Skogssällskapet

Slutrapport

Projektrubrik: Tree species mapping using harvester and remote sensing data

Huvudsökande: Liviu Theodor Ene

Projektets löptid: 2021-04-01 – 2023-01-31

Populärvetenskaplig sammanfattning

Product recovery predictions are important tools in forest planning, allowing managers to make informed decisions about how to sustainably manage and utilize the forest resources, for efficient planning of forest operations and for optimizing wood flows in the forest industry. Following the rapid developments in the digitalization of the forest industry, some of the largest Swedish forestry companies and forest owner associations have already adopted custom prediction systems that utilize harvester production data, forest registry information, and GIS products.

The project is taking a step forward with an improved way to product recovery predictions and estimating the uncertainties in operational forest planning. The approach combines multitemporal, multi-source GIS data and standardized information from harvester production files to create a more accurate and mathematically sound predictive system.

The results provide a better understanding of the data requirements for improving the prediction accuracy for species-specific forest attributes such as volumes and diameter distributions, and it also suggest what imputation approaches are most suitable for product recovery predictions. Moreover, statistical methods for quantifying the imputations uncertainty in terms of prediction intervals aiming to contain the unobserved forest stand parameters (total and species specific volumes per forest tract) with specified probabilities were successfully developed.

The analyses were conducted on dataset containing standardized harvester production files from over 3800 forest tracts located in Southern Sweden, in combination with multi-source, multi-temporal GIS data such as airborne laser scanning data, aerial imagery and satellite imagery (Sentinel-2), terrain raster, and maps of the vegetation type cover and forest state estimates.

With low implementation costs and a focus on data reliability and transparency, the proposed imputation system has the potential to impact in the forestry industry by setting the baseline for assessing the efficiency of future methodologies. Not least, the proposed approach is able to provide the required inputs for robust multicriteria optimization methods for operational planning and logistics.

Resultat

The most relevant findings are described below. More details (including figures and tables) and references are provided in the attached document (Appendix A).

Data requirements

An overview of the proposed workflow for data ingestion and imputations is presented in Figure 5.

The first step was to identify the most relevant data sources (§1 in Appendix A) to be used for training the machine learning models. 12 subsets of auxiliaries were produced, consisting of various feature extracted from the airborne laser scanning data and aerial imagery acquired during the country-wide surveys, satellite imagery (Sentinel-2), terrain rasters, and maps of the vegetation type cover and forest state estimates.

The kMSN algorithm was assessed against statistical multivariate methods that could be used for product recovery predictions, namely Partial Least Squares (PLSR) and Multivariate Random Forests regression (MRFR). Both PLSR and MRFR are popular machine learning algorithms, robust to noise and can handle a large numbers of correlated predictors.
Although the best results were obtained when combining all auxiliary sources of information, the gains were minimal relative to using only a selected subset of auxiliaries. For instance, using tree species information as species-specific volume estimates provided by the historic SLU Forest Map (SLU) from 2015 could be more efficient than using the dominant species or forest type classes from 'Nationella marktäckedata' (NMD) raster maps produced by the Swedish Environmental Protection Agency between 2017-2019. Arguably, this is because the species-specific volume maps provide continuous information on all main tree species simultaneously, without thresholding the occurrence of certain species. This helps solving the ambiguities that occur in mixed forests, where usually deciduous species are a minority. Moreover, the results indicate that direct imagery data did not produce major improvements in accuracy when used together with either SLU or NMD maps. Nevertheless, the airborne laser scanning data remained the most important source of information for predicting the quantitative forest attributes.
It was concluded that a combination of predictors derived from multiple airborne laser scanning returns (height and intensity metrics derived from point clouds and textural information from vegetation height models), ancillary information (depth to water maps, spatial coordinates of the forest areas), hot-encoded forest cover classes form NMD maps and historical SLU estimates provided the best trade-off with regard to accuracy, computation efficiency and data availability. This finding was consistent for all the tested prediction methods (Figure 7).
Imputation accuracy
The results indicate that kMSN provided the most accurate results with regard to species-specific volume and diameter distribution predictions for all 12 subsets of auxiliaries, followed by the PLSR, while MRFR came last. Arguably, a possible explanation could be that PLSR and MRFR require hyperparameter tunning (for increasing their predictive performance) that relies on heuristic approaches such as internal cross-validations for (PLSR) or a grid search (for MRFR) and may sometimes misbehave during the computational routines for method validation process, while kMSN is a robust algorithm that requires little or no tunning.
The average kMSN performance for total and species specific volume prediction using the selected subset of predictors (#8) during the simulation studies is presented in Table 3 with particular examples in Figure 8. The diameter distribution accuracy is presented in Table 4, with details in Figures 9-10.
To evaluate the potential for method improvement, the k-MSN imputations were assessed against the hypothetical case when the true species-specific volumes were replacing the SLU estimates, concluding that the recommended action for further improving the results the is to work towards increasing the accuracy of mapping the Pine sp. Moreover, improving the accuracy of species-specific volumes did not markedly influence the total volume predictions, and neither notably reduced the monetary losses.
Uncertainty estimation

The uncertainty estimation focused on the kMSN algorithm since it performed best under all cases. The methodology uses a variant of the locally adaptive conformal prediction based on multivariate quantile regression that allows constructing prediction intervals characterizing the variability of the imputation. The numerical properties of 4 types of coverage types (Appendix A §2.5) were tested using computer simulations. The results (Figure 11, Tables 5-6) indicate that the conformal inference approach produced coverage rate estimates that are very close to the nominal levels, with the conditional coverage types being in general more efficient. Details on conformal prediction intervals for overall and species specific volumes at forest tract level are presented in Figures 12-15.

Målbeskrivning

The use of harvester data for feedbacks during forest operations and for product recovery predictions has been gradually adopted by the Swedish forestry. This methodology helps improving the information accuracy in the forest stand databases, and consequently supports a more accurate and dynamic industry supply of timber and pulpwood with the requested dimensions and species specifications.
The project investigated the feasibility of an improved imputation system that goes beyond the current state and uses a combination of multitemporal, multisource GIS data and standardized information from harvester production files. This new approach targeted two main objectives of high relevance in operational forest planning:
1.the improvement of tree species-species specific product recovery predictions on forest tracts planned to harvest, with main focus on species specific volumes and diameter distributions, and
2.to provide a mathematically sound framework for analyzing the uncertainties related to such predictions.

The developed methodology comes with low implementation costs in terms of ensuring the data support and reliability (i.e., availability of high-quality multitemporal, multisource GIS and ground-truth information from harvesters), computational feasibility and deployment, and transparency of the algorithmic models.

Having a mathematically sound methodology for describing the accuracy of predicting multiple forest stand attributes in common statistical terms helps understanding the quality of the results, provides the means for comparing alternative methods and supports the use of robust multicriteria optimization methods in operational planning and logistics.

Budget execution:
In the project, the budget was travel costs (10,000 SEK) were redistributed to the salaries, since the participation to the planned conferenced was not possible. The planned outreach is described in the communication and dissemination section below.

Kommunikation och nyttiggörande av resultat

Partial project results on the methodology for species-specific product recovery predictions were presented to the major Swedish forest companies at the Forestry planner's collaboration group (PSG) organized by Skogforsk in October 2022.
The final project results will be summarized and discussed at the next PSG meeting planned for June 2023, as well as at the RIU conference planned for November 2023 in Skinnskatteberg, Sweden.

A work report will be published on-line after passing the internal review at Skogforsk, followed by a popular scientific article disseminated on Skogforsk's website and in Skogforsk's magazine Vision. Primary target groups for these articles are the forest professionals.

An abstract on methods for uncertainty assessment has been submitted for review to oral presentation at 'The International Boreal Forest Research Association Conference IBFRA2023' to be held in Helsinki Finland, during 28-31 August 2023. A full papers will be submitted in a peer-reviewed special issue of the Silva Fennica scientific journal that follows the conference.

An abstract on the comparison of imputation methods for species-specific forest attributes is in submission for review at the 'Growth and Yield Innovations Conference 2023' organized by the Western Mensurationists and the Forest Growth Organization of Western Canada (FGrOW) to be held in Canmore, Alberta, June 18-21.

Bilagor

Bilaga 1. Sluttrapport_Appendix_A.pdf