Author(s): Cheikh Mohamedou, D.Sc., Annika Kangas, Alireza Hamedianfar and Jari Vauhkonen

Title: Potential of Bayesian formalism for the fusion and assimilation of sequential forestry data in time and space

Year: 2022

Version: Preprint version

Review article

Title

Potential of Bayesian formalism for the fusion and assimilation of sequential forestry data in time and space

Authors

Cheikh Mohamedou [1], Annika Kangas [2], Alireza Hamedianfar [1], Jari Vauhkonen [1,*]

Affiliations

[1] University of Helsinki, Department of Forest Sciences, Latokartanonkaari 7 (P.O. Box 27), FI-00014 Helsinki, Finland

[2] Natural Resources Institute Finland (Luke), Bioeconomy and environment unit, Yliopistokatu 6 B, FI-80100 Joensuu, Finland

* Corresponding author, e-mail jari.vauhkonen@helsinki.fi

Abstract

Forest resource assessments based on multi-source and multi-temporal data have become more common. Therefore, enhancing the prediction capabilities of forestry dynamics by efficiently pooling and analyzing time-series and spatial sequential data is now more pivotal. Bayesian filtering and smoothing provide a well-defined formalism for the fusion or assimilation of various data. We ascertained how often the generic, standardized Bayesian framework is used in the scientific literature and whether such an approach is beneficial for forestry applications. A review of the literature showed that the use of Bayesian methods appears to be less common in forestry than in other disciplines, particularly remote sensing. Specifically, time-series analyses were found to favor *ad hoc* methods. Our review did not reveal strong numeric evidence for better performance by the various Bayesian approaches, but this result may be partly due to the challenge in comparing a variety of methods for different prediction tasks. We identified methodological challenges related to assimilating predictions of forest development; in particular, combining modeled growth with disturbances due to both forest operations and natural phenomena. Nevertheless, the Bayesian frameworks provide possibilities to efficiently combine and update prior and posterior predictive distributions and derive related uncertainty measures that appear under-utilized in forestry.

Keywords: Forest inventory; Hierarchical Bayes Model; Kalman filter; Markov Chain Monte Carlo (MCMC); Credible interval

1. Introduction

1.1. Motivation and objectives

Systematic forest inventories have been carried out for over a century (Kangas et al., 2018a) and have utilized remote sensing and other digital map data already for decades to estimate forest variables (Katila and Heikkinen, 2020). While historical and new data have considerable potential to improve forestry-related predictions, this is not self-evident and may not be realized unless sampling, modelling and estimation methods are used appropriately with respect to the different properties of the data sources (Kangas et al., 2018b, 2019). It becomes valid to ask whether lessons learned from data fusions in other fields could be applied to rationalize forest inventory data processing. In particular, the field of engineering has developed a standardized, formal approach for estimating the state of the system through noisy observations (Särkkä, 2013). Many of the observations made by Särkkä (2013) on generic measurement systems can easily be extended to forestry data: even with the most carefully measured field plots, much of the signal may remain hidden (i.e., the forestry dynamics that we attempt to model). Instead, we must deal with “noise” (Särkkä, 2013) in the form of measurement, model and sampling errors (cf., Kangas et al., 2019).

The sequential filtering, smoothing and prediction process described by Särkkä (2013) is based on the Bayesian approach that exploits the posterior distribution of model parameters in contrast to analyses that optimize a predefined objective function. Based on reviews and applications of Bayesian data fusion in forest inventories (Varvia, 2018) and ecosystem modeling (Van Oijen, 2017; Mäkelä, 2020), the Bayesian approach provides a well-defined formalism to 1) define a problem in a practical probabilistic framework, 2) incorporate *a priori* (expert) knowledge to the observed data, 3) incrementally update the posterior distribution as more data become available, and 4) express and incorporate the uncertainty of estimated model parameters and predictions. These properties suggest that fusing data, while quantifying the related uncertainties based on exact schematics of the Bayesian approach could be beneficial for forest inventories.

In this study, we reviewed the concepts of Bayesian data fusion, especially with regard to the applicability of these concepts to sequential forestry data. We defined the concepts of sequential Bayesian estimation (Section 1.2) and specified the interests for the review (Section 1.3). The following sections specifically review:

- How often algorithms related to Bayesian smoothing, filtering, and prediction have been used in forestry to date, compared to other scientific fields, and what features do the reported key applications of these concepts provide to forestry (Section 2)?
- To what extent do forestry applications that use either *ad hoc* or Bayesian methods, and in the latter case, take advantage of the features related to Bayesian smoothing, filtering, and prediction(Section 3)?
- Currently realized numeric and potential future benefits from standardizing the formulation of these problems using Bayesian approaches and proposals (Sections 4–5).

1.2. Key generic concepts

As in Knödel et al. (2007), we consider “fusion” and “assimilation”, as well as “combination”, “integration”, “merging”, “synergy” and “interaction” as possible synonyms for the purpose of using information from various sources with an objective to improve extraction of relevant information from the data. We describe Bayesian estimation as a means to these purposes, i.e., to estimate the value of an unknown random variable θ given the series of observations $y_{1:T} = \{y_1, y_2, …, y_T\}$, by updating the posterior probability distribution $p(\theta | y_{1:T})$ using the Bayes’ rule:

$$p(\theta \mid y_{1:T}) = \frac{p(y_{1:T} \mid \theta)\, p(\theta)}{p(y_{1:T})} \quad (1)$$

where $p(\theta)$ is the prior distribution assumed of the phenomena before actual measurements, and $p(y_{1:T})$ is a normalization constant. As the latter is independent of θ, it is often ignored for the posterior distribution or is approximated, the consequence of which is derived on p. 18 or pp. 118–120, respectively, in Särkkä (2013). The posterior distribution is obtained by assimilating observations and prior knowledge. The point estimate is, for example, the maximum (termed “Maximum A Posteriori”, MAP estimate) or other similar statistic of the distribution.

Because of the intractability of computing the posterior distribution for the full history of observations, our interest here is in the marginal posterior distribution of the given state. The states and related measurements form sequences, for which reason the application of Eq. 1 is called sequential Bayesian estimation. Depending on the marginal posterior distribution of interest (or measurements available for the state to be estimated), the estimation can be divided into sub-categories of smoothing, filtering and prediction (Särkkä 2013, p. 11). An analogy between these categories and the typical steps related to the processing observed and predicted forest inventory data can easily be elucidated (Figure 1):

- Smoothing distributions — when the interest is in the state before the current measurement;
- Filtering distributions — when the interest is in the current state; and
- Prediction distributions — when the interest is in a future state,

and the current and previous measurements are taken into account in all three categories.

[FIGURE 1 AROUND HERE]

The sequential Bayesian estimation is carried out using recursive equations, in which the posterior probability distribution of interest is initialized with the prior and then estimated by repeating the update and prediction steps to:

- update the state by combining the prior and observation likelihoods by the Bayes rule (Eq. 1),

- predict the next state-distribution by propagating the state of the previous measurement according to the specific transition model.

The updated probability distribution is obtained by multiplying the prior and observation likelihoods, whereas the predicted distribution is the integral of the products of the probability distributions associated with the current state and the transition from the previous to the current state. As elaborated by Särkkä (2013), different Bayesian filtering and smoothing algorithms adapt these general equations: for instance, the updating and prediction equations in the well-known Kalman filter (Kalman 1960) assume linear Gaussian models for both the measurements and the transition.

1.3. Specifications for the review

While the analogy between Bayesian concepts and forestry applications has been noted, the juxtaposition of generic concepts and forest inventory applications, as reviewed by Kangas et al. (2019) for example, highlighted a number of issues that guided our approach here. In our review, we re-considered *a priori* information (i.e., prior in time (Särkkä 2013)) with regard to the options (filtering, predicting and smoothing) related to time sequences (Figure 2a). Typical forest inventory applications (Figure 2) add at least another dimension to *a priori* information: the spatial element of the forest data that introduces sampling error to the estimates in addition to the temporal noise in observations over time. Thus, in addition to a time-series, forestry applications may account for spatially structured errors in sequential data. For example, kriging (e.g., Cressie 1993) can be seen as a special case of Bayesian inference with the prior acting as the spatial correlation. A simpler model with a constant correlation assumption, such as Empirical Best Linear Unbiased Prediction or the EBLUP estimator, is a special case of kriging. All these are referred to as spatial models.

However, if we consider "prior understanding of the phenomena modelled", for instance, realized as the model form, then *a priori* information must be, at least, three-dimensional. First, this type of prior knowledge can be related to neighboring trees or stand(s); for instance, the application of regression models results in additional information that is essentially new information for one time point in space, although the estimation can also employ data from multiple time points (Figure 2b). Second, all Bayesian models are weighted models and use different inferential methods, and the model forms produce additional information to estimate the state at a time point (Figure 2c). Third, further information could be based on the prediction of results for smaller computation units using the whole dataset of a large area-of-interest, i.e., using a “small area estimation” type of method (Figure 2d). Building upon these considerations and attempting to verify our stated hypothesis on the temporal vs. spatial dimensions of *a priori* information, we searched for published papers that considered Bayesian estimation for sequential observations in time and space.

[FIGURE 2 AROUND HERE]

In addition, we found that it was useful to contrast between approaches that were formally defined as Bayesian data fusion and those termed *ad hoc* methods. For the latter, we refer to the Cambridge Academic Content Dictionary (Cambridge University Press) that defines *ad hoc* as a method "for a particular purpose or need, especially for an immediate need". By this distinction, we acknowledged that the tasks outlined in Figures 1–2 can be accomplished by means of a method (chain) that is mathematically less reasoned but was selected due to its simplicity to solve the particular task. We wanted to determine the possible benefits of formally defined Bayesian methods that associate the data points with the probability or uncertainty of the event. Therefore, we categorized studies that, without any supporting information, weighted data points equally as *ad hoc* methods. The latter category includes the maximum likelihood of the joint probability of measurements, which can be seen as a MAP-estimate with uniform prior (Särkkä 2013).

2. Systematic review

In this review, we used a systematic keyword search to approximate the extent that the algorithms proposed for the Bayesian approach are used in forestry vs. other disciplines. The construction of the literature search and its findings are described in Table S1 and Figure S2, respectively, of the Supplementary material. Based on the search, we found a total of 21 articles for qualitative analyses that were required to be related to forest attributes with ground truthing, i.e., we excluded those articles related to land use/land cover or similar classifications based on remotely sensed data. The resulting studies ranged from the estimation of forest stand characteristics to forest inventories and classification of species, including a varying number of data points and sources. Data types varied between simulated and observed data but were mainly based on inventory sample plots where the parameters of interest were measured. The studies were mostly focused on boreal forest but all biogeographical zones (Boreal, Temperate, and Tropical) were covered.

The list of studies that could be compared in terms of improvement or the rate of error change in percentage points, computed as the difference between two relevant error rates thereby demonstrating the improvement from the reference method or single point in time by means of Bayesian or similar methods is shown in Table 1. The root mean squared error (RMSE) was considered as the main error criterion; in cases where it was not stated in a publication, an alternate error measure (e.g., coefficient of variation, variance, error ratio, or error increase or decrease based on some benchmark criterion) was recorded. If multiple forest variables were evaluated within a particular study, we either selected the inventory variable of main interest, the variable considered most representative among the multiple variables studied, or the variable that was most comparable between studies (usually the growing stock volume (GSV) or above-ground biomass (AGB)).

In the Sequence-column of Table 1, we indicated the main dimension (temporal or spatial) of the variability accounted for by the studied methods. Hou et al., (2019) could not be unambiguously classified, as the improvement noted for that study was essentially based on information from another model rather than from these dimensions. In addition to the studies listed in Table 1, there were similar cases that could not be strictly assigned as temporal or spatial considerations, which is reflected in the sub-title structure of the qualitative review (Section 3). We found studies that used the Bayesian approach in an abstract form without comparing it to other statistical approaches (de Groot et al., 2019; Mölder et al., 2019), reporting an improvement (compared with another study or within their study), or whether the method was effective. Some studies did not report a measure that could be compared to those in Table 1, although improvements due to the use of the Bayesian method or data fusion were reported (Uusitalo et al., 2006; Picard et al., 2012; Lu et al., 2019). Further, the results of two studies were reported at the individual tree-level (Picard et al., 2012; Van Oijen et al., 2005), whereas Table 1 covered studies reported at the area-level. The studies listed in this paragraph were not included in the summary table, but are nevertheless qualitatively reviewed below.

[TABLE 1 AROUND HERE]

The error measures shown in Table 1 varied from 2.3% to 38.4%, with a median of approximately 13%. The rate of error change (improvement) ranged from 1% to ≈57%, with a median of 4.7%. The number of time points (Tp#) varied by up to 25. The time span in studies with multiple time points varied from 2 years to a maximum of 40 years (observed, Babcock et al., 2016) or 50 years (simulated, Ehlers et al., 2013) between data acquisitions. For growing stock volume estimations, Katila and Heikkinen (2020) utilized the greatest number of sample plots (42,541) to cover the largest area of interest. The computation units in Ver Planck et al. (2018) were stands that varied from 0.6 ha to 47 ha, with an average size of 6.6 ha.

In attempting to schematize the time, space and improvement factors in Table 1, we noted that the largest (>10 percentage points) improvements were reported based on relatively small datasets (number of plots × plot size <15 ha and total inventory area ≤10,000 ha). In contrast, the least improvements were reported for the National Forest Inventory (NFI) or similar inventory contexts where the area covered hundreds of thousands or millions of hectares. Between these scales, there was a notable absence of studies that considered regional scales or larger datasets that modelled small areas. The rate of the error change seemed to benefit from expansion of the temporal dimension by increasing the number of time points, but especially by increasing the interval of the updates. We came to this conclusion by comparing the rates of error change in studies #5 and #11 (with approximately 4.5 and 9 years between data acquisitions, respectively) to studies #8, #10, and #13 which reported more frequent data acquisitions. However, the remarks above should be treated with caution due to the small number of studies that measured different aspects.

3. Qualitative review of Bayesian methods and applications for forest variable estimation

3.1. Filtering and data assimilation

Bayesian or Kalman filtering (Kalman, 1960), also termed Data Assimilation (DA), was introduced for forest variable estimation by Dixon and Howitt (1979) and for forest inventory data updating and forest monitoring by Czaplewski et al. (1988). Gertner (1984) used filtering to merge forest growth estimates with observations. Interestingly, the motivation for his study was the relative inexpensiveness of utilizing growth projections compared to the collection of new observations, and the need for a method that was not as restrictive as the Kalman filter in terms of the model forms and data; both aspects were re-invented in later studies (Section 3.1.1). The work was further refined by the identification of samples to be collected in the future (Gertner, 1987), by updating the model parameter estimates of the different growth projections (Gertner, 1987; Gertner et al., 1999), and by dealing with uncertainty in both *a priori* or sampled information (Gertner and Zhu, 1994). Van Deusen (1987) employed the Kalman filter to analyze dendrochronological time-series. Kangas (1991) also presents one of the early filtering approaches, while Green and Strawderman (1992) utilized posterior distributions in Bayesian analyses. The Bayesian filtering and smoothing methods have been used in many fields of science (e.g., health sciences, ecology, learning, or adaptive systems) to employ prior and future observations (Särkkä 2013).

In some studies, Kalman filter and Bayesian methods were used in a manner that would suggest they were two different approaches (see Ehlers et al., 2013; Nyström et al., 2015; Fortin et al., 2020). However, according to Särkkä (2013), these can both be classified as Bayesian inference-based methods, where the Kalman filter falls under general Bayesian theory (see Section 1.2). It is important to note that Smoothing and Filtering are concepts of sequential Bayes estimation over time using incremental measurements (time-series data taking into consideration past and current observations) and should not be confused with the same terms as used in image processing or similar contexts. It is also worth noting that the use of formalism of time-varying measurements is for illustrative purposes and is not strictly adhered to or applicable to other sequential data types.

3.1.1 Applications of time-series filtering to stand-level inventories

Work by Ehlers et al. (2013), Nyström et al. (2015) and Lindgren et al. (2017) illustrated the potential benefit of DA for stand-level forest inventories with remotely sensed support data. Ehlers et al. (2013) tested two methods, a general Bayesian (providing distributions) and an Extended Kalman Filter (EKF; did not provide distributions but estimated mean and variance) with simulated data. When the two methods were compared, generally higher predicted variance values were reported with the Bayes method compared to EKF. The mean values were nearly equal, possibly due to the linearization of the growth model by the EKF. However, the variance depended on the prediction error of the growth model used. The methodology in this particular case study performed best in low-precision volume estimates for short time periods and with the use of an accurate growth prediction model.

Nyström et al. (2015) tested the EKF approach on empirical data. Updating past estimates of forest variables with growth models and integrating those with current models that use DA led to an improvement in the precision compared to using the target time-point estimates as such, although the increase was not high compared to the most recent estimate. The study highlighted that DA can be based in multiple Bayesian filtering approaches, but that the properties of the method affect the applicability. As explained in Section 1.2, the standard Kalman filter assumes Gaussian (normal) distribution both for the predictions and measurements, and for linear forecasting models. As many forest variables cannot be assumed to be normally distributed, the EKF approach based on the Taylor approximation is applicable to non-linear forest growth models (Lindgren et al., 2017).

The follow-up study by Lindgren et al. (2017) was motivated by improved data availability as they used alternative remote sensing data (Synthetic Aperture Radar) and Bayesian updating to predict the interval before the next optical remote sensing dataset was available. The time-series of Nyström et al. (2015) consisted of 6 observations over eight years. While 19 observations over 4 years (Lindgren et al. 2017) can be considered an improvement, it is still far less than in other study fields that use DA (e.g., engineering, signal processing and meteorological fields).

3.1.2 Applications of time-series filtering in National Forest Inventory (NFI) contexts

Fortin (2020), building upon the stand-level experiences of the Bayesian methods reviewed above, increased the NFI sample size by updating the sample plots from past inventory campaigns, which were then used in combination with the new plots. Use of the updated plots increased the precision of the estimates, and the Bayesian method yielded similar improvements (best-case coefficient-of-variation ~ 2.28) compared to a multiple imputation method. A similar value (2% error reduction) was reported by Katila and Heikkinen (2020) who fused historical data to Finnish MS-NFI estimates of growing stock volumes. This is the number of municipalities with significant estimate error reduction based on a GLS estimator, which used the covariance matrix to determine the weighted average from three time-points. The performance was possibly hindered by the relatively short time-series of only three time-points. Although the data fusion concept was not explicitly Bayesian, it follows the same concept of using prior data to improve existing estimates. Hou et al. (2021) list several methodological benefits of a procedure based on Bayesian DA with linear mixed models to combine results from a rotating panel inventory, measured in a cycle of 5–10 years, to a single date under a given sampling error requirement.

3.2. Time-series smoothing

As strictly defined by Särkkä (2013), diverse types of smoothing, such as specifically fixed interval smoothing (that uses all observations available for a specific target to make an estimate), fixed lag smoothing (implements latency in the steps, and uses the current values to update the earlier steps, and so on), and fixed point smoothing (starts as a Kalman filter, but at a specific point begins to backward update all previous measurements), were absent from the review. When the concept of smoothing is considered more generically, the study by Mäkinen et al. (2010) can be fitted to this category. In that study, forest data mining techniques were used to detect outliers in compartment-wise field inventory data. This approach is closely associated with machine learning concepts in identifying hidden patterns and undiscovered structures within a dataset. Suty et al. (2013) used past measurement data to investigate the bias introduced by field protocols to stem volume increment estimates for the Swedish NFI. The simulations in the study indicated that both the permanent and temporary types of inventory sample plots were insensitive to random measurement errors, although a theoretical chance of bias for larger trees was identified and attributed to the scarcity of these observations in the empirical inventory data. Neither Mäkinen et al. (2010) nor Suty et al. (2013) could be categorized as employing a Bayesian approach, but both shared similar aims with the smoothing concept in that they attempt to improve data reliability by investigating current and prior observations. Therefore, the concept of Bayesian smoothing could possibly be studied further to fit these types of applications.

3.3. Time-series prediction

Future predictions are an important use of forest inventory data, in addition to the various statistics related to the current state of forest resources (Kangas et al., 2019). Although it was difficult to isolate the exact prediction and updating steps in the frameworks reviewed, forecasting state time behavior that has not yet been measured can be obtained by iterating the prediction step of the optimal filter (Section 1.2; see also Särkkä, 2013). Below, we review the studies that used Bayesian approaches in a somewhat similar way to the generic concept described above.

An exhaustive European-wide study of growth models (Van Oijen et al., 2005) benchmarked Bayesian calibration, Bayesian model comparison, and Bayesian model averaging (BMA; see also Leamer, 1978; Fragoso et al., 2014) to account for either the parametric or structural uncertainty of the growth models. In particular, BMA provides a robust approach to predict forest growth, as it assimilates predictions by the empirical or process models as weighted averages with weightings that relate to posterior probabilities (see also Section 3.4). Lu et al. (2019) used the BMA method to model tree mortality in relation to environmental factors. According to Lu et al. (2019), stepwise regression was found to predict tree mortality less accurately than BMA; the latter exhibited a more narrow and reliable confidence interval, and greater accuracy associated with parameter estimation, which was clearly shown by the posterior probability.

Minunno et al. (2019) applied DA to calibrate process-based forest growth predictions based on NFI and permanent growth experiment data. The calibration guided the parameterization of the process model closer to experimental conditions, thus reducing uncertainties related to model outputs. In an area-based matrix model, European Forestry Dynamics Model (EFDM; Packalen et al., 2014), the transition matrix for forest stands between the states that represent forest development was estimated through a Bayesian procedure by two consecutive observations. The Bayesian approach of EFDM is connected to recursive filtering in the case of insufficient NFI plot data – in that case, the prior, computed from the observations or assumptions, is applied to fill the transition probabilities. Aside from making predictions, their up- or downscaling may be desired. Tian et al. (2020) theoretically demonstrated the breakdown of stand growth to individual trees using Bayesian calibration of a whole-stand growth series with diameter distribution of one time point.

Bayesian methods become useful for future predictions that involve uncertainty in the model predictions and inventory observations. Nyström and Ståhl (2001) showed that Monte-Carlo simulation could estimate error propagation in growth models that often need extensive simulation to obtain reliable estimates. Quantifying and reducing uncertainty (Section 3.5) requires computational considerations, as reviewed by Van Oijen (2017) and summarized in Section 3.6.

3.4. Using spatial sequential data and other information

Many of the Bayesian approaches applied in the reviewed studies did not fit within the structural definition of system state and time as defined by Särkkä (2013). In addition, our review highlighted many other Bayesian approaches that have been used for different types of inferences, which could not be categorized as accounting for temporal or spatial variability. Below, we review the approaches based on spatial data and those categorized as benefiting from additional information from other models.

3.4.1 Using spatial sequential data

Finley et al. (2008, 2011, 2013, 2014) and Ver Planck et al. (2018) sought an initiative template for spatially explicit modeling of forest variables at the landscape-scale through remotely sensed covariates. For instance, Finley et al. (2014) modelled spatially misaligned light detection and ranging (LiDAR) data and sample plots to yield predictive maps for biomass-related attributes. The Bayesian hierarchical approach allowed the uncertainty in forest canopy height metrics and variables measured from inventory plots to be associated with the candidate models. Predictions based on the posterior predictive distribution sampling averaged parameter estimates over uncertainty. The increasing prevalence of correlation structures between response variables, which could otherwise lead to poorer data fits, was successfully addressed. This concept, however, was limited by the computational workload of a complete multivariate geostatistical model, which resulted in the use of only 50% of the dataset (see Finley et al., 2008).

Babcock et al. (2016) used a Hierarchical Bayesian Modeling concept to couple LiDAR and long-term forest inventory data. That study, as well as the studies described in the previous paragraph, was based on the Markov Chain Monte Carlo (MCMC; see also Section 3.6.) approach as a numeric method to sample from the predicted posterior distributions of AGB to compute statistics related to mean, variance, and credible intervals of the distributions. The method is relatively easy to implement for sequential DA with the Bayesian system and it allows for appropriate complex parameter associations and the propagation of uncertainty on through to prediction.

3.4.2 Using additional information from another model or inferential method

A number of variants that worked under the same principles but were not strictly in Bayesian formalism were identified in our review. First, the studies of Van Oijen et al. (2005) and Picard et al. (2012) can be regarded as data fusion of multiple models to retrieve a fused, single estimate of the variable of interest. In their abstract, Minunno et al. (2019) referred to this principle as "model data assimilation". In the BMA approach (Section 3.3.), a weighted average of probability density functions, based on the individual predictions, is the predictive function of the quantity of interest. The weightings of the models that produce the predictions are equal to the posterior probabilities. The BMA predictive variance can be split into two parts: one corresponding to the variability between the models and the other to the variability within the models. Notably, Katila and Heikkinen (2020) used a similar concept for improving the estimate, and both studies reported a small but consistent improvement in the forest variables of interest.

As an alternative to text book Kalman filtering, Hou et al. (2019) employed DA by incorporating seemingly unrelated regressions (SUR) and best linear unbiased prediction (BLUP). The justification to develop this approach was to circumvent the need for the continuous collection of observations before updating by means of Kalman filtering, which might not be operationally feasible with the non-permanent network of sample plots. Junttila et al. (2008) and Zhao et al. (2020) used a Bayesian regression technique called Sparse Bayesian Modeling (Tipping, 2004), where the model was designed to compare various weighted combinations of feature values with each other to obtain optimum weight distribution and an optimum collection of features. When used in different forest zones (as in Zhao et al., 2020), the method showed equal efficiency, especially when limited sample plots were available. Finally, Bayesian spike-and-slab regression (Mölder et al. 2019) is an option to utilize *a priori* information in modelling. Spike and slab (Mitchell and Beauchamp, 1988) refer to the type of prior regression coefficient used in linear regression models. These terms assume that the regression coefficients are mutually independent with a two-point mixture distribution that consists of a uniform flat distribution (slab) and a degenerate distribution at zero (spike). This method is particularly useful when the number of possible predictors is greater than the observations.

3.5. Managing uncertainty

Bayesian methods allow more flexible management of uncertainties in parameter estimates than the more frequently used statistical approaches. These features are specifically reviewed here, because comprehensive uncertainty analysis would decrease the chance of non-optimal decisions in various applications, such as ecological risk assessment, forest planning, inventory sampling design, and environmental decision support. The Bayesian school offers a quantitative measure of uncertainty (on the basis of available evidence) as the probability around an estimate. Conceptually, the notation is close to the frequently used confidence interval concept, which is interpreted (e.g., with 95% confidence level) such that in 95% of hypothesized repeats of the experiment, the true (unknown) estimate would lie within the lower and upper limits of the interval (the parameter is a fixed value and the limits are random values). Interpretation of the Bayesian credible interval in a corresponding case would be that the true (unknown) estimate lies with 95% probability within the interval (the estimated parameter is a random value, while the limits are fixed). The credible interval is dependent on the evidence provided by the observed data, and it corresponds to the confidence interval in case of uninformative (uniform) prior (cf., discussion of ML and MAP estimates on p.18 of Särkkä 2013).

Theoretically, the inclusion of a large number of data streams into the assimilation may enhance the data fusion result, although the addition of data at different scales may cause bias or inconsistency between the content of various data observations, or among the input data and the processing model (MacBean et al. 2016). Thum et al. (2016) found an inconsistency when assimilating both annual increment and total biomass data to improve the broader period of mortality and turnover processes. Minunno et al. (2019) demonstrated a reduction in the uncertainty of parameters after applying the calibration. Therefore, it is essential to detect whether DA of one data stream produces a better or worse fit, and whether this should lead to the determination of an optimal fit among the datasets.

Varvia et al. (2017) used the Bayesian approach to investigate the feasibility of posterior probability density and point estimate measures for area-based forest attribute estimation at the plot-level. Posterior variances and credible intervals were generated for species-wise growing stock volumes and used for uncertainty analyses. The study found that the Bayesian 95% CI provided a reliable measure for the estimated uncertainty when the training datasets were well distributed with a species-wise compartment. The point estimate of various species-specific forest variables was less accurate compared with a benchmark k-nearest neighbor estimate, although the Bayesian point estimate yielded a more accurate estimate for the total figure (i.e., all forest species summed-up) and, overall, the ability to report CI could be considered an asset. The methods underestimated the abundant tree species, such as pine and overestimated the less frequent tree species (e.g., deciduous). Regardless, the method exhibited equal robustness compared to other state-of-the-art methods for forest inventories.

Similarly, Mäkelä (2020) addressed the uncertainty in ecosystem modeling through the Bayesian approach of canonical correlation analysis (CCA; Hotelling, 1937), which is a technique for detecting correlations between two multivariate or random variables and extracting linear components that represent the correlation. The method was equally useful in identifying the uncertainty caused by varied factors on ecosystem modeling. For instance, it found that forest management was the dominant factor that contributed to the uncertainty in the study. The idea of identifying uncertainty elements is intriguing, especially with Bayes, as illustrated by Varvia (2018) and Mäkelä (2020).

3.6. Computational aspects

In summary, the adoption of a Bayesian approach requires that the modelling framework is defined as probability distributions of parameter uncertainty. Skewed and multi-dimensional distributions result in a high computational demand, which requires advanced algorithms (van Oijen, 2017). Aside from extensions (e.g., EKF described above), alternative approaches are available. When propagating errors in growth models, Nyström and Ståhl (2001) assumed Gaussian distributions, although other forms of distributions could lead to intractable calculations following their approach. Motivated by these drawbacks, Gove (2009) re-formulated the approach based on sequential Monte Carlo filters (particle filters) to allow for nonlinear, non-Gaussian assumptions, as well as the integration of new inventory information with model predictions. The presented filter is close to a regular Kalman filter but differs in the sampling mechanism, in which many particles generated by Monte Carlo methods represent random variation, while a small deterministic sample of the stated space is taken to estimate the mean and covariance of each state.

Somewhat cognate to the above-reviewed methods, Gibbs sampling reduces the impractical restrictions of the Kalman filter for real-world data analyses and was used by Green and Strawderman (1992). It was also employed by Liénard et al. (2015) to parameterize biomass transition matrices from forest inventories in order to predict forest development under disturbances (see also Liénard and Strigul, 2016). Itter et al. (2017, 2019) proposed a hierarchical model structure for the radial growth of individual trees. Their model hierarchy was built from fixed stand and tree parameters and climate parameters that evolve over time, thereby affecting the stand and, subsequently, tree growth. The model parameters were solved and updated by means of MCMC and Gibbs sampling, which is described in detail in Appendix S1 in Itter et al. (2017).

As elaborated by Van Oijen (2017), the application of Bayesian methods to more complex problems over time has shifted thinking from exact solutions of a single parameter vector to algorithms that approximate these parameters. The idea of MCMC, in brief, is to explore the parameter space toward the highest posterior probability, where the prior multiplied by the likelihood for that parameter vector guides the representative sampling from the posterior distribution. Yet, our review also noted more deterministic alternatives to MCMC. For instance, de Groot (2019) modeled spatially explicit forest management history and pest control by means of Integrated Nested Laplace Approximations (INLA; Rue et al., 2007), which is an alternative to the MCMC-based statistical inference in latent Gaussian models. The key benefit of INLA is that it has simpler computation based on individual posterior marginal model parameters, thereby avoiding posterior predictive simulations, and so permits rapid and accurate computations (Nothdurft, 2020). Indeed, Nothdurft (2020) estimated annual radial increments with a hierarchical model motivated by Itter et al. (2017, 2019), but that was solved by the INLA approach. We include this here as an example of how the Bayesian method can reduce structural uncertainty and simplify the process involved.

4. Discussion

4.1. Summarized key findings and limitations of the review

We used exact search terms based on Särkkä (2013) to discover relevant literature for the quantitative analysis (Figure 3). The success of a literature review depends on whether the concept and algorithm names were used in the published studies. We believe that our choice here was successful, since the search results had to be augmented by only a few papers in the qualitative analysis. According to our literature search results, there is a clear trend that studies employing Bayesian methods are slowly increasing. However, the adoption of these methods appears to be less common in forestry than in other disciplines, in particular remote sensing. It was previously identified that the possibilities to use Bayesian filters to evaluate the past and to predict the future are not commonly recognized in forest inventories or related studies (Kangas et al., 2019).

The median error change due to the introduction of a Bayesian or *ad hoc* approach for data fusion was 4.7 percentage points (Table 1). There was a strong variation between studies and the degree of improvement depended, for example, on the type of data, the forest variables of interest, target species, sample size, and study design. No more detailed numeric recommendation (such as a model explaining the performance) could be developed based on a meta-analysis, as the results varied considerably with regard to the aforementioned aspects. Therefore, it is important to give due consideration to the various characteristics of the experiment that is under study.

A detailed examination of Vastaranta et al. (2018) provides an example on how difficult it was to assign some of the studies in Table 1. In their work, they combined new inventory data from 2016 with older data collected in 2014 and updated to 2016 by a growth model. Their reported RMSE values improved from 26.0% (based on the use of 2016 data alone) to 20.4% with best-case weightings for the combination of the two datasets. This equates to an error change of 5.6 percentage points and was, therefore, included in Table 1. However, according to Vastaranta et al. (2018), merely updating the 2014 dataset would have yielded a change in 11.4 percentage points (from an RMSE value of 33.2% based on validation of the 2014 models with 2016 data to 21.8% based on the 2014 data updated with the growth model). Therefore, the improvement in RMSE attributable to the acquisition of new data, from 21.8% (best-case result with one time point) to 20.4% (best-case with two time points), is equal to 1.4 percentage points. This is in line with the other studies presented in Table 1 and indicates that the values included in our table may depend to some extent on the reporting practices of the individual studies.

In our review, we ignored many factors related to forest status, such as site index, management and development stage. On the one hand, we noted that the comparison of relatively similar studies (cf., Junttila et al., 2008, and Varvia et al., 2017, in terms of forest conditions and sample plot data) based on different Bayesian approaches yielded very different results. Junttila et al. (2008) reported an initial RMSE value of 19.9%, while Varvia et al. (2017) reported a value of 17.7%, and improvements of 1 and almost 10 percentage points, respectively. On the other hand, Hou et al. (2019) reported a 33% reduction in the RMSE value compared with non-calibrated values in juvenile forests, which are generally considered a difficult target. It is possible that predictions of forest stand characteristics, such as growing stock volume based on high-resolution auxiliary data, such as LiDAR are already highly accurate and may be difficult to improve upon, although the methods could differ in the prediction of species-specific and minor species' properties (although see Varvia et al. 2019 for contradictory results). Yet, the results described above also indicate that there are considerable differences within the Bayesian procedures that should be further explored.

Compared to some *ad hoc* methods included in the review, we did not find strong numeric evidence for better performance in the prediction of forest variables by the various Bayesian approaches. This result may be partly because we compared a variety of methods for different prediction tasks. However, our review highlighted the interpretation of the prior information compared to the theoretical representation (Sections 1.2. and 1.3; Särkkä, 2013). In Table 1, greater improvements were more frequently observed by the Bayesian approach with data from a single time point rather than employing a time-series. In these cases, the prior information either originated from spatial data outside the computation unit, from another model (BMA, SUR) or from other relevant information (location accuracy). Lower gains were observed for inventories that were employed at larger scales, where the sampling error may be the dominant source of error. Lower gains were also observed with large datasets and higher gains were noted with small datasets, where the original uncertainties can obviously be greater. This finding may suggest that the additional spatial information from neighboring trees or stands at one time point is more useful than actual prior information in time. However, this may require confirmation from studies where these aspects were specifically considered. The result could be explained by additional sources of variation that accrue from the requirement to use several data acquisitions over time, rather than just one used in space. At the national level, several acquisitions include remote sensing campaigns with different parameterizations that may result in additional variation.

4.2. Utilizing the full potential of Bayesian methods

Even though the adoption of a Bayesian approach resulted in unexpectedly small numeric benefits, it may rationalize the analyses by accounting for the features of the Bayesian approaches that we identified in the introduction section. In summary, one of the more promising disciplines of sequential Bayesian estimation is when evaluating observations at varying points in time; Särkkä (2013) approached optimal filtering and smoothing as the least-squares optimality of the posterior distribution of states of a system observed through marginal distributions of noisy, time-varying measurements. The estimate of the state space is affected by the prior probability distribution, transition probability distribution (a Markov chain), and measurement model. Särkkä (2013) further lists several numerical approximation methods (or categories) that can be operated based on different assumptions. In the following section, we consider how some of these properties are exhibited in related forestry studies.

4.2.1. Prior

The prior distribution, or assumptions on information used as the prior, was identified to have a major role in delivering accurate results. For instance, validations carried out over very different geospatial scales (e.g., Nyström et al., 2015; Fortin, 2020) all denote the significant contribution of an accurate growth model toward good performance. In DA studies, the variance related to growth predictions has generally been smaller than the sampling variance related to the acquisition of new plots, which has led to an emphasis in the utility of updating the plots with the growth model over the acquisition of new data. Yet, when focused only on the growth model, the meaning of prior is considerably narrower, and analyses could potentially be improved by re-thinking its meaning and role. Firstly, because of the long history of forest inventories, we always have some prior information to narrow down the interval of possible results. Secondly, even a minor degree of prior information can be useful when it is used judiciously: for instance, in imputing-like fashion with observations to fill gaps due to sparse field sample in NFI-based forest development matrices, which has been proposed by Packalen et al. (2014) and was applied as part of the EFDM software, for instance, by Vauhkonen et al. (2019).

4.2.2. Change model

In the context of estimating forest variables, the findings of Fortin (2020) and Kangas et al. (2020) suggest that the growth model should actually be considered as the “change model”, as it is used to predict all possible changes in the forest stand during the growing period; therefore, abrupt changes in forest conditions need to be addressed. Notably, both Fortin (2020) and Kangas et al. (2020) suggest disturbances (both natural and forestry operations) are the most difficult changes to model. If harvests in the past sample-plot data could be detected, and the difficulties associated with this task were acknowledged, then the use of prior data could be expected to yield more accurate results (Kangas et al., 2020; Fortin 2020). A sudden change in the forest stand (e.g., clear-cutting) can be relatively easy to detect from plotted time-series, whereas it can be difficult to model for future development within the overall utilized modeling scheme. For that reason, Fortin (2020) used a harvest probability model rather than observations obtainable from a GIS database. Driven by the possibility to continuously update the posterior distribution by new observations from harvesters (e.g., Uusitalo et al. 2006), more observations from actual harvests could be expected. Yet, the inability by most current forest development models to consider disruptive events can be seen as a limit to their use in updating forest inventory data (Fortin, 2020). Indeed, Fortin (2020) and Kangas et al. (2020) were among the first studies to take changing forest conditions (including harvests) into account in DA, and it is reasonable to expect further developments on these aspects in the future.

4.2.3. Weighting of the prior and new observations

With filtering, it has been accepted that the combined use of past and current inventory data produces more accurate estimations (Ehlers, 2013; Lindgren et al., 2017; Nyström et al., 2015), although Kangas et al. (2020) did not find this to be self-evident. To gain benefits, the estimates from the different time points or data sources must be properly weighted, which is typically caried out by assigning weightings (inversely) proportional to the variance of the different estimates. Work by Lindgren et al. (2017) assimilating multiple time points (and the follow-up by Ehlers et al., (2018) that was based on analyzing the correlations of the non-independent errors of these estimates) provide ideas and “rules-of-thumb” as to the value of the weightings. In Lindgren et al. (2017), the predictions that used past data received a greater weighting than the new data after 2–3 assimilations (depending on the forest attributes), and the weighting placed on new acquisitions was < 10% after ≈7 acquisitions. Although the performance varied in estimating forest attributes based on single data acquisitions, the variation associated with the assimilated result decreased and stabilized after the first iterations. Yet, all benefits of assimilating past data were not necessarily visible due to an underestimation in the variances of inter-correlated estimates (Ehlers et al., 2018). According to Ehlers et al. (2018), independent observations based on a different acquisition technique or estimation method should receive a greater weighting as they are less correlated. A graphical analysis of the values in Ehlers et al. (2018) suggests that even two time-points based on independent sensors or modeling should not be treated equally, i.e., combined with the same weighting, but that the correlation of the errors should be accounted for in the variance estimates. In cases of availability and the use of accurate auxiliary data, the resulting overall error might not be significantly lowered compared to the use of prior inventory data or accurate auxiliary data (Nyström et al., 2015; Kangas et al., 2020), although the internal consistency of the estimates could be improved to a degree that is beneficial for later applications.

4.2.4. Posterior distributions and uncertainty quantification

The use of posterior probability densities, which are also an inherent property of many Bayesian methods, creates possibilities for approaches such as uncertainty assessments. One example where uncertainty metrics were determined based on the posterior density and point estimates is illustrated by Varvia (2018). In this study, which was based on a leave-one-out subset of data, the accuracy of the Bayesian 95% credible intervals was found to be ideal for uncertainty analyses. One important initial finding of this study was that the prediction accuracy of the variables may improve when the variable pools are increased, although this may reduce the possibilities for uncertainty quantification. On the other hand, as outlined above, the full utilization of the posterior distribution may require the abandonment of parametric and Gaussian assumptions that lead to difficulty in solving the related modeling tasks analytically. As such, there is a subsequent need for computationally efficient techniques, such as MCMC (e.g., Babcock et al., 2016; Varvia et al., 2019). The instructive example of Varvia et al. (2019), who obtained considerable computational advantages by avoiding exact computations and pre-computing some parameters, suggests that some of the techniques reviewed above may require similar computational considerations in order to become operational. Establishing rules as to when to use a Bayesian or other paradigm would be especially useful for improving methodological choices for forest variable estimation. This will be a challenge as the choice depends on the applications and data usage. Moreover, RMSE or bias is not always an appropriate measure of accuracy. Loss attributable, for example, to misrepresentation of net present value requires consideration of the monetary element of these errors. Therefore, the question could possibly be approached through the concept of value of information (e.g., Kangas, 2010) by computing the (computational) costs and losses associated with the adoption of a Bayesian vs. an alternative approach.

5. Conclusion

Bayesian approaches have been increasingly utilized in many applications during the last decade, although less so in forestry than in the other reviewed scientific disciplines. However, state-of-the-art data processing frameworks and cross-disciplinary technologies allow for the integration of remotely sensed and forest inventory data, which are available at more frequent intervals and produce a longer time-series. The Bayesian frameworks provide the possibility to incorporate prior information, utilize the posterior distribution, as well as update it incrementally with more data, and efficiently measure and quantify uncertainties. These properties appear under-utilized in various data fusion approaches that characterize forest stand conditions. Thus, more significant contributions from the adoption of Bayesian approaches could likely be reached by developing know-how and competence in using standard formal methods instead of *ad hoc* methods.

We have identified the following as important aspects that affect the choice of whether to build a time-series analysis concept upon Bayesian formalism: 1) whether a forest variable estimate is treated as a time-series or as a time epoch, 2) access to a change model that, in addition to forest growth, is able to model sudden changes due to disturbances, 3) number of data points and time span of the time series, acquisition means and optimal interval for the time points during the total time span, and 4) assumed distributions of the data (in particular, Gaussian vs. other). Whether the choice is a Bayesian method, we further suggest that 5) Bayesian filtering should be combined with Bayesian smoothing whenever applicable, 6) uncertainty quantification should be built upon an analysis of credible intervals of posterior distributions to assess which sources of uncertainty predominate in forest variable estimation, and 7) advanced algorithms can be used for the calculation of numerical approximations of multi-dimensional integrals to reduce the computational time due to the adoption of the Bayesian method.

Acknowledgements

Funding: This work was supported by the Academy of Finland [grant number 324193].

References

Babcock, C., Finley, A.O., Bradford, J.B., Kolka, R., Birdsey, R., and Ryan, M.G. 2015. LiDAR based prediction of forest biomass using hierarchical models with spatially varying coefficients. Remote Sens. Environ. 169, 113-127.

Babcock, C., Finley, A.O., Cook, B.D., Weiskittel, A., and Woodall, C.W. 2016. Modeling forest biomass and growth: Coupling long-term inventory and LiDAR data. Remote Sens. Environ. 182, 1-12.

Cressie, N.A.C. 1993. Statistics for spatial data – revised edition. John Wiley & Sons. ISBN 0-471-00255-0, 902 p.

Czaplewski, R.L., Alig, R.J., and Cost, N.D. 1988. Monitoring land/forest cover using the Kalman filter: A proposal. In: Ek, A.R.; Shifley, S.R., Burk, T.E. (Eds.), Forest growth modelling and prediction: Volume 2. Gen. Tech. Report NC-120. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN, USA, p. 1089-1096, https://www.fs.usda.gov/treesearch/pubs/download/32871.pdf

de Groot, M., Diaci, J., and Ogris, N. 2019. Forest management history is an important factor in bark beetle outbreaks: Lessons for the future. For. Ecol. Manage. 433, 467-474.

Dixon, B., and Howitt, R. 1979. Continuous forest inventory using a linear filter. For. Sci. 25, 675-698.

Ehlers, S., Grafström, A., Nyström, K., Olsson, H., and Ståhl, G. 2013. Data assimilation in stand-level forest inventories. Can. J. For. Res. 43, 1104-1113.

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

Finley, A.O., Banerjee, S., and Cook, B.D. 2014. Bayesian hierarchical models for spatially misaligned data in R. Methods Ecol. Evol. 5, 514-523.

Finley, A.O., Banerjee, S., and MacFarlane, D.W. 2011. A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas. J. Amer. Stat. Ass. 106, 31-48.

Finley, A.O., Banerjee, S., Cook, B.D., and Bradford, J.B. 2013. Hierarchical Bayesian spatial models for predicting multiple forest variables using waveform LiDAR, hyperspectral imagery, and large inventory datasets. Int. J. Appl. Earth Obs. Geoinfo. 22, 147-160.

Finley, A.O., Banerjee, S., Ek, A.R., and McRoberts, R.E. 2008. Bayesian multivariate process modeling for prediction of forest attributes. J. Agr. Biol. Env. Stat. 13, 60-83.

Fortin, M. 2020. Updating plots to improve the precision of small-area estimates: the example of the Lorraine region, France. Can. J. For. Res. 50, 648-658.

Fragoso, T.M., Bertoli, W., and Louzada, F. 2018. Bayesian Model Averaging: A systematic review and conceptual classification. Int. Stat. Rev. 86, 1-28.

Gertner, G.Z. 1984. Localizing a diameter increment model with a sequential Bayesian procedure. For. Sci. 30, 851-864.

Gertner, G. 1987. A procedure to evaluate sampling schemes for improving already calibrated models. For. Sci. 33, 632-643.

Gertner, G.Z., Fang, S., and Skovsgaard, J.P. 1999. A Bayesian approach for estimating the parameters of a forest process model based on long-term growth data. Ecol. Mod. 119, 249-265.

Gertner, G.Z., and Zhu, H. 1996. Bayesian estimation in forest surveys when samples or prior information are fuzzy. Fuzzy Sets Syst. 77, 277-290.

Gove, J.H. 2009. Propagating probability distributions of stand variables using sequential Monte Carlo methods. Forestry, 82, 403-418.

Green, E.J., and Strawderman, W.E. 1992. A comparison of hierarchical Bayes and empirical Bayes methods with a forestry application. For. Sci. 38, 350-366.

Hotelling, H. 1936. Relations between two sets of variates. Biometrika, 28, 321-377.

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 Sens. Environ. 234, 111431.

Hou, Z., Domke, G.M., Russell, M.B., Coulston, J.W., Nelson, M.D., Xu, Q., and McRoberts, R.E. 2021. Updating annual state- and county-level forest inventory estimates with data assimilation and FIA data. For. Ecol. Manage. 483, doi: 10.1016/j.foreco.2020.118777

Itter, M.S., D'Orangeville, L., Dawson, A., Kneeshaw, D., Duchesne, L., and Finley, A.O. 2019. Boreal tree growth exhibits decadal-scale ecological memory to drought and insect defoliation, but no negative response to their interaction. J. Ecol. 107, 1288-1301.

Itter, M.S., Finley, A.O., D'Amato, A.W., Foster, J.R., and Bradford, J.B. 2017. Variable effects of climate on forest growth in relation to climate extremes, disturbance, and forest dynamics. Ecol. Appl. 27, 1082-1095.

Junttila, V., Maltamo, M., and Kauranne, T. 2008. Sparse Bayesian estimation of forest stand characteristics from airborne laser scanning. For. Sci. 54, 543-552.

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82, 35-45.

Kangas, A. 1991. Updated measurement data as prior information in forest inventory. Silva Fenn. 25, 181-191.

Kangas, A.S. 2010. Value of forest information. Eur. J. For. Res. 129, 863-874.

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. Scand. J. For. Res. 33, 397-412.

Kangas, A., Korhonen, K.T., Packalen, T., and Vauhkonen, J. 2018b. Sources and types of uncertainties in the information on forest-related ecosystem services. For. Ecol. Manage. 427, 7-16.

Kangas, A., Räty, M., Korhonen, K.T., Vauhkonen, J., and Packalen, T. 2019. Catering information needs from global to local scales-potential and challenges with national forest inventories. Forests 10, 9.

Kangas, A., Gobakken, T., and Næsset, E. 2020. Benefits of past inventory data as prior information for the current inventory. For. Ecosyst. 7(1), 20.

Katila, M., and Heikkinen, J. 2020. Reducing error in small-area estimates of multi-source forest inventory by multi-temporal data fusion. Forestry 93, 471-480.

Knödel, K., Toloczyki, M., Bohn, A., Abel, T., Lange, G., and Tejedo, A. 2007. Data fusion. p. 1054 in Environmental Geology - Handbook of Field Methods and Case Studies. Edited by K. Knödel, G. Lange, H.-J. Voigt. Springer-Verlag Berlin Heidelberg, ISBN 978-3-540-74669-0

Leamer, E.E. 1978. Specification searches: Ad hoc inference with nonexperimental data. Wiley, New York.

Liénard, J.F., and Strigul, N.S. 2006. Modelling of hardwood forest in Quebec under dynamic disturbance regimes: a time-inhomogeneous Markov chain approach. J. Ecol. 104, 806-816.

Liénard, J.F., Gravel, D., and Strigul, N.S. 2015. Data-intensive modeling of forest dynamics. Environ. Mod. Soft. 67, 138-148.

Lindgren, N., Persson, H.J., Nyström, M., Nyström, K., Grafström, A., Muszta, A., Willén, E., Fransson, J.E.S., Ståhl, G., and Olsson, H. 2017. Improved prediction of forest variables using data assimilation of Interferometric Synthetic Aperture Radar data. Can. J. Rem. Sens. 43, 374-383.

Lu, L., Wang, H., Chhin, S., Duan, A., Zhang, J., and Zhang, X. 2019. A Bayesian Model Averaging approach for modelling tree mortality in relation to site, competition and climatic factors for Chinese fir plantations. For. Ecol. Manage. 440, 169-177.

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., and Schürmann, G. 2016. Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. Geosci. Model Dev. 9, 3569-3588.

Mäkelä, J. 2020. Bayesian methods applied for ecosystem model calibration and uncertainty source estimation. Finnish Meteorological Institute Contributions 160, http://hdl.handle.net/10138/310949

Mäkinen, A.M., Kangas, A.S., and Tokola, T. 2010. Identifying the presence of assessment errors in forest inventory data by data mining. For. Sci. 56, 301-312.

Minunno, F., Peltoniemi, M., Härkönen, S., Kalliokoski, T., Makinen, H., and Mäkelä, A. 2019. Bayesian calibration of a carbon balance model PREBAS using data from permanent growth experiments and national forest inventory. For. Ecol. Manage. 440, 208-257.

Mitchell, T.J., and Beauchamp, J.J. 1988. Bayesian variable selection in linear regression. J. Amer. Stat. Ass. 83, 1023-1032.

Mölder, A., Sennhenn-Reulen, H., Fischer, C., Rumpf, H., Schönfelder, E., Stockmann, J., and Nagel, R.V. 2019. Success factors for high-quality oak forest (*Quercus robur*, *Q. petraea*) regeneration. For. Ecosyst. 6(1), 49.

Nothdurft, A. 2020. Climate sensitive single tree growth modeling using a hierarchical Bayes approach and integrated nested Laplace approximations (INLA) for a distributed lag model. For. Ecol. Manage. 478, doi: 10.1016/j.foreco.2020.118497

Nyström, K., and Ståhl, G. 2001. Forecasting probability distributions of forest yield allowing for a Bayesian approach to management planning. Silva Fenn. 35, 185-201.

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

Packalen, T., Sallnäs, O., Sirkiä, S., Korhonen, K., Salminen, O., Vidal, C., Robert, N., Colin, A., Belouard, T., Schadauer, K., Berger, A., Rego, F., Louro, G., Camia, A., Räty, M., and San-Miguel, J.

2014. The European Forestry Dynamics Model. Concept, design and results of first case studies. Publications office of the EU, Report EUR 27004 EN, doi: 10.2788/153990

Picard, N., Henry, M., Trotta, C., and Saint-Andre, L. 2012. Using Bayesian Model Averaging to predict tree aboveground biomass in tropical moist forests. For. Sci. 58, 15-23.

Rue, H., Martino, S., and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. Royal Stat. Soc. B, 71, 319-392.

Särkkä, S. 2013. Bayesian filtering and smoothing. Cambridge University Press, 232 p. ISBN 9781139344203, doi: 10.1017/CBO9781139344203

Suty, N., Nyström, K., and Ståhl, G. 2013. Assessment of bias due to random measurement errors in stem volume growth estimation by the Swedish National Forest Inventory. Scand. J. For. Res. 28, 174-183.

Thum, T., MacBean, N., Peylin, P., Bacour, C., Santaren, D., Longdoz, B., Loustau, D., and Ciais, P. 2017. The potential benefit of using forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites. Agr. J. For. Meteorol. 234-235, 48-65.

Tian, X., Sun, S., Mola-Yudego, B., and Cao, T. 2020. Predicting individual tree growth using stand-level simulation, diameter distribution, and Bayesian calibration. Ann For. Sci. 77(57), doi: 10.1007/s13595-020-00970-0

Tipping, M.E. 2004. Bayesian inference: An introduction to principles and practice in machine learning. In Carbonell, U.J.G., Siekmann, G.J. (Eds.), Advanced Lectures on Machine Learning, Vol. 3176, p. 248. Springer-Verlag, Berlin Heidelberg. doi: 10.1007/978-3-540-28650-9_3

Uusitalo, J, Puustelli, A., Kivinen, V.P., Nummi, T., and Sinha, B.K. 2006. Bayesian estimation of diameter distribution during harvesting. Silva Fenn. 40, 663-671.

van Deusen, P.C. 1987. Testing for stand dynamics effects on red spruce growth trends. Can. J. For. Res. 17, 1487-1495.

van Oijen, M. 2017. Bayesian methods for quantifying and reducing uncertainty and error in forest models. Curr. For. Rep. 3, 269-280.

van Oijen, M., Rougier, J., and Smith, R. 2005. Bayesian calibration of process-based forest models: Bridging the gap between models and data. Tree Physiol. 25, 915-927.

Varvia, P. 2018. Uncertainty quantification in remote sensing of forests. Dissertations in Forestry and Natural Sciences, Publications of the University of Eastern Finland 313, 44 p, http://urn.fi/URN:ISBN:978-952-61-2867-2

Varvia, P., Lähivaara, T., Maltamo, M., Packalen, P., Tokola, T., and Seppänen, A. 2017. Uncertainty quantification in ALS-based volume estimation. IEEE Trans. Geosci. Remote Sens. 55, 1671-1681.

Varvia, P., Lähivaara, T., Maltamo, M., Packalen, P. and Seppänen, A. 2018. Gaussian process regression for forest attribute estimation from airborne laser scanning data. IEEE Trans. Geosci. Remote Sens. 57, 3361-3369.

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. Int. J. Remote Sens. 39, 8766-8783.

Vauhkonen, J., Berger, A., Gschwantner, T., Schadauer, K., Lejeune, P., Perin, J., Pitchugin, M., Adolt, R., Zeman, M., Johannsen, V.K., Kepfer-Rojas, S., Sims, A., Bastick, C., Morneau, F., Colin, A., Bender, S., Kovacsevics, P., Solti, G., Nagy, D., Kolozs, L., Nagy, K., Twomey, M., Redmond, J., Gasparini, P., Notarangelo, M., Rizzo, M., Makovskis, K., Lazdins, A., Lupikis, A., Kulbokas, G., Anton-Fernandez, C., Castro Rego, F., Nunes, L., Marin, G., Calota, C., Pantic, D., Borota, D., Roessiger, J., Bosela, M., Seben, V., Skudnik, M., Adame, P., Alberdi, I., Canellas, I., Lind, T., Trubins, R., Thürig, E., Stadelmann, G., Ditchburn, B., Ross, D., Gilbert, J., Halsall, L., Lier, M. and Packalen, T. 2019. Harmonised projections of future forest resources in Europe. Ann. For. Sci. 76(3), 79.

Ver Planck, N.R., Finley, A.O., Kershaw, J.A., Weiskittel, A.R., and Kress, M.C. 2018. Hierarchical Bayesian models for small area estimation of forest variables using LiDAR. Remote Sens. Environ. 204, 287-295.

Zhang, T., Su, J., Liu, C., and Chen, W.H. 2019. Bayesian calibration of AquaCrop model for winter wheat by assimilating UAV multi-spectral images. Comput. Electronics Agr. 167, 105052.

Zhang, X., Duan, A., and Zhang, J. 2013. Tree biomass estimation of Chinese fir (*Cunninghamia lanceolata*) based on Bayesian method. PLoS ONE 8(11), e79868.

Zhao, P., Gao, L., and Gao, T. 2020. Extracting forest parameters based on stand automatic segmentation algorithm. Sci. Rep. 10, 1-13.

FIGURE CAPTIONS

Figure 1. Schematic diagram of a forest time-series between current (t) and previous time points (t-1) that will be predicted into the future (t+1). Example forestry dynamics to be predicted include forest operations (clearing of trees on the left side of the plot), mortality (brown tree crown in the middle), survival and growth of remaining trees, and ingrowth of small trees. If the trajectory is considered generically as a time-varying system, the concepts of smoothing, filtering, and prediction (Särkkä, 2013) can be identified as shown in Figure 1. The up- and downward brackets in the diagram indicate the time points used as inputs and obtained as outputs, respectively.

Figure 2. Schematic diagram of **(a)** the filtering and smoothing concept adapted from Särkkä (2013), where the sequence of hidden states at time points t_{-1}, t and t_{+1} is inferred through noisy observations with a (linear) temporal element but no spatial element(s). Three sub-figures illustrate the spatial element that should be considered when employing these concepts in typical forest inventory applications. Arrows indicate the possible desired outputs of using circular plots at varying time points and located in an area of interest outlined by a solid black line in **(b)** regression modeling of a variable of interest *y* using covariate(s) *x*, coefficients β and error term *e,* **(c)** large area estimation for population total or mean (μ) based on inclusion probabilities π and observations of *y* from plots {*i*} in the area, and **(d)** small area estimation, where the aim is to predict for subpopulations of interest *i* using the observations and estimators derived from those subpopulations, In (d), kriging would be a point estimate based on spatial correlation.

Table 1. Synthesis of the reviewed studies. Study denotes literature reference by author name and year; Sequence denotes the type of sequential information (temporal, spatial, other; asterisks in this column indicate that the method did not belong to the Bayesian school); Variable denotes the forest variable of main interest; AOI denotes inventory area for which the results were generalized (** indicate that the results were not reported in the study, but were assumed to be related to the national forest inventory context, i.e., at a scale of millions of hectares); CU denotes the area of the initial computation unit; Plots denotes the number of sample plots; Tp# denotes the number of time points; TSp denotes the time span of the acquisition from the first to last datapoint in years; Error denotes the initial error rate; Rate denotes the improvement due to the inclusion of additional time points or methodological changes. The latter parameter was computed as the difference (in percentage points) between the error rate before and after the application of data fusion.

Study	Sequence	Variable	AOI (ha)	CU (m^2)	Plots	Tp#	TSp (a)	Error (%)	Rate (pp)
1. Junttila et al., 2008	Spatial	GSV (m^3 ha^{-1})	1,200	254	472	1	0	19.90	1.00
2. Finley et al., 2013	Spatial	GSV growth (Mg ha^{-1})	15,782	80	451	1	0	2.33	1.20
3. Babcock et al., 2015	Spatial	AGB (Mg ha^{-1})	10,472	200	62	1	0	38.43	1.77
4. Katila & Heikkinen, 2020	Temporal *	GSV (m^3 ha^{-1})	18×10^6	100	42,451	3	4	3.02	2.00
5. Fortin, 2020	Temporal	GSV (m^3 ha^{-1})	537×10^3	113	180	4	9	2.28	2.09
6. Ehlers et al., 2013	Temporal	GSV (m^3 ha^{-1})	** $\times10^6$	400	8,793	25	50	11.16	3.13
7. Finley et al., 2008	Spatial	AGB (Mg ha^{-1})	1,053	1,000	437	1	0	27.71	4.43

8. Nyström et al., 2015	Temporal	GSV (m^3 ha^{-1})	1,500	314	15,131	6	8	13.50	4.70
9. Vastaranta et al., 2018	Temporal *	AGB (Mg ha^{-1})	2,000	256	332	2	2	20.40	5.60
10. Lindgren et al., 2017	Temporal	GSV (m^3 ha^{-1})	1,200	314	137	19	4	30.00	6.00
11. Kangas et al. 2020	Temporal	AGB (Mg ha^{-1})	853	200	174	2	11	3.10	6.22
12. Varvia et al., 2017	Spatial	GSV (m^3 ha^{-1})	10,000	255	492	1	0	17.70	10.30
13. Babcock et al., 2016	Temporal Spatial	AGB (Mg ha^{-1})	1,600	78	604	40	40	17.52	25.36
14. Hou et al., 2019	Other (model) *	Mean height (m)	56	22	200	1	0	9.59	33.00
15. Ver Planck et al., 2018	Spatial	AGB (Mg ha^{-1})	1,500	66×10^3	195	1	0	10.60	56.90

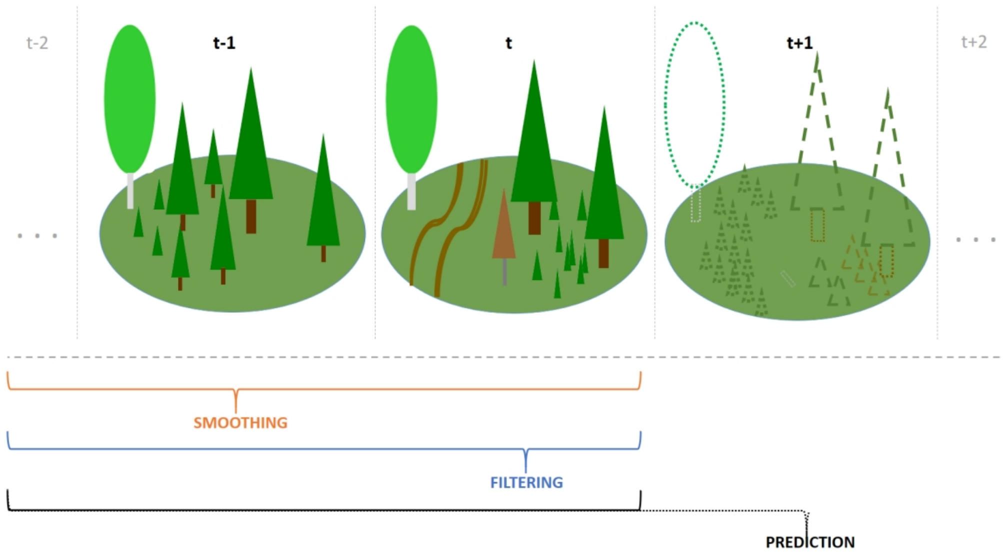

Figure 1. Schematic diagram of a forest time-series between current (t) and previous time points (t-1) that will be predicted into the future (t+1). Example forestry dynamics to be predicted include forest operations (clearing of trees on the left side of the plot), mortality (brown tree crown in the middle), survival and growth of remaining trees, and ingrowth of small trees. If the trajectory is considered generically as a time-varying system, the concepts of smoothing, filtering, and prediction (Särkkä, 2013) can be identified as shown in Figure 1. The up- and downward brackets in the diagram indicate the time points used as inputs and obtained as outputs, respectively.

643x363mm (72 x 72 DPI)

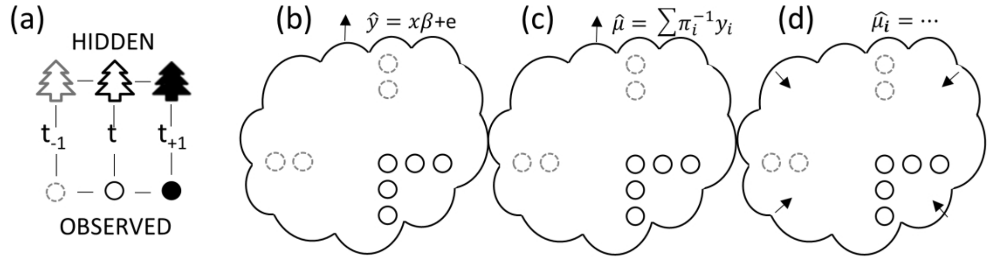

Figure 2. Schematic diagram of (a) the filtering and smoothing concept adapted from Särkkä (2013), where the sequence of hidden states at time points t-1, t and t+1 is inferred through noisy observations with a (linear) temporal element but no spatial element(s). Three sub-figures illustrate the spatial element that should be considered when employing these concepts in typical forest inventory applications. Arrows indicate the possible desired outputs of using circular plots at varying time points and located in an area of interest outlined by a solid black line in (b) regression modeling of a variable of interest y using covariate(s) x, coefficients β and error term e, (c) large area estimation for population total or mean (μ) based on inclusion probabilities π and observations of y from plots {i} in the area, and (d) small area estimation, where the aim is to predict for subpopulations of interest i using the observations and estimators derived from those subpopulations, In (d), kriging would be a point estimate based on spatial correlation.

189x57mm (300 x 300 DPI)