was to investigate the potential of DA for predicting core forest inventory variables using the extended Kalman filter, and variations of this filter, based on time series of remote sensing data. Specifically, the objectives of the included papers were:

1. To make a first evaluation of the usefulness of DA for forest inventory purposes based on a time series of real remote sensing data. In this study, DP predictions were made for large sample plots.
2. To assess the potential of the Extended Kalman Filter applied to a time series of InSAR data (TanDEM-X) over sample plots estimating H_L, BA and volume. The study was based on a dense time series of remote sensing data across a short period.
3. To investigate the magnitude of correlated prediction errors when basing predictions on different sources of remote sensing data. The sensors applied were ALS, OS and InSAR.
4. To evaluate if classical calibration has a potential to reduce error correlations caused by the tendency of regression models to overestimate low true values and underestimate high true values, which reduces the efficiency of DA. The study was conducted through simulations.
5. To assess the efficiency of DA when applied to a time series of data from a mix of remote sensing sensors, accounting for error correlations in the assimilation method through a modified filter. The sensors used were ALS, DP, InSAR and OS and the validation was done at the level of sample plots.

3. Materials and methods

3.1 Study area

The test site for the studies in this thesis was the Remningstorp research estate in southwestern Sweden (58°27′N, 13°39′E, Figure 2). The estate covers roughly 1500 ha of mainly highly productive managed forest. The forestry practice at the estate follows a cycle common in the Nordic countries, with planting, pre-commercial thinning, thinning and final felling. Norway spruce *(Picea abies)* is the dominating species at the estate, followed by Scots pine (*Pinus sylvestris*) and deciduous forests consisting of birch (*Betula spp*) with some oak as well as other deciduous species. Field measurements have been carried out on this research estate for a long time.

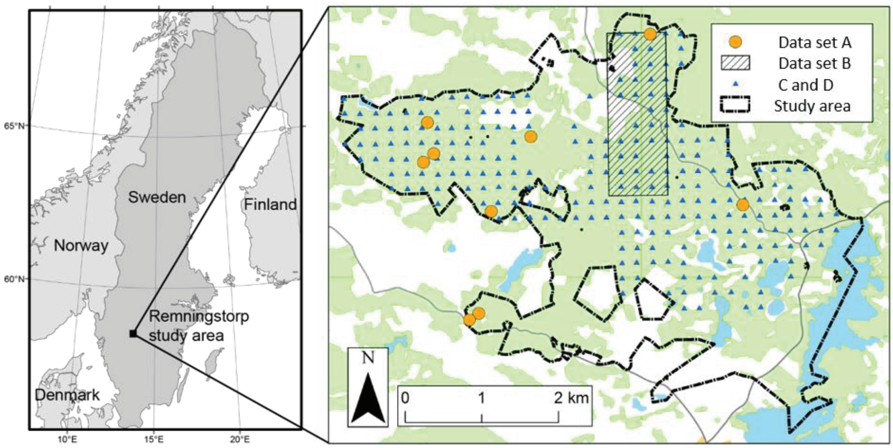

Figure 2. The study area for the thesis was the Remningstorp forest estate in southwestern Sweden. The map to the right shows the location of the sample plots (see Table 3).

3.2 Sample plot datasets

Field sample plots from several different campaigns were used for the studies in this thesis, each with somewhat different designs but all sharing a similar measurement protocol at the sample plot level. All trees above a threshold diameter were calipered and the stem diameter was registered together with the tree species. Heights were measured for a subsample of the trees; height for the remainder of the trees was predicted with models based on the relationship between diameter and height. The subsample trees were also cored for age. Site Index (SI) was assessed according to the Swedish system for site index assessment based on vegetation characteristics (Hägglund and Lundmark 1977). Plot level forest variables (H_L, BA, Volume, Mean Diameter, tree species composition, basal area weighted age) were computed, and in case of BA and Volume aggregated to per hectare units. The sample plot data sets are presented in table 3.

Table 3. An overview of the sample plot data used in the studies in this thesis. Number of plots that were used depends on the study as well as the time point within each study.

Year	Dataset	Radius	Site	Papers	No. Plots in study
2011-2013	A	40 m	Remningstorp	I	9
2004	B	10 m	Remningstorp	I	258-416
2010	C	10 m	Remningstorp	I, II, III, IV,V	117-214
2014	D	10 m	Remningstorp	II, III, IV,V	117-148
1988-2010	E	10 m	NFI	I,II,V	15131

Dataset A consists of large sample plots with 40m radius, measured in 2011-2013, all placed subjectively in homogenous stand areas in different types of forest. All trees above 4 cm diameter at breast-height were calipered.

A centrally located area of the estate was sampled with a dense grid of sample plots in 2004 (dataset B). In total 849 plots were measured, with 10 m radius. In 2010, a second sample of 10 m radius plots was inventoried across the major part of the property, spread out across the entire estate in a regular grid, initially measuring 263 plots (dataset C). These plots were measured again in 2014 (dataset D).

Permanent NFI sample plots, spread out across Sweden, were used for developing growth forecast models. In the Swedish NFI sample design, there are both temporary plots, that are never revisited, and permanent plots, that are measured every five years. The measurement protocol of the NFI resembles the protocol used at the Remningstorp test site (Fridman et al. 2014).

All studies were limited to forests with normal growth patterns, i.e. plots with disruptions due to management actions such as harvest or major damages were left out. This reduced the number of plots available for the studies.

3.3 Remote sensing data used in the studies

ALS data were acquired for scientific purposes in 2010 and 2014 during leaf-on season. In addition Lantmäteriet, acquired data in 2011 during leaf-off conditions (Lantmäteriet 2019). Processing of the ALS point clouds involved flight line overlap and removal of isolated points before the point cloud was height-normalized to height above ground.

DP data were acquired from Lantmäteriet (2019). Images were delivered with camera orientation parameters, and processing into point clouds was done using the SURE software (Rothermel and Wenzel 2012). Area-based metrics were calculated using FUSION (McGaughey 2014) for both DP and ALS.

The optical satellite data in this thesis were obtained from the now decommissioned SPOT-5 satellite, which provides images with 10 x 10 m pixel size. Until the de-orbiting of SPOT-5 in 2015, Lantmäteriet made images from it available for the public in a geo-corrected format, with one image per year covering Sweden (Lantmäteriet 2020). Metrics over the sample plots were extracted from the rasters at the sample plots locations.

The TanDEM-X mission has scientific purposes and thus its settings vary (Moreira et al. 2004; Krieger et al. 2007), which means that the InSAR products have varying suitability for forest variable estimation (Soja and Ulander 2013). DSM rasters were generated with standard InSAR processing and the heights were normalized with an ALS-based DEM (Persson and Fransson 2016). Remote sensing data were extracted for areas corresponding to the field sample plots and regression-based prediction models were developed, linking remote sensing metrics with field measurements.

An overview of the remotely sensed data applied in the thesis is given in Table 4.

Table 4. Remote sensing acquisitions used in this thesis

Sensor type	Date	Included in papers	Sensor
DP	2003-10-13	I	Z/I DMC01
	2005-06-28	I	
	2007-05-26/ 2007-06-03	I	
	2009-09-01	I	
	2010-05-02	I	
	2012-05-23	I, V	
	2014-07-26	V	Vexcel UltraCam-Xp WA
ALS	2010-08-29	III, V	TopEye s/N 700
	2011-04-21	III	Leica ALS60/23
	2014-09-14	III, V	Riegl LMS Q680i
OS	2010-06-04	III	SPOT-5 HRG
	2011-06-06	III, V	
	2013-07-17	III, V	
InSAR	2011-06-04	II, III	TanDEM-X
	2012-02-01	II	
	2012-02-12	II	
	2012-02-23	II, V	
	2012-06-01	II, III	
	2012-08-28	II,V	
	2012-11-09	II	
	2013-02-27	II	
	2013-03-21	II	
	2013-07-02	II	
	2013-07-24	II	
	2013-08-04	II	
	2013-08-11	II	
	2013-09-13	II	
	2013-11-18	II , V	
	2014-03-08	II	
	2014-06-08	II, III	
	2014-06-26	II	
	2014-08-02	II	

3.4 Data assimilation: The extended Kalman filter

A large set of DA methods is available, with various approaches to the problem of merging information from several sources. In this thesis I focus on the well-known Extended Kalman Filter (EKF; Kalman 1960; Kalman and Bucy 1961). The Kalman filter is an iterative method that consists of two major parts: Updating and Merging, repeatedly applied to the time series of measurements. The EKF was applied as in the study by Ehlers et al. (2013) in papers I, II and V (papers III and IV omit the updating step). This involves the following steps:

1. At the starting point of the time series, **initial predictions** are made from remote sensing data and field reference data. Along with the predictions, there are also estimates of uncertainty.
2. The predictions are **updated** through a growth forecast model until the next time point when remote sensing data are available. The variance of the updated predictions is estimated along with the predicted values.
3. Predictions from the new remote sensing data are made. The forecasted and the new predictions are **merged** with weights assigned inversely proportional to the variance of the predictions.
4. The growth forecast model is applied, starting from the merged predictions of step 3. The uncertainty measures are updated along with the predicted values.
5. Yet another set of remote sensing-predictions is available, and a new merger can take place as described in step 3.

The steps 4 and 5 are repeated, as new data sets are made available.
The merger is a weighted average, with weights inversely proportional to the respective variance of the two predictions. The weighted average (at time point *t2*) is:

(1) $$\hat{y}_{i,t2,DA} = k_{i,t2}\hat{y}_{i,t2,u} + \left(1 - k_{i,t2}\right)\hat{y}_{i,t2}$$

Where $\hat{y}_{i,t2,DA}$ is the merged prediction for time point two for sample plot *i* based on the updated prediction from time step one, $\hat{y}_{i,t2,u}$ and the new prediction at time point two, $\hat{y}_{i,t2}$.

The weight, $k_{i,t2}$, for EKF is:

(2) $$k_{i,t2} = \frac{var(\hat{y}_{i,t2})}{var(\hat{y}_{i,t2,u}) + var(\hat{y}_{i,t2})}$$

The variance of the merged prediction is:

(3) $$var(\hat{y}_{i,t2,DA}) = k_{i,t2}^2 var(\hat{y}_{i,t2,u}) + (1 - k_{i,t2})^2 var(\hat{y}_{i,t2})$$

During the update step, the variance must also be updated along with the target variable, as it is required as an input in the next time step. Thus the forecast model must be of a type that allows the variance of the prediction to be updated. In the studies, a non-linear growth function was applied, of similar kind as in many other fields of science and real world applications, and thus the EKF rather than the standard Kalman filter had to be used. In EKF the updating function is linearized through Taylor Expansion to approximate the variance of the predictions after the update. This approximation works best if the linearized function is relatively close to being linear.

The standard Kalman filter and the EKF further stipulate that errors of predictions and forecasts have a zero mean, i.e. that predictions and forecasts are unbiased. Further, in applications variances normally need to be estimated from data and thus the variances in (2) and (3) need to be replaced by variance estimators, in which case (3) is an approximation.

3.5 Growth forecast models

The same set of growth forecast models were used for studies I, II and V; the models are described in the appendix to paper I. Studies III and IV did not involve any forecasting steps. Separate growth models were estimated for five classes according to dominating tree species on the sample plot, namely spruce, pine, mixed coniferous, deciduous and mixed coniferous/deciduous forest. The forecast models were of the form:

(4) $$g_{i,t} = exp(b_0 + b_1 x_{i,1,t} + \cdots + b_j x_{i,j,t}) + e_i$$

Where $g_{i,t}$ is net growth for five growing seasons, and $x_{i,j,t}$ is the j^{th} predictor variable for plot i at time point t Models were developed using sample plots from the Swedish NFI (dataset E in Table 3).

We used SI, age and tree species proportions from field data for all papers and assumed them to be known without errors to simplify calculations. An input starting value for the forecast period was, however, taken from the predictions.

The notation used in this summary part of the thesis corresponds to the notation used in the last paper. The other papers use different notation, and the gradually changing notation throughout the course of the studies can be seen as a receipt of a gradually increased understanding of what is required for DA to work for merging remote sensing-based predictions of plot and stand level forest characteristics.

3.6 Prediction from remote sensing data

All papers followed an empirical tradition and the ABA approach to predict forest attributes. In the empirical tradition the variable of interest is related to the measurements through a model, in this case regression models of somewhat varying type. Selection of variables as predictors in the models was based on conceptual understanding of the relations in the data and by studying residual plots of preliminary and final models, fit through regression analysis. Slightly different methods of regression analysis were used in the different papers.

A well-known property of predictions from regression analysis is the central tendency, which implies that high true values are underestimated and low true values overestimated. Regression models are specified to deliver unbiased predictions conditional on the predictor variables (the remote sensing metrics in this thesis), but not conditional on the true values. This is a problem in DA applications, since predictions ideally should be unbiased conditional on the true value for every unit predicted. Classical calibration is a means to obtain approximately unbiased predictions throughout the entire range of true values (Osborne 1991). The first step of classical calibration is to specify an error characterisation model, which describes the relationship between a sample plot's true value $y_{i,t}$ and the remote sensing-based prediction $\hat{y}_{i,t}$ (Tian et al. 2016; Persson and Ståhl 2020). The error characterisation model was assumed to be linear:

(5) $$\hat{y}_{i,t} = A_t + B_t y_{i,t} + \epsilon_{i,t}$$

Here, A_t and B_t are parameters, and $\epsilon_{i,t}$ a random error term with zero expectation. Estimating the parameters and rearranging the terms leads to the formula for calibrated predicted values, $\hat{y}_{i,t,c}$:

(6) $$\hat{y}_{i,t,c} = \frac{(\hat{y}_{i,t} - \hat{A}_t)}{\hat{B}_t}$$

Note that calibration works only of $\hat{B}_t$is non-zero. Since $\hat{B}_t$ typically is smaller than 1, the variance of the calibrated prediction increases compared to the variance of the original prediction.

3.7 Validation

Bias and RMSE, and RMSE relative to the mean of the true variable of interest (rRMSE) were used to evaluate the performance of the different DA filters in the studies. They were defined as:

(7) $$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} r_{i,t}^2}$$

(8) $$rRMSE = \frac{RMSE}{\bar{y}_t} * 100$$

(9) $$Bias = \frac{1}{n}\sum_{i=1}^{n} r_{i,t}$$

(10) $$rBias = \frac{Bias}{\bar{y}_t} * 100$$

Here $\bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$, and $r_{i,t} = y_{i,t} - \hat{y}_{i,t}$, for the variable of interest on sample plot *i* at time point *t*.

3.8 Materials and methods for the individual articles

3.8.1 Paper I

To our knowledge, this is the first empirical study of DA applied to a time series of remote sensing predictions of forest inventory variables. Six DP acquisitions were included. The first prediction was from 2003 and the last from 2012, covering eight growth periods. The aerial images were processed into 3D point clouds. A standard ABA processing was used, in which the metrics from the point cloud were calculated for both the plots and grid-cells (pixels) equally large as the sample plots. Non-linear prediction models were developed, predicting H_L, BA and volume at plot level (sample plot datasets B and C forecasted to the acquisition year). The prediction models were also applied to the grid-cells. EKF was then used to assimilate the grid-cell level predictions within the validation stands, for each pixel separately. At the time point for the final validation (2012) the grid-cell values were aggregated into stand-wise mean values and compared with the volume measured in the field (sample plot dataset A). The EKF method was compared with the last single prediction, and with the first prediction (from 2003) updated with growth forecasts to the state of 2012, without DA.

3.8.2 Paper II

In paper II, we investigated the use of a dense time series of InSAR data for DA. Nineteen predictions over four growth seasons, from 2010 to 2014, were assimilated using EKF. We used sample plots of 10 m radius (sample plots C and D), rather than stands as in paper I. Field plot values for the time in between the inventory years were interpolated. Plots were excluded in case of harvest or severe damage during the study period. To exclude a plot, we set a threshold of 10% maximum loss of the BA from 2010 to the re-inventory in 2014.

Each pair of images was processed using a standard interferometric processing, which gives a height map that is normalized to height above ground using an ALS-based DEM (Lantmäteriet 2019). Linear regression models predicting H_L, BA and Volume were estimated and evaluated with Leave-One-Out Cross-Validation (LOOCV).

3.8.3 Paper III

This study investigated the correlation of errors across time points for remote sensing-based predictions. Error correlation between predictions at different time points reduces the efficiency of the EKF, and thus the variance of the assimilated prediction decreases at a slower rate. This article investigated the size of the correlations and gave an example of the effect of overlooking correlations in the filter.

Acquisitions from three sensors were used in this study: ALS, OS and InSAR, with three acquisitions each, covering the test site at Remningstorp at the time in between the field inventory campaigns in 2010 and 2014 (sample plot datasets C and D, interpolated). Similar as in paper II, plots with decreasing BA (due to harvest or other disturbance) were removed, like forests younger than 20 years. This left 117 plots for model development for prediction of H_L BA, volume, and basal area weighted mean diameter.

Correlation coefficients between predictions from all pairs of sensors were calculated. This study also demonstrated the effects of correlated errors in a simulated example, where weighting was performed according to:

(11) $$k_{i,t2} = \frac{var(\hat{y}_{i,t2}) - cov(\hat{y}_{i,t2,u}, \hat{y}_{i,t2})}{var(\hat{y}_{i,t2,u}) + var(\hat{y}_{i,t2}) - 2cov(\hat{y}_{i,t2,u}, \hat{y}_{i,t2})},$$

which includes the covariance. The variance of the assimilated prediction will then be:

(12) $$var(\hat{y}_{i,t2,DA}) = k_{i,t2}^2 var(\hat{y}_{i,t2,u}) + (1 - k_{i,t2})^2 var(\hat{y}_{i,t2})$$
$$+2k_{i,t2}(1 - k_{i,t2}) cov(\hat{y}_{i,t2,u}, \hat{y}_{i,t2})$$

3.8.4 Paper IV

This study evaluates the importance of calibrating remote sensing-based predictions before utilising them in DA. A population of forest plots was simulated as well as remote sensing metrics obtained from two different types of sensors (resembling ALS and OS). The EKF was applied either with, or without, calibration of remote sensing-based predictions. The variable studied was volume and the population comprised 10 000 sample plots.

Ten acquisitions with each sensor were simulated. The simulated metrics were used for predicting volume through ordinary least-squares regression.

Thus, we had two types of sensors and calibrated as well as non-calibrated data for each plot, summing up to four types of predictions.

EKF was used to assimilate the data for three different cases (time series): (i-ii) ten predictions with each of the sensors separately and (iii) an initial ALS-based prediction followed by nine OS-based predictions.

3.8.5 Paper V

In paper V, predictions from various remote sensing sensors and time points were assimilated: two ALS, two DP, three InSAR and two OS acquisitions over the four growth seasons between 2010 and 2014. The time series started and ended with an ALS-acquisition. The study variable was volume per hectare. Datasets C and D were used, interpolating field reference values in between 2010 and 2014. We applied classical calibration to the predicted values and compared with non-calibrated predictions. The predictions were assimilated using EKF and a new filter, which we name EKFm, which takes error correlations into account. Weights were used to merge the predictions in the same way as in EKF but in calculating weights the covariance between predictions was taken into account, like in paper III (eq. 11). The EKFm filter was compared to the standard EKF, using either calibrated or non-calibrated predictions as input. Thus, we compared four cases of DA. We also compared the DA predictions with the predictions from individual time points and with the first ALS prediction, updated only with a growth forecast model across the study period.

4. Results and discussion

DA requires adaption to the specific field of application. In meteorology, DA has been adapted and applied for several decades, a work that has only recently started for DA in forest inventories. This thesis has contributed to some important development and identified areas that need to be investigated further.

Paper I presents, to our knowledge, the first empirical study where DA is applied together with remote sensing-based predictions of common forest variables. Paper I concluded that the theoretical gain from DA was far from being realised in practice, despite the substantial potential of DA demonstrated in an earlier simulation study (Ehlers et al. 2013). Assimilation of six acquisitions over eight growth periods resulted in only a small gain in accuracy for all variables, compared with the last prediction, but a notable gain for Volume and BA when compared with the forecast from the first prediction. H_L was almost equally accurate irrespective of method (Table 5).

Table 5. Results from paper I, with EKF, last prediction and forecast from first prediction

	rRMSE		
	Volume	**BA**	**H_L**
DA	13.5%	12.0%	9.3%
Last prediction	15%	12.8%	9.6%
Forecast from first prediction	19.7%	14.2%	9.5%

The validation in paper I was done on dataset A (Table 3), measured once in the field at the end of the time series. The results indicate the strength of DP to predict H_L at a single time point, as well as the relatively high accuracy of forecast models for H_L. In cases where predictions and forecasts are already sufficiently accurate, there seems to be little practical gain in including new data.

EKF of 19 InSAR predictions of H_L, BA and volume was investigated in paper II. A difference compared to paper I was that the validation method was LOOCV. Using LOOCV meant that the sampling distribution of the validation data did not differ from the training data, as in paper I. This assured that the overall bias was small. Further, validation was made on sample plots, i.e. the same type of data was used for estimating and evaluating models. In addition, validations after each assimilation iteration made it possible to analyse the results at all time points. Results showed that the rRMSE dropped during the first few iterations for all variables (Figure 3). Data assimilation with EKF consistently improved the predictions of H_L, compared to the individual predictions. For BA and volume, the assimilated predictions were better or roughly equal to the predictions based on individual data acquisitions at most time points. Following a short initial rRMSE decrease, however, the trend turned to being increasing. Since the weight for new InSAR acquisitions was very low after the first few iterations cf. eq. 2), the results of DA turned out to be almost like simple growth forecasts of the merged data after the first few iterations.

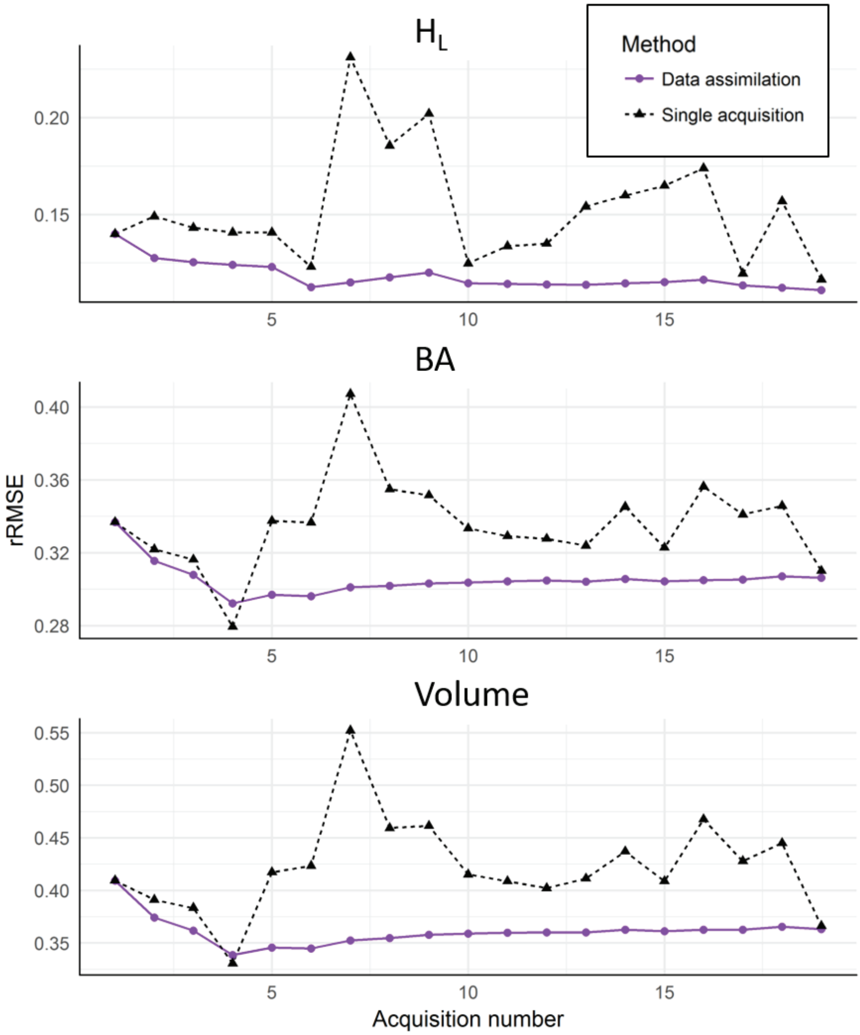

Figure 3 Results from paper III, for predictions of H_L, BA and volume. Purple lines represent the rRMSE of assimilated predictions and black dashed lines the rRMSE of predictions from single acquisitions.

Even though EKF gave an improved accuracy the potential of DA, as previously shown for simulated data (Ehlers et al. 2013), was not fully realized in papers I and II. The theory for standard EKF assumes uncorrelated prediction errors between predictions, for the weighting to be optimal. If this does not hold the variance of the EKF prediction will be underestimated. As the next iteration of EKF then is based on an underestimated variance for the next computation of weights, the updated prediction will be given a too high weight in comparison with the new prediction to be assimilated (cf. Figure 4). The results from paper II suggests that many non-optimal iterations of EKF notably decrease the accuracy. A very important part of DA is therefore to assess the accuracy of predictions properly. The importance of investigating the error correlation across predictions based on different remote sensing acquisitions was an evident insight from the results of paper I and II.

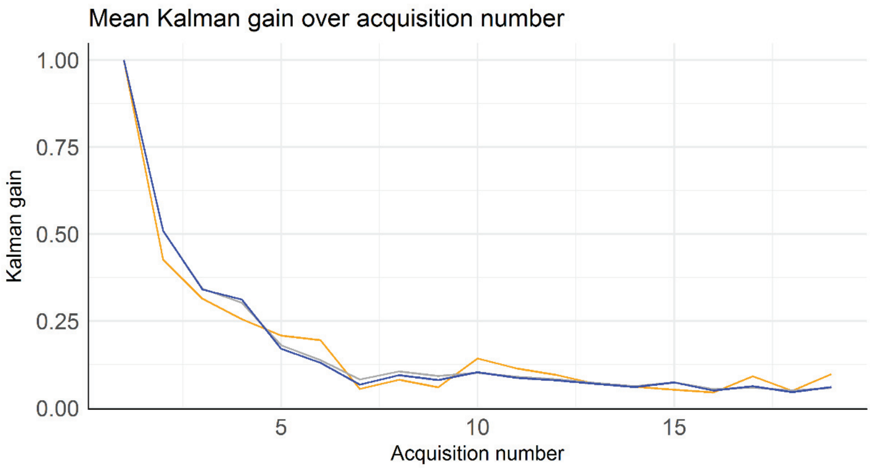

Figure 4. The weight, i.e. the Kalman gain, to new data in the merged prediction across iterations, for the InSAR predictions in paper II. EKF gives a very low weight to new predictions after a few iterations. The yellow line shows Kalman gain for H_L, the grey for BA, and the blue for volume.

The study by Vastaranta et al (2018) have similarities to mainly paper I and II. World View-2 stereo images were processed into 3D point clouds, from which predictions of biomass were derived at two time points with two years in between The first prediction time point were updated with forecast models to the second time point, and a simple average were used to combine the two predictions. This procedure, which resembles a DA method, improved rRMSE by 0.9%, from 21.8% to 20.9%, and reduced the bias, indicating that the small but present gain of combining data across time may hold outside our test site and studies.

In paper III, we quantified the error correlations of ALS, OS and InSAR predictions. In general, the error correlations were positive and rather strong. They also tended to be stronger between two predictions from the same sensor, while lower correlations were found for two acquisitions from different sensors. In particular, acquisitions from OS had low error correlation with acquisitions from all the other sensors. When it comes to season of the year, it was noticeable that the ALS from 2011, acquired during leaf-off conditions, had lower error correlation with the other ALS acquisitions, acquired during leaf-on conditions.

Paper III also gives examples of the theoretical effect of error correlations on DA results. Ten equally precise predictions were iteratively assimilated. Error correlations were accounted for in the assimilation (eq. 11) and were assumed to be of three different levels. Results after ten iterations, as percent of the initial standard deviation, are presented in Table 6.

Table 6. Results from paper III. The standard deviation after assimilation of ten equally precise predictions were calculated assuming different levels of error correlation.

	Error correlation		
	0	**0.4**	**0.8**
Standard deviation after ten iterations	32%	68%	91%

Error correlations thus impede on the potential of DA, although the results in Table 6 assumes that there are no growth updates between the data acquisitions. For practitioners it will therefore not be very useful to assimilate highly correlated data. Practical data assimilation would also add further error sources, like imperfect error statistics, which would make the estimation of weights uncertain and possibly remove all benefits from assimilating highly correlated data.

Paper IV concerns one of the reasons why prediction errors correlate: the tendency that predictions from regression-based models overestimate small true values and underestimate large true values. A large population of sample plots and remote sensing metrics was simulated to study the effects of this tendency and classical calibration was applied to counteract it. Results showed that regression-based prediction contributes to correlated errors and that this effect can be removed through classical calibration. The reduced correlation of errors improved results from assimilation with EKF (Figure 5).

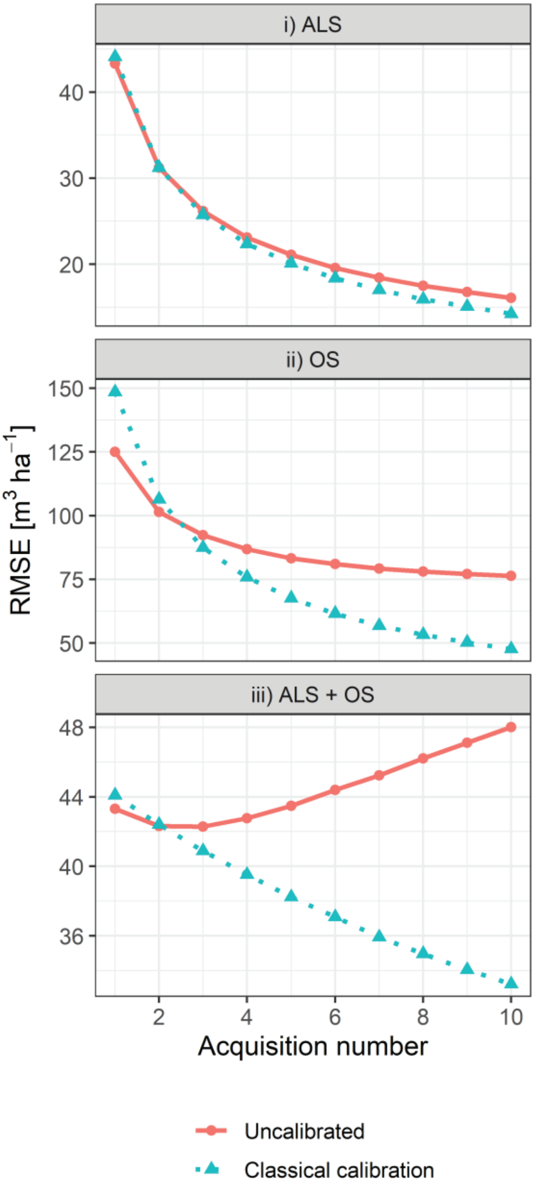

Figure 5. Results from paper IV. RMSE for 10 simulated (i) ALS predictions and (ii) OS predictions, assimilated with EKF. In (iii) the time series starts with an ALS-based prediction, which is followed by 9 OS-based predictions. The DA algorithm uses either non-calibrated or calibrated predictions.

The improvement obtained by reducing the error correlation through calibration is encouraging; however, there are many other reasons for errors to correlate across acquisitions apart from the regression towards the mean. For example, OS data lacks height related information (e.g. Magnusson and Fransson 2005). Another example is DP, where several studies point out the lack of density related metrics (Bohlin et al. 2012, 2017; Ali-Sisto and Packalen 2017); thus sparse forest is likely to be overestimated and the opposite applies for dense forests. Across time, sparse forest will likely repeatedly be overestimated in all predictions from DP, leading to errors that correlate. This happens regardless of whether the data are calibrated, as in paper IV, or not. Such lack of sensitivity is present in predictions from all studied sensors, causing error correlation. Regarding ALS, many studies point out the effect of dominating tree species, especially in deciduous forest during leaf-on conditions (Naesset 2005; Nilsson et al. 2017). InSAR from TanDEM-X has shown to be affected by whether or not the forest is deciduous or coniferous during leaf-on conditions (Soja and Ulander 2013), which has also been noticed for DP data (Bohlin et al. 2017). These effects due to the combination of sensor and forest type make the errors correlated across sensors. Although residuals from predictions based on various sensors can have trends over the same variable, different methods of measuring the forest as well as varying timing when it comes to phenology (season) will have an impact on the correlations. A mix of sensors in DA might therefore be beneficial. An additional reason for correlated errors, which will not be remedied by mixing sensors, is erroneous geolocation of field plots. Such errors may potentially mislead analysts to assert that the error correlations are stronger than they, in fact, are. This follows since the field reference state for a plot is different from the state on the plot for which predictions were made. However, the impact of geolocation errors on the results from the studies included in this thesis were judged to be minor, due to the use of GPS equipment with sub-meter accuracy.

Paper V utilises several of the findings from paper I-IV: the importance of error correlation discovered in paper I-III, along with the classical calibration approach suggested in paper IV. Error correlations requires adaption of the assimilation method, and/or the predictions as in paper IV. In paper III, we also found error correlations to be lower for predictions from different sensors compared to predictions based on the same sensor, and there might thus be an advantage to use several sources of remote sensing data in DA schemes. The logical step in paper V was then to evaluate assimilation of a time series of predictions from various remote sensing sensors. Paper V provides an empirical validation of an adapted filter (EKFm) which accounts for error correlations (eq. 11) and compares it with results from the standard EKF. Predictions with or without classical calibration were assimilated. The results from paper V (Figure 6) showed a steadily improved rRMSE across iterations of DA using EKFm. Such success was not obtained for EKF applied to predictions without calibration, which was the method applied in paper I and II. The modified DA filter (EKFm) combined with calibrated predictions was much more efficient.

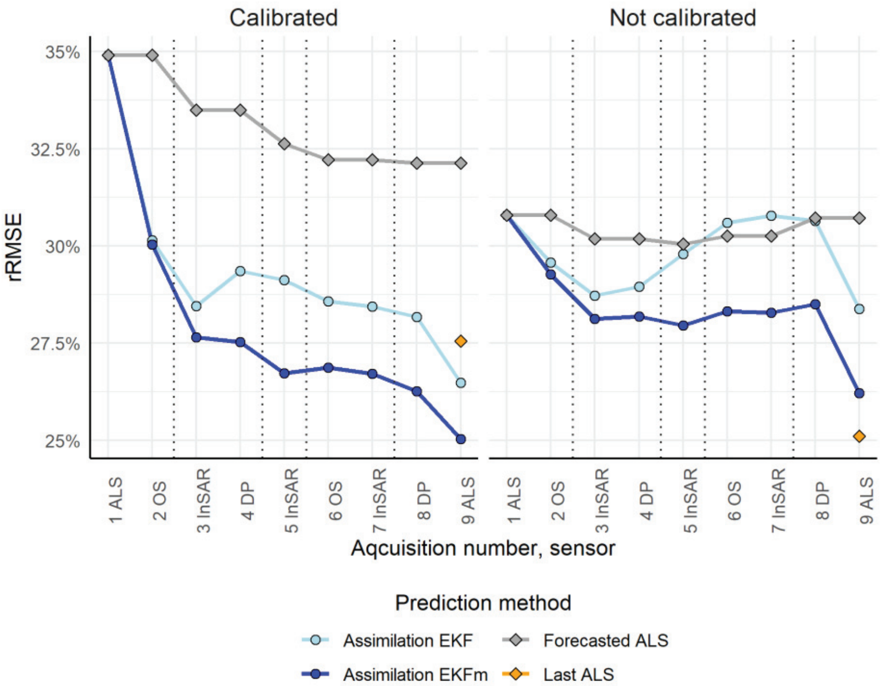

Figure 6. Results from paper V, where nine predictions of volume based on ALS, OS, DP and InSAR where assimilated using EKF or a modified filter (EKFm) that accounts for error correlations. Classical calibration applied to the predictions improved the results.

Paper V also gives an empirical example of the effect of classical calibration on the error correlation. Figure 7 shows the correlation of prediction errors for different combinations of sensors.

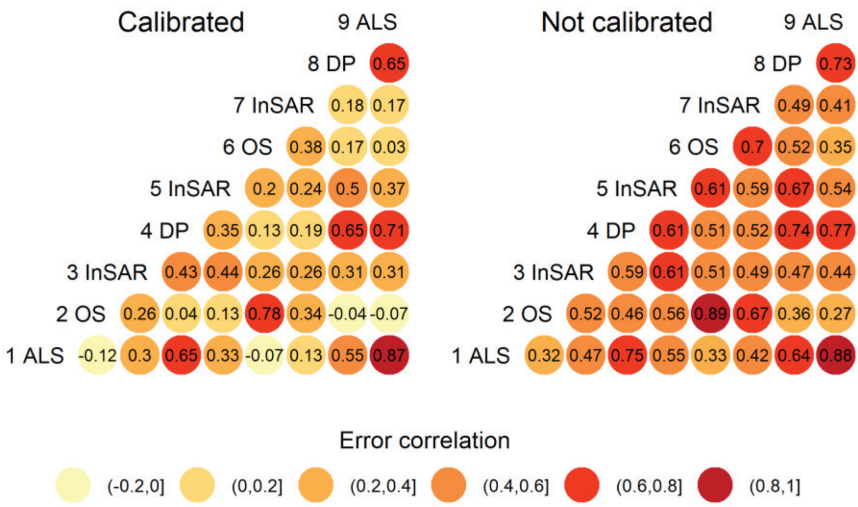

Figure 7. Error correlation coefficients for the predictions of volume in paper V. Number and abbreviated sensor refers to acquisitions in paper V in the order they were assimilated in the study.

The ALS data used in paper V were leaf-on. Leaf-on ALS acquisitions are recognised as less optimal for datasets with both deciduous and coniferous dominated sample plots. In such cases sample plots with deciduous forest tend to be overestimated (Nilsson et al. 2017). The main gain by adding an OS acquisition is reducing the difference between deciduous and coniferous plots, which often gives a large reduction in the overall rRMSE as well. The two OS predictions had the highest rRMSEs of all sensors in the study but still contributed to a large improvement in rRMSE in the DA scheme. This indicates the value of incorporating data of varying kind; even if the data have poor correlation with the target variable, they may be useful in combination with other sources. If the ALS acquisition had been leaf-off, the results would likely have been a less dramatic improvement. Even though the included ALS data were leaf-off, Kukkonen et al. (2018) found that metrics from auxiliary optical data improved the accuracy when added in the same model as ALS. This indicates that a gain could also be achieved if DA is used in a time series including leaf-off data and OS data. Based on the findings in other studies, and the positive effect of assimilating an OS acquisition after a leaf-on ALS scanning, it could have been possible to obtain even better accuracy at the end of the time series in paper V. As the

ALS-scanning was made during leaf-on season, a negative bias is expected for deciduous forests. The OS acquisitions could contribute with some information about tree species, and thus assimilated predictions could potentially be more accurate than ALS alone.

The test site Remningstorp has high growth rates compared to the average conditions in Sweden, and the predicted growth seemed too low already in paper I. This led to positive bias of the forecasts in paper I and V, and also in paper II where, however, bias was not presented in the article. In this thesis, I used classical calibration to reach approximate unbiasedness for remote sensing-based predictions across the entire range of true values, but EKF assumes zero mean for the errors of the growth models as well. Classical calibration could potentially be a way to remove bias also for predictions of growth.

The importance of incorporating error correlation was evident in studies III and V. The assumption was that parts of the random prediction error for a given plot would remain relatively constant. Over longer time periods this approximation is probably not going to hold and further research needs to be conducted to gain knowledge of how error correlations between remote sensing acquisitions should be incorporated in DA systems across longer time periods.

The scenario in paper V is of practical relevance; an ALS based prediction initiates the time series and predictions from other more frequently available sensors follow. An advantage of OS is the high temporal availability of free data (Table 1), but the high residual error correlation between OS predictions (Figure 7) reduces the potential gain (Table 6). It is therefore perhaps only worthwhile to include a selection of OS images in an operational DA method. There is thus a need to develop selection criteria regarding what new acquisitions to include, based on what data are already assimilated. The major advantage of OS is the access of frequent data and thus the possibility to detect changes. Persson et al. (2018) improved tree species classification using four sets of Sentinel-2 data , referring to phenological differences between acquisitions as a reason for the improved results over using data from a single time point. The frequent flow of data from Sentinel-2 might thus be more important for predictions of tree species composition than for volume predictions, but also for detecting changes. Seasonal changes between acquisitions have been suggested as an advantage also for ALS; a study by Räty (2019) predicted volume with data from leaf-off and leaf-on conditions used together, which improved the accuracy. Sensor type is important for error properties, but the forest conditions, including phenology, are important too.

Given the results of paper III, IV and V, we have achieved a better understanding of the results of paper I and II. The method used in paper I and II was standard EKF of predictions without classical calibration, which gave poor results in paper V. In addition, the validation data in paper I were stands, subjectively selected and with another range of the forest variables than the training data. Provided the findings in paper IV, there could have been a gain in calibrating the predictions. Paper V presents high error correlation coefficients for volume predictions from DP (Figure 7). Similar error correlations could be expected also from the predictions in paper I, which likely impeded on the results of DA. The error correlation between predictions from InSAR data varies more than error correlations from other sources of data, probably due to the settings of the TanDEM-X satellites, as well as season and weather conditions at the particular acquisitions. In paper III, all three InSAR acquisitions were from the summer season, while for the three in paper V the season varied, but so did also the satellite settings. Combined experiences from the studies stress the importance of considering error correlation in DA applications, which likely had an impact on the results in both papers I and II.

5. Conclusions

Many studies combine metrics from several sensors with promising results. What distinguishes most of these studies from DA approaches are short time periods, so that growth and changes can be disregarded. The results point at the potential in combining data from different sensors, but growth forecasts and updates after management actions and damages needs to be included in an operational system. Such a system would be a DA system.

The results presented in paper V are promising for DA-based prediction of forest variables from remote sensing data. The scenario in paper V is of interest for operational forestry, implying that an ALS prediction can be improved upon and kept up-to-date with the aid of other sensors until a new ALS prediction is available. The time span studied in paper V was four growing seasons, while the currently planned interval for ALS acquisitions by the Lantmäteriet is approximately seven years (Lantmäteriet 2019). A study that covers a slightly longer time span than paper V could thus be an interesting next step. In the studies in this thesis, it is not always clear to what extent the errors derive from forecasts or from predictions, but the relatively short periods in most papers suggests that prediction contributes more than growth forecasts to the total error. If longer periods are studied, sensors that capture height development (e.g. DP) could be more valuable.

Poor estimation of uncertainties will impede on DA through improper Kalman gains. As highlighted in mainly paper III and V this is important, especially after a few iterations of DA. Running the filter through many iterations will therefore require very accurate error propagation. Many data sets will, on the other hand, improve the accuracy, at least in theory. It should be noted, however, that theory also suggests that the more data that are added, the less an extra iteration will add to the accuracy. In addition, the more the errors correlate, the less the gain.

6. Further research

A DA system adapted for practical usage has been the scope of this thesis. The papers have taken DA a few steps closer to practical implementation and have left other areas for future research.

A step closer to an operational system would be to include handling of disturbances in the DA system. Change detection is a well-researched area of remote sensing, which could be well integrated into a DA framework.

Forest management planning in the Nordic countries have traditionally been done with stands as the spatial unit. Aggregation through simple averaging of the pixels in a stand is straightforward, but as e.g. Kangas (2018a) points out, error statistics are of interest as well and deserve more research.

Accounting for error correlations brings practical issues as well. Not only does it require the remote sensing acquisitions to cover the same sample plots, but keeping field measurements up to date is also necessary. Further studies are needed about how to possibly overcome this in order to make DA more feasible for practical forestry. Perhaps error correlations could be modeled based on the acquisition properties, such as sensors, phenology etc.

The EKF filter relies on Taylor linearization to propagate the variances through the growth forecast model in the updating step. A Bayesian filter could make better approximations, but could also be computationally expensive. Particle filters uses a set of particles, samples from the predictions´ error distribution, to resemble the distribution following update in a less computationally demanding way. While EKF is restricted to models that can be linearized, particle filter handles any type of forecast method. In this work, we have applied DA for forest variables separately, but making DA multivariate is a desired development.

The papers base the weighting on variances after a DA iteration on theoretical estimates (eq. 3 and 12). An alternative could be to replace this procedure with empirical estimates solely based on field plot data after each iteration. In the current setting, field plots are anyway required for training the predictions at all time steps. But in practice many remote sensing datasets would have different areal extents and the intersecting areas, with similar remote sensing data history, might be small, and thus schemes to derive all uncertainty measures empirically from field data might require large and costly field datasets. Thus, reliable theoretical variance and covariance estimation procedures has a potential to reduce the need for costly field sample plot data.

Overall, DA for forestry looks promising given the development in this thesis. But many details remain to be solved before a reliable system for practical forestry can be developed.

References

Ali-Sisto, D., and Packalen, P. (2017). Forest Change Detection by Using Point Clouds from Dense Image Matching Together with a LiDAR-Derived Terrain Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10(3): 1197–1206. doi:10.1109/JSTARS.2016.2615099.

Bohlin, J., Bohlin, I., Jonzén, J., and Nilsson, M. (2017). Mapping forest attributes using data from stereophotogrammetry of aerial images and field data from the national forest inventory. *Silva Fennica* 51(2): 1–18. doi:10.14214/sf.2021.

Bohlin, J., Wallerman, J., and Fransson, J.E.S. (2012). Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. *Scandinavian Journal of Forest Research* 27(7): 692–699. doi:10.1080/02827581.2012.686625.

Czaplewski, R. (1990). Kalman filter to update forest cover estimates. *In* State-of-the-art methodology of forest inventory: a symposium proceedings. *Edited by* LaBau, Vernon J.; Cunia, Tiberius. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR. pp. 457-465. Available from https://www.fs.usda.gov/treesearch/pubs/43904.

Czaplewski, R., and Thompson, M. (2008). Opportunities to Improve Monitoring of Temporal Trends with FIA Panel Data. *In* Proceedings of the Forest Inventory and Analysis (FIA) Symposium 2008. *Edited by* W. McWilliams, G. Moisen, and R. Czaplewski. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, Park City, UT. pp. 1–55. Available from http://http//www.citeulike.org/user/drczap/article/4025969.

Deutscher, J., Granica, K., Steinegger, M., Hirschmugl, M., Perko, R., and Schardt, M. (2017). Updating lidar-derived crown cover density products with sentinel-2. International Geoscience and Remote Sensing Symposium (IGARSS) 2017-July(August): 2867–2874. doi:10.1109/IGARSS.2017.8127597.

Dixon, B.L., and Howitt, R.E. (1979). Continuous forest inventory using a linear filter. *Forest Science* 25(4): 675–689.

Ehlers, S., Grafström, A., Nyström, K., Olsson, H., and Ståhl, G. (2013). Data assimilation in stand-level forest inventories. *Canadian Journal of Forest Research* 43(12): 1104–1113. doi:10.1139/cjfr-2013-0250.

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., Ringvall, A.H., and Ståhl, G. (2014). Adapting National Forest Inventories to changing requirements - The case of the Swedish National Forest Inventory at the turn of the 20th century. *Silva Fennica* 48(3): 1–29. doi:10.14214/sf.1095.

Ghil, M., and Malanotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. *Advances in Geophysics* 33(C): 141–266. doi:10.1016/S0065-2687(08)60442-2.

Grabska, E., Hawrylo, P., and Socha, J. (2020). Continuous detection of small-scale changes in scots pine dominated stands using dense sentinel-2 time series. *Remote Sensing* 12(8). Multidisciplinary Digital Publishing Institute (MDPI). doi:10.3390/RS12081298.

Hägglund, B., and Lundmark, J.-E. (1977). Site index estimation by means of site properties of Scots pine and Norway spruce in Sweden. *Studia Forestalia Suecica* 138: 38.

Hannrup, B., Bhuiyan, N., and Möller, J. (2015). Nationwide evaluation of a system for automated follow-up of thinning. Uppsala.

Holmgren, J., and Persson, Å. (2004). Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment* 90(4): 415–423. doi:https://doi.org/10.1016/S0034-4257(03)00140-8.

Holopainen, M., and Talvitie, M. (2006). Effect of data acquisition accuracy on timing of stand harvests and expected net present value. *Silva Fennica* 40(3): 531–543. doi:10.14214/sf.335.

Holopainen, M., Vastaranta, M., Haapanen, R., Yu, X., Hyyppä, J., Kaartinen, H., Viitala, R., and Hyyppä, H. (2010). Site-type estimation using airborne laser scanning and stand register data. *The Photogrammetric Journal of Finland* 22(1): 16–32.

Hou, Z., Mehtätalo, L., McRoberts, R.E., Ståhl, G., Tokola, T., Rana, P., Siipilehto, J., and Xu, Q. (2019). Remote sensing-assisted data assimilation and simultaneous inference for forest inventory. *Remote Sensing of Environment* 234(August): 111431. doi:10.1016/j.rse.2019.111431.

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S., and Zhu, Y.H. (2000). Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management* 128(1–2): 109–120. doi:10.1016/S0378-1127(99)00278-9.

Kaartinen, H., Hyyppä, J., Yu, X., Vastaranta, M., Hyyppä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B.M., and Wu, J.C. (2012). An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sensing* 4(4): 950–974. doi:10.3390/rs4040950.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering* 82(Series D): 35–45.

Kalman, R.E., and Bucy, R.S. (1961). New results in linear filtering and prediction theory. *Transactions of the AS M E, Journal of Basic Engineering* 83: 95–108.

Kangas, A. (1991). Updated measurement data as prior information in forest inventory. *Silva Fennica* 25(3). doi:10.14214/sf.a15611.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K.T., Maltamo, M., Nilsson, M., Nord-Larsen, T., Næsset, E., and Olsson, H. (2018a). Remote sensing and forest inventories in Nordic countries–roadmap for the future. *Scandinavian Journal of Forest Research* 33(4): 397–412. doi:10.1080/02827581.2017.1416666.

Kangas, A., Gobakken, T., and Næsset, E. (2020). Benefits of past inventory data as prior information for the current inventory. *Forest Ecosystems* 7(1). Forest Ecosystems. doi:10.1186/s40663-020-00231-6.

Kangas, A., Gobakken, T., Puliti, S., Hauglin, M., and Næsset, E. (2018b). Value of airborne laser scanning and digital aerial photogrammetry data in forest decision making. *Silva Fennica* 52(1). doi:10.14214/sf.9923.

Katila, M., and Heikkinen, J. (2020). Reducing error in small-area estimates of multi-source forest inventory by multi-temporal data fusion. *Forestry: An International Journal of Forest Research* 93(3): 471–480. doi:10.1093/foresj/cpz076.

Kennedy, R.E., Yang, Z., and Cohen, W.B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. *Remote Sensing of Environment* 114(12): 2897–2910. doi:https://doi.org/10.1016/j.rse.2010.07.008.

Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., and Zink, M. (2007). TanDEM-X: A satellite formation for high-resolution SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 45(11, 1): 3317–3341. doi:10.1109/TGRS.2007.900693.

Kukkonen, M., Korhonen, L., Maltamo, M., Suvanto, A., and Packalen, P. (2018). How much can airborne laser scanning based forest inventory by tree species benefit from auxiliary optical data? *International Journal of Applied Earth Observation and Geoinformation* 72(July 2018): 91–98. Elsevier. doi:10.1016/j.jag.2018.06.017.

Lantmäteriet. (2019). Geodata products. Available from https://www.lantmateriet.se/en/maps-and-geographic-information/geodataprodukter/produktlista/ [accessed 6 February 2020].

Lantmäteriet. (2020). SACCESS. Available from https://saccess.lantmateriet.se/portal/saccess_se.htm [accessed 24 November 2020].

Lefsky, M.A., Cohen, W.B., Parker, G.G., and Harding, D.J. (2002). Lidar remote sensing for ecosystem studies. *Bioscience* 52(1): 19–30. doi:10.1641/0006-3568(2002)052[0019:lrsfes]2.0.co;2.

Liang, X., Kukko, A., Hyyppä, J., Lehtomäki, M., Pyörälä, J., Yu, X., Kaartinen, H., Jaakkola, A., and Wang, Y. (2018). In-situ measurements from mobile platforms: An emerging approach to address the old challenges associated with forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing* 143(August 2017): 97–107. doi:10.1016/j.isprsjprs.2018.04.019.

Magnusson, M., and Fransson, J.E.S. (2005). Estimation of forest stem volume using multispectral optical satellite and tree height data in combination. *Scandinavian Journal of Forest Research* 20(5): 431–440. doi:10.1080/02827580500339864.

McGaughey, R.J. (2014). FUSION -Providing fast, efficient, and flexible access to LIDAR, IFSAR, and terrain datasets. Available from http://forsys.cfr.washington.edu/fusion/fusionlatest.html.

Moreira, A., Krieger, G., Hajnsek, I., Hounam, D., Werner, M., Riegger, S., and Settelmeyer, E. (2004). TanDEM-X: A TerraSAR-X add-on satellite for single-pass SAR interferometry. *In* IGARSS 2004: IEEE International Geoscience and Remote Sensing Symposium Proceedings, Vols 1-7: Science for Society: Exploring and Managing a Changing Planet. IEEE, Anchorage, AK, USA. pp. 1000–1003.

Naesset, E. (2005). Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment* 98(2–3): 356–370. doi:10.1016/j.rse.2005.07.012.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80(1): 88–99.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., and Söderman, U. (2004). Laser scanning of forest resources: The Nordic experience. *Scandinavian Journal of Forest Research* 19(6): 482–499. doi:10.1080/02827580410019553.

Nilsson, M., Nordkvist, K., Jonzén, J., Lindgren, N., Axensten, P., Wallerman, J., Egberth, M., Larsson, S., Nilsson, L., Eriksson, J., and Olsson, H. (2017). A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory. *Remote Sensing of Environment* 194: 447–454. doi:10.1016/j.rse.2016.10.022.

Nilsson, P., Cory, N., Kempe, G., and Fridman, J. (2015). Skogsdata 2015/Forest statistics 2015. Swedish University of Agricultural Sciences, Umeå.

Noordermeer, L. (2020). Large-area forest productivity estimation using bitemporal data from airborne laser scanning and digital aerial photogrammetry. Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway.

Osborne C (1991) Statistical calibration: A review. *International Statistical Review / Revue Internationale de Statistique* 59(3): 309 doi: 10.2307/1403690

Persson, H.J., and Fransson, J.E.S. (2016). Comparison between TanDEM-X and ALS based estimation of above ground biomass and tree height in boreal forests. *Scandinavian Journal of Forest Research* 32(4): 1–14. doi:10.1080/02827581.2016.1220618.

Persson, H.J. and Ståhl, G. (2020). Characterizing uncertainty in forest remote sensing studies. *Remote Sensing*. 12(3) doi: 10.3390/rs12030505

Persson, M., Lindberg, E., and Reese, H. (2018). Tree species classification with multi-temporal Sentinel-2 data. *Remote Sensing* 10(11): 1–17. doi:10.3390/rs10111794.

Puliti, S., Ørka, H.O., Gobakken, T., and Næsset, E. (2015). Inventory of small forest areas using an unmanned aerial system. *Remote Sensing* 7(8): 9632–9654. doi:10.3390/rs70809632.

Rahlf, J., Breidenbach, J., Solberg, S., Næsset, E., and Astrup, R. (2014). Comparison of four types of 3D data for timber volume estimation. *Remote Sensing of Environment* 155: 325–333. doi:10.1016/j.rse.2014.08.036.

Räty, J., Packalen, P., and Maltamo, M. (2019). Nearest neighbor imputation of logwood volumes using bi-temporal ALS, multispectral ALS and aerial images. *Scandinavian Journal of Forest Research* 34(6): 469–483. doi:10.1080/02827581.2019.1589567.

Reese, H., Nilsson, M., Pahlén, T.G., Hagner, O., Joyce, S., Tingelöf, U., Egberth, M., and Olsson, H. (2003). Countrywide Estimates of Forest Variables Using Satellite Data and Field Data from the National Forest Inventory. *Ambio* 32(8): 542–548. doi:10.1579/0044-7447-32.8.542.

Rothermel, M., and Wenzel, K. (2012). SURE - Photogrammetric Surface Reconstruction from Imagery. Proceedings LC3D Workshop.

Saad, R., Eyvindson, K., Gong, P., Lämås, T., Ståhl, G. (2017). Potential of using data assimilation to support forest planning. *Canadian Journal of Forest Research*. 47(5): 690-695. doi: 10.1139/cjfr-2016-0439

Soja, M.J., and Ulander, L.M.H. (2013). Digital canopy model estimation from TanDEM-X interferometry using high-resolution lidar DEM. International Geoscience and Remote Sensing Symposium (IGARSS): 165–168. doi:10.1109/IGARSS.2013.6721117.

St-Onge, B., Vega, C., Fournier, R.A., and Hu, Y. (2008). Mapping canopy height using a combination of digital stereo-photogrammetry and lidar.

International Journal of Remote Sensing 29(11): 3343–3364. doi:10.1080/01431160701469040.

Ståhl, G. (1992). A study on the quality of compartmentwise forest data acquired by subjective inventory methods (24). Department of Biometry and Forest Management, Swedish University of Agricultural Sciences, Umeå, Sweden.

Ståhl, G., Carlsson, D., and Bondesson, L. (1994). A Method to Determine Optimal Stand Data Acquisition Policies. *Forest Science* 40(4): 630–649. doi:10.1093/forestscience/40.4.630.

Svenska FSC. (2020). FSC-standard för skogsbruk i Sverige. Bonn, Germany.

Tian Y., Nearing G.S., Peters-Lidard C.D., Harrison, K.W., Tang, L. (2016). Performance metrics, error modeling, and uncertainty quantification. Monthly Weather Review 144(2): 607–613. doi: 10.1175/MWR-D-15-0087.1

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., and Katila, M. (2008). Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment* 112(5): 1982–1999. doi:10.1016/j.rse.2007.03.032.

Vastaranta, M., Wulder, M.A., White, J.C., Pekkarinen, A., Tuominen, S., Ginzler, C., Kankare, V., Holopainen, M., Hyyppä, J., and Hyyppä, H. (2013). Airborne laser scanning and digital stereo imagery measures of forest structure: comparative results and implications to forest mapping and inventory update. *Canadian Journal of Remote Sensing* 39(5): 382–395. doi:10.5589/m13-046.

Vastaranta, M., Yu, X., Luoma, V., Karjalainen, M., Saarinen, N., Wulder, M.A., White, J.C., Persson, H.J., Hollaus, M., Yrttimaa, T., Holopainen, M., and Hyyppä, J. (2018). Aboveground forest biomass derived using multiple dates of WorldView-2 stereo-imagery: quantifying the improvement in estimation accuracy. *International Journal of Remote Sensing* 39(23): 8766–8783. doi:10.1080/01431161.2018.1492176.

Véga, C., and St-Onge, B. (2009). Mapping site index and age by linking a time series of canopy height models with growth curves. *Forest Ecology and Management* 257(3): 951–959. doi: 10.1016/j.foreco.2008.10.029.

Wallerman, J., and Holmgren, J. (2007). Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data. *Remote Sensing of Environment* 110(4): 501–508. doi:10.1016/j.rse.2007.02.028.

Wikström, P., Edenius, L., Elfving, B., Eriksson, L.O., Lämås, T., Sonesson, J., Öhman, K., Wallerman, J., Waller, C., and Klintebäck, F. (2011). The Heureka Forestry Decision Support System: An Overview. *Mathematical and Computational Forestry & Natural-Resource Sciences* 3(2): 87–95.

Xu, C., Morgenroth, J., and Manley, B. (2015). Integrating data from discrete return airborne lidar and optical sensors to enhance the accuracy of forest description: A review. *Current Forestry Reports* 1(3): 206–219. doi:10.1007/s40725-015-0019-3.

Yu, X., Hyyppä, J., Karjalainen, M., Nurminen, K., Karila, K., Vastaranta, M., Kankare, V., Kaartinen, H., Holopainen, M., Honkavaara, E., Kukko, A., Jaakkola, A., Liang, X., Wang, Y., Hyyppä, H., and Katoh, M. (2015). Comparison of laser and stereo optical, SAR and InSAR point clouds from air- and space-borne sources in the retrieval of forest inventory attributes. *Remote Sensing* 7(12): 15933–15954. doi:10.3390/rs71215809.

Åge, P.-J. 1985. Forest inventory - Photo interpretation, Report 13. Lantmäteriet, Gävle, Sweden.

Popular science summary

Forest management planning use maps with the forest delineated into stands, and associated databases with the stands forest properties. This information has traditionally been gathered by visiting all stands in the field at regular intervals, whereby the forest is measured and assessed with manual methods. Recently, an automated data collection by means of remote sensing has been introduced. The main technology is airborne laser scanning, which provides measurements for ground elevation and the height and density of the canopy. With aid of field-measured sample plots and statistical models, these measurements can be translated into forest variables of interest. Predictions of the forests height, basal area, average diameter and volume can thereby be done with at least the similar accuracy as the traditional field assessments. Several other remote sensing techniques, e.g. canopy height models from aerial images or radar data, or color from satellite image pixels are also available. These are often available at lower cost and more frequently than laser scanning, but have so far not provided as good forest data. Forests change over time, through growth and tree mortality or being harvest. Therefore, a prediction of the forest conditions at a time point must be updated. So far this has been made with forecast models until mew data collection, at which the old data have been discarded.

In this thesis, data assimilation is investigated to use the information from several types of remote sensing data to update and maintain the quality of forest data without discarding old data. The investigated method, Kalman filter, means that data of the forest is forecasted and that new remote sensing predictions is combined with it by weighted average in proportion to its information content. After a number of adaptations of the basic Kalman filer, this has proven to work. The developed methods thus lay the foundation for a new way of automatically use several types of remote sensing data to automatically keep forest data up to date.

Populärvetenskaplig sammanfattning

Vid skötsel och förvaltning av skog används kartor där skogen är indelad i bestånd, med tillhörande databas som anger skogens egenskaper. Denna information har traditionellt samlats in genom att alla bestånd på en fastighet besökts återkommande, varvid skogen mätts och bedömts med manuella metoder. På senare tid har datainsamlingen börjat automatiseras med fjärranalys. Den främst använda tekniken är flygburen laserskanning, vilket ger mätvärden på såväl markens läge som trädskiktets höjd och täthet. Med stöd av fältmätta referensytor och statistiska modeller kan sedan dessa mätvärden översättas till skogliga variabel. Skattningar av trädens höjd, grundyta, medeldiameter och volym har med laserskannings kunnat göras automatiskt minst lika bra om med de traditionella fältmetoderna. Flera andra fjärranalystekniker, t.ex. trädhöjdsmodeller beräknade från flygbilder eller radardata, eller skogens färg enlig satellitbildspixlar finns också tillgängliga, ofta till lägre kostnad och mer frekvent än laserskanning, men har hittills inte givit lika bra skogliga data. Skogar ändras över tiden, med tillväxt och träd som dör eller avverkas. En uppskattning av skogstillståndet vid en tidpunkt måste därför uppdateras för att vara aktuellt, vilket hittills gjorts med framskrivningsfunktioner till dess att nya data samlats in och de gamla kastats bort.

I denna avhandling undersöks om dataassimilering kan användas för att ta tillvara informationsinnehållet i flera typer av fjärranalysdata och kontinuerligt uppdatera och bibehålla kvaliteten i den skogliga databasen utan att gamla data kastas bort. Den undersökta metoden, Kalman filtrering, innebär att skattningar av skogstillståndet framskrivs med tillväxtfunktioner och att varje ny fjärranalysregistrering viktas ihop med den framskrivna skattningen i proportion till sitt informationsinnehåll. Efter ett antal anpassningar av den grundläggande metoden för Kalman filtrering så har detta visat sig fungera. De utvecklade metoderna lägger därmed grunden för ett nytt sätt att automatiskt använda flera typer av fjärranalysdata för att automatiskt hålla skogliga data ajour.

Acknowledgements

Many people have helped me along the way and in different ways contributed to this thesis, from the very start to the finish. The journey towards writing this thesis started during the GIS courses during my masters' education here at SLU that led me into asking for a topic for my M.Sc. thesis. Thanks' to you at the remote sensing section for excellent teaching and contagious interest, spreading the enthusiasm for remote sensing! Thanks' to Karin and Håkan back then in 2012 when I wrote my M.Sc. thesis, you really gave me the best support and ultimately led me into looking for a PhD position. The Swedish national space board and the Swedish research council FORMAS funded my PhD studies, and I am grateful for the opportunity that gave me to dig deep into remote sensing.

Especially, I would like to thank my main supervisor Håkan Olsson, who believed in me and helped me into my PhD-student position, and supported me all the way through. Thanks for all your time and efforts, for guiding me and keeping me on track. I would also like to thank my assistant supervisors. Mattias, Henrik, Eva and Anton, you have been a great support and company.

Göran Ståhl, you have despite your crowded calendar somehow found time for me. Thanks' for all the support, and for the patient guiding through the marvels of the statistical world.

To all my coworkers at the remote sensing section and the department of forest resource management, past and present, it has been great to work at SLU, mainly thanks to you. Thank you for sharing your knowledge and for all the "fika" together. Thank you Emma, Mona, Jonas J, Bisse, Inka and many more for being the best company when I needed to talk about parenting and all the crazy things my kids have been up to!

To Ida, thanks for the support to get me through this. Anna-Lena, Thomas, Pelle and more people, you have supported me when I needed it, big thanks for that.

Carola and Mikael, and all other bikers and dog training people, thanks for all the lunch training and all the unforgettable moments, around Vättern and many other places! It has really helped me to go through with this thesis.

Thank you my dear sister, you are rock to lean on. Mom and Dad, you gave me the curiosity, I never built an aircraft that actually took off from the ground but working in remote sensing is not far from.

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2021:6

This thesis addresses the application of data assimilation methods to remote sensing predictions of forest variables. Data assimilation are methods to combine measurements over time, to improve prediction accuracy. The methods are used extensively in meteorology and other fields but are not well investigated for forestry. This thesis presents five studies where data assimilation of forest variable predictions have been evaluated with promising results.

Nils Lindgren received his doctoral education at the Department of Forest Resource, Swedish University of Agricultural Sciences, Umeå, and his Master of Science in Forestry at the same university.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of thesis summary: http://pub.epsilon.slu.se/

ISSN 1652-6880
ISBN (print version) 978-91-7760-690-1
ISBN (electronic version) 978-91-7760-691-8