APPLIED PERIOD REPORT: Satellite image-based estimation of forest canopy cover

AMAITUM JOSHUA ELUKUT

2216274

MSc European Forestry

School of Forest Sciences

University of Eastern Finland

SUPERVISORS: Prof. Lauri Korhonen & Shaohui Zhang

UNIVERSITY OF
EASTERN FINLAND

ABSTRACT

This report describes my internship experience with Satellite-image based mapping of forest cover between June to August 2023. It was performed at the Forest Inventory Research Group at the University of Eastern Finland which specializes in forest information technology research covering topics ranging from applied geoinformatics to information technology (IT) methods applied to forestry. The training started off with data collection in 20 heterogenous field plots which were purposively selected to represent forests of varying ages and species composition. In addition to tree and plot level parameters, digital hemispherical photographs (DHPs) and digital canopy photographs (DCPs) were obtained to support later efforts towards predicting Lead Area Index (LAI) and canopy cover (CC) respectively. After data cleansing, binarized images were obtained from DHPs and later Matlab was used to determine LAI. For CC, this was calculated directly from the DCPs using Matlab with average canopy cover and closure obtained for each plot. These were later aggregated with plot data and parameters like gap fraction for further analysis. In addition to manipulating data from Joensuu, I was tasked with creating binarized images from DHPs taken in Merikarvia in 2022. These would support my supervisors' future work efforts. This was followed by the modelling task, which was done with the beta regression technique, as it is useful when dealing with data or variables constrained between 0 and 1, as is the case with canopy cover. In addition to the summary of the data from the plots in Joensuu, a sentinel-2 image from Corpenicus Open Access Hub, a forest mask from Natural Resources Institute Finland (Luke), a base map, and orthophotographs were also utilized. The beta regression analysis was performed in RStudio software to predict canopy cover for the study area. The final model developed had 3 bands namely: nir, swir1, and nirnar. Furthermore, it yielded Pseudo R-squared = 0.7481 & RMSE = 0.1476758. Based on this, a canopy cover map was developed by making predictions on the sentinel-2 image, while using a forest mask to avoid predictions in the non-forest areas such as water bodies. The final map development was performed with QGIS software. In addition, I performed other tasks such as self-studying the Advanced Remote Sensing course and attending a summer school on remote sensing of forests. Overall, the internship offered good learning opportunities about satellite-image based remote sensing, starting from the theory to data collection, processing, regression analysis, and map development. I am very satisfied as it will provide a strong foundation for my career and overall contribution to forest science.

DESCRIPTION OF WORK AND ACTIVITIES

WEEKLY DIARY

The goal of the Applied period at Forest Inventory Research group at the University of Eastern Finland was to offer training on Satellite image-based mapping of forest leaf area index (LAI) and canopy cover (CC) using Joensuu, Finland as the case study. Training on data collection as well as analysis of both LAI and CC was provided, although this report primarily focuses on the latter. In summary, the tasks comprised the following: measuring forest attributes in field plots; hemispherical photography and canopy cover measurement and image processing; acquisition of satellite images of the study area; Satellite image-based CC modelling and cross prediction between sites; and Construction of the area CC map.

Summary of Activities

Period	**Activities**
1-6 June	*Hemispherical photography and canopy cover measurements:* Acquisition of LAI and CC photos using two different cameras. *Forest measurements:* From 20 plots, forest attributes such as coordinates, species, basal area, diameter, and crown base height.
7-9 June	*Ground control points and forest plots data collection in Hiddenportti National Park:* Collection of tree and plot level characteristics together with accurate positional information
11-16 June	*Processing of digital hemispherical photographs (DHP):* Received training on the use of HSP software for image processing *Processing of digital canopy photographs (DCP):* Received training on use of Matlab for image processing *Data cleaning & aggregation:* Detection of incorrect records and summarizing processing results for further analysis in remote sensing tasks
18-23 June	*Further processing of DHP & DCP images from Joensuu and Merikarvia:* In addition to images collected for the internship, I was handed additional task for those done in 2022
25-30 June	*Self-study of the Advanced Remote Sensing course (AdReSen):* It was encouraged to self-study some sections of the course to improve competencies required for the internship tasks
3-7 July	*Acquisition of satellite image and other map layers:* Retrieved Sentinel-2 image, base maps, otho image, forest mask, and post correction of field plots *Further processing of DHP & DCP images from Joensuu and Merikarvia:*
9-14 July	*Further processing of DHP & DCP images from Joensuu and Merikarvia:*

16-21 July	*Self-study of the Advanced Remote Sensing course* *Further processing of DHP & DCP images from Joensuu and Merikarvia*
24-28 July	*Data summary/aggregation:* Involved finalizing the computation of parameters needed for remote sensing tasks such as canopy closure, canopy cover, canopy gap fraction and crown gap fraction *Remote sensing analysis:* Tasked to adapt own codes from the AdReSen course to conduct Satellite image-based mapping of CC
7-11 August	*Self-study of the Advanced Remote Sensing course*
14-18 August	*Remote sensing analysis:* R modelling with different variables
21-25 August	*Remote sensing analysis:* Finalised the development of the most appropriate model based on root-mean-square-error (RMSE), coefficient of determination (R^2), plots of residuals vs. fitted values and predicted vs. observed *Map development:* Made a final map of canopy cover distribution in Joensuu *Report writing:* Develop a written output detailing learning achievements

Problems encountered

- Initial challenges with the setup and use of equipment such as the vertex, cameras for taking the digital hemispherical and canopy photographs. It was difficult to get the correct exposure settings.
- Nighttime data collection activity provided unfavorable cold and mosquito infested conditions in Joensuu.
- Originally unfamiliar with the use of HSP and Matlab software. Some images such as those with clouds and open skies were particularly problematic.
- Some challenges interpreting the R codes during the self-study of the advanced remote sensing course.
- Limited knowledge on how to search for appropriate satellite images, forest masks, otho and base maps for the study area.

Solutions provided

- Provision of an accurate field guide with detailed instructions on how to setup the field plots, the correct camera configurations under different light intensity condition, together with the assistance of a field supervisor who was always available to provide guidance.
- Continuous exposure to the tasks and subsequent data collection in different plots and ground control points became easier as we familiarized with the tasks.

- Guidance was provided by the supervisors (on attributes such as searching distance and Zenith). As I interfaced with the software, I became more familiar and able to accomplish the tasks within required timeframes and processing result quality improved.
- Selecting images with the best exposure settings, as well as consulting supervisors and my colleague. Overall, the tasks became more adaptable with experience.
- Problems with use of R for data analysis were addressed through the continuous research and seeking guidance from the supervisors and colleagues.
- Addressed issues of secondary data collection through the continuous research and seeking guidance from the supervisors and colleagues

Improvements

- Ensure interns are paired on roles as this can improve on the learning experience, as was the case for this internship where I learned about both LAI and CC.
- Offer structured training and orientation programs to allow for students to integrate well into their roles in the company. This can include training on data collection and analysis prior to the start of the applied period.
- Develop a feedback mechanism through which interns receive and analysis of how they are fairing in their roles.

PROJECTS CONDUCTED

DATA COLLECTION

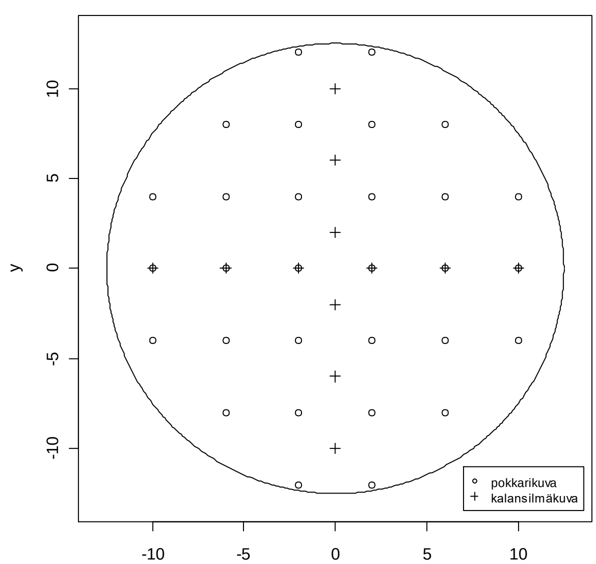
Plot and tree measurements

This was performed in the 20 plots in Joensuu together with the 4 plots from Hiidenportti national park with the aim of gathering information on tree and stand level characteristics. For the Joensuu plots, radius was 12.5m, and information such as plot coordinates obtained using a trimble, basal area by tree species, and Height of the basal area median tree for each species were collected. For Hiidenportti, plot size of either 10m and 15m were used and each tree with DBH above 5cm was enumerated. Attributes such as height, species, dead class, and detectability were recorded. Plot level characteristics included soil type and fertility class. An example of the data collection forms for Joensuu is displayed below.

Plot	MainSp	SubGroup	SiteIndex	Species	G	DBH	H	CBH

Hemispherical photography and canopy cover measurements

This was performed on 20 subjectively placed plots in Joensuu to support LAI and CC estimates. To maximize the variability of the data, various plot types were considered such as mature and young pines, bogs, dense spruce stands, as well as stands with average canopy cover. A Canon D2000 camera was used to capture 12 images per plot for LAI estimation and a pocket digital camera were used to capture 30 images for CC estimation. Images were taken only after sunset due to sensitivity to light conditions. Care was taken to ensure the background (sky) was standardized as possible so the canopy can be distinguished without bias. The structure of the plot and sample digital hemispherical and canopy images are shared below.



Structure of the plot. Crosses are fisheye images (start 2 m from center, 4 m between images), circles pocket camera images (4 m grid). Image height about 1.3 m above ground level

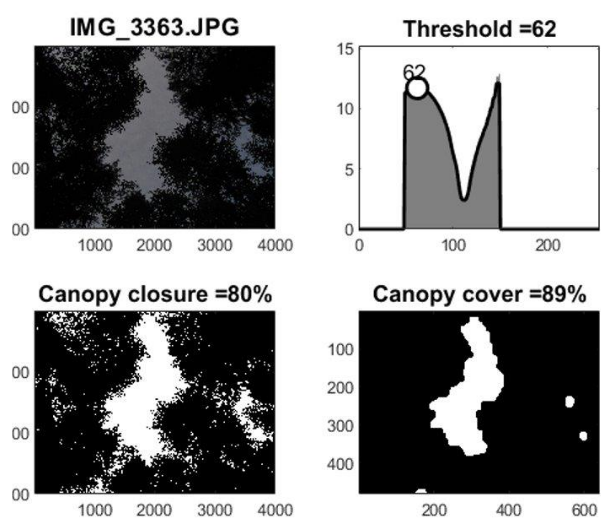
A digital canopy photograph

Ground control points collection

These are defined as points on the ground with known coordinates and are important for establishing the spatial reference and accuracy of geospatial data. The activity involved setting up a Trimble GPS and antenna together with a tripod with GNSS receiver. This was placed at predefined locations such as ditch intersections and rocks which are accurately visible from satellite images. Wait time was about 5 minutes before an accurate position was established, recorded, and saved. About 14 points were collected altogether in areas around Hiidenportti national park.

Image processing

This was performed using Matlab, a programming and numerical computing platform used by millions of engineers and scientists to analyze data, develop algorithms, and create models. Matlab codes were provided beforehand and for each plot, the digital canopy photographs were processed with the average canopy cover and canopy closure obtained and input into the summary table for further remote sensing analysis. Care was taken to ensure the threshold value was appropriate during the computation.



Output of the Matlab image processing with canopy closure and canopy cover

Acquisition of satellite image and other map layers

To enable the development of the Canopy Cover map for the Joensuu area, it was important to obtain different map layers that could be manipulated in R and ArcGis.

Firstly, the Sentinel-2 image was obtained from Corpenicus Open Access Hub. This encompassed the areas around Joensuu within which the 20 field plots were obtained and it contained several bands whose combinations were useful for monitoring various attributes.

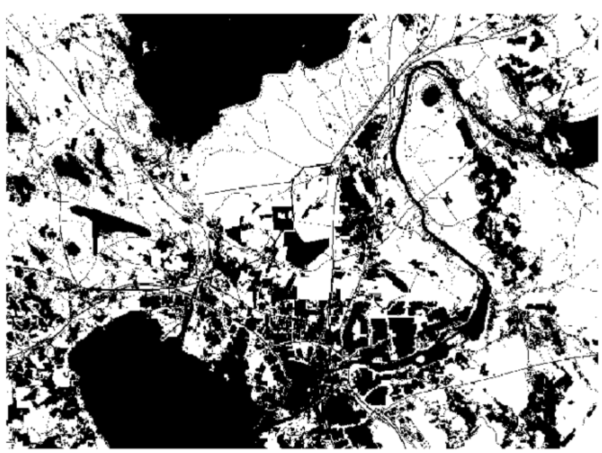
A base map for the area was obtained and it was particularly useful when identifying the locations of features from other map layers when they were overlaid on it.

Orthophotos were obtained from the National Land Survey of Finland. These are combinations of several individual aerial photos whose geometry corresponds to a map and are utilized for planning and monitoring the environment for example through either as background material for various map presentations or for forest inventory.

Finally, a forest mask was downloaded from the Natural Resources Institute Finland to mask out non-forest areas. It delineates areas of forest cover from non-forest areas thereby enabling the prediction of tree canopy cover in the correct locations while avoiding features like water bodies, rocks, and buildings.

Some map layers are displayed below.

Sentinel-2 image



Forest mask

DATA ANALYSIS: Remote Sensing

R, a programming language and open-source software environment primarily used for statistical computing and data analysis, was utilized for the remote sensing tasks. R offers a wide range of packages and functions that can be used to process, analyze, and visualize remote sensing data. Beta regression analysis was employed in model building because is a useful statistical tool in remote sensing when dealing with response variables such as forest canopy cover that are bounded between 0 and 1. The analysis was started by data preprocessing, followed by predictor preselection, and thereafter the analysis of goodness of fit of the chosen model. Below is the final selected model with associated parameters.

- **CC**= 0.604768 + 0.001856***nir** + 0.012244***swir1** - 0.011423)***nirnar**
- Pseudo R-squared = **0.7481** & RMSE = **0.1476758**

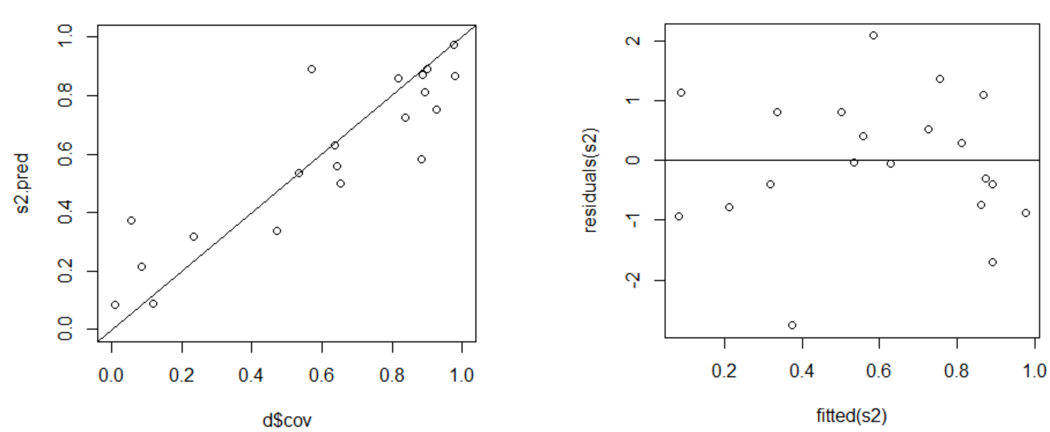
To obtain the final prediction, apply inverse of the logit link function.

Canopy Cover Final = exp(CC)/(1+exp(CC))

The RMSE is used to assess the accuracy or performance of the model, by measuring the average magnitude of the errors between predicted and actual observed values. The value 0.1477 obtained above indicates the average deviation from which I deduce that the model is a good fit for the data. This value was however affected by the inaccurate forest mask which caused prediction of high canopy cover values even on non-forest areas such as rocks, buildings, and paved surfaces.

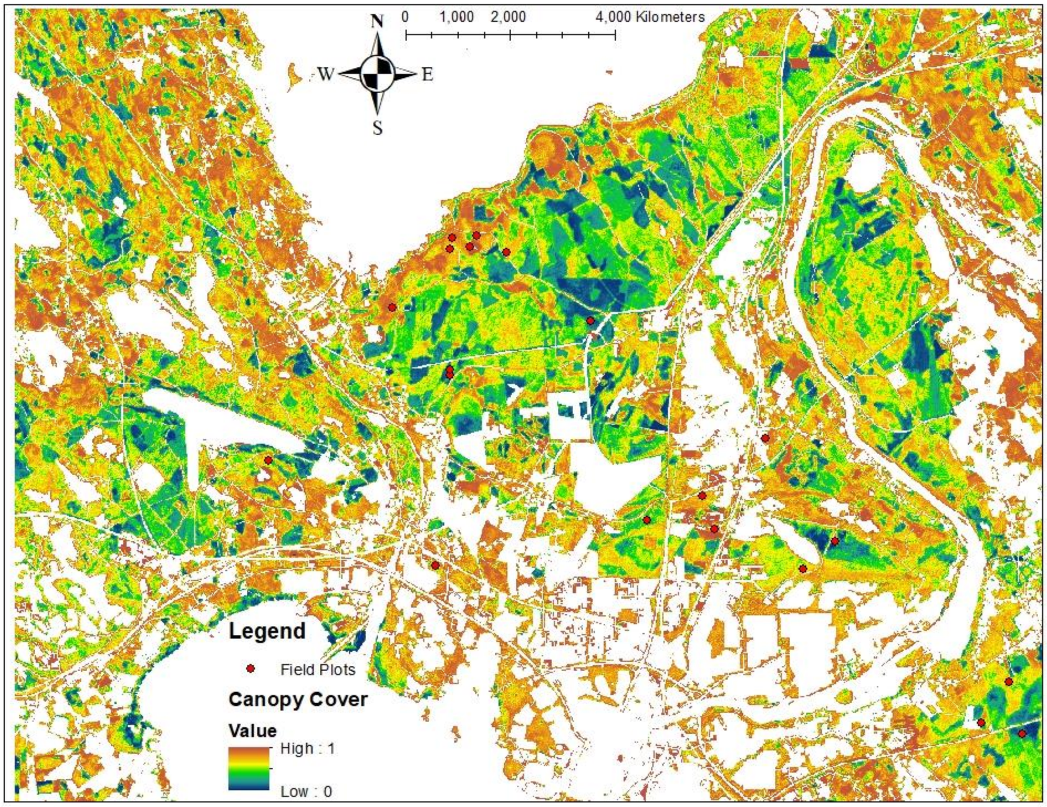
The R-squared measures the proportion of the variance in the dependent variable that is explained by the predictor variables in the model. Thus, from the result, up to 75% of the variability. Although not as high as in regression models for volume prediction which can be up to above 90%, it showed high practical significance in predicting canopy cover compared to some other models, for example, those with the blue band that had higher Pseudo R-squared values but poorer/less accurate predictions of canopy cover. This is likely because it is susceptible to atmospheric scattering and absorption by aspects such as aerosols which leads to significant noise in the data and making it less reliable. Furthermore, a 3-predictor variable model proved more efficient than that with 4 predictors. Below are the plots of the predicted versus observed values as well as the residuals versus fitted values.

A plot of predicted vs. observed values based on the model



Map development

CANOPY COVER MAP OF JOENSUU



Following confirmation of the final model from the beta regression analysis, raster reclassification was performed to distinguish the forested areas from non-forested areas using predefined classification rules. The creation of the forest mask enabled non prediction in areas like buildings and water bodies. The prediction of canopy cover by the model on the sentinel-2 image of the Joensuu area yielded a canopy cover mask which was further manipulated in QGIS to remove the no data areas and develop the final map that is displayed above. When the map was overlaid on the Google Earth image, it showed relatively consistent predictions of the canopy cover as was observed even in the field plot locations.

ANALYSIS OF HOST ORGANIZATION

About University of Eastern Finland Forest Inventory Research Group

The University of Eastern Finland (UEF) is an international, participatory and inclusive scientific community. It has four faculties: the Philosophical Faculty, the Faculty of Science, Forestry and Technology, the Faculty of Health Sciences, and the Faculty of Social Sciences and Business Studies. The campuses are in Joensuu and Kuopio, Finland.

Forest Inventory Research Group at UEF specializes in forest information technology research which covers topics ranging from applied geoinformatics to information technology (IT) -methods applied to forestry.

According to the University of Eastern Finland website, modern forestry is multi-functional and therefore management objectives may include non-timber benefits such as amenities, biodiversity conservation, and some environmental effects possible through forestry operations. Various optimization methods can be used to schedule the best treatments for forest stands.

The research group is located within the School of Forest Sciences whose mission is to educate specialists for jobs in the forestry and environmental field in an interdisciplinary and multicultural context with various exchange and international degree students. The School's main fields of research are forest ecology, forest management, forest planning, remote sensing forests with laser scanning, logistics of logging, and wood materials science.

The research group is led by Professors Matti Maltamo and Lauri Korhonen who oversee work performed by various Doctoral and Post-doctoral Researchers. Some specialisations include: forestry applications of airborne laser scanning, satellite image-based remote sensing, and ecological modelling.

SWOT ANALYSIS

STRENGTHS

- Strong expertise in remote sensing: The team has several skilled individuals with educational and work experiences in the field. Their knowledge of the field, and in particular analysis techniques and software enhance their contributions to the forest sector at the European/International scale.
- Modern equipment: There is availability of state-of-the-art data collection and analysis tools for example hypsometers, GPS devices, other mensuration tools, as well as computers. This enables a smooth and robust data collection process.

The organization has a good research record on various topics on forest remote sensing based on the numerous publications of the different members of the research group, many of which have been made during their time with the team.

WEAKNESSES

- Limited staffing: This makes internships through the year less likely. Furthermore, during the internship period, there was a period when in-person support was limited due to vacation break. Fortunately, a research assistant writing their Master's thesis was able to offer guidance throughout the internship period.

Avoid having interns without available in-person support as this would make the learning experience very challenging.

OPPORTUNITIES

- Software alternatives e.g. python: It would be good to build capacity in use of other programming languages that are becoming increasingly common. This would enable further collaborations with organizations locally such as with Luke as well as international ones to have multidisciplinary teams and share from their available resources. This is based off the fact that research publications already feature partners from broad fields of remote sensing thus it would be beneficial to continue these pursuits.

Main objectives to set

- Increase research productivity by publishing more papers, applying for several grants, and completing projects efficiently.
- Expanding research scope by building capacity in fields like ground-based vehicle lidar in addition to current competencies in airborne and satellite based remote sensing platforms.
- Improve research quality through new methodologies and advanced technologies such as developing in house programming language packages such as in R.
- Expanding knowledge transfer such as translating research findings for stakeholders like private companies and decision makers such as those in government.

THREATS

- Progression to newer technologies that may not be adopted: Many remote sensing practitioners are adopting Python language and there could be challenges collaborating with other teams of no capacity is built, despite having the same foundations as R. This can threaten the position of the organization as a partner.
- High costs: Data collection is expensive and time consuming. Furthermore, it can be difficult to recruit new staff and researchers.
- Restrictions that could prevent access to updated data.

These threats, however, may not threaten the future of the organization as it has competent professionals who keep refining their knowledge, possibilities for further collaborations through which funding can be attained, and a lot of data is open-source, thus it is possible to work around the prohibitive restrictions.