Potential of using data assimilation to support forest planning

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Forest Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>cjfr-2016-0439.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Concept papers</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>12-Jan-2017</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Saad, Rami; Swedish University of Agricultural Sciences, Eyvindson, Kyle; University of Jyväskylä, Department of Biological and Environmental Science Gong, Peichen; Swedish University of Agricultural Sciences, Centre for Environmental and Resource Economics Lämås, Tomas; Swedish Univ. of Agricultural Sciences, Forest Resource Management Ståhl, Göran; Swedish University of Agricultural Sciences, Forest Resource Management</td>
</tr>
<tr>
<td>Keyword:</td>
<td>uncertainty, suboptimal loss, remote sensing, Bayesian statistics, stochastic optimization</td>
</tr>
</tbody>
</table>
Potential of using data assimilation to support forest planning

Authors:

Rami Saad(i)*, Kyle Eyvindson(ii), Peichen Gong(iii), Tomas Lämås(i) and Göran Ståhl(i)

(i) Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden

(ii) Department of Biological and Environmental Science, P.O. Box 35, 40014 University of Jyväskylä, Finland.

(iii) Department of Forest Economics, Swedish University of Agricultural Sciences, Umeå, Sweden; European Forest Institute, North European Regional Office, SLU, SE-901 83 Umeå, Sweden.

*Corresponding author. Email: rami.saad@slu.se

Address:

Swedish University of Agricultural Sciences (Sveriges lantbruksuniversitet)
Department of Forest Resource Management
Division of Forest Planning
SE-901 83 Umeå, Sweden
Visiting address: Skogsmarksgränd
Phone: +46 90 786 8108
website: www.slu.se/srh

https://mc06.manuscriptcentral.com/cjfr-pubs
Abstract

Uncertainty in forest information typically results in economic and ecological losses as a consequence of suboptimal management decisions. Several techniques have been proposed to handle such uncertainties. However, these techniques are often complex and costly. Data assimilation (DA) has recently been advocated as a tool which may reduce the uncertainty and thereby improve the quality of forest planning results. It offers an opportunity to make use of all new sources of information in a systematic way, and thus provide more accurate and up-to-date information to forest planning. In this study we refer to literature on handling uncertainties in forest planning as well as related literature from other scientific fields in order to assess the potential benefits of using DA in forest planning. We identify five major potential benefits: (i) The accuracy of the information will be improved; (ii) The information will be kept up-to-date; (iii) The DA process will provide information with estimated accuracy; (iv) Stochastic decision making can be applied, whereby the accuracy of the information can be utilized in the decision making process; and (v) DA data allows for the analysis of optimal data acquisition decisions.

Keywords: uncertainty; suboptimal loss; remote sensing; Bayesian statistics; stochastic optimization.
1. Introduction

Wise management of forest resources requires accurate estimates of what is contained within the forest. These estimates can be considered forest information, a collection of forest variables which can be estimated through various inventorying techniques. This information is often organized in databases, where relatively homogenous parcels of forested land area are aggregated as stands. Through thematic maps, the various attributes can be shown spatially. Depending on the level of detail, the data will contain specific information on the growing stock volume, height, tree species, age, area size and site index of the stand. Traditionally, forest information has been acquired in the field through ocular estimation or through objective samples, updated every 5-10 years. Recent developments in remote sensing have allowed for the possibilities of acquiring forest information from distance at reduced cost (Næsset 2002; Gobakken & Næsset 2004; Saad et al. 2015). Regardless of how forest information is acquired, it is not free from errors and these errors are one of the many sources of uncertainty in forest planning. Following the data acquisition, the old forest information would no longer be used in the forest planning process, and the potential remaining value of the old information is thus ignored.

Uncertainty in forest information occurs due to random or systematic errors in the inventory estimates. Systematic errors may occur as a result of subjective judgments or problems with measurement devices that lead to consistent over- or underestimates of the true value (Ståhl 1992). Random errors are unpredictable deviations, introduced by (random) measurement errors or through measuring only a sample of the population of interest. The uncertainty of the initial state is propagated through the growth models used to predict the future forest state (Mowrer 2000; Nyström & Ståhl 2001; Eid 2000). Uncertainty in forest information typically leads to suboptimal decisions in forest planning (Duvemo & Lämås 2006, Pukkala 1998). Therefore, considering and reducing uncertainty in forest information is of
major importance for forest planners; this is particularly relevant since forestry involves large economic, ecological, and social values. Aside from inventory errors there are other sources of uncertainty which can affect the optimal forest management plan, e.g., growth models (Nyström & Ståhl 2001), market prices (Gong 1994), fire risk (Savage et al. 2010, González-Olabarria & Pukkala 2011), wind risk (Heinonen et al. 2009) and climate change (Crowe & Parker 2008; Kangas & Kangas 2004; Pasalodos-Tato et al. 2013; Yousefpour et al. 2012; Ferreira et al. 2016). As this study focuses on the link between forest inventory data and forest planning, the only source of uncertainty that will be considered is the uncertainty in the initial state of the forest resulting from errors in forest inventory data.

Data assimilation (DA) is an approach to merge temporally separated data about some feature of interest. The data may be acquired using different techniques. In the realm of forest inventory, DA has the potential of improving the accuracy of the information and also to provide an estimate of the uncertainty of the information (Czaplewski & Thompson 2008; Ehlers et al. 2013). The development and use of DA has its history in, e.g., meteorology, where large amounts of spatiotemporal data are used to forecast the weather (Ghil & Malanotte-Rizzoli 1991; Lahoz et al. 2010). In essence, DA is a process which can merge data from different sources into a single usable source. One feature of this process is the ability to combine the estimates of uncertainty from each data source to provide updated estimates of the uncertainty for the information (Ehlers et al. 2013, Nyström et al. 2015). In a forestry context, a typical setup could be to keep the information up to date by integrating growth models in the DA process (Nyström et al. 2015) and to use remote sensing to obtain new estimates of the target forest information at regular intervals at low cost (McRoberts et al. 2010).
The technical implementation of DA can be done through a variety of approaches. Two commonly applied approaches are the Kalman filter (Welch & Bishop 2006) and Bayesian statistics (Dowd 2007). Comparatively, the Kalman filter is simple to apply while the Bayesian approach is relatively demanding. In both cases, existing (prior) information is forecasted to the time point when new data are acquired. The forecasted and new information are then merged. An updated (posterior) estimate is obtained as a weighted average (e.g., Ehlers et al. 2013; Figure 1). Through this process the quality of the information will be improved by assigning less importance to the information with lower quality, updating both the estimate and the estimates of uncertainty are also updated. This provides the forest planner with information on both the point estimate of the study variable and its corresponding uncertainty. An additional feature of the Bayesian approach is the estimation of probability distributions of the study variables. Special features of the processes studied require special attention. For instance, large changes (i.e. harvesting actions or storm fellings), require additional change detection procedures (e.g. using multi-temporal remotely sensed data) to identify in what areas DA cannot be routinely applied.

While the current use of DA in forestry inventory applications is rather limited, research to the proper application is ongoing. One concrete example of applying DA to improve forestry inventory estimations is the work of Nyström et al. (2015). The DA process applied the extended Kalman filter (Welch & Bishop 2006) with univariate models and a few simplifying assumptions. The forest state information was updated by the inclusion of both forecasting models and estimates of the forest state obtained through laser scanning combined with field reference data from sample plots. The results suggest the assimilation process improves the estimates of forest information over either only forecasted estimates.

https://mc06.manuscriptcentral.com/cjfr-pubs
and the most recent estimates from the remotely sensed data. For more detailed description of the applications readers are directed to Nyström et al. (2015).

Forest planning uses information about the current state of the forest to predict future results of different forest management alternatives. Forest planning is motivated by the specific objectives of the decision maker (Davis & Liu 1991; Edwards & Steins 1999; Kazana et al. 2003; Leskinen et al. 2009; Pukkala 1998; Randhawa et al. 1996; Borges & Hoganson 2000). One simple economically orientated objective is to maximizing profit or net present value (NPV). With more complicated objectives (i.e. multiple goals) involving several spatiotemporal scales (Duvemo & Lämä 2006; Duvemo et al. 2014; Kangas 2010), the need for accurate forest information increases. The information that relates to the objectives of the decision maker have more importance; frequently these variables relate to the value of timber and pulp wood in the forest stands (Bettinger et al. 2009; Davis et al. 2001; Eriksson 2008).

Forest planning occurs on a variety of temporal and spatial scales. The selection of scale may imply the selection of specific management goals. For instance, in landscape level long term planning the goals may involve nature conservation, carbon sequestration and balancing the volume, species composition, and the distribution of harvest assortments. Alternatively, short term local scale harvest scheduling may focus solely on facilitating timber procurement and logging procedures, with an aim to provide the raw materials required for industrial demands (Bettinger et al. 2009; Davis et al. 2001; Eriksson 2008). Considerations in the importance of uncertainty can differ between different spatial and temporal scales.
Even though uncertain information may lead to suboptimal decisions, the magnitude of the suboptimal loss occurring as a consequence of imperfect forest information is never known (Kangas 2010); however, it is often estimated (Holmström et al. 2003, Saad et al. 2014). Several studies (Holmström et al. 2003; Kangas et al. 2014; Kangas 2010; Ståhl et al. 1994) suggest that cost-plus-loss analyses can be conducted as a means to assess the appropriate level of information quality for certain cases. Incorporating uncertainty into the planning process can be difficult due to intricate mathematical algorithms and the limited ability of traditional mathematical programming methods, such as linear programming, to account for uncertainty (Hoganson & Rose 1987; Pukkala 1998; Pasalodos-Tato et al. 2013, Ferreira et al. 2016). In complex forest planning and decision situations, to explicitly incorporating uncertainty into the optimization model may be very difficult (Mowrer 2000). Therefore, forest planners in practice often ignore uncertainty for simplicity.

Uncertainty in inventory information increases through time as growth models are used to update the forest information (Nyström & Ståhl 2001; Fig 2). While the growth models may be of high quality, predictions are simplifications, and there are no techniques available to remove the uncertainty of predictions of the future forest state (Pietilä et al. 2010). The tool for controlling this uncertainty is to collect new information.

The objective of this study is to explore, highlight and discuss the potential benefits in forest management planning of using DA processes in forest inventories. The information provided by the DA
process contains novel features, but there are also challenges in applying DA. We highlight the potential benefits of DA information for different forest planning contexts and discuss the challenges.

2. The potential for using DA in forest planning

DA has the potential to reduce uncertainties and provide estimates of uncertainties which can be used to improve forest planning. With DA, new information is combined with the prior information producing an improved estimate (called posterior information in Bayesian statistics) (Dane & Horowitz 1965; Erdem & Keane 1996; Prueitt & Park 1997). Rather than discarding information following the acquisition of new information, DA offers a framework for continuously building the new information on the old information, which is updated and merged with new observations (Figure 1). This will allow the decision maker to revise his choice of action and thus to overcome uncertainty.

The potential benefits of DA depend upon the use of the improved information in the forest planning process. This depends upon the nature of optimization models or decision methods underlying the applied forest planning tool. For forest planning tools based on deterministic optimization models, point estimates of relevant forest variables are used to reflect the current state and future development of the forest. The main benefit of using DA in planning originates from the improved accuracy in the initial state of the forests. Through a sensitivity analysis, the robustness of the planning results can be evaluated. The estimates of uncertainty obtained using DA are helpful in defining the relevant intervals of uncertain forest variables to be tested in sensitivity analysis. However, the sensitivity analysis may
reveal that the planning results vary dramatically with changes in the values of the uncertain variables. In that case, there is no obvious approach to evaluate which forest plan is optimal based on the results of sensitivity analysis. In other words, if one applies a forest planning tool based on deterministic optimization models, one cannot (fully) utilize the estimates of uncertainty. However, if one has access to a planning tool with a stochastic optimization model, then both the point estimates and estimates of uncertainty from DA can be used in the planning process, which can provide larger benefits than in the case where only the point estimates are used (Birge & Louveaux 2011).

There are a variety of optimization methods available which can integrate estimates of uncertainty into the planning process. The choice of which method to use depends upon on the requirements of the optimization model and its tractability. For instance, robust optimization (Bertsimas & Sim 2004) can be used to protect against the infeasibility of the constraints caused by the potential of uncertainty. In a road building and harvest scheduling problem, Palma & Nelson (2014) introduced uncertainty into the timber estimates, and found that the solution of the robust optimization model to solve road building and harvest scheduling is less sensitive to uncertainty in timber volume information as compared to the deterministic model commonly used in forestry. DA can potentially improve the quality of solution by improving the quality of the timber estimates. One benefit of robust optimization is the limited increase of problem size in comparison to the linear equivalent (Bertsimas & Sim 2004), thus if the linear equivalent is tractable, the robust version should also be tractable.

While robust optimization protects against the infeasibility of specific constraints caused by uncertainty, stochastic optimization integrates the uncertainty into the entire problem formulation (Birge & Louveaux 2011). Stochastic programming problems can be formulated in many different ways, as
uncertainty can be considered in the objective function as well as in the constraints. In forest planning, stochastic optimization is being researched to demonstrate the potential for the implementation into practice (Garcia-Gonzalo et al. 2016; Eyvindson & Kangas 2016b). Stochastic programs have the potential to answer different questions than their deterministic counterpart. For instance, issues of individual risk preferences can be accounted for at the holding level. Related to the issue of DA, the specific timing of when the data should be updated can be formulated as a multi-stage stochastic optimization problem. Based on the preferences of the decision maker, and optimization model used, the improvement of forest information could be timed specifically (Eyvindson et al. 2017).

One major hurdle to implement stochastic programming is the issue of problem size, which could easily become too large to be tractable. If the entire stochastic problem needs to be formulated, issues of tractability can be a major concern in forest planning problems (Eriksson 2006). One way to maintain the tractability of stochastic optimization problems is to include a finite number of the possible values of uncertain variables, e.g., through a set of scenarios (Birge & Louveaux 2011). The set of scenarios should be large enough to appropriately reflect the uncertainties being considered and should be small enough to keep the model tractable. The optimization problem should direct the discretization of the set of scenarios, rather than simply trying to create a strong approximation to the original distribution (King & Wallace 2012). For each optimization problem, the selected scenario set should be tested for stability and solution quality, a variety of tools have been developed for this purpose (e.g. Kleywegt et al. 2001; Bayraksan & Morton 2011). It is intuitively clear that the smaller the uncertainty is, the smaller number of scenarios are needed to produce a good approximation of the uncertain variables. Therefore, the use of DA can promote the tractability of the problems by reducing the uncertainty, which will be reflected in the appropriate scenario set size used (Eyvindson & Kangas 2016a).
Whichever optimization method one uses to integrate estimates of uncertainty in forest planning, the planning process typically becomes more complex and more costly. To highlight the benefit of using methods which integrate estimates of uncertainty in forest planning, the value of the solution can be evaluated. The value depends upon the specific problem and the preferences of the decision maker. When dealing with individuals, risk preferences vary considerably, and the value of improved information depends on the individual decision maker’s acceptance of risk. At this level, the potential benefits of integrating estimates of uncertainty can be evaluated through the value of the information. This can be calculated directly by comparing the optimized results from different levels of data quality (Kangas et al. 2014). In more complex decision situations involving several decision makers or intangible benefits, the value of the improvement is more subjective and difficult to estimate. It depends upon the subjective valuations of the decision maker(s) of the increased quality of the management plan.

We would like to emphasize that the costs associated with implementing DA can be justified only if the use of information from DA can result in adequately large improvements of the management plan. One way to value the improvements in management plans is through the cost-plus-loss technique (Holmström et al. 2003; Eid 2000). At stand levels, these studies identified the potential value from obtaining perfect information by preventing losses. However, perfect information is not possible to obtain, so the comparison could be made between with and without the use of DA. Similarly, Ståhl et al. (1994) proposed a Bayesian approach to evaluate if an updated inventory should be conducted to maximize the expected NPV. Both methods suggest that new information should be collected if the new information has the potential to change the decision taken. For the case when the objective is to
maximize NPV, Holmström et al. (2003) suggests that new information should be collected only for stands which are near the potential for management actions. However, in short term planning where industry supply is addressed, the case may be different (Duvelmo et al. 2014). With the advent of new low cost information (i.e. from remote sensing) at scales larger than individual stands, the DA process holds substantial advantages to forest planning.

Implementing techniques which incorporate all of the information provided by DA will require significant changes to the current decision support system (DSS) tools and additional education for forest planners. The current DSS tools are designed to simulate forest growth and development through deterministic models, with forest information expressed as point estimates. Depending on the optimization tool being used, adjustments can be made to current DSSs to generate the required information for the optimization models. For instance, simulators can integrate inventory and growth model errors and produce a large number of scenarios for use in stochastic programming. Once integrated into the DSS, forest planners will need to understand the changes, and be able to inform decision makers of the potential impact on the planning process. Thus, to integrate DA into current DSSs will require additional development of the tools, on both the data processing side and the optimization side.

Below we list and discuss five reasons why DA processes have potential to improve forest planning. These possibilities are important to consider in forest planning, as DA processes are likely to be implemented in forest inventories in the future.
1- The accuracy of the information will be improved since new data are continuously merged with old forecasted information. DA processes can incorporate series of remote sensing data that may otherwise be difficult to use. This will improve accuracy and increase the probability of making correct decisions and thus improve the forest planning and decision making processes. DA also offers a cost-efficient means to utilize all new sources of information (i.e., at the given cost of the inventories) and will ensure that the posterior information always has the highest possible accuracy.

2- The information will be kept up-to-date even though no new measurement is made. A backbone of the DA process is the forecasting mechanism, which can be applied even if no new measurement is made in a certain time period. Thus the existing information will always be up-to-date, which improves the planning possibilities. Also, whenever new data arrive these will be assimilated with the existing information whereby up-to-date posterior information is obtained. Changes in the forest due to forest management, such as thinnings, will be continuously monitored (Kangas 1991).

3- The DA process will provide information with estimated accuracy. Contrary to the current situation where databases typically contain only point estimates DA databases will comprise uncertainty estimates as well. Estimated accuracy of the estimates is important since the decision maker may, at least intuitively, utilize this knowledge in the decisions. With this kind of knowledge the scenario analysis technique could be applied as an add-on to existing DSSs. For example, different starting values could be simulated and the effects on the decisions evaluated. If this is repeated many times utilizing the estimated accuracy of the information, the consequences of using data with the given level of uncertainty can be evaluated.

4- Stochastic decision making methods can be applied, which can integrate the estimated uncertainty of the information into the decision making process. Several DA processes provide
entire (joint) probability distributions of true values, which can be used in stochastic optimization methods. In addition, Bayesian decision theory (Hirshleifer & Riley 1979) might be applied as decisions are selected based on evaluations over the entire range of potential true values of the state variables. To utilize this possibility, DSSs would need to be further developed to account for probabilistic state descriptions and forecasts rather than basing the calculations on point estimates. This would imply a paradigm shift in planning and it must be carefully evaluated to what extent it would be possible to apply this in short and long term planning, due to the substantially larger problem spaces that will be encountered. With stochastic decision making, many important features of decision making under uncertainty can be incorporated, such as the risk preferences of the decision maker.

5- DA data allows for the analysis of optimal data acquisition decisions. As an extension to the fourth point, use of DA data has potential to not only consider the uncertainty in the information for traditional forest management decisions (such as thinning and clear-felling) but also to analyze whether or not it would be cost-efficient to acquire new information. This was demonstrated by Ståhl et al. (1994) in a research study where Bayesian decision making was incorporated in a dynamic programming setting. It was shown by Kangas et al. (2014) that acquiring new information while optimizing the harvest decision is profitable. In this case, the challenges linked to developing DSSs for practical uses would be even larger than in the fourth point. On top of the general Bayesian decision making algorithms there would also be a need for algorithms that evaluate data acquisition alternatives. This would increase the dimension of the problem even further.
3. Concluding remarks

Improving forest information through DA processes offers several benefits to forest planners. The primary benefits are the improved accuracy of the current forest information and the uncertainty estimates surrounding this information. To utilize the benefits of DA, current DSS tools require the ability to explicitly incorporate information about the uncertainty of forest information and make modifications so that stochastic optimization tools can be used. There are several techniques applied in research which can handle uncertainty, but that implementation in DSSs in practice seems to be missing except in SIMO (e.g., Rasinmäki et al. 2009); however, the application that consider uncertainty in SIMO is not yet widely used. Thus there is a need to develop DSSs that can incorporate uncertainty in the decision making process, e.g., through Bayesian approaches where the probability distribution of true values can be utilized. Furthermore, DA systems in forestry need to be further investigated and developed in order to be implemented properly in forestry. Only a few empirical studies of using DA for forest information (e.g., Nyström et al. 2015) have been conducted so far and it is recommended to further assess the benefits of DA in forest inventories.
References


Eid, T. 2000. Use of uncertain unventory data in forestry scenario models and consequential incorrect harvest decisions. Silva Fenn. 34:89-100.


Saad, R., Wallerman, J., Holmgren, J., and Lämås, T. 2015. Local pivotal method sampling design combined with micro stands utilizing airborne laser scanning data in a long term forest management planning setting. Silva Fenn. 50(2). DOI: http://dx.doi.org/10.14214/sf.1414


Welch, G., and Bishop, G. 2006. An Introduction to the Kalman Filter. Chapel Hill, NC, USA.
Yousefpour, R., Bredahl Jacobsen, J., Thorsen, BJ., Meilby, H., Hanewinkel, M., and Oehler, K. 2012. A
review of decision-making approaches to handle uncertainty and risk in adaptive forest management

Figure 1. The forecasted information of the timber volume (dotted line), i.e., prior distribution, is
combined with the new information (dashed line) in order to obtain the posterior distribution (solid
line), which results in an updated estimate of the timber volume. As shown the posterior distribution is
narrower compared to the prior distribution.

Figure 2. An illustration of the timber volume development’s distribution when the growth function
evolves over time. As shown the variation in the timber volume increased.
The forecasted information of the timber volume (dotted line), i.e., prior distribution, is combined with the new information (dashed line) in order to obtain the posterior distribution (solid line), which results in an updated estimate of the timber volume. As shown the posterior distribution is narrower compared to the prior distribution.
An illustration of the timber volume development's distribution when the growth function evolves over time. As shown the variation in the timber volume increased.

301x228mm (72 x 72 DPI)