Data assimilation in forest inventory: First empirical results using ALS data

Mattias Nyström, Nils Lindgren, Jörgen Wallerman, Anton Grafström, Anders Muszta, Kenneth Nyström, Göran Ståhl, Håkan Olsson
Swedish University of Agricultural Sciences (SLU), Department of forest resource management, SE 901 83 Umeå, Sweden; e-mail: firstname.lastname@slu.se

Highlights: A first data assimilation case study using a time series of ALS for updating forest stand data is presented. Forest stand data are predicted from each ALS acquisition. Kalman filtering and growth models are then used to combine each new ALS based prediction with forecasts from the previous data acquisition.

Key words: *data assimilation; ALS; LiDAR; forest inventory*

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 airborne laser scanning (ALS). 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 predictions based on empirical data from ALS data from six occasions obtained over an 8 year period (2003–2011). The target variables were Lorey's mean height, basal area and growing stock volume. The results from data assimilation were compared with two established methods; prediction using ALS data from the most recent time point or forecasting of ALS predictions using growth models.

Material and 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). The forest is dominated by Norway spruce (*Picea abies*) and Scots Pine (*Pinus sylvestris*), with some deciduous forest of mainly birch (*Betula spp.)*.

Remote sensing data

Airborne laser scanning (ALS) data from six years: 2003, 2004, 2007, 2008, 2010 and 2011 were used for the assimilation. The scannings were acquired in the autumn under leaf-on conditions except for 2007 and 2011 which were acquired in the spring under leaf-off conditions. Metrics were extracted with LAStools using a height threshold of 2 m [3]. The sample plot area was used for extracting metrics when developing the prediction models and raster cells of 18 m × 18 m when applying the prediction models for the evaluation plots.

Field reference data

The field reference training data set used to predict the three target variables consisted of two field campaigns. The first field campaign was year 2004 and 2005 and collected sample plots with 10 m radius in a regular pattern with 40 m distance between the plot centers. A subset of these plots, in total 258, was used as field reference for the ALS data acquired in 2003, 2004, 2007 and 2008. The field data was for- or back-casted to obtain as good temporal match as possible. This was done using the Heureka forestry decision support system [4].

The second field reference training data set was collected in year 2010 and consisted of 263 sample plots with 10 m radius in a regular pattern with 200 m distance between the plot centers. This reference data was used for the two latest ALS acquisitions.

Prediction of target variables

Prediction of the target variables, Lorey's mean height, basal area and growing stock volume, was done using the area-based method [5]. To predict Lorey's mean height, the 95th height percentile was used. In the prediction model for basal area and growing stock volume, the vegetation ratio and the quadratic average of the returns were used. These variables were selected with support of best subset regression to find two predictors describing height and density. The selected metrics were forced to be the same for all acquisitions.

Growth models

Growth models were developed to estimate the growth between one assimilation and the date of the next ALS acquisition. These functions were developed through regression analysis and used permanent plots from the Swedish National Forest Inventory [6].

Evaluation plots

The plots used to evaluate data assimilation consisted of ten 40 m radius sample plots in normally developed stands without cuttings in the time period between 2003 and 2011. All plots were however not covered by all laser scanning's and the number of actual scannings used in the assimilation of an evaluation plot varied between two and six, with an average of 3.7. Validation data for year 2011 were obtained by calipering all trees at these evaluation plots and a sample of trees were height measured.

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 [7] 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 and assumed known without errors.

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 prediction from ALS data, $\tilde{z}_t$. Further, the variance of the assimilated variable, $\hat{x}_t$, can be calculated as $p_t^2 = (1 - K_t)\tilde{p}_t^2$.

The assimilation was conducted for each raster element (18 m × 18 m). For validation on the evaluation plots, the mean value was calculated of the raster elements with center point within each evaluation plot's boundary.

Results

Data assimilation was compared to the two established methods: prediction using ALS data from the most recent time point and forecasting of ALS predictions from the first time point. Field measured evaluation plots consisting of 40 m radius sample plots were used as ground truth. Figure 1 shows the deviation from the field measured value for each of the methods and the ten evaluation plots, for the variable basal area. Table 1 shows the RMSE of the deviation from the field measurements for all three estimated variables. It can be seen that the mean deviations are smaller using data assimilation compared to forecasting the value from the first ALS data prediction. The prediction using the most recent ALS data resulted in lower RMSE for Lorey's mean height though.

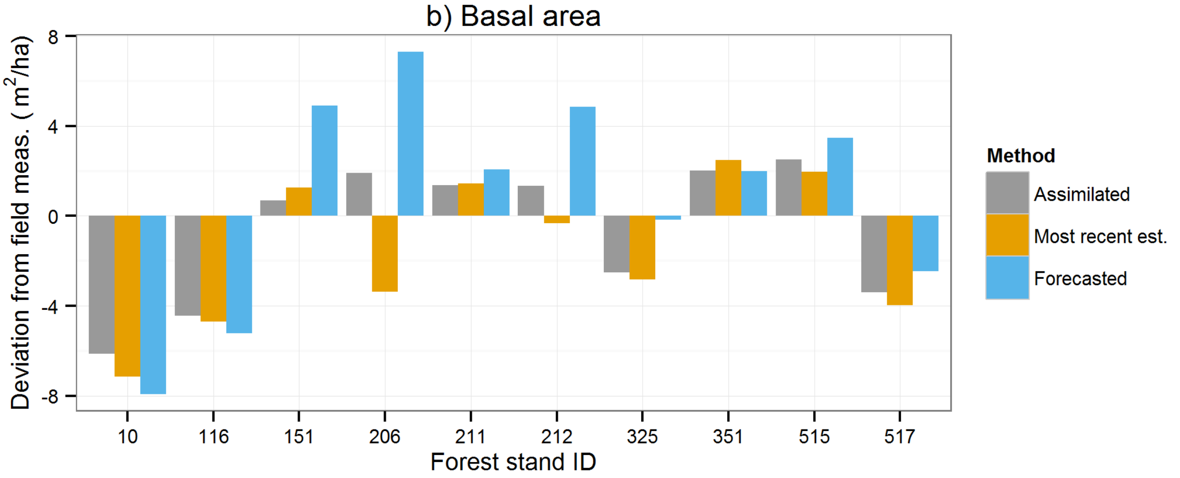

Figure 1: Deviation of basal area calculated as each methods value 2010 minus the field measured value 2010 for the ten evaluation plots.

Table 1: RMSE of the ten assimilated plots. In parenthesis the relative RMSE. The comparison is to the field measured values for the plots.

Target variable	Unit	Assimilated	Last ALS	Forecast from first ALS
Lorey's mean height	m	1.7 (8.5%)	1.6 (8.0%)	2.0 (9.9%)
Basal area	m^2/ha	3.1 (9.5%)	3.5 (10.8%)	4.7 (14.5%)
Stem volume	m^3/ha	41.0 (13.3%)	44.1 (14.3%)	63.5 (20.6%)

Discussion

In this study the potential of data assimilation is verified. The strength of data assimilation will probably first be seen when combining data from several remote sensing techniques. For example if data first is acquired using ALS and the next data acquisition with a cheaper technique and probably lower accuracy, we will be able to update the high accuracy acquisition from the ALS with new data. As we only use ALS data in this early study, there might be temporal autocorrelations between the estimation errors. Methods to compensate for this will be developed in future studies.

The first acquisition is from year 2003 and the last from 2011. This means that it is a rather short time period for forestry. In an operational case we would have a model that continuously is updated when new data become available and probably span over much longer time than eight years.

Further research is needed to investigate at which level the assimilation should be performed. It could be more convenient to assimilate directly on the whole plot instead of splitting the plot into raster cells.

In this study we relied on only two field training data sets and forecasting or back-casting was done to the years' of the ALS acquisitions. The use of only two field surveys is probably the most limiting factor which might explain why the assimilation results is only marginally better than the use of only the latest time-point. In a best case we would have field data from every year where remote sensing data is available. The starting point for the assimilation is an estimate using the first ALS data. Another possibility would be to start with the last known state for each stand, for example from the forest stand register.

Conclusions

This study presents among the first results of a data assimilation of forest stand variables. The results verify the potential of data assimilation, but the study also identifies several practical obstacles that must be addressed before data assimilation can be successfully applied in practice.

Acknowledgement

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] Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. Remote Sens. Environ., 56, 1–7.

[4] Wikström, P.; Edenius, L.; Elfving, B.; Eriksson, L. O.; Lämås, T.; Sonesson, J.; Öhman, K.; Wallerman, J.; Waller, C.; Klintebäck, F. (2011). The Heureka forestry decision support system: An overview. Math. Comput. For. Nat. Sci., 3, 87–94.

[5] 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.

[6] 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.

[7] Welch, G.; Bishop, G. (2006). An introduction to the Kalman filter; Chapel Hill, NC, USA.