ASSIMILATING REMOTE SENSING DATA WITH FOREST GROWTH MODELS

Mattias Nyström[1], Nils Lindgren[1], Jörgen Wallerman[1], Sarah Ehlers[1], Anton Grafström[1], Anders Muszta[1], Kenneth Nyström[1], Erik Willén[2], Johan E. S. Fransson[1], Jonas Bohlin[1], Göran Ståhl[1], Håkan Olsson[1]

1. Swedish University of Agricultural Science, Department of Forest Resource Management, Umeå, Sweden; firstname.lastname@slu.se
2. Skogforsk, Uppsala, Sweden; firstname.lastname@skogforsk.se

ABSTRACT

Data assimilation techniques were used to estimate forest stand data in 2012 by sequentially combining remote sensing based estimates with predictions from growth models. Estimates of stand data, based on canopy height models obtained from image matching of digital aerial images at six different time-points between 2003 and 2012 served as input to the data assimilation. The assimilation routines were built on the extended Kalman filter. The study was conducted in hemi-boreal forest at the Remningstorp test site in southern Sweden (lat. 13° 37' N; long. 58° 28' E). The assimilation results were compared with two conventional approaches: the first was to only use the most recent estimate obtained from remotely sensed data (2012) and the second was to forecast the first estimate (2003) to the endpoint (2012). All three approaches were validated on 40 m radius stands. The results showed that the data assimilation approach provided similar results as use of the estimates based on the most recent set of remotely sensed data but better results than the forecasts from the first estimate. Although the assimilation results did not provide any large improvement compared to only using data from the most recent remote sensing acquisition, a framework that can be used for sequentially utilizing all types of remote sensing data was successfully established. Our expectation is that the assimilation results will improve in future studies when data from different sensors will be combined and when more frequent field reference data will become available.

INTRODUCTION

Data assimilation is a technique that offers great potential for combining all new sources of data of relevance for forest estimates [1]. The success of data assimilation in other areas, such as meteorology, is well documented [2]. However, in order to realize the potentials in the context of forestry the statistical methods need to be adapted to this field of application. In brief, a system for data assimilation in forestry should be based on a geographical model of the forest in which forest data are forecasted using growth models. All new data from remote sensing and from measurements in field, should then be used to adjust the forecasted forest information to the extent motivated by the accuracy of the new data compared to the accuracy of the forecasts.

The objective of this study was to present first empirical results of the application of data assimilation to forest stand data using remote sensing data from image matching of aerial images. In a previous study made by our research group, the results were based on theoretical assumptions [1]. In the present study we applied data assimilation to estimates based on empirical data from image matching data from six occasions obtained over a 9 year period (2003–2012). The study variables were Lorey's mean height (H_L), basal area (BA) and stem volume (V). The results from data assimilation were compared with two established methods; estimates from the most recent time-point and forecasting of the first estimate (2003) using growth models.

METHODS

Study area

The study was carried out at the forest estate Remningstorp in south-western Sweden (lat. 13° 37' N; long. 58° 28' E, Figure 1). The forest is dominated by Norway spruce (*Picea abies*) and Scots Pine (*Pinus sylvestris*), with some deciduous forest of mainly birch (*Betula spp.)*.

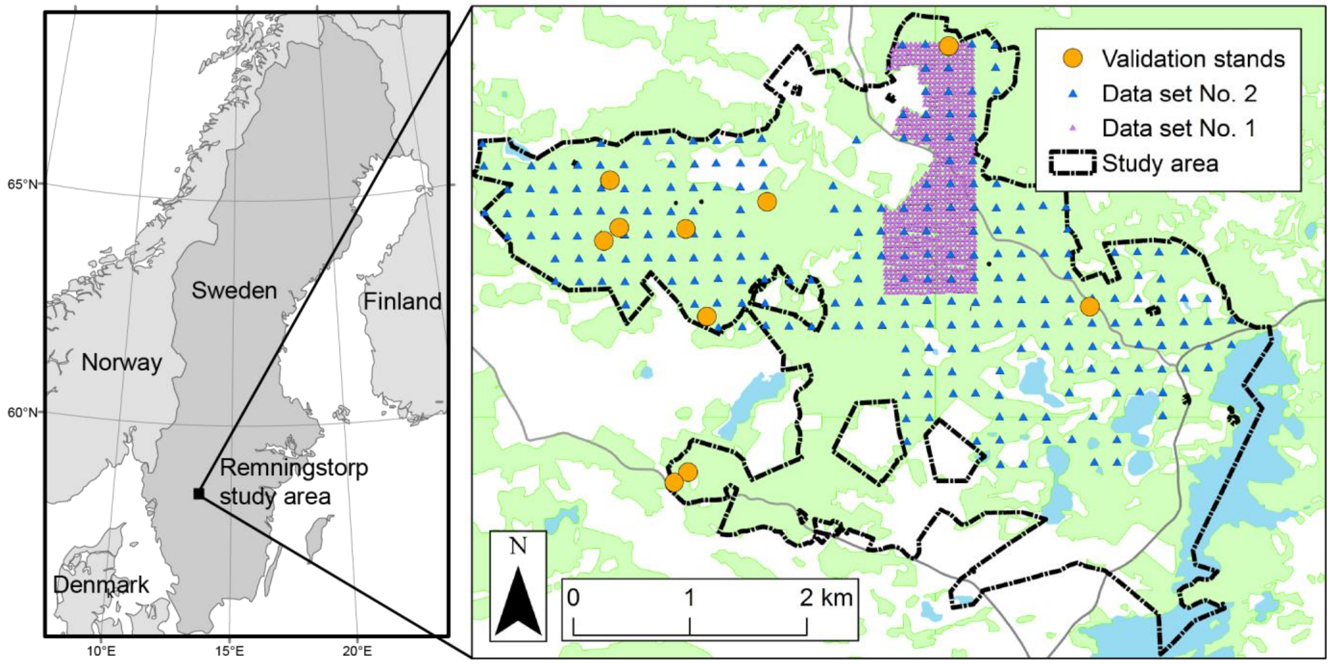

Figure 1. *Overview of the study area, the location of the field inventoried sample plots, and the validation stands used for assimilation. © Lantmäteriet I 2014/00764.*

Remote sensing data

Aerial images were acquired with the Z/I DMC digital mapping camera system [3] operated by Lantmäteriet (Swedish National Land Survey) at six different survey campaigns with partial or complete cover of the study site (Table 1). Image matching of the digital aerial images was performed using the SURE software [4,5] to produce point cloud data for each dataset. SURE generates a height value for each pixel using cost-based matching, similar to the semi-global matching method [6]. Finally, the height values of the point cloud were transformed from height above sea level to height above ground level by subtracting the digital terrain model (DTM) created using airborne laser scanning.

An area based approach [7] was used to calculate metrics describing the point cloud data, such as height distribution and spatial density characteristics for every field sample plot using the Fusion software [8]. A height threshold [9] of 2 m was used. Raster elements covering the test site of all metrics were calculated on a cell size of 18 m × 18 m, which corresponds to the field reference plot size of 314 m^2.

Field reference data

In this study, field data were applied for three different purposes: *(i)* for developing models linking the matched aerial image data with the ground conditions for the study variables, *(ii)* for developing growth models, and *(iii)* for validating the results at the endpoint of the data assimilation period.

For the development of models estimating the state of the study variables from the remote sensing data we applied field data from sample plots distributed across the Remningstorp study area. Plots were available from two different field campaigns during the study period (data set 1 and 2 in Figure 1 and Table 2). All trees on the plots were calipered, and a sub-sample of the trees was selected for measurements of height and age. Site index (SI) was also assessed on the plots. In Table 1, a summary of the sample plot data available for the development of models is given.

Table 1. *Digital aerial images and reference field plots used for modeling the study variables.*

Acquisition date	**Growth season**[1]	**Leaf-on/ leaf-off**	**Altitude** (m a.g.l.)	**Field ref. data set**[2]	**No. of ref. plots used**
2003-10-13	2003	leaf-on	3000	1	361
2005-06-28	2005	leaf-on	4800	1	416
2007-05-26, 2007-06-03	2006	leaf-on	4800	1	258
2009-09-01	2009	leaf-on	4800	2	214
2010-05-02	2009	leaf-off	4800	2	214
2012-05-23	2011	leaf-on	4800	2	166

[1] Refers to if the acquisition was performed before or after the 15th of June.
[2] Refers to Table 2.

Table 2. *Field inventories used for the development of predictive models. V = Stem volume, BA = Basal area, H_L = Lorey's mean height.*

No	**Inv. year**	**No of plots**	**Radius** (m)	**V** (m^3/ha) min/mean/max	**BA** (m^2/ha) min/mean/max	$\mathbf{H_L}$ (m) min/mean/max
1	2004	849	10	0/277/1050	0/28/80	0/19/34
2	2010	247	10	0/202/697	0/22/60	0/16/33

Table 3. *Summary of the validation stands, i.e. the ten large plots with 40 m radius that were used to validate the results of the data assimilation, at growth season 2011.*

ID	**Site index** (m/100 years)	**Age** (years)	**V** (m^3/ha)	**BA** (m^2/ha)	$\mathbf{H_L}$ (m)	**Forest type class**
10	25	39	389	39.3	21.5	Spruce stand
116	30	55	501	43.4	25.3	Spruce stand
151	31	44	283	29.3	19.8	Spruce stand
206	26	105	520	45.2	29.4	Pine stand
211	32	41	308	30.5	21.6	Spruce stand
212	32	40	241	26.4	20.2	Mixed stand
325	31	43	386	37.0	22.2	Spruce stand
351	31	46	265	27.1	20.5	Spruce stand
515	25	26	115	20.6	12.1	Mixed conifer stand
517	30	30	185	30.8	13.1	Mixed conifer stand

Growth models suited for the data assimilation approach were developed using data from the Swedish National Forest Inventory. The validation stands used for validating the results of the data assimilation and the two conventional methods was located in the interior of fairly homogeneous stands and had a 40 m radius, corresponding to about 0.5 hectare. The ten validation stands (Table 3, Figure 1) were classified into four different forest type classes for which different growth models were used. The forest type class was based on the conditions at the time-point of validation (2012). The composition was recorded as **spruce stand** if more than 65% of the total volume was Norway spruce, **pine stand** if more than 65% of the total volume was Scots pine, **mixed conifer stand** if more than 65% of the total volume was conifer, and **mixed stand** if the total volume of broadleaved trees was between 35 and 65% and less than 65% of the total volume was coniferous. None of the validation stands had been subject to management (such as thinning or clear-felling) during the period 2003–2012. All the validation stands (except two) were inventoried in the same growth season as the last aerial images were acquired. The two other stands were back-casted (1 and 2 years) to the time-point of the image acquisition, using the growth models developed in this study.

Estimation of the study variables

The state of each validation stand was estimated for each image using the corresponding raster with metrics. Estimates were trained with reference plots, data set 1 and 2, described in Table 1. In order to obtain as good temporal matches as possible, sample plot data were either fore- or back-casted for short time periods to correspond to the time-points of the image acquisitions. Linear regression was applied for modeling Lorey's mean height and non-linear regression for modeling stem volume and basal area, in order to capture the relationship between each study variable and the information provided by the aerial imagery. Furthermore, the residual variances of stem volume and basal area showed increasing trends with respect to the predicted values. These trends were captured using additional linear models, relating residual standard deviation to the predicted value, fitted to standard deviations calculated for the residuals of ten equal intervals of predicted values.

Data assimilation

Existing information about a forest area is forecasted using a model that provides an estimate at the time of the next data acquisition and an estimate of the precision of the forecasted information. Thus, the precision of the forecasted information can be compared with the precision of the new information. In the assimilation step, the two sources of information are combined through weights that are inversely proportional to their variances. The combined estimate is then forecasted to the time point of the next data acquisition, etc.

In this study we applied the extended Kalman filter [11] for the data assimilation. Only one variable at the time was addressed. The other variables used in the growth models were site index, tree species composition and age, which were known from field surveys.

Modeling the development over time of the target variable is a core part of the data assimilation system. Using the extended Kalman filter, the assimilated variable, $\hat{x}_t$, can be calculated at time point t as

$$\hat{x}_t = (1 - K_t)\tilde{x}_t + K_t\tilde{z}_t \quad (1)$$

where K_t is the Kalman gain calculated as $K_t = \frac{\tilde{p}_t^2}{\tilde{p}_t^2 + r_t^2}$ where $\tilde{p}_t^2$ is the variance of the forecasted value, $\tilde{x}_t$, and r_t^2 is the variance of the estimate from image matching, $\tilde{z}_t$. Further, the variance of the assimilated variable, $\hat{x}_t$, can be calculated as $p_t^2 = (1 - K_t)\tilde{p}_t^2$.

Validation of the three methods (assimilation, most recent estimate, and forecast) was made by calculating the deviation (e_i) from the field measured value for the ten validation stands at the last time-point. For each method, the deviations were calculated by subtracting the field measured value from the method's predicted value. In addition, the root mean squared error ($RMSE$) was calculated for the three study variables and methods as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \quad (2)$$

RESULTS

The initial state (2003) was estimated from point clouds obtained from image matching. The validation was made with field data for the 40 m radius validation stands year 2012. Figure 2 shows the deviation from the field measured value for Lorey's mean height for the three methods of the ten validation stands. Table 4 shows the $RMSE$ of the deviation from the field measurements. It can be seen that the $RMSEs$ are smaller using data assimilation compared to forecasting the value from the first remote sensing prediction. However, the most recent estimate resulted in slightly lower $RMSE$ for the three study variables.

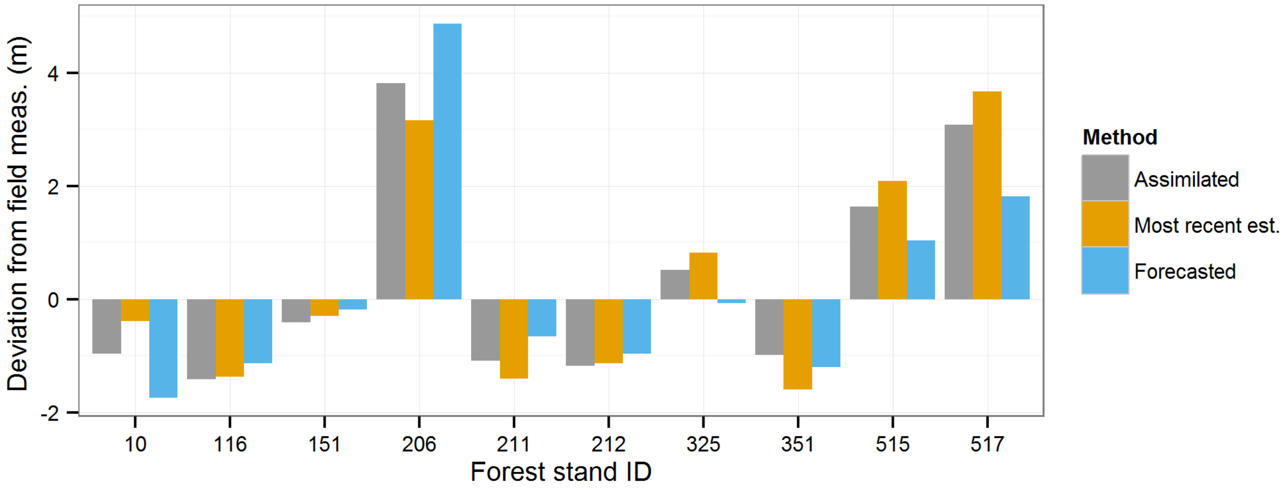

Figure 2. *Deviation at year 2012 between field measurement and assimilation, most recent estimate and forecast of Lorey's mean height. A positive value means that the stand has an overestimated value.*

Table 4. *RMSE of the deviation from the field measurement 2012 for the ten assimilated stands.*

Study variable	Assimilated	Most recent estimate	Forecasted
V (m^3/ha)	59.0 (18.5%)	54.7 (17.1%)	104 (32.5%)
BA (m^2/ha)	4.0 (12.2%)	3.9 (11.9%)	5.4 (16.4%)
H_L (m)	2.1 (10.4%)	2.0 (10.0%)	2.3 (11.5%)

DISCUSSION

In this study, the most recent estimate from remote sensing data and the last set of reference plots were from the same growth season as the field measured validation stands. This, in combination with the limited availability of reference plots from earlier years, probably contributed to the result that the most recent estimate performed about the same as data assimilation. The full strength of data assimilation will probably first be seen when combining data from multiple sources and when a more suitable set of reference data is available. For example, if data is first acquired using airborne laser scanning (ALS) and later acquired with a technique that has lower accuracy, we will be able to update the high accuracy acquisition from the ALS with new data. In addition, the data assimilation framework can be used for maintaining information from earlier high quality measurements, for example from field visits, and combining it, rather than replacing it, with new information from remote sensing.

For some of the stands, there is a consistent over- or underestimation of the three study variables. This can be seen for all three methods evaluated here. A reason for the large deviations from the field measurements (Figure 2) could be that image matching contains relatively strong information on height, but information correlated with stand density has proven to be harder to achieve [12]. Thus, both the errors in estimates of V and BA should correlate with H_L. The lack of density information in point clouds from image matching might also explain why data assimilation did not increase accuracies in this study. One of the validation stands (206) deviates more than the other stands. It can be noted that this stand is much older than the others (see Table 3), which might affect the estimates from the image matching.

CONCLUSIONS

This study presents empirical results of data assimilation applied to estimates of forest variables. A system for data assimilation was developed and implemented for assimilation of forest variables. The study shows that data assimilation of stem volume, basal area and Lorey's mean height results in similar accuracy as estimates from the last available aerial images. However, the results from data assimilation were superior to the results obtained when estimates from the first available aerial images were forecasted using traditional growth models.

ACKNOWLEDGEMENTS

This study was financed by the Forest Society of Sweden and the Kempe foundations.

REFERENCES

1. Ehlers S, Grafström A, Nyström K, Olsson H, Ståhl G, 2013. Data assimilation in stand-level forest inventories. Can. J. For. Res., 43: 1104–1113.
2. Rabier F, 2005. Overview of global data assimilation developments in numerical weather-prediction centres. Q. J. R. Meteorol. Soc., 131: 3215–3233.
3. Hinz A, Dörstel C, Heier H, 2001. DMC – The digital sensor technology of Z / I-Imaging. In Photogrammetric Week 01, Wichmann Verlag: Heidelberg, Germany, pp. 93–103.
4. Rothermel M, Wenzel K, Fritsch D, Haala N, 2012. SURE - Photogrammetric surface reconstruction from imagery. In LC3D Workshop, Berlin, Germany, pp. 1–9.
5. Wenzel K, Rothermel M, Haala N, Fritsch D, 2013. SURE – The ifp software for dense image matching. In Photogrammetric week 2013, Heidelberg, Germany, pp. 59–70.
6. Hirschmüller H, 2008. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell., 30: 328–341.
7. Næsset E, 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ., 80: 88–99.
8. McGaughey R J, 2014. FUSION/LDV: Software for LiDAR data analysis and visualization. FUSION version 3.42, 175.
9. Nilsson M, 1996. Estimation of tree heights and stand volume using an airborne lidar system. Remote Sens. Environ., 56: 1–7.
10. Fridman J, Holm S, Nilsson M, Nilsson P, Ringvall A H, 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 Fenn., 48: 1–29.
11. Welch G, Bishop G, 2006. An introduction to the Kalman filter, Chapel Hill, NC, USA.
12. Bohlin J, Wallerman J, Fransson J E S, 2012. Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. Scand. J. For. Res., 27: 692–699.