Aalto-yliopisto
Aalto-universitetet
Aalto University

Master's Programme in Geoinformatics

Estimation of boreal forest attributes using hyperspectral remote sensing data

Henriikka Karhuvaara

Master's thesis
2025

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Henriikka Karhuvaara

Title of thesis Estimation of boreal forest attributes using hyperspectral remote sensing data

Programme Master's Programme in Geoinformatics

Major Geoinformatics

Thesis supervisor Prof. Miina Rautiainen

Thesis advisor(s) Dr. Aarne Hovi and Dr. Jussi Juola

Date 3.2.2025 **Number of pages** 65 **Language** English

Abstract

Multispectral and hyperspectral satellite remote sensing can be applied to observe the Earth and detect changes caused by human actions or the climate. Currently there are global multispectral satellite missions that can observe a large variety of different targets on the Earth's surface and atmosphere by measuring their reflectance spectra with a few selected spectral bands. By 2030, satellite mission providers such as NASA and ESA plan to offer enhanced observation of different targets by measuring their reflectance spectra in hundreds of narrow spectral bands with global hyperspectral satellite missions.

In this work, hyperspectral satellite images collected by EnMAP (The Environmental Mapping and Analysis Program) mission were applied to estimate stem volumes and tree species proportions in five different boreal forest locations of Finland. Estimations were performed with two machine learning regression methods, and the estimation accuracies were assessed based on available field data collected by the Finnish Forest Centre. In general, the best estimation accuracies were reached for total stem volumes with RRMSE values between 52–93 %. Estimation accuracies of total stem volumes were also the most consistent between the study locations. Estimation accuracies of stem volumes and proportions related to single species were more dependent on the species distribution of the study locations, and the best accuracies were reached for the most dominant species of each location with RRMSE values between 44–143 %. Higher mean species volumes and proportions in the field data were related to better estimation accuracy also between the study locations.

This study gave information on the applicability of hyperspectral satellite data to estimate selected boreal forest attributes in a few targeted locations. More reliable conclusions on the applicability of the estimations, for example, to monitor forest biodiversity can be made once the availability of hyperspectral satellite data increases.

Keywords Passive optical remote sensing, hyperspectral satellite remote sensing, machine learning, estimation of biophysical variables, boreal forest attributes, biodiversity indicators

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Henriikka Karhuvaara

Työn nimi Boreaalisten metsien puustotunnusten estimointi hyperspektrisestä kaukokartoitusaineistosta

Koulutusohjelma Geoinformatiikan maisteriohjelma

Pääaine Geoinformatiikka

Vastuuopettaja/valvoja Prof. Miina Rautiainen

Työn ohjaaja(t) MMT Aarne Hovi ja TkT Jussi Juola

Päivämäärä 3.2.2025 **Sivumäärä** 65 **Kieli** Englanti

Tiivistelmä

Multispektrisillä ja hyperspektrisillä satelliittisensoreilla voidaan kerätä tietoa maapallosta sekä ihmisen ja ilmaston maapallolle aiheuttamista muutoksista. Tällä hetkellä globaalien multispektristen satelliittihankkeiden avulla voidaan tarkkailla erilaisia kohteita maapallon pinnalla ja ilmakehässä mittaamalla niiden heijastamaa auringon säteilyä muutamalla valitulla aallonpituusalueella. Vuoteen 2030 mennessä muun muassa NASA ja ESA aikovat tarjota globaalisti myös hyperspektristä satelliittidataa, mikä mahdollistaa eri kohteiden yksityiskohtaisemman tarkkailun mittaamalla niiden heijastamaa säteilyä sadoilla kapeilla aallonpituusalueilla.

Tässä työssä estimoitiin boreaalisten metsien puuston runkotilavuutta ja puulajiosuuksia viidestä EnMAP-satelliitin tuottamasta hyperspektrisestä kuvasta Suomesta. Estimointi tehtiin kahdella eri koneoppimismenetelmällä ja estimointitarkkuutta arvioitiin saatavilla olevan Metsäkeskuksen koeala-aineiston perusteella. Yleisesti ottaen paras estimointitarkkuus saavutettiin puuston kokonaisrunkotilavuudelle RRMSE vaihteluvälillä 52–93 %. Puuston kokonaisrunkotilavuuden estimointitarkkuus oli myös yhtenäisin eri tutkimusalueiden välillä. Lajikohtaisen runkotilavuuden ja lajiosuuden estimointitarkkuus oli riippuvainen alueiden lajijakaumasta, sillä paras estimointitarkkuus saavutettiin kunkin tutkimusalueen yleisimmälle puulajille RRMSE vaihteluvälillä 44–143 %. Puulajin suurempi keskiarvorunkotilavuus ja -osuus koeala-aineistossa paransi estimointitarkkuutta myös eri tutkimusalueiden välisessä vertailussa.

Tämä tutkimus antoi tietoa hyperspektrisen satelliittidatan soveltuvuudesta boreaalisten puustotunnusten estimointiin muutamalla valitulla tutkimusalueella. Puustotunnusten estimoinnin soveltuvuudesta muun muassa boreaalisten metsien biodiversiteetin kartoittamiseen voidaan tehdä luotettavampia johtopäätöksiä, kun hyperspektrisen satelliittidatan saatavuus paranee.

Avainsanat Passiivinen optinen kaukokartoitus, hyperspektrinen satelliittikaukokartoitus, koneoppiminen, biofysikaalisten muuttujien estimointi, boreaalisten metsien puustotunnukset, biodiversiteetti-indikaattorit

Contents

Preface and acknowledgements

I want to thank Miina Rautiainen, Aarne Hovi and Jussi Juola for the interesting topic of this master's thesis and guidance in the process. Preparing this work was a good learning opportunity and the writing process was immersive.

Otaniemi, 3 February 2025
Henriikka Karhuvaara

1 Introduction

1.1 Background

Observing and analysis of the Earth is essential to understand its response to the changing climate and human actions. Forests as the largest terrestrial carbon sink, cover approximately 30 % of the Earth's land area and have an important role in the global carbon cycle (FAO 2020). The importance of forests is shown in several programs that aim to protect the forest ecosystem and biodiversity. For example, in the EU Biodiversity Strategy for 2030 it is separately mentioned that all remaining primary and old-growth forests should be monitored and strictly protected as these forests are one of the richest ecosystems on the Earth (European Commission, Directorate-General for Environment 2021). Additionally, the United Nation's Sustainable Development Goals (SDGs) that are defined to guide the functions of member states between 2016 and 2030 mention the important role of forests in the terrestrial ecosystems' sustainability (FAO 2020).

Remote sensing can be applied to monitor and model different targets on Earth's surface and atmosphere by measuring their reflectance spectra with selected spectral bands. In forested areas, the reflectance spectrum is affected by properties such as leaf pigmentation and cell structure, canopy structure and density of leaves and branches (Fassnacht et al. 2016; Richards 2022). Especially hyperspectral satellite instruments have been successfully applied to quantify this variation between forest ecosystems and can possibly be applied to monitor crucial proporties of forests, such as biodiversity (Chabrillat et al. 2022).

Currently, the amount of hyperspectral satellite missions is growing, and mission providers such as ESA (2021) and NASA (2018) are planning to offer hyperspectral satellite data systematically and globally by 2030 with their CHIME (The Copernicus Hyperspectral Imaging Mission) and SBG (The Surface Biology and Geology) missions. The forecasted growth in the availability of hyperspectral satellite data has resulted in development of efficient data processing methods, and especially different machine learning regression methods have given promising results in estimation of biophysical variables from hyperspectral data (Verrelst et al. 2019). However, to increase the applicability and reliability of these methods, understanding of the factors that affect the estimation accuracies is important (Fassnacht et al. 2016).

The planned global hyperspectral satellite data combined with efficient and reliable data processing methods could offer a powerful tool to guide and monitor sustainable forest management and protect forest biodiversity

(Corona et al. 2011; Gauthier et al. 2015). In addition to forestry, these technologies could serve a wide range of sectors related to the current environmental questions and challenges (Chabrillat et al. 2022).

1.2 Objectives of this master's thesis

The overall goal of this work was to estimate different boreal forest attributes from hyperspectral satellite data collected by EnMAP (The Environmental Mapping and Analysis Program) mission. The estimations were performed with two machine learning regression methods that have previously been successfully applied to estimate biophysical variables from hyperspectral data. Hyperspectral satellite images from five different boreal forest locations were selected for the analysis to grow understanding on possible factors that affect the estimation accuracies. The studied boreal forest locations were in Finland and focus of the estimations was on attributes that are related to boreal forest biodiversity. The estimated boreal forest attributes were the total stem volumes, species-specific stem volumes and species proportions.

Research questions of this master's thesis were following:

- What kind of results were reached in the estimations of total stem volumes, species-specific stem volumes and species proportions from the hyperspectral satellite data?
- Were there differences in estimation accuracies between the boreal forest study sites in Finland? What kind of factors were related to the differences?

2 Literature review

This section presents the main topics related to this work. Section 2.1 shortly presents the science of passive optical remote sensing in general and then focus is moved to hyperspectral satellite missions. Section 2.2 presents statistical methods that can be applied in estimation of biophysical variables from the growing stream of hyperspectral data. Section 2.3 presents previous studies related to estimation of boreal forest attributes with suggestions of estimated attributes in context of boreal forest biodiversity.

2.1 Hyperspectral satellite remote sensing

2.1.1 Passive optical remote sensing

Passive optical remote sensing sensors measure electromagnetic radiation from solar origin. The sensors are typically mounted on aircraft and satellite platforms to measure the reflectance spectra of observed targets. Multispectral instruments make measurements in a few selected spectral bands whereas hyperspectral instruments can make measurements in hundreds of narrow spectral bands, resulting into a contiguous reflectance spectrum. The spectral bands in these kinds of optical instruments are typically between wavelengths 400–2500 nm. Different remote sensing instruments can be compared, for example, based on their spectral, spatial and temporal resolution. Spectral resolution is determined by the bandwidth of measured wavelengths, spatial resolution by the area from which a single measurement can be made from and temporal resolution is related to the repetitive measurement coverage of the moving platform (Camps-Valls et al. 2022).

Currently and in the past, multispectral satellite remote sensing has offered an efficient method to observe vast areas of the Earth in different locations. Satellite missions such as Landsat by NASA (2024) and Sentinel by ESA (2024) provide multispectral satellite data systematically and globally, and the missions have comprehensively improved the understanding of Earth's processes and resources. Hyperspectral instruments have been mostly applied from aerial platforms, but they allow more advanced analysis of Earth targets as they preserve their characteristic reflectance spectra better than multispectral instruments (Camps-Valls et al. 2022; Goetz 2009; Schaepman et al. 2009).

The applicability of hyperspectral satellite instruments has increased in the 21st century thanks to advancements in computational capacity and imaging technology (Goetz 2009). For example, with push-broom technology the recording time of electromagnetic radiation is longer for each image pixel and

measurements can be done in multiple dimensions, allowing more information to be collected (Richards 2022). The development of hyperspectral satellite imaging is expected to enable global monitoring of the Earth with an unforeseen accuracy and detail, improving the understanding of Earth processes and environmental challenges (Chabrillat et al. 2024). The continuous reflectance spectrum measured with a hyperspectral instrument for each satellite image pixels, gives detailed information on the biophysical and chemical properties of Earth's atmospheric, aquatic and terrestrial processes (Goetz 2009; Schaepman et al. 2009).

2.1.2 Hyperspectral satellite missions

Experiments on the hyperspectral satellite imaging started with Hyperion mission. The Hyperion Imaging Spectrometer aboard EO-1 spacecraft was launched in 2000 as part of a demonstrative mission by NASA to gain information on applicability of hyperspectral imaging in satellite remote sensing (Folkman et al. 2001). The Hyperion mission proved that hyperspectral satellite monitoring is possible, and it is the background for the technical requirements in many of the current available and planned hyperspectral satellite missions (Pearlman et al. 2003). The push-broom Hyperion instrument had a spectral coverage of 400–2500 nm, spatial resolution of 30 m, spectral resolution of 10 nm and temporal resolution of 16 days (ESA 2023).

Currently, there are a few hyperspectral satellite missions that offer targeted hyperspectral data for public use with similar technical properties as the Hyperion instrument. However, the temporal resolution of these missions is still limited. Some examples of these kind of missions are EnMAP (The Environmental Mapping and Analysis Program) by the German Aerospace Center (DLR), PRISMA (PRecursore IperSpettrale della Missione Applicativa) by the Italian Space Agency (ASI) and ZiYuan 1 and GaoFen 5 by the China National Space Administration (CNSA).

EnMAP mission has a primary goal of providing hyperspectral satellite data to improve the understanding of different environmental processes on the Earth (Guanter et al. 2015; Storch et al. 2023). The mission has shared hyperspectral data since November 2022 to various applications (Chabrillat et al. 2024). EnMAP instrument has a spectral coverage of 420–2450 nm, spectral resolution of 6.5–10 nm, spatial resolution of 30 m and temporal resolution of 27 days (Kaufmann et al. 2016). Main applications of the EnMAP mission are vegetation analysis, geology, waters, cryosphere, urban areas, atmosphere and hazards (Chabrillat et al. 2024). PRISMA mission was launched in 2019, and it provides hyperspectral satellite data of selected targets, for example, to resource management, environmental monitoring and disasters management (Loizzo et al. 2018). PRISMA instrument has a spectral

coverage of 400–2505 nm, spectral resolution of 12 nm, spatial resolution of 30 m and temporal resolution of 29 days (Guarini et al. 2018). ZiYuan 1 and GaoFen 5 have similar properties and goals as the EnMAP and PRISMA missions (CRESDA 2022a; 2022b).

Hyperspectral measurements have also been made with instruments that apply the International Space Station (ISS) as a measurement platform. Examples of these kinds of instruments are EMIT (Earth Surface Mineral Dust Source Investigation) by NASA Jet Propulsion Laboratory and DESIS (The DLR Earth Sensing Imaging Spectrometer) by the German Aerospace Center, and they serve more targeted scientific research goals (DLR 2019; Green et al. 2020). The main goal of EMIT instrument is to measure the mineral composition of Earth's arid dust source regions (Green et al. 2020). Data collected by DESIS instrument has been applied, for example, to support methodological development in different applications related to hyperspectral satellite monitoring (DLR 2019).

In addition to the current and past missions, there are planned hyperspectral satellite missions with goals of providing public hyperspectral satellite remote sensing data globally and systematically by 2030. For example, CHIME (The Copernicus Hyperspectral Imaging Mission for the Environment) is a hyperspectral satellite mission by the European Union (EU) and the European Space Agency (ESA) to address environmental challenges (ESA 2021). SBG (The Surface Biology and Geology) is a hyperspectral satellite mission by NASA with similar goals (NASA 2018). The planned global hyperspectral satellite missions are developed based on the information gained in the previous aerial and spaceborne hyperspectral imaging experiments (ESA 2021; NASA 2018).

There are also available and planned commercial hyperspectral satellite missions from companies such as GHGSat, Kuva Space and Planet. These kinds of providers offer hyperspectral satellite data for commercial use to applications such as monitoring industrial sites, decision-making and reporting of emissions (GHGSat 2024; Kuva Space 2023; Planet 2024). Similarly as the public missions, the commercial providers have goals of improving sustainability and enforcing green policies (Tikka et al. 2023). Public and private data providers also collaborate, and for example, Planet is part of a public-private partnership, including organizations such as NASA and Carbon Mapper (Duren et al. 2021).

2.2 Statistical analysis of hyperspectral data

Earth processes can be observed by defining parameters to represent different properties of land, water and the atmosphere (Bioucas-Dias et al. 2013). For example, statistical methods can be applied to efficiently estimate these different parameters from the growing stream of hyperspectral satellite data. Statistical approaches can be divided into regression and classification methods. Traditionally, regression methods answer to problems with a quantitative response and classification methods to problems with a qualitative response (James et al. 2013).

In estimations made from hyperspectral data, the classifiers and regression models are typically formed based on training data. Training data consists of observation pairs with measured reflectance spectrum and related field data value of the estimated attribute (Verrelst et al. 2019). Classification methods have been applied, for example, to estimate different tree species classes (Fassnacht et al. 2016). Regression methods, on the other hand, have been applied to estimate numerical descriptors of forest areas, such as tree volumes and heights (Halme et al. 2019; Hyyppä et al. 2000).

Especially non-parametric regression methods, also referred to as machine learning regression algorithms, are well suitable for analysis of hyperspectral data. These kinds of methods can utilize the whole measured reflectance spectrum and estimate a large variety of different variables (Verrelst et al. 2019). Non-parametric regression methods form the estimation functions without predefined limitations, and they offer more flexibility when compared to parametric regression methods that optimize the free parameters of an estimation function with a predefined shape (James et al. 2013).

Random Forest Regression (RF), Support Vector Regression (SVR), Gaussian Process Regression (GPR) and different artificial neural networks (ANNs) are examples of machine learning regression methods that have been applied to estimate biophysical variables from hyperspectral data (Feilhauer et al. 2017; Huang 2004; Tuia et al. 2018; Verrelst 2012). SVR forms a kernel-function to estimate the selected variable from the spectral data with an estimation error within a predefined range (Smola & Schölkopf 2004; Verrelst et al. 2019). GPR is also a kernel-based regression method, and it results into a probabilistic output by functioning in the Bayesian framework (Rasmussen & Williams 2006). RF applies a combination of different decision trees to find the best combination of features for estimating the selected target (Breiman 2001; Shafri 2017). A basic example of ANN is formed by connected layers of artificial neurons with weighted links related to the estimation accuracies of the network paths (Haykin 1999). RF and SVR are quite simple to implement and well available in different software packages, whereas ANNs

and GPR are more complex, which can be seen as a limitation for a wider use (Shafri 2017).

In addition to the presented machine learning regression methods, there is a large amount of other statistical methods that offer unique benefits and disadvantages to analysis of hyperspectral data. The selection of applied methods should be made based on the goals, and it is typically made as a compromise between the model interpretability and flexibility (James et al. 2013).

2.3 Estimation of boreal forest attributes

2.3.1 Estimation accuracies of different boreal forest attributes

Because of the current data availability, most of the previous studies focused on estimation of boreal forest attributes have applied multispectral satellite data and hyperspectral aerial data. Applicability of this kind of data to estimate boreal forest attributes has been studied, for example, by Hyyppä et al. (2000), Lee et al. (2004) and Halme et al. (2019). The multispectral data applied by Hyyppä et al. (2000) and Lee et al. (2004) was measured with 6–7 bands between wavelengths 450–2350 nm by Landsat TM/ETM+ satellites. Halme et al. (2019) applied multispectral data measured with 9 bands between wavelengths 443–945 nm by Sentinel-2 satellite. As the hyperspectral aerial data, Lee et al. (2004) applied the largest amount of bands, total of 193 bands between wavelengths 370–2500 nm followed by Halme et al. (2019) with 128 bands between wavelengths 400–1000 nm. Hyyppä et al. (2000) applied the lowest amount of bands, total of 30 bands between wavelengths 466–870 nm.

In more detail, Lee et al. (2004) compared applicability of the multispectral satellite data and hyperspectral aerial data to track vegetation leaf area index (LAI) in four different biomes, including boreal forest zones in North America. The best estimations for LAI were reached when regression analysis was applied to individual bands in the hyperspectral data. Wavelengths especially in the vegetation red-edge (710–750 nm) and shortwave-infrared region (900–1700 nm) were identified important, although the most important benefit of hyperspectral data in comparison to multispectral data was the larger amount of bands.

Hyyppä et al. (2000) estimated stem volume ($m^3\ ha^{-1}$), basal area ($m^2\ ha^{-1}$) and mean tree height (m) in a boreal forest site in Finland with different machine learning regression algorithms. Estimations performed from the hyperspectral aerial data reached better estimation accuracies than estimations made from the multispectral satellite data for all the estimated attributes. The calculated coefficient of determination (R^2) varied between 0.26–0.31

for estimations made from the multispectral data and between 0.47–0.58 for estimations made from the hyperspectral data. The relative standard error (SE) varied between 37–56 % for estimations made from the multispectral data and between 31–45 % for estimations made from the hyperspectral data.

Also, Halme et al. (2019) estimated mean tree height (m), basal area ($m^2\ ha^{-1}$) and leaf area index (LAI). Additionally, stem biomass (t ha^{-1}) and main tree species from classes of pine, spruce and broadleaved species were estimated. Estimations were performed with machine learning regression SVR and GPR methods from a boreal forest location in Finland. Similarly as in the previously presented studies, estimations made from the hyperspectral aerial data were generally more accurate than estimations made from the multispectral satellite data. The accuracy of main tree species estimations was measured with an overall accuracy, and it varied between 83–88 % for estimations made from the multispectral data and between 87–93 % for estimations made from the hyperspectral data. For rest of the estimated attributes, R^2 varied between 0.56–0.75 for estimations made from the multispectral data and between 0.59–0.83 for estimations made from the hyperspectral data. The relative root-mean-square error (RRMSE) varied between 17–29 % for estimations made from the multispectral data and between 15–29 % for estimations made from the hyperspectral data.

Currently, there are some studies that have compared estimations of boreal forest attributes made from multispectral satellite data and the available hyperspectral satellite data of targeted locations. However, the amount of these kind of studies is limited. Examples of estimation accuracies reached for boreal forest attributes from hyperspectral satellite data can be found from a study made by Juola et al. (2024). They compared the applicability of multispectral and hyperspectral satellite data to estimate leaf area index (LAI) and canopy closure (CC) from a conifer-dominated boreal forest location in Finland. The applied multispectral data was collected with 12 bands between wavelengths 442–2186 nm by Sentinel-2 satellite and the hyperspectral data was collected with 219 bands between wavelengths 418–2445 nm by EnMAP satellite. Estimations were made with univariate and multivariate generalized additive models (GAMs). The results were aligned with the previously presented comparison between estimations made from multispectral satellite data and hyperspectral aerial data. In general, also the estimations made from hyperspectral satellite data had better accuracies than the ones made from multispectral satellite data. The best estimations made from hyperspectral data had approximately 1–6 % lower RRMSE values than the best estimations made from multispectral data.

Previous studies have shown that hyperspectral data improves especially the estimation accuracy of boreal forest attributes that include species-specific

information (Halme et al. 2019; Juola et al. 2024; Lee et al. 2004). Comparison of estimation accuracies reached for boreal forest attributes related to different tree species can be found from a study made by Tuominen et al. (2017). They estimated the total volume of growing stock ($m^3\ ha^{-1}$) and species-specific volumes of growing stock ($m^3\ ha^{-1}$) for pine, spruce and broadleaved species. Estimations were performed with the k-nearest neighbor (k-NN) method from multispectral satellite data of a boreal forest location in Finland. The multispectral data was collected with 7 bands between wavelengths 433–2300 nm by Landsat 8 satellite. From all the estimated attributes, the total volume of growing stock reached the best accuracy with RRMSE value at 60.95 %. From the species-specific estimations, pine as the most dominant species reached the best accuracy with RRMSE value at 81.67 % and spruce as the least dominant species reached the lowest estimation accuracy with RRMSE value at 189.74 %. Estimations of broadleaved species reached RRMSE value at 162.08 %.

The presented studies concluded that use of hyperspectral data allows enhanced estimation accuracies to the estimated boreal forest attributes when compared to the multispectral data. However, the benefits of hyperspectral satellite remote sensing over the current multispectral satellite missions can be more reliably assessed once the availability of hyperspectral satellite data grows. More studies are also needed especially to understand the factors that cause differences in the estimation accuracies of different boreal forest attributes and locations (Fassnacht et al. 2016). The larger amount of hyperspectral satellite data will also ensure a better representability of the training data required in development of the predictive models. Use of more independent datasets will reduce the possible overfitting of developed models and enable better estimation accuracies on a larger scale (James et al. 2013; Lee et al. 2004).

2.3.2 Approaches for the selection of estimated attributes

The selection of estimated boreal forest attributes can be made to support different goals. For example, Tuominen et al. (2017) estimated the growing stock volume that is typically applied to monitor forest's resources and functionality as a carbon sink (FAO 2020). Lee et al. (2004) and Juola et al. (2024) estimated forest attributes that can be applied to model the forest canopy, whereas Hyyppä et al. (2000) and Halme et al. (2019) estimated variables that can be applied to monitor forest structure and biomass. These kinds of estimations could possibly be applied to support sustainable forest management and forest biodiversity.

Also, the goal of this work was to estimate boreal forest attributes that are related to forest biodiversity. The definition of forest biodiversity is complex

but generally, it can be described as the diversity of life in all its levels and forms within forested areas, including trees and other inhabitants (Hunter 1999). Biodiversity can be measured with different spatial scales such as diversity of a single ecosystem, changes in diversity between different ecosystems and overall diversity of ecosystems within a region (Whittaker 1972). Also, variety of elements, ecological processes and structural organization of elements within a region can be used to measure biodiversity (Noss 1990).

Especially old growth forests have rich biodiversity as they typically have variability in species and structural diversity in trees (Kneeshaw & Gauthier 2003). Both species richness and structural complexity are, in fact, approaches used in the definition of biodiversity indicators also in other types of forests (Corona et al. 2011). In addition to species richness and structural complexity, also proportion of broadleaved species is linked to biodiversity in boreal forests (Metsäkeskus 2024b). Forest attributes applicable to measure these known biodiversity indicators of boreal forests offer a good approach for the estimations.

3 Material and methods

This section presents the research material and methods applied in this work. Section 3.1 presents the overview of the study material and locations. Sections 3.2 and 3.3 present the applied hyperspectral satellite data and field data in more detail. Section 3.3 additionally defines the estimated boreal forest attributes based on the field data. Section 3.4 gives a step-by-step description of the data processing flow and estimation methods.

3.1 Overview

Five hyperspectral satellite images collected by the EnMAP mission were used to estimate the selected boreal forest attributes related to biodiversity (EnMAP data ©DLR 2023-2024). Images were collected near the peak of growing season from Finland where boreal forests cover approximately 86 percent of the total land area (Räty & Vaahtera 2023). The study locations were defined based on the applied satellite images and named based on the cities nearest to the satellite images: Sodankylä, Perho, Kuopio, Äänekoski and Virrat (Figure 1).

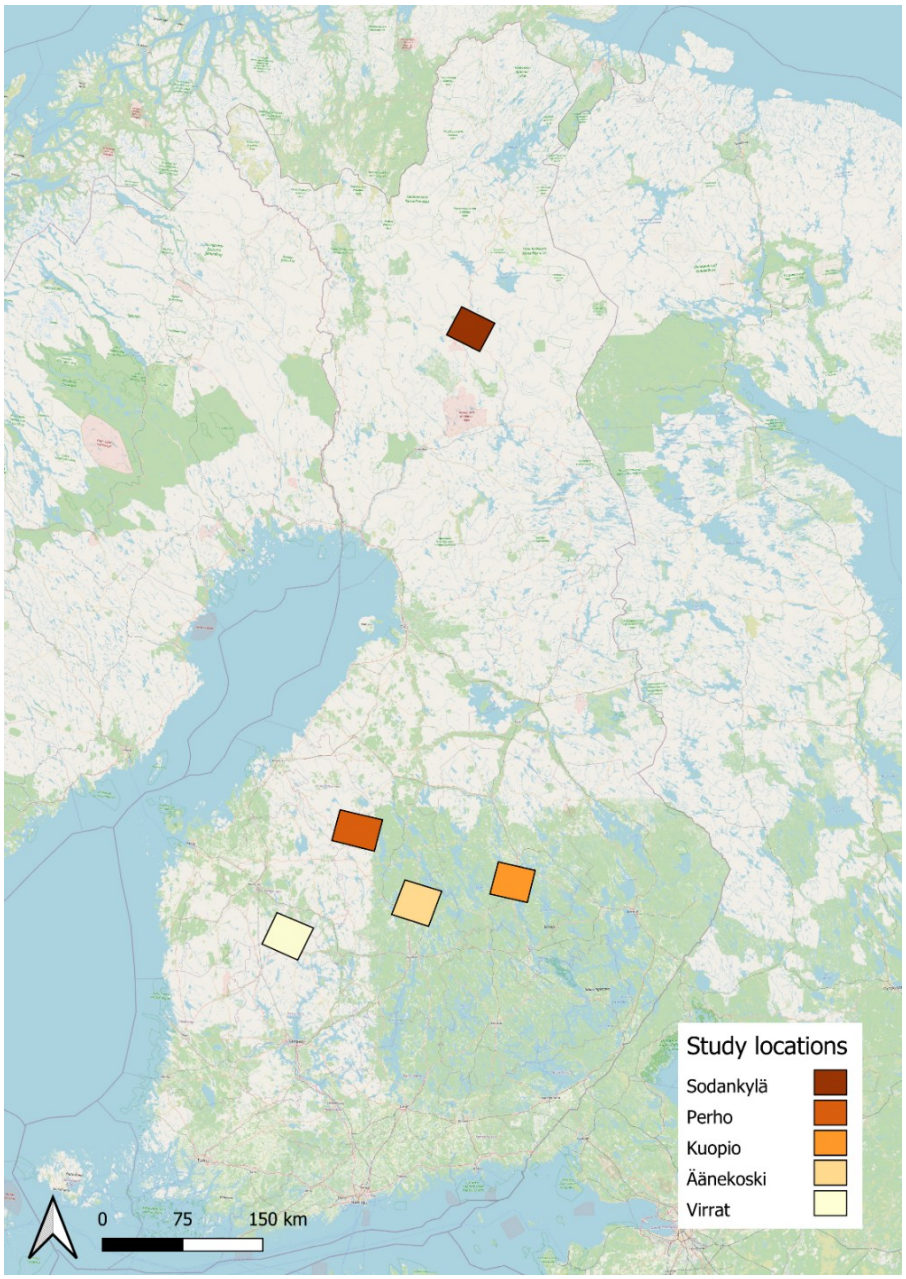

Figure 1. Study locations defined based on the applied satellite images and named based on the nearest city.

The Finnish Forest Centre maintains an open dataset that is described to cover the variability of Finnish forests, and it was used as field data in this study (Metsäkeskus 2024a). The field data has a national coverage, and it is mentioned to have a great potential for future remote sensing applications for example by Halme et al. (2019). The applied field data sample plots were selected based on the extent of applied hyperspectral satellite images and data collection dates (Table 1). Additionally, the selected field plots were presented on top of RGB visualizations of the satellite images to ensure an even distribution within the image extents (Figures 2–6).

Table 1. Information on satellite images and field data.

Study location name	Satellite image acquisition date	Field data name	Field data collection year	Number of field plots within satellite image
Sodankylä	7.8.2024	Peurasuvanto	2023	287
Perho	27.6.2024	Perho	2023	218
Kuopio	11.6.2023	Pohjois-Savo	2021	98
Äänekoski	12.6.2023	Keski-Suomi	2020	214
Virrat	25.6.2024	Alavus, Kurikka	2022, 2023	197

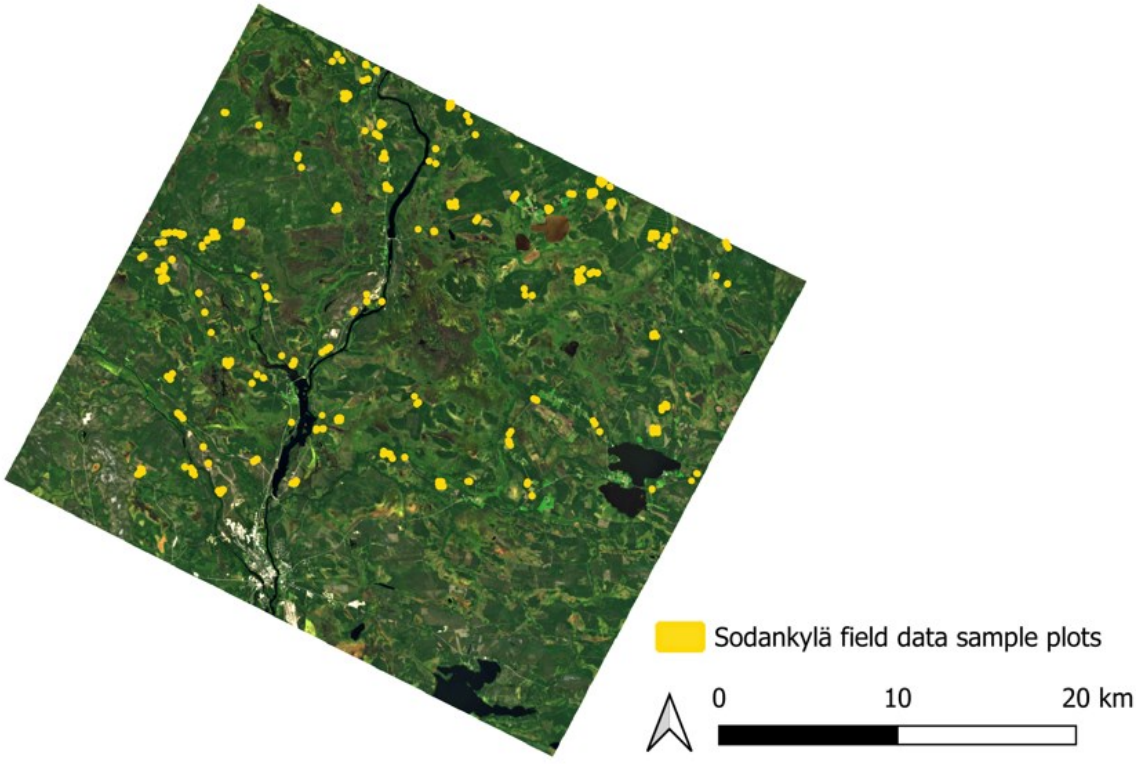

Figure 2. Locations of field plots in the RGB visualization of the hyperspectral satellite image from Sodankylä.

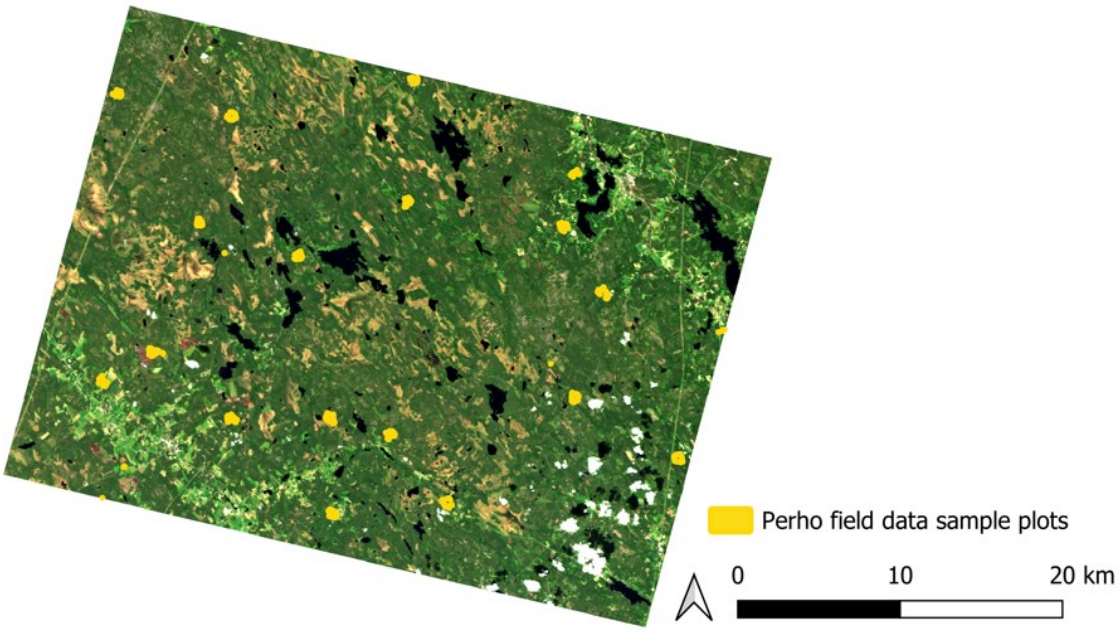

Figure 3. Locations of field plots in the RGB visualization of the hyperspectral satellite image from Perho.

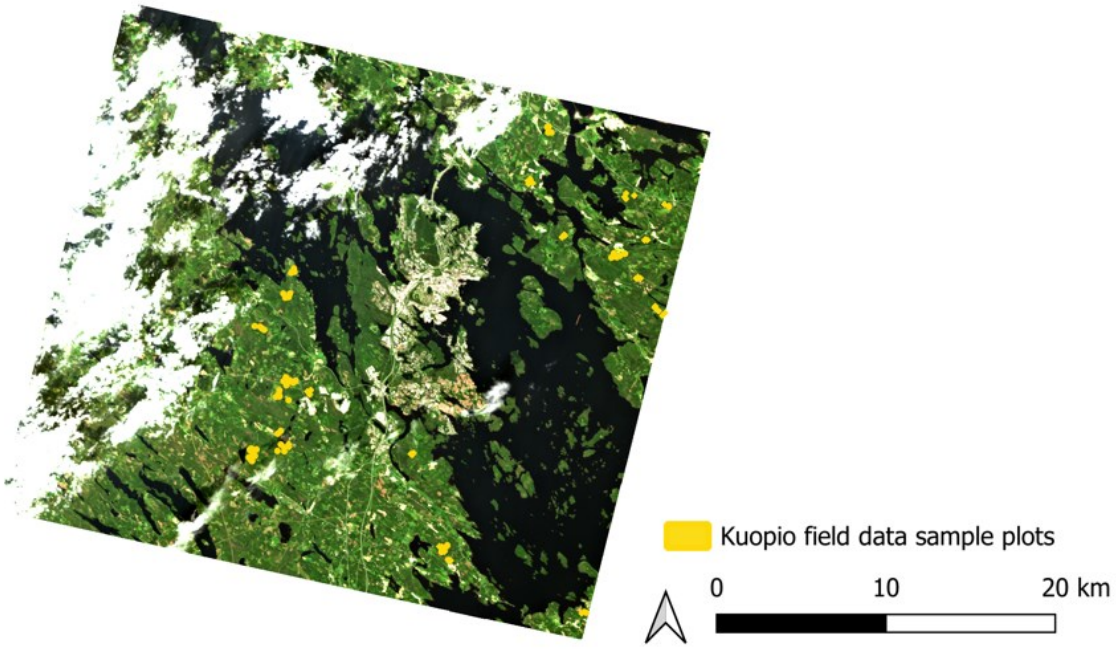

Figure 4. Locations of field plots in the RGB visualization of the hyperspectral satellite image from Kuopio.

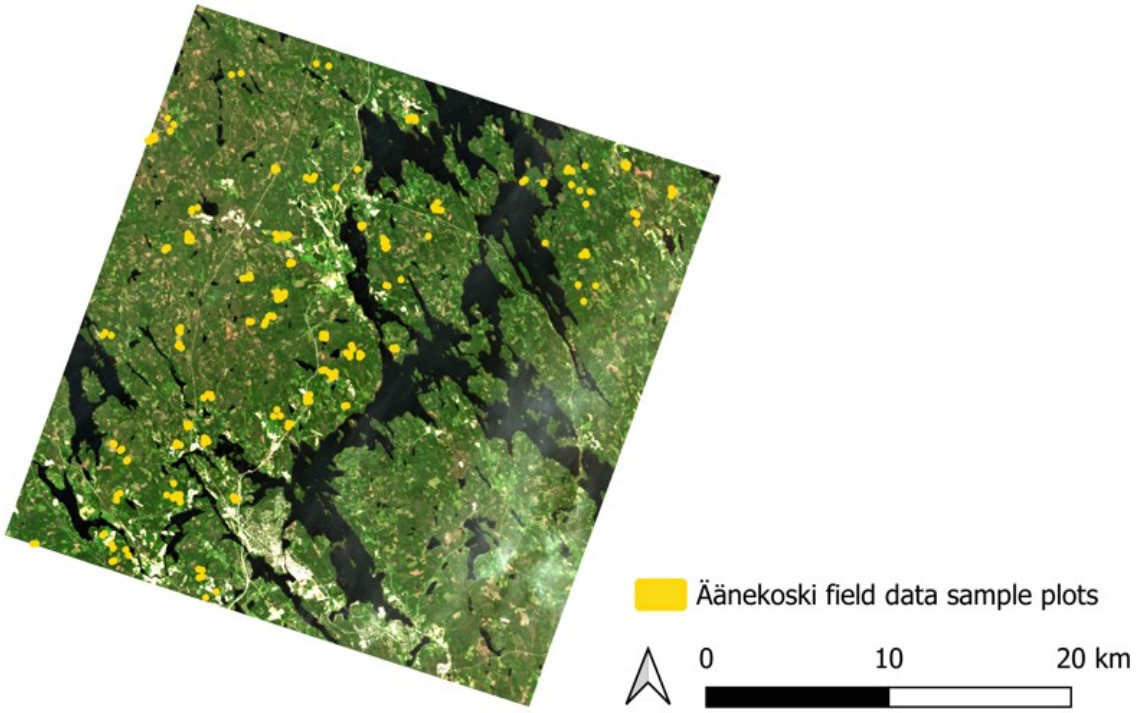

Figure 5. Locations of field plots in the RGB visualization of the hyperspectral satellite image from Äänekoski.

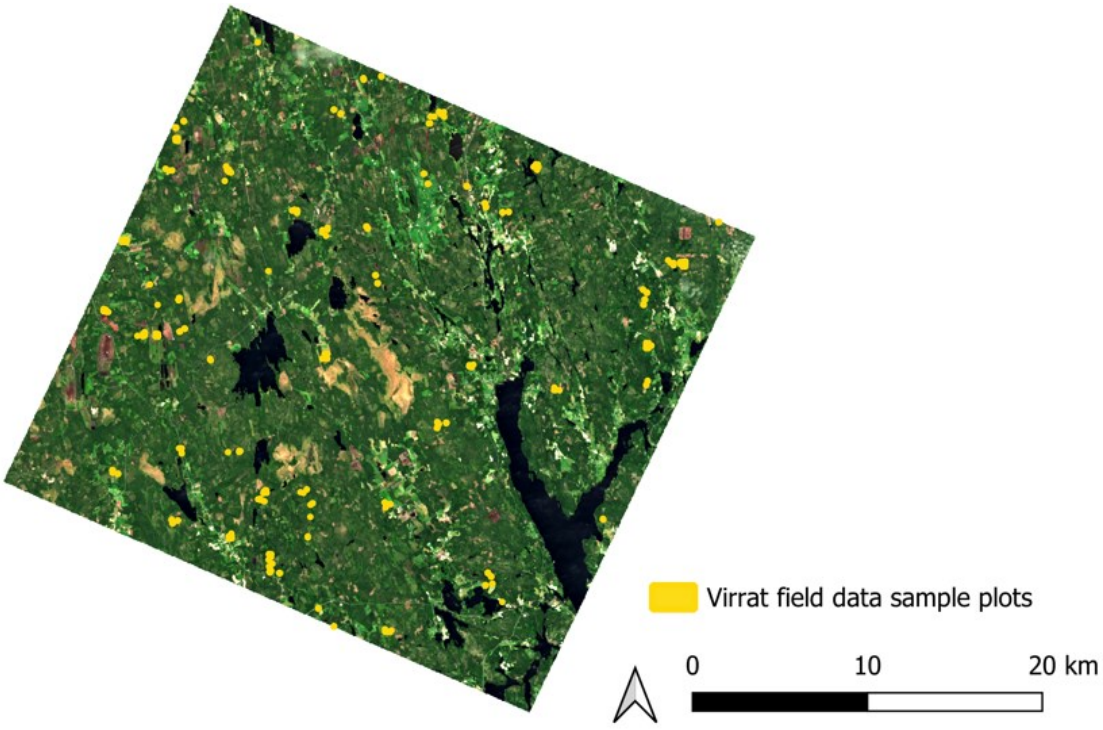

Figure 6. Locations of field plots in the RGB visualization of the hyperspectral satellite image from Virrat.

3.2 Hyperspectral satellite data by the EnMAP mission

Hyperspectral satellite images from Finland were searched from the EnMAP Data Access Portal (DLR 2024a). Images were searched based on the acquisition date and cloud coverage. The interest was in images collected during the Finnish summertime between June and August with a cloud coverage less than 20 percent. Only images from summertime were used as the reflectance spectrum measured near the peak of growing season was wanted for the analysis. Additionally, a large enough number of available field plot locations was required. Applicable satellite images were ordered as Level 2A products with orthorectification and atmospheric correction for land with the recommended default parameters (Table 2). Also, image ID, sun elevation angle and image acquisition start and stop time were saved for each of the ordered satellite images (Table 3). Images collected from study locations of Kuopio, Äänekoski and Virrat were processed with EnMAP processor version V01.04.02 and images collected from study locations of Sodankylä and Perho with version V01.05.01.

Table 2. Processing parameters for EnMAP Level 2A products.

Download parameter	**Selected value**
Processing level	L2A
Format	GeoTIFF and Metadata
Projection	UTM Zone Of Scene Center
Resampling	Bilinear Interpolation
Atmospheric Processing Type	Land Mode
Terrain Correction	Automatic
Band Interpolation	No
Cirrus and Haze Removal	No
Ozone Column	Automatic
Season	Automatic

Table 3. Detailed information of the used EnMAP satellite images.

Study location name	Image ID	Sun elevation angle	Image acquisition start and stop time
Sodankylä	ENMAP01-____L2A-DT0000086428_20240807T105513Z_001_V010501_20241028T194318Z	38.3°	2024-08-07T10:55:13.005, 2024-08-07T10:55:17.453
Perho	ENMAP01-____L2A DT0000079836_20240627T095559Z_001_V010501_20241028T194258Z	49.7°	2024-06-27T09:55:59.975, 2024-06-27T09:56:04.423
Kuopio	ENMAP01-____L2A-DT0000021509_20230611T095216Z_003_V010402_20240905T212739Z	50.1°	2023-06-11T09:52:07.980, 2023-06-11T09:52:21.332
Äänekoski	ENMAP01-____L2A-DT0000021653_20230612T101731Z_002_V010402_20240929T143021Z	50.4°	2023-06-12T10:17:26.998, 2023-06-12T10:17:35.898
Virrat	ENMAP01-____L2A-DT0000079509_20240625T104254Z_001_V010402_20241006T032309Z	50.9°	2024-06-25T10:42:54.025, 2024-06-25T10:42:58.473

3.3 Field data and the estimated boreal forest attributes

The applied field data is measured during the Finnish spring and summer with a minimum positioning accuracy of 1 m, and it is published under Creative Commons Attribution 4.0 International (CC BY 4.0) license (Metsäkeskus 2024d). Field data until 2021 was available from the Finnish Forest Centre data portal (Metsäkeskus 2024c). Newer data was available with an order from the Finnish Forest Centre. The field data was delivered based on measurement locations and data collection years. Each measurement location included sample plots with information such as id, coordinates and measurement date. Each sample plot was related to separate tree species-specific tables that included information such as stem volume, basal area, mean height and mean age of the species within the sample plot.

The main tree species in Finland is Scots pine (*Pinus sylvestris* L.), followed by Norway spruce (*Picea abies* (L.) Karst), silver birch (*Betula pendula* Roth), downy birch (*Betula pubescens* Ehrh.) and other broadleaved species (Räty & Vaahtera 2023). As mentioned in Section 2.3.2, species-richness,

proportion of broadleaved species and structural complexity are indicators of boreal forest biodiversity (Corona et al. 2011; Metsäkeskus 2024b). Additionally, it was mentioned that growing stock of a forest can be applied to measure the forest's functionality as a terrestrial carbon sink (FAO 2020). Based on this and the information available in the field data, the estimated attributes were selected to be the total stem volume ($m^3\ ha^{-1}$) of trees, the stem volumes ($m^3\ ha^{-1}$) of pine, spruce and broadleaved tree species and the proportion of pine, spruce and broadleaved tree species.

Main tree species in the broadleaved class were silver birch and downy birch. Some field plots included also small volumes of other broadleaved species such as European aspen (*Populus tremula* L.), grey alder (*Alnus incana* (L.) Moench) and black alder (*Alnus glutinosa* (L.) Gaertn). As the volumes of these broadleaved species were smaller, all the broadleaved species were combined to one class. Field plots that included other coniferous tree species than Scots pine and Norway spruce were excluded as the amount of these kind of plots was small.

Proportions of the tree species were estimated with two different approaches. In the first approach the proportions of tree species were calculated from the field data before estimations and the estimated attributes were these calculated tree species proportions. In the second approach, proportions of tree species were calculated by dividing the estimated stem volume of each tree species with the sum of estimated stem volumes of all tree species.

3.4 Data processing flow

3.4.1 Pre-processing

Pre-processing started by combining the satellite image pixels with the field data sample plot locations. As a result, each of the sample plot locations was connected to one pixel reflectance spectrum measured by the EnMAP hyperspectral instrument.

Next, the quality of each selected pixel was investigated. An automatic quality check was done based on the satellite image quality layers. The automatic quality check was performed to check that selected pixels were on land, had no cloud, cloud shadow, haze or snow (DLR 2024b). As the quality layers had a limited accuracy, additional manual quality improvements were performed. First, pixels that were in cloud based on visual inspection were removed. Then, pixel locations were checked from Sentinel-2 time series to ensure that no changes, such as felling, had occurred within the time gap between satellite image acquisition date and field data measurement date. Changes between the dates were analysed by visually observing changes in

the normalized difference vegetation index (NDVI). If a clear change was detected, pixel was removed. Finally, the reflectance spectra of selected pixels were analysed to define boundaries for meaningful vegetation reflectance values. Pixels with a reflectance value above 0.1 in the vegetation green peak wavelength of 535.422 nm and pixels with a reflectance value below 0.08 in the near-infrared wavelengths of 993.145 nm and 1211.05 nm were removed.

When the final selection of field plots was ready, information of the estimated attributes was calculated for each of them. Stem volumes of pine and spruce were saved from the species-specific tables provided in the field data. Stem volumes of broadleaved species were summed from the tables related to broadleaved species. If stem volume information was not available for some of the estimated species in a field plot, the volume of this species was set to zero. Field plots that did not have stem volume information for any of the estimated species were excluded. Total stem volume of a field plot was calculated as a sum of stem volumes related to all the estimated species. Proportion of a species in a field plot was calculated by dividing the species-specific stem volume with the total stem volume of the field plot.

Mean, minimum and maximum values of the stem volumes and species proportions were calculated for each study location to reveal possible differences in their data characteristics. Additionally, the calculated species proportions were applied to classify field plots based on dominant tree species. Field plots were divided to four classes of >75 % proportion of pine, >75 % proportion of spruce, >75 % proportion of broadleaved species and mixed species. This classification was applied to calculate mean reflectance spectrum for each of the classes, however, it was not applied in the estimations.

3.4.2 Estimations of stem volumes and species proportions

Estimations of total stem volumes, species-specific stem volumes and species proportions were performed with Random Forest Regression (RF) and Support Vector Regression (SVR). As mentioned in Section 2.2, both models have previously been applied in estimation of biophysical variables from hyperspectral data, and they are simple to implement (Verrelst et al. 2019; Shafri 2017). The models were implemented with scikit-learn library (version 1.5.2) that offers open-source machine learning in Python (Scikit-learn 2024a).

The applied SVR and RF models were separately formed for each of the estimated attribute and location. To minimize the amount of negative estimations, a logarithmic transformation was applied before and after the estimations. Hyperparameters for the models were tuned with grid search, also provided by scikit-learn (Scikit-learn 2024d). SVR was applied with a Radial

Basis Function (RBF) kernel and the tuned hyperparameters were a regularization parameter C and gamma that defines the influence of a single training sample in the estimation (Scikit-learn 2024b; 2024e). For RF, the tuned hyperparameters can be selected from a wider range. In this case only the number of decision trees in the model was selected as a tuned hyperparameter. Other parameters were applied with default values as no clear improvements were received by tuning the values (Scikit-learn 2024c).

As the amount of data was limited, both hyperparameter tuning and estimations were performed with 5-fold cross validation. In 5-fold cross validation, the field plots of each study location were divided into five subsets, and estimations of stem volumes and species proportions were performed for each subset separately. Estimations of the whole dataset were then formed as a combination of estimations for the subsets (Figure 7).

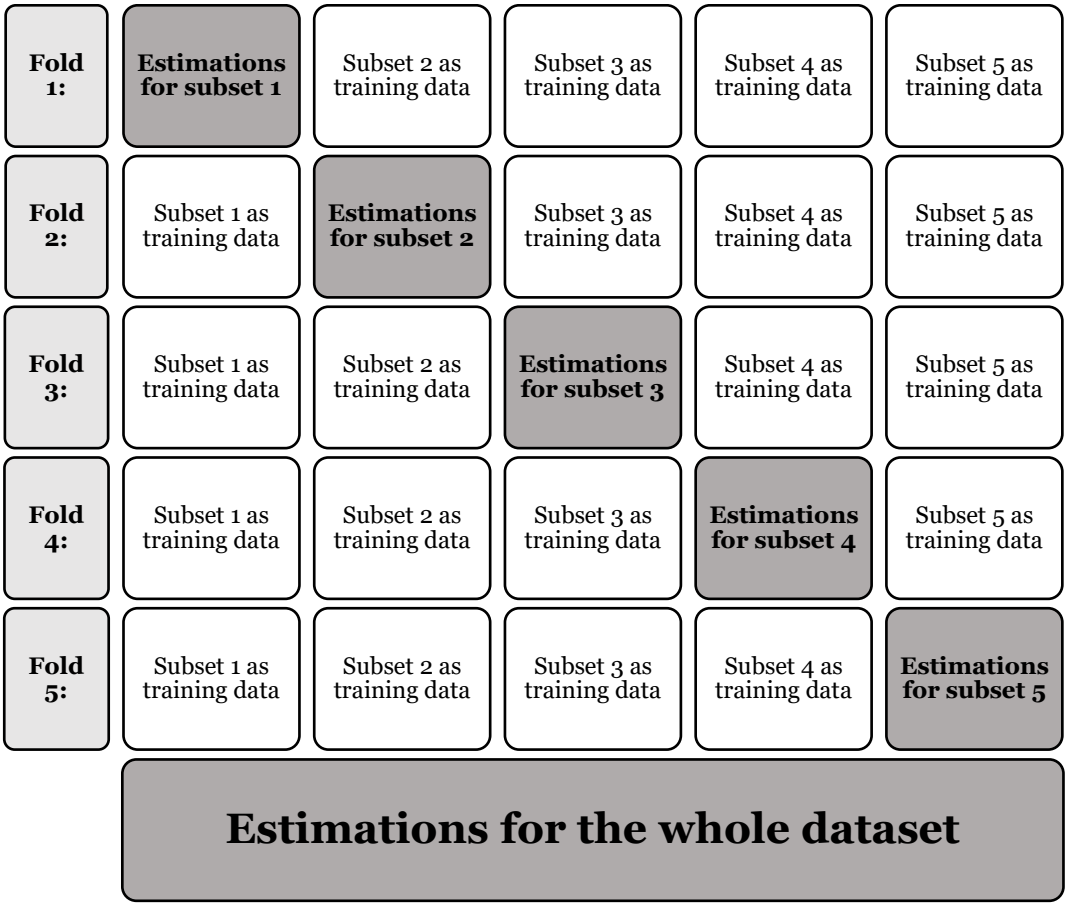

Figure 7. Representation of performed stem volume and species proportion estimations with 5-fold cross validation.

For example, in the total stem volume estimations of Kuopio, each of the subsets were formed by 19–20 field plots. The training data of each fold was formed by four subsets, consisting of measured reflectance spectrum and

total stem volume of 78–79 field plots. Trained SVR and RF models were applied to estimate the total stem volume for each field plot in the estimation subset, based on the measured reflectance spectrum. Additionally, the field data values of total stem volumes were saved to assess the accuracy of the estimations. When the estimations were performed for all subsets, a root-mean-square error (RMSE) and a coefficient of determination (R^2) were calculated to estimate the model fitting and estimation accuracy. Calculations were made by comparing the estimated value and the field data value of each field plot. Additionally, a relative root-mean-square error (RRMSE) was calculated by dividing RMSE value with the mean of the field data values.

The data processing flow, consisting of pre-processing and performing estimations for the total stem volumes, species-specific stem volumes and species proportions, was separately performed for each of the study locations (Figure 8).

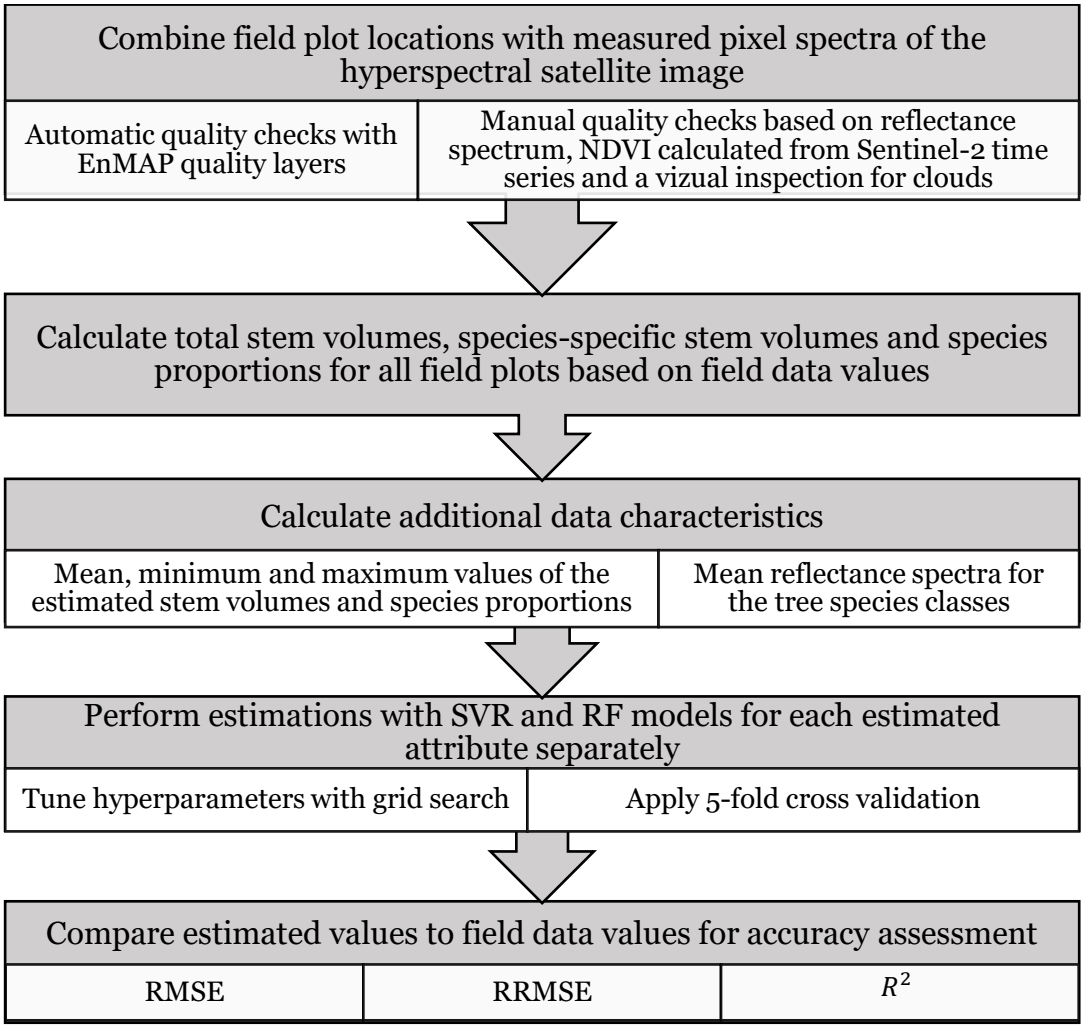

Figure 8. Data processing flow that was performed for each study location separately.

4 Results

This section presents the results of this study. Section 4.1 presents the characteristic of the data with information related to the field data values of the estimated stem volumes and species proportions. Also, the dominance of pine, spruce and broadleaved species in the study locations is presented with related mean reflectance spectra. Section 4.2 presents the estimation accuracies of total stem volumes, species-specific stem volumes and species proportions for each study location separately.

4.1 Characteristics of the data

4.1.1 Stem volumes and species proportions in the field data

To compare the field plot values of total stem volume in each study location, the minimum, maximum and mean values were calculated from the field data (Table 4).

Table 4. The minimum, maximum and mean field data values for total stem volumes ($m^3\ ha^{-1}$) in each study location.

	Total stem volume		
	Min	Max	Mean
Sodankylä	0.29	262.26	69.76
Perho	0.33	525.11	136.50
Kuopio	0.8	665.5	151.06
Äänekoski	0.1	839.0	168.24
Virrat	0.16	531.96	89.42

The field plots located in Sodankylä had the lowest maximum and mean total volumes at 262.26 and 69.76 $m^3\ ha^{-1}$. Virrat had the second lowest mean total volume at 89.42 $m^3\ ha^{-1}$, although the maximum total volume was significantly larger than in Sodankylä at 531.96 $m^3\ ha^{-1}$. Äänekoski had the highest maximum and mean total volumes at 839.0 and 168.24 $m^3\ ha^{-1}$ followed by Kuopio and Perho with quite similar values.

Additionally, the minimum, maximum and mean values were calculated from the field data for stem volumes and proportions of pine (Table 5), spruce (Table 6) and broadleaved species (Table 7).

Table 5. The minimum, maximum and mean field data values for pine stem volume ($m^3\ ha^{-1}$) and proportion (%) in each study location.

	Pine stem volume			**Pine proportion**		
	Min	Max	Mean	Min	Max	Mean
Sodankylä	0	207.85	41.68	0	100	64
Perho	0	445.09	76.15	0	100	58
Kuopio	0	337.1	28.92	0	100	17
Äänekoski	0	643.5	76.11	0	100	43
Virrat	0	383.96	49.32	0	100	53

Table 6. The minimum, maximum and mean field data values for spruce stem volume ($m^3\ ha^{-1}$) and proportion (%) in each study location.

	Spruce stem volume			**Spruce proportion**		
	Min	Max	Mean	Min	Max	Mean
Sodankylä	0	207.03	14.15	0	91	13
Perho	0	408.40	32.82	0	100	19
Kuopio	0	506.8	84.47	0	100	52
Äänekoski	0	494.42	63.95	0	100	37
Virrat	0	390.07	20.06	0	100	22

Table 7. The minimum, maximum and mean field data values for broad-leaved stem volume ($m^3\ ha^{-1}$) and proportion (%) in each study location.

	Broadleaved stem volume			**Broadleaved proportion**		
	Min	Max	Mean	Min	Max	Mean
Sodankylä	0	148.04	13.93	0	100	23
Perho	0	304.33	27.53	0	100	23
Kuopio	0	431.0	37.67	0	100	31
Äänekoski	0	294.7	28.17	0	100	20
Virrat	0	241.36	20.04	0	100	25

Lowest mean tree species proportion was in Sodankylä for spruce with a value at 13 % followed by mean pine proportion in Kuopio and mean spruce proportion in Perho with values below 20 %. Highest mean tree proportion was in Sodankylä for pine with a value at 64 % followed by pine in Perho, pine in Virrat and spruce in Kuopio with mean proportion above 50 %. Highest mean volumes were in Perho and Äänekoski for pine with values approximately at 76 $m^3\ ha^{-1}$ and in Kuopio for spruce with a value at 84.47 $m^3\ ha^{-1}$. Spruce and broadleaved tree species in Sodankylä had the lowest mean volumes with values below 15 $m^3\ ha^{-1}$.

To compare the species distribution of the study locations, mean proportions of pine, spruce and broadleaved species were also visualized (Figure 9). Spruce was the main tree species in Kuopio and pine was the main tree species in rest of the study locations. Main tree species was the least distinctive in the study location of Äänekoski, where the difference between pine and spruce mean proportion was only 6 %. Broadleaved species were quite evenly distributed in the study locations with an average mean proportion at 24.4 %.

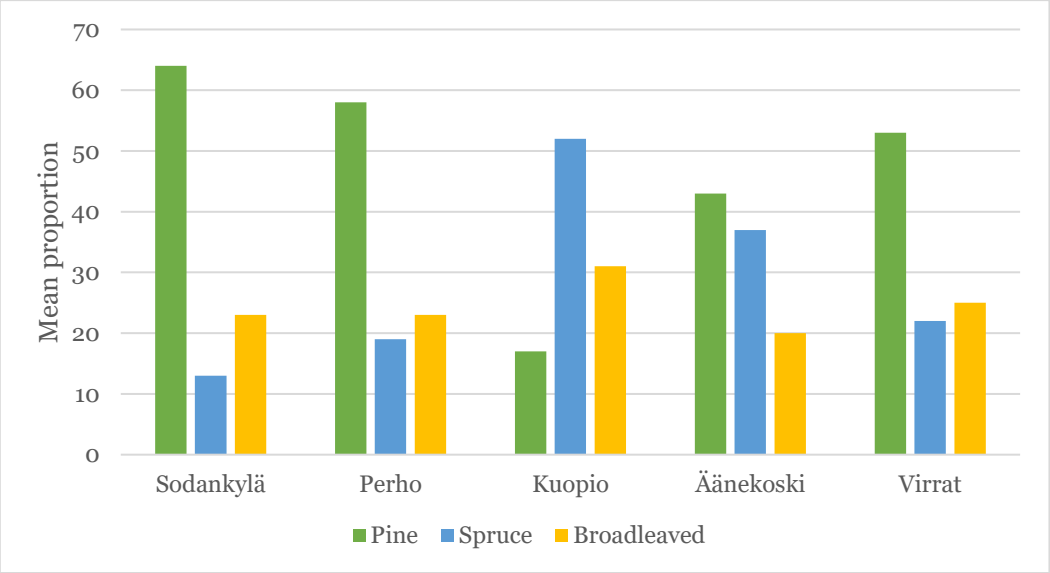

Figure 9. Distribution of pine, spruce and broadleaved mean proportions in the study locations.

4.1.2 Dependence of mean reflectance spectra on tree species

Although the distribution of mean species proportions varied between the study locations, field plots with a proportion above 75 % of pine, spruce and broadleaved species were found in all locations (Table 8).

Table 8. Number and percentage of field plots with mixed species or above 75 % dominance of pine, spruce or broadleaved species.

	Pine		**Spruce**		**Broad-leaved**		**Mixed**	
	Number of sample plots	%	Number of sample plots	%	Number of sample plots	%	Number of sample plots	%
Sodankylä	158	55	11	4	20	7	98	34
Perho	107	49	25	11	33	15	53	24
Kuopio	13	13	46	47	20	20	19	19
Äänekoski	72	34	48	22	15	7	79	37
Virrat	94	48	34	17	38	19	31	16

In Sodankylä 55 % of field plots had pine as the dominant tree species. Also, in Perho and Virrat pine had an above 75 % dominance in almost half of the field plots. In Kuopio, spruce had dominance in 47 % of the field plots. In Äänekoski, mixed sample plots had the largest proportion at 37 %. Broad-leaved species were less dominant in the study locations, as the highest dominances were in Kuopio and Virrat, where approximately a fifth of sample plots had a dominance of broadleaved species.

After the field plots were classified based on dominant tree species, a mean reflectance spectrum was calculated for each class and study location (Figure 10).

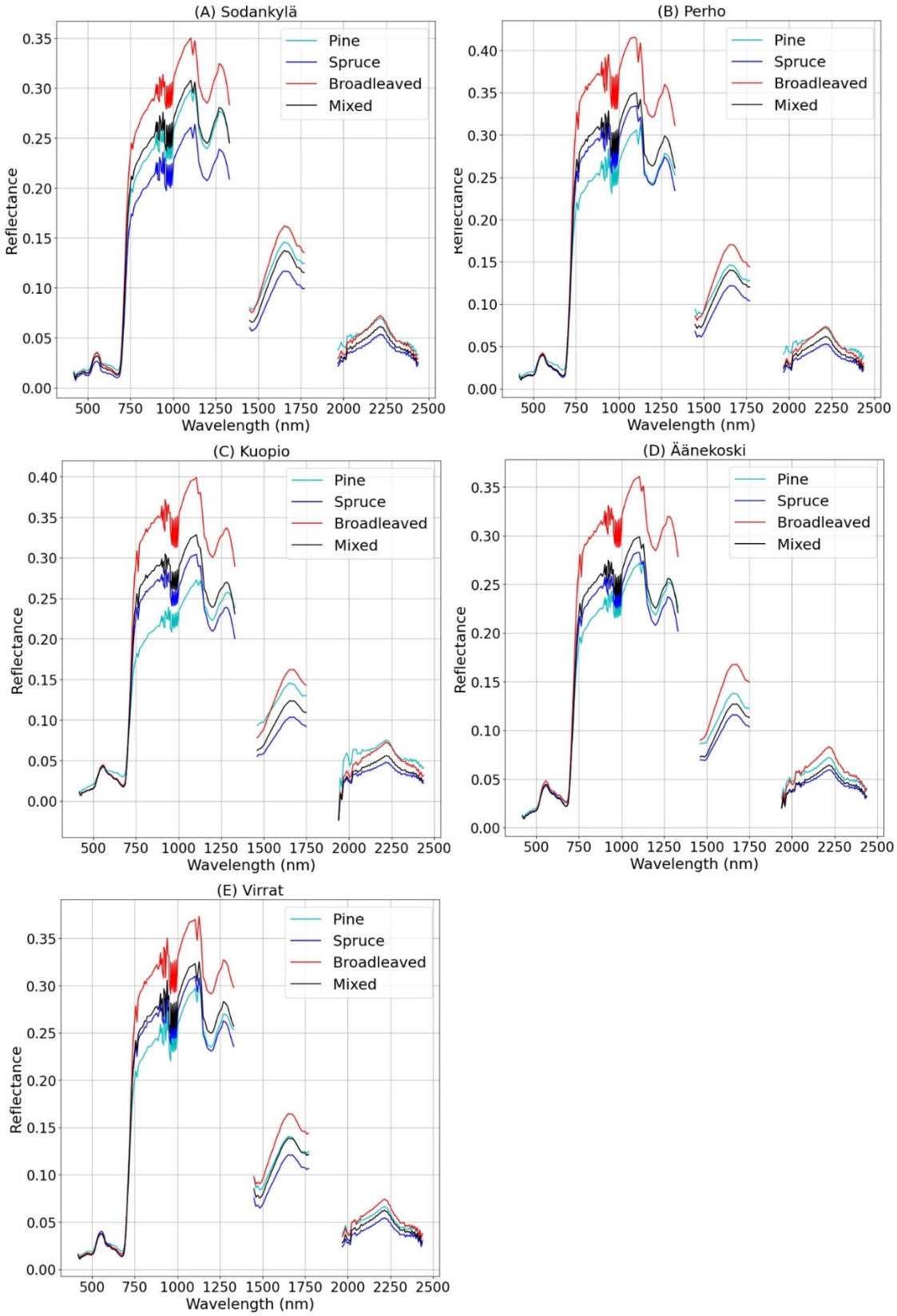

Figure 10. Mean spectra of dominant tree species in (A) Sodankylä, (B) Perho, (C) Kuopio, (D) Äänekoski and (E) Virrat.

In all locations, the mean reflectance values were most similar within the visible wavelengths of 400–750 nm and most different within the near-infrared and shortwave-infrared wavelengths of 750–1750 nm. Field plots with broadleaved trees as the dominant species had the highest mean reflectance values in all study locations between wavelengths 750–1750 nm.

Within a wavelength range of 1500–1750 nm field plots with spruce as the dominant tree species had the lowest mean reflectance values in all locations. In the same range, the mean reflectance values of field plots with pine as the dominant species and mixed species showed intermediate values. In Virrat the mean reflectance values of pine and mixed species were almost identical (Figure 10E) whereas in Kuopio, these were the most different (Figure 10C).

Within a wavelength range of 750–1400 nm, field plots with mixed species had the second highest mean reflectance values in all locations. In Perho, Kuopio, Äänekoski and Virrat mean reflectance values of field plots with pine and spruce as the dominant tree species changed order approximately at 1100 nm (Figures 10B-E). In the range of 750–1100 nm pine had the lowest mean reflectance values, and in the range of 1100–1400 nm spruce had the lowest mean reflectance values. In Sodankylä field plots with spruce as the dominant tree species had the lowest mean reflectance values within the whole range of 750–1400 nm (Figure 10A).

In wavelengths larger than 2000 nm the order of mean spectra was quite similar as within wavelengths 1500–1750 nm. However, especially in Kuopio, field plots with pine as dominant tree species had higher mean reflectance values than the broadleaved species (Figure 10C). Similar behaviour was visible near water absorption bands of 1400 nm and 1900 nm, where mean reflectance values of pine were higher than for broadleaved species in Sodankylä, Perho and Kuopio (Figures 10A-C).

4.2 Accuracy of stem volume and species proportion estimations

Model fitting and estimation accuracies of total stem volumes, species-specific stem volumes and species proportions were described with a coefficient of determination (R^2), a root-mean-square error (RMSE) and a relative root-mean-square error (RRMSE) (Tables 9–18). Additionally, the estimated values reached with SVR model were visualized with the field data values as scatterplot graphs (Figures 11–20).

4.2.1 Sodankylä

From the stem volume estimations of Sodankylä, total volume had the best accuracy with both SVR and RF models, and pine had the best estimation accuracy from the species (Table 9). Similar differences in estimation accuracies of stem volumes were shown in the visualization of estimated values compared to field data values (Figure 11).

Table 9. Stem volume estimations of Sodankylä.

	Total volume		**Pine volume**		**Spruce volume**		**Broad-leaved volume**	
	SVR	RF	SVR	RF	SVR	RF	SVR	RF
R^2	0.61	0.58	0.37	0.47	0.23	0.46	0.41	0.32
RMSE	36.33	37.59	37.32	34.21	29.34	24.51	16.12	17.40
RRMSE	52%	54%	90%	82%	207%	173%	116%	125%

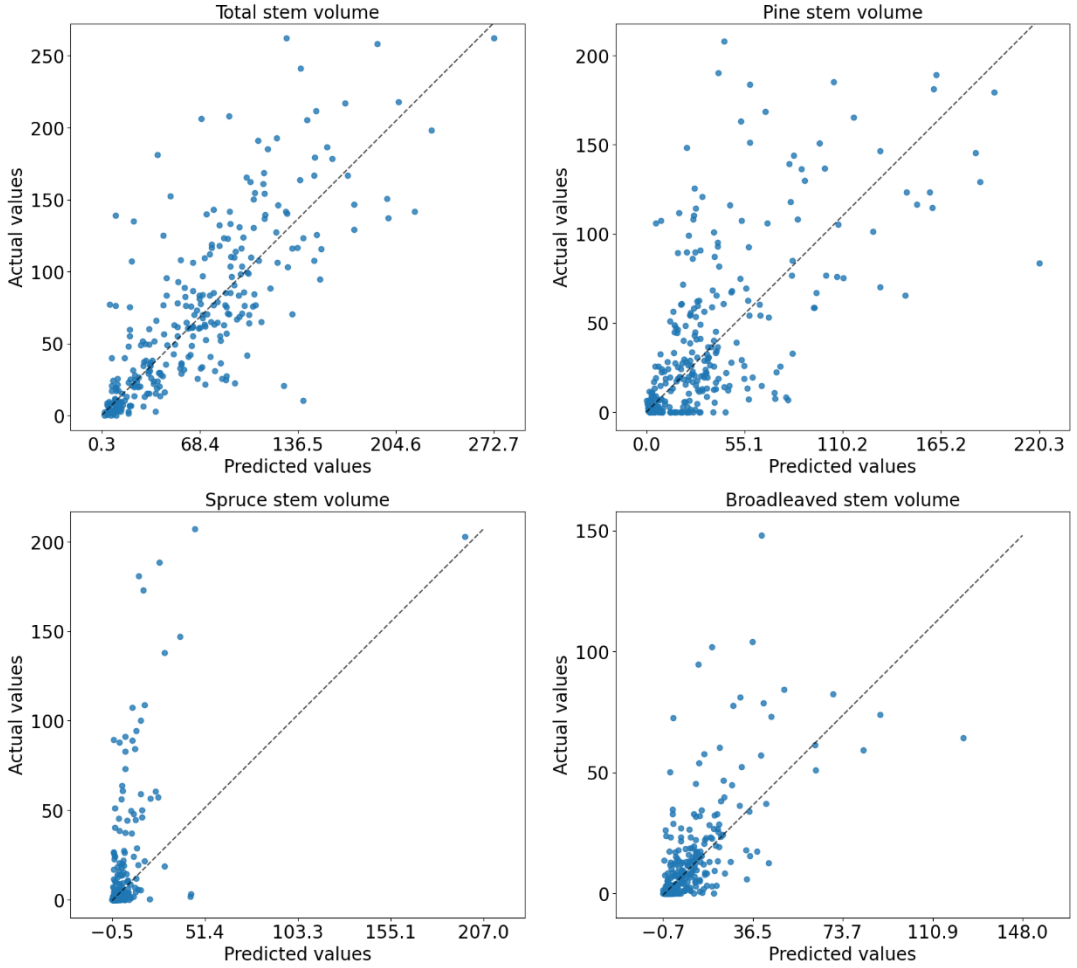

Figure 11. SVR estimated stem volumes compared to field data values in Sodankylä.

Total volume reached R^2 values between 0.58–0.61 and RRMSE values between 52–54 %. In R^2 values the species-specific volume estimations had more variability between the applied SVR and RF models. The highest R^2 value at 0.47 was reached in estimation of pine volume, although also spruce volume and broadleaved volume reached values above 0.40. Pine volume estimations reached the smallest relative estimation errors with RRMSE values between 82–90 %, followed by broadleaved volume with RRMSE values between 116–125 %. Spruce volume estimations had the highest relative estimation errors with RRMSE values between 173–207 %.

In the estimation of proportions, there was more variability between the best R^2 and RRMSE values (Table 10). A similar finding can be made from the visualization of the estimated values and field data values (Figure 12).

Table 10. Species proportion estimations of Sodankylä.

	Pine proportion		**Spruce proportion**		**Broadleaved proportion**	
	SVR	RF	SVR	RF	SVR	RF
Based on estimated volumes						
R^2	0.33	0.40	0.20	0.31	0.34	0.32
RMSE	0.30	0.28	0.21	0.19	0.22	0.22
RRMSE	46%	44%	155%	144%	94%	95%
Estimation as a percentage						
R^2	0.39	0.35	0.28	0.35	0.50	0.35
RMSE	0.28	0.29	0.20	0.19	0.19	0.22
RRMSE	45%	46%	147%	140%	82%	93%

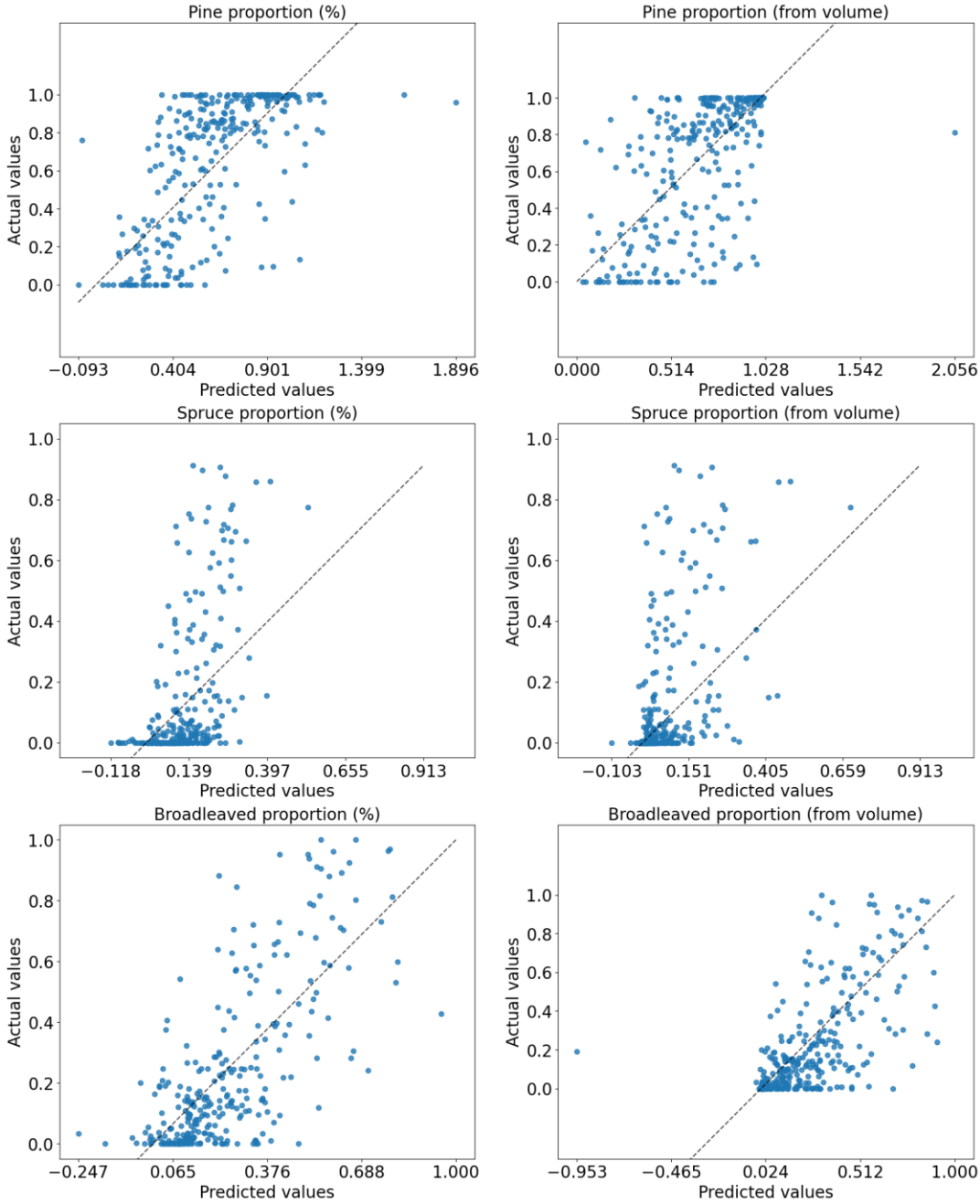

Figure 12. SVR estimated species proportions compared to field data values in Sodankylä.

The highest R^2 value at 0.50 was reached when estimating broadleaved proportion as a percentage. The second highest R^2 value at 0.40 was reached for pine proportion calculated from the estimated volumes. In general, pine estimations had the best RRMSE values, with all values between 44–46 %. The lowest estimation accuracy was received for spruce with R^2 values between 0.20–0.35 and RRMSE values between 140–155 %.

4.2.2 Perho

If the SVR and RF models were compared separately, total volume had the best accuracy from the volume estimations, followed by pine volume estimations (Table 11). The differences in estimation accuracy of total volume and pine volume compared to spruce volume and broadleaved volume were visible in the visualization of estimated values and field data values (Figure 13).

Table 11. Stem volume estimations of Perho.

	Total volume		**Pine volume**		**Spruce volume**		**Broad-leaved volume**	
	SVR	RF	SVR	RF	SVR	RF	SVR	RF
R^2	0.38	0.31	0.33	0.29	0.20	0.05	0.21	0.27
RMSE	91.50	97.08	73.62	75.33	62.23	67.95	47.13	45.15
RRMSE	67%	71%	97%	99%	190%	207%	171%	164%

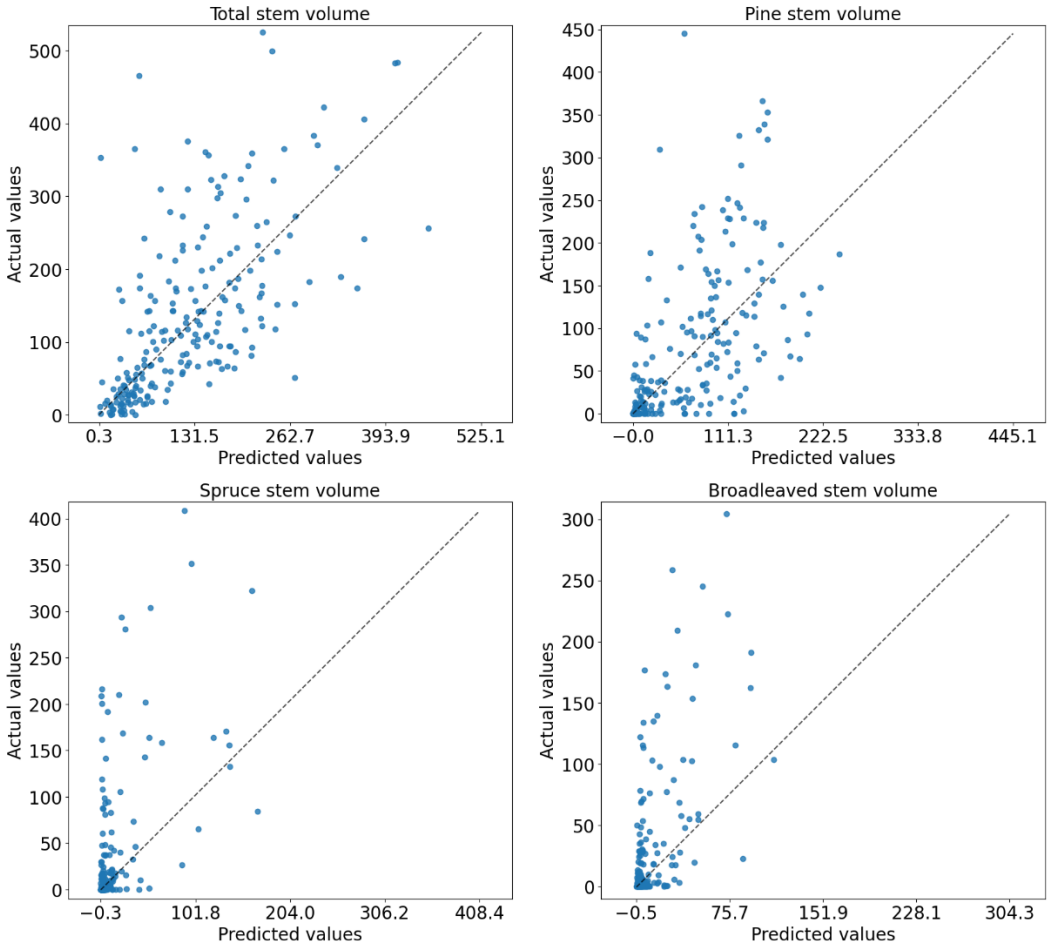

Figure 13. SVR estimated stem volumes compared to field data values in Perho.

Total volume had the highest R^2 value at 0.38 and the smallest RRMSE value at 67 %. Pine volume estimations had quite similar R^2 values with the highest value at 0.33 combined with higher RRMSE values above 90 %. Spruce volume estimations had the lowest accuracy with R^2 values between 0.05–0.20 and RRMSE values between 190–207 %. Broadleaved species reached an intermediate accuracy with R^2 values between 0.21–0.27 and RRMSE values between 164–171 %.

The species proportion estimations had more variability (Table 12), which was also visible in the visualizations of the estimated values and field data values (Figure 14).

Table 12. Species proportion estimations of Perho.

	Pine proportion		**Spruce proportion**		**Broadleaved proportion**	
	SVR	RF	SVR	RF	SVR	RF
Based on estimated volumes						
R^2	0.45	0.45	0.23	0.17	0.53	0.59
RMSE	0.31	0.31	0.27	0.28	0.24	0.22
RRMSE	53%	53%	140%	145%	103%	97%
Estimation as a percentage						
R^2	0.52	0.47	0.25	0.21	0.64	0.52
RMSE	0.29	0.30	0.26	0.27	0.21	0.24
RRMSE	50%	52%	138%	141%	90%	104%

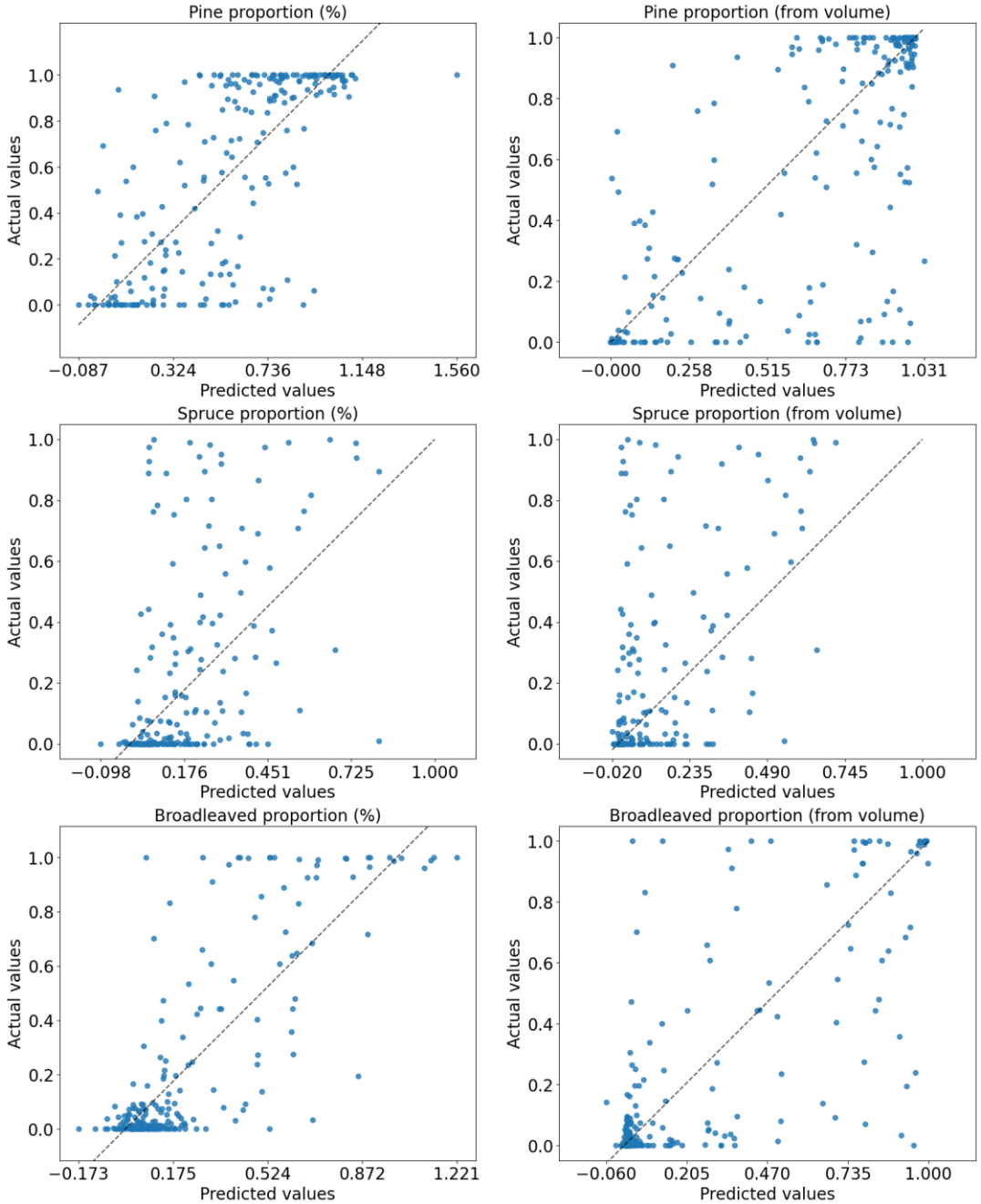

Figure 14. SVR estimated species proportions compared to field data values in Perho.

Estimation of broadleaved species had the highest R^2 values between 0.52–0.64 combined with RRMSE values between 90–104 %. The best RRMSE values were received for pine proportion estimations with the values between 50–53 % combined with R^2 values between 0.45–0.52. Spruce proportion had the lowest estimation accuracies with R^2 values between 0.17–0.25 and RRMSE values between 138–145 %.

4.2.3 Kuopio

From stem volume estimations total volume and spruce volume had the best estimation accuracies followed by pine and broadleaved species with less accurate estimations (Table 13). Also, the visualization of the estimated values and field data values showed the better estimation accuracy of total volume and spruce volume when compared to pine volume and broadleaved volume (Figure 15).

Table 13. Stem volume estimations of Kuopio.

	Total volume		**Pine volume**		**Spruce volume**		**Broad-leaved volume**	
	SVR	RF	SVR	RF	SVR	RF	SVR	RF
R^2	0.39	0.42	0.36	0.21	0.51	0.37	0.23	0.05
RMSE	103.02	100.25	52.24	58.16	80.76	91.75	60.88	67.48
RRMSE	68%	66%	181%	201%	96%	109%	162%	179%

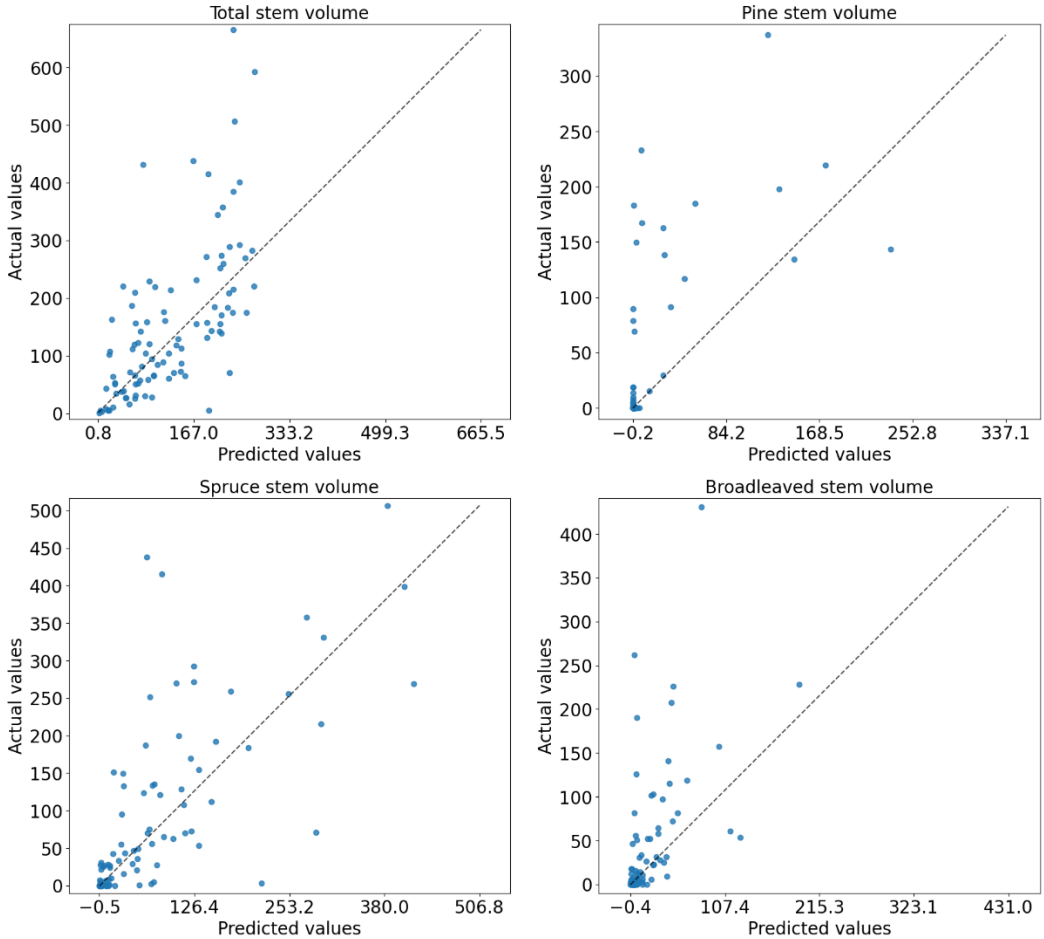

Figure 15. SVR estimated stem volumes compared to field data values in Kuopio.

Spruce reached the highest R^2 value at 0.51, although total volume estimations reached the second highest values between 0.39–0.42 combined with better RRMSE values. For total volume estimations RRMSE values were between 66–68 % whereas for spruce volume estimations RRMSE values were between 96–109 %. Pine volume estimations had better R^2 values between 0.21–0.36 when compared to broadleaved volume estimations with R^2 values between 0.05–0.23. Broadleaved volume estimations, on the other hand, had better RRMSE values between 162–179 % when compared to pine volume estimations with RRMSE values between 181–201 %.

In species proportion estimations, the estimation accuracies of all species were quite good (Table 14). This was shown as a quite even distribution of estimated values and field data values between the species (Figure 16).

Table 14. Species proportion estimations of Kuopio.

	Pine proportion		**Spruce proportion**		**Broadleaved proportion**	
	SVR	RF	SVR	RF	SVR	RF
Based on estimated volumes						
R^2	0.62	0.60	0.60	0.49	0.69	0.60
RMSE	0.21	0.22	0.27	0.30	0.21	0.24
RRMSE	119%	123%	51%	58%	70%	79%
Estimation as a percentage						
R^2	0.65	0.53	0.64	0.38	0.72	0.49
RMSE	0.20	0.23	0.25	0.33	0.20	0.28
RRMSE	114%	133%	49%	64%	66%	90%

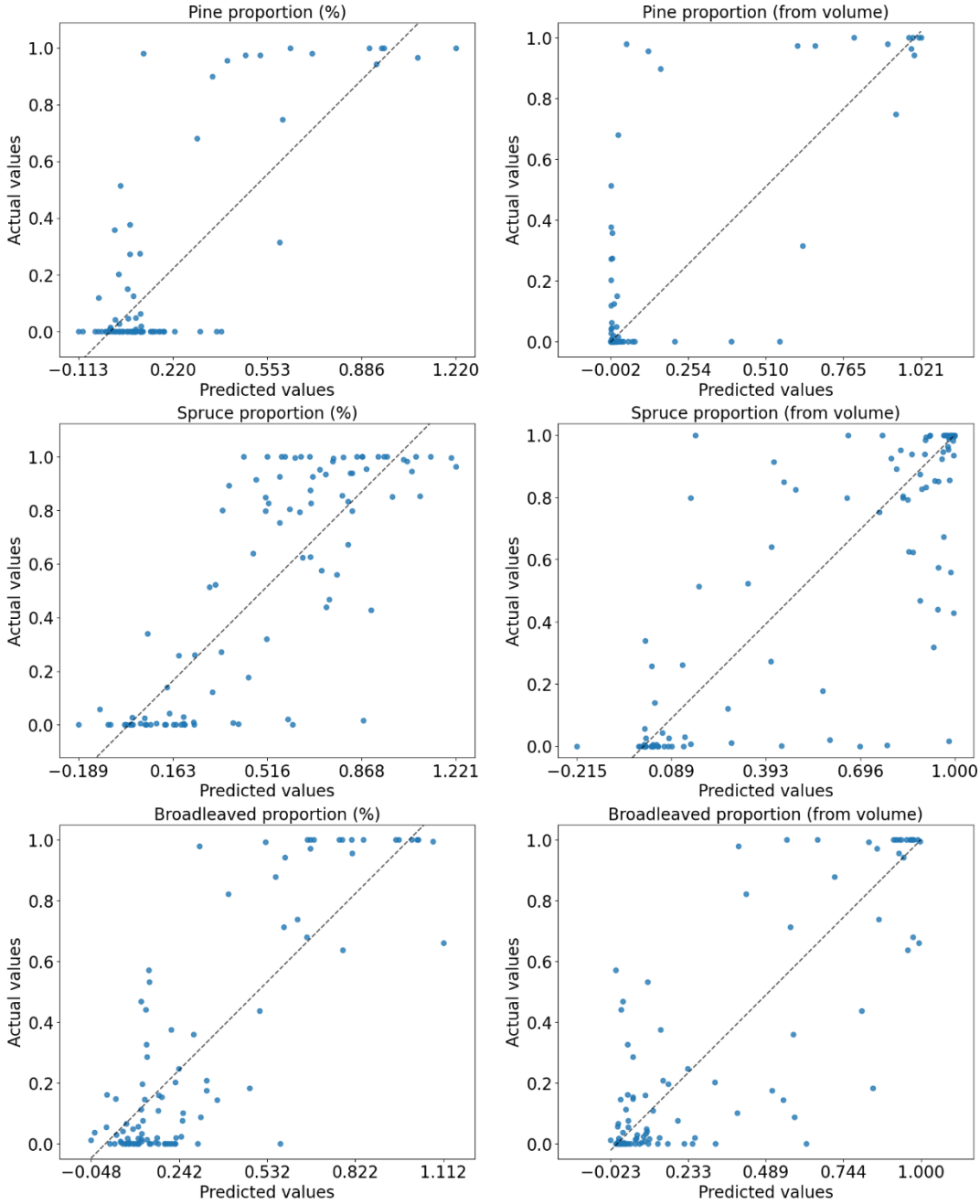

Figure 16. SVR estimated species proportions compared to field data values in Kuopio.

The R^2 values of all estimated species percentages were quite high, with the best values between 0.64–0.72 for all species. Also, the R^2 values of species proportions calculated from estimated volumes were high, with the best values between 0.60–0.69 for all species. The highest R^2 values were reached in broadleaved proportion estimations for both strategies with related RRMSE values between 66–70 %. The smallest estimation errors were reached in the spruce proportion estimations with RRMSE values between 49–64 %.

4.2.4 Äänekoski

When comparing the SVR and RF models separately, total volume had the best accuracy of stem volume estimations followed by pine and spruce with quite similar accuracies (Table 15). Especially the worse estimation accuracy of broadleaved volume was shown in the visualization of estimated values and field data values (Figure 17).

Table 15. Stem volume estimations of Äänekoski.

	Total volume		**Pine volume**		**Spruce volume**		**Broad-leaved volume**	
	SVR	RF	SVR	RF	SVR	RF	SVR	RF
R^2	0.40	0.22	0.24	0.07	0.25	0.10	-0.12	-0.07
RMSE	136.79	155.72	97.95	108.59	93.39	102.55	56.92	55.66
RRMSE	81%	93%	129%	143%	146%	160%	202%	198%

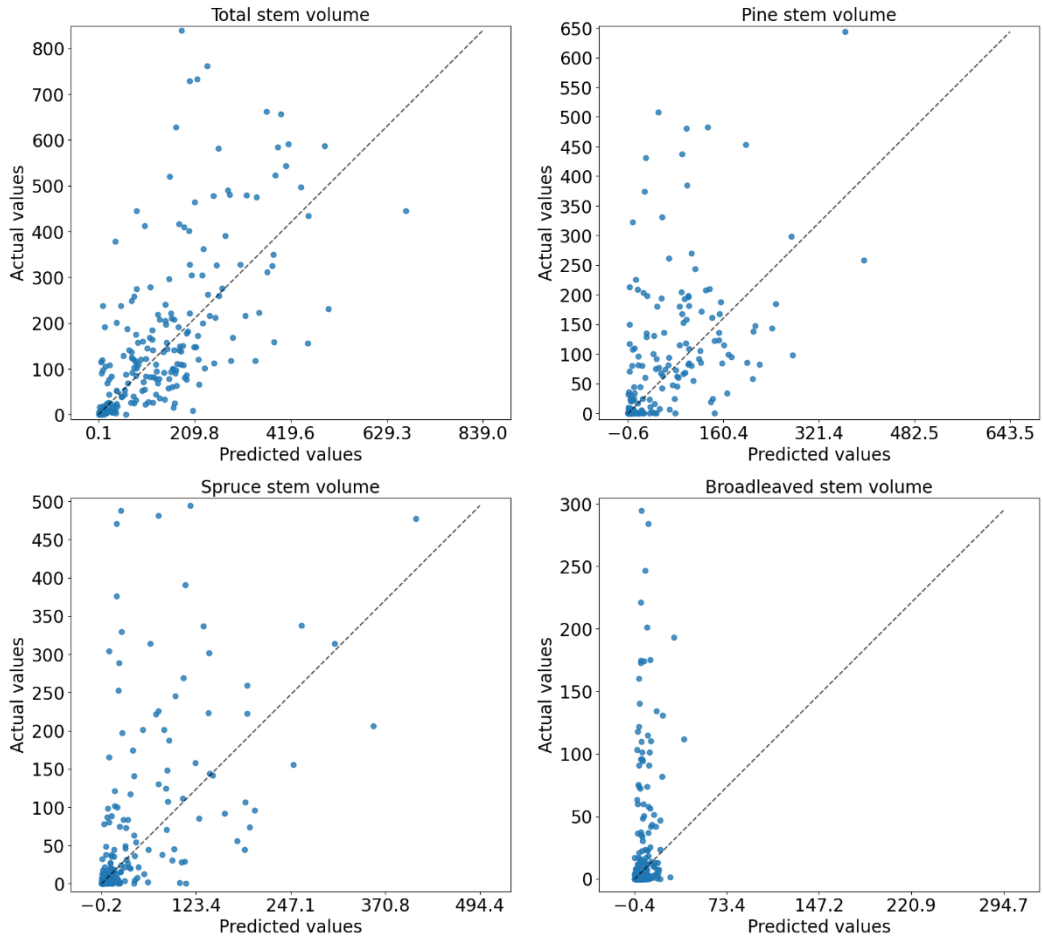

Figure 17. SVR estimated stem volumes compared to field data values in Äänekoski.

The relative estimation errors of total volume were the smallest with RRMSE values between 81–93 %. Total volume reached also the best R^2 value at 0.40. The highest R^2 values reached by pine and spruce stem volume estimations were between 0.24–0.25 and related RRMSE values were between 129–146 %. Estimations of broadleaved volume had a poor estimation accuracy with R^2 values below zero and RRMSE values close to 200 %.

Also in the proportion estimations, pine and spruce had quite similar accuracies, and improvements in estimation accuracy of broadleaved species were shown when compared to the volume estimations (Table 16). Especially this improvement in the estimation accuracy related to broadleaved species was shown in the visualization of estimated values and field data values (Figure 18).

Table 16. Species proportion estimations of Äänekoski

	Pine proportion		**Spruce proportion**		**Broadleaved proportion**	
	SVR	RF	SVR	RF	SVR	RF
Based on estimated volumes						
R^2	0.43	0.42	0.43	0.37	0.23	0.36
RMSE	0.31	0.31	0.27	0.29	0.24	0.22
RRMSE	72%	73%	74%	78%	118%	108%
Estimation as a percentage						
R^2	0.60	0.42	0.59	0.38	0.38	0.28
RMSE	0.26	0.31	0.23	0.28	0.22	0.23
RRMSE	60%	72%	63%	77%	106%	114%

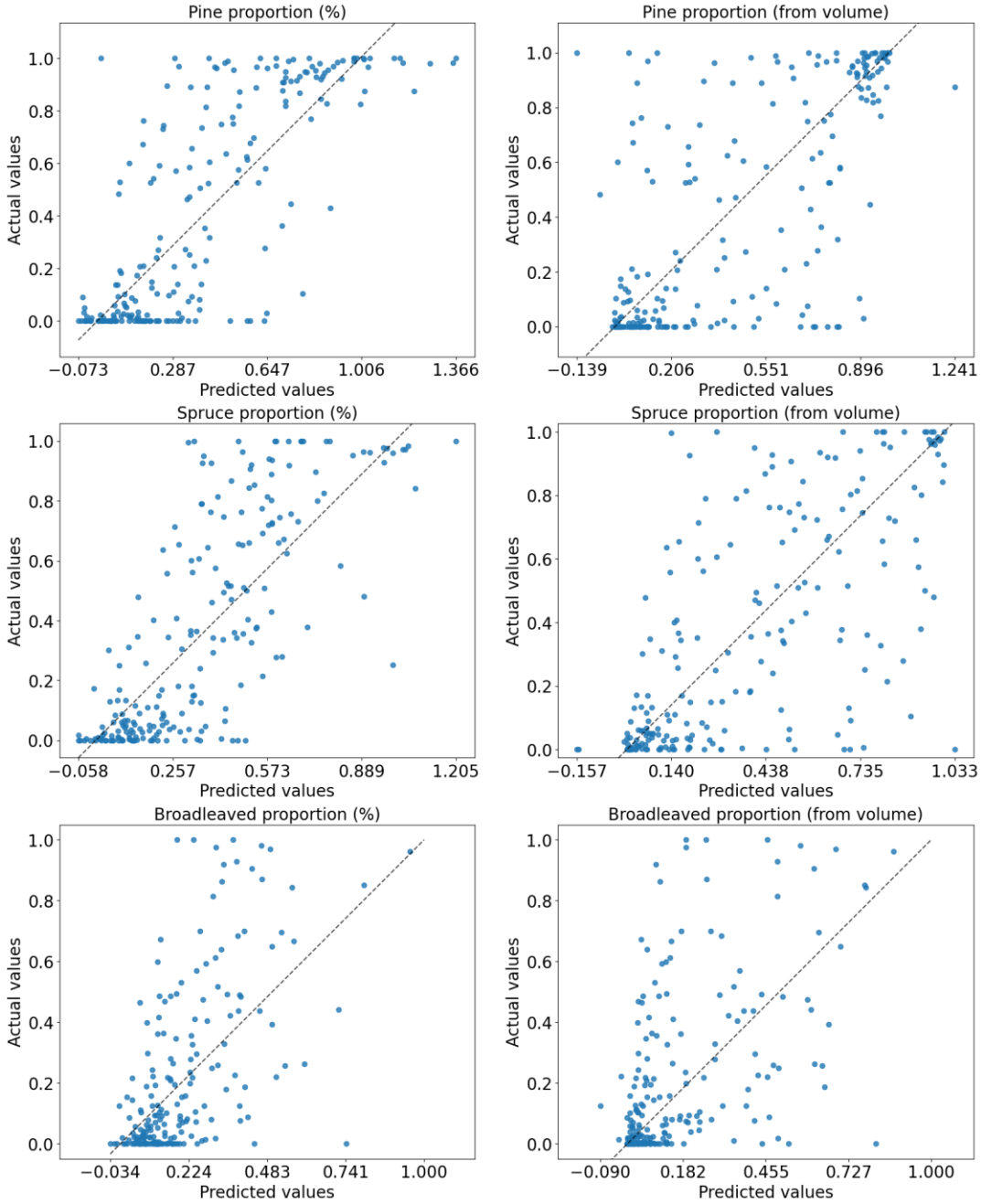

Figure 18. SVR estimated species proportions compared to field data values in Äänekoski.

When pine and spruce proportion were estimated as a percentage, the best R^2 values were between 0.59–0.60 with related RRMSE values between 60–63 %. When pine and spruce proportion were calculated from the estimated volumes, the best R^2 values were at 0.43 with related RRMSE values between 72–74 %. In broadleaved proportion estimations the best R^2 values were between 0.36–0.38 with both approaches, and the related RRMSE values were between 106–108 %.

4.2.5 Virrat

From the estimation of stem volumes, total volume reached the smallest relative estimation errors followed by pine volume estimations, although spruce volume estimations reached the best R^2 values (Table 17). Estimated values of total stem volume were most similar to field data values also in the scatterplot visualizations (Figure 19).

Table 17. Stem volume estimations of Virrat

	Total volume		**Pine volume**		**Spruce volume**		**Broad-leaved volume**	
	SVR	RF	SVR	RF	SVR	RF	SVR	RF
R^2	0.47	0.51	0.39	0.44	0.59	0.62	0.23	0.38
RMSE	71.83	69.12	58.48	56.29	39.24	38.06	38.67	34.68
RRMSE	80%	77%	119%	114%	196%	190%	193%	173%

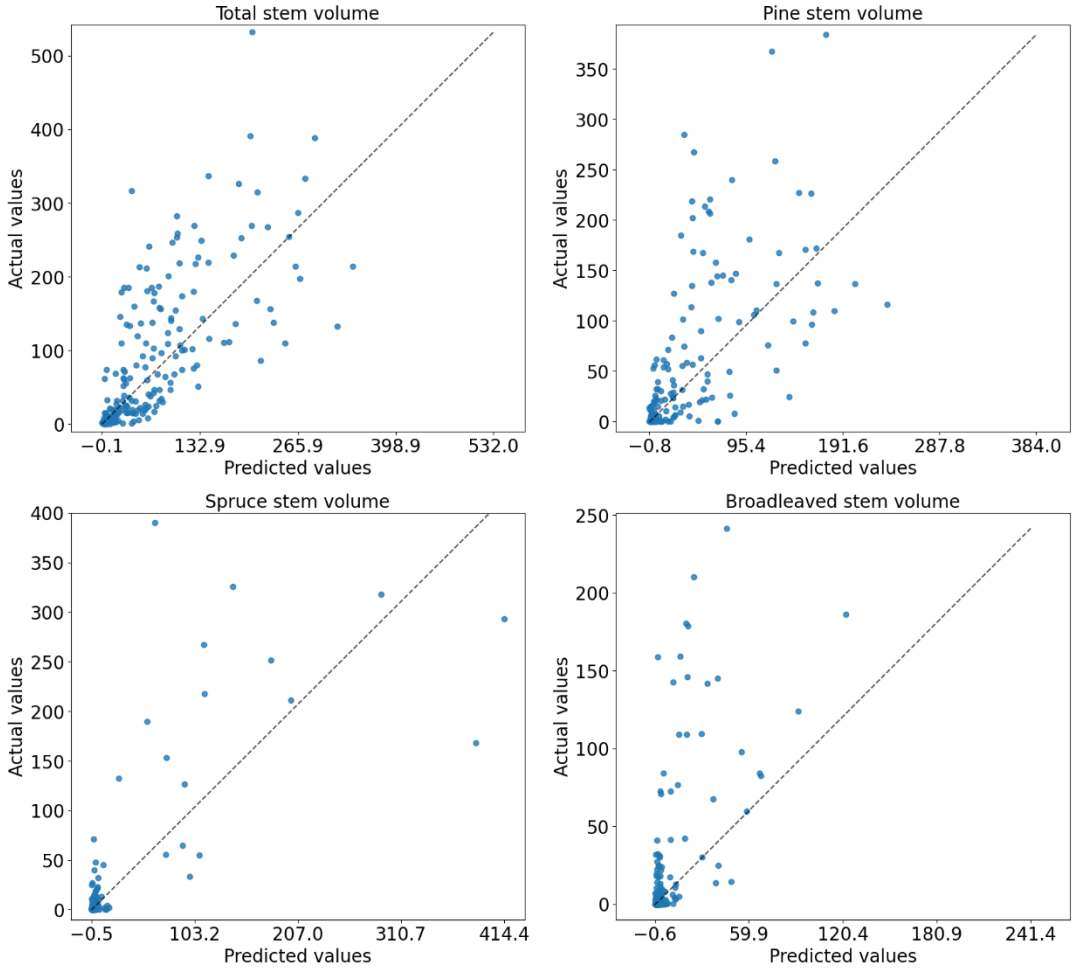

Figure 19. SVR estimated stem volumes compared to field data values in Virrat.

Total volume reached RRMSE values between 77–80 % with related R^2 values between 0.47–0.51. Pine volume reached RRMSE values between 114–119 % combined to R^2 values between 0.39–0.44. Spruce volume estimations had the highest R^2 values between 0.59–0.62, however, with large estimation errors shown as RRMSE values between 190–196 %. Broadleaved volume had slightly lower R^2 values between 0.23–0.38 combined to quite high estimation errors with RRMSE values between 173–193 %.

In proportion estimations, the RRMSE values followed a similar distribution between the species as in stem volume estimations, however, the R^2 values were quite similar between the species (Table 18). The quite even distribution between species in R^2 values were shown also in the visualization of estimated and field data values (Figure 20).

Table 18. Species proportion estimations of Virrat

	Pine proportion		**Spruce proportion**		**Broadleaved proportion**	
	SVR	RF	SVR	RF	SVR	RF
Based on estimated volumes						
R^2	0.42	0.45	0.38	0.48	0.31	0.46
RMSE	0.33	0.32	0.28	0.26	0.30	0.27
RRMSE	62%	60%	131%	120%	120%	107%
Estimation as a percentage						
R^2	0.47	0.43	0.54	0.44	0.54	0.46
RMSE	0.31	0.33	0.24	0.27	0.25	0.27
RRMSE	59%	61%	113%	124%	99%	106%

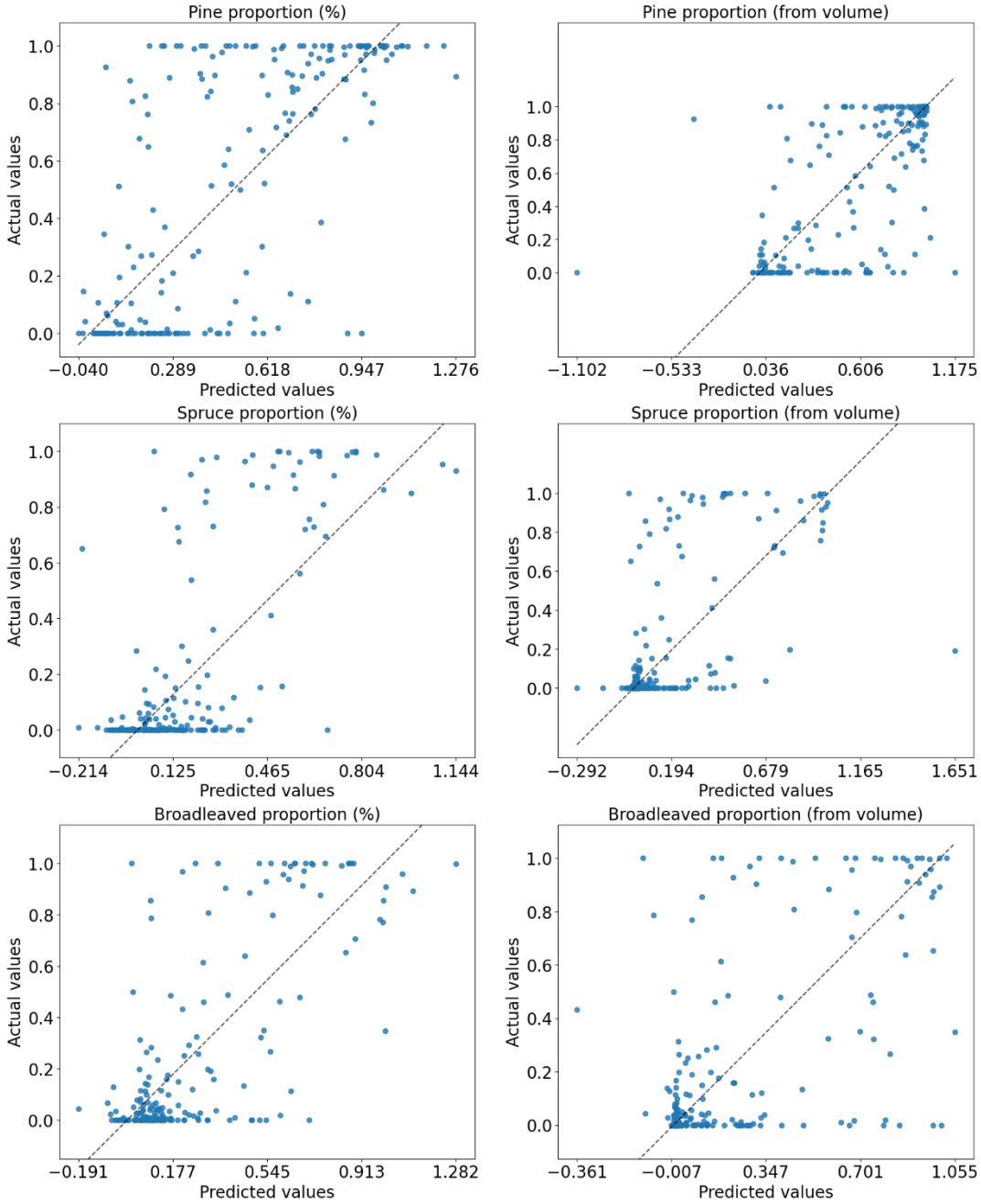

Figure 20. SVR estimated species proportions compared to field data values in Virrat.

The smallest relative estimation errors were reached in pine proportion estimations with RRMSE values between 59–62 % in both estimation approaches, combined to R^2 values between 0.42–0.47. The highest R^2 values at 0.54 were reached in spruce and broadleaved percentage estimations with related RRMSE values between 99–113 %.

When considering all study locations, the best fitting machine learning model reached an R^2 value at 0.72. This was reached for SVR model in Kuopio, when broadleaved proportion was estimated as a percentage. The smallest relative estimation error with RRMSE value at 44 % was reached for RF model in Sodankylä, when pine proportion was calculated from estimated stem volumes. In all five study locations, total stem volume estimations reached R^2 values between 0.22–0.61 and RRMSE values between 52–93 %. The RRMSE values reached in the species-specific stem volume estimations varied between 82–207 % and the R^2 values varied between -0.12–0.62. In the species proportion estimations RRMSE values varied between 44–155 % and R^2 values varied between 0.17–0.72, when considering both approaches.

5 Discussion

This section presents the most relevant results in terms of the defined research questions. In section 5.1 the estimation accuracies of total stem volume, species-specific stem volume and species proportion is discussed and compared to related background information in the previous studies. In section 5.2 possible factors related to differences in the estimation accuracies between the study locations are discussed. Based on the results, section 5.3 presents possible improvements and suggestions for following research.

5.1 Estimations of stem volumes and species proportions

As presented in Section 2.3.1, most of the previous studies related to estimation of boreal forest attributes from optical remote sensing data have applied multispectral satellite instruments and aerial hyperspectral instruments. These studies offered some background information for the previously achieved estimation accuracies of boreal forest attributes, although they were not directly comparable with the estimation accuracies reached from the applied hyperspectral satellite data.

The R^2 values between 0.22–0.61 and RRMSE values between 52–93 % reached in the total stem volume estimations were quite well aligned with the previous studies. The values were compared to stem volume estimations performed by Hyyppä et al. (2000), stem biomass estimations performed by Halme et al. (2019) and total growing stock estimations performed by Tuominen et al. (2017). Hyyppä et al. (2000) reached R^2 value of 0.55 with the hyperspectral aerial instrument and 0.31 with the multispectral satellite instrument. The relative standard errors were between 45–56 %. Halme et al. (2019) reached R^2 value of 0.61 and RRMSE value of 29% with both multispectral satellite data and hyperspectral aerial data. Tuominen et al. (2017) reached RRMSE value at 60.95 % when performing estimations from multispectral satellite data.

In general, the total stem volume estimations had smaller relative estimation errors and more consistent R^2 values than the species-specific stem volume estimations, that reached R^2 values between -0.12–0.62 and RRMSE values between 82–207 %. Tuominen et al. (2017) noticed similar differences in the RRMSE values when comparing the estimation accuracy of total growing stock to the estimation accuracy of species-specific growing stocks. The estimations of species-specific growing stock reached larger RRMSE values between 81.67–189.74 %.

The accuracy of pine, spruce and broadleaved stem volume estimations varied within the study locations. The smallest relative estimation errors were reached for the most dominant species of each study location. The species

dominance was visible as the largest mean volumes and proportions in the field data. Pine as the most dominant species in Sodankylä, Perho, Äänekoski and Virrat reached the best RRMSE values in species-specific volume estimations at these locations. In Kuopio, spruce as the most dominant species reached the best RRMSE values. The RRMSE values of the most dominant species in each location varied between 82–143 % whereas for the less dominant species of each location, the RRMSE values varied between 116–207 %. Also, Tuominen et al. (2017) compared the RRMSE values reached for growing stock of pine, spruce and broadleaved species and received similar results. Pine as the most dominant species reached the best RRMSE value, whereas larger RRMSE values were reached for the less dominant broadleaved species and spruce.

Estimations of species volumes reached slightly worse estimation accuracies than estimations of species proportions, that reached R^2 values between 0.17–0.72 and RRMSE values between 44–155 %. From the two approaches applied to estimate species proportions, estimation as a percentage showed slightly better accuracies with a better RRMSE value in 73 % of the cases and a better R^2 value in 67 % of the cases, when compared to the proportions calculated from estimated volumes. However, the differences in the estimation accuracy between the two approaches were generally small. Also in the species proportion estimations, the smallest relative estimation errors were reached for the most dominant species of each location. The RRMSE values varied between 44–73 % for the most dominant species of each location and between 63–155 % for the less dominant species.

Opposed to the RRMSE values, the differences in R^2 values between pine, spruce and broadleaved species did not seem to be explained by the species dominance. When only the most dominant species of each location were considered, R^2 values of volume estimations varied between 0.07–0.51 and R^2 values of proportion estimations varied between 0.33–0.64. For the less dominant species of each study location, R^2 values of volume estimations varied between -0.12–0.62 and R^2 values of proportion estimations varied between 0.17–0.72.

Overall, the R^2 values applied to describe the SVR and RF model fitting were quite low. The calculated R^2 values describe the proportion of variability in the estimations that can be explained by the measured reflectance spectrum (James et al. 2013). As the reflectance spectrum measured for each satellite image pixel is affected by a large variety of features, it can not fully explain the variability in the estimated attributes. Due to this limitation, the R^2 values are typically quite low for estimations performed from hyperspectral satellite data (Bioucas-Dias et al. 2013).

Based on the results, estimation of stem volumes from hyperspectral satellite data offered a good alternative to estimations made from multispectral satellite data. For example, at best the R^2 values reached for total stem volume estimations were as good as previously reached from hyperspectral aerial data and the RRMSE values were similar to ones reached from multispectral satellite data. Estimations of species proportion, especially as a percentage, gave a good alternative to estimation of species-specific stem volumes by offering a slightly better accuracy.

5.2 Estimation accuracies between the study locations

Estimation accuracies of the total stem volumes, species-specific stem volumes and species proportions varied between the study locations and many factors could have affected these differences. The effect of less quantifiable factors, such as differences in the developed SVR and RF models or possible forest changes between the data collection dates, were more difficult to validate. For this reason, the discussed factors were limited to the differences in data characteristics between the study locations. From the available data characteristics, differences in field data values of stem volumes and species proportions were considered, as well as the differences in the amount of field plots within the study locations.

When the relative estimation errors of species volumes and proportions were compared within each study location, the most dominant species of each location reached the smallest errors. A similar finding was found between the study locations, when large enough differences occurred in field data values of species volumes and proportions.

Pine was the most dominant species in Sodankylä, Perho, Äänekoski and Virrat, where mean pine proportions were between 43–64 % and mean pine volumes between 41.68–76.15 $m^3\ ha^{-1}$. In these locations pine volume and proportion estimations reached better RRMSE values than in Kuopio, where mean pine proportion and volume were lower with values at 17 % and 28.92 $m^3\ ha^{-1}$. The RRMSE values reached in pine volume estimations of Kuopio varied between 181–201 % whereas in the other locations RRMSE values varied between 82–143 %. The RRMSE values reached in pine proportion estimations of Kuopio varied between 114–133 % whereas in the other locations RRMSE values varied between 44–73 %.

The best RRMSE values in spruce volume and proportion estimations were reached in Kuopio, where spruce was the most dominant species with mean proportion at 52 % and mean volume at 84.47 $m^3\ ha^{-1}$. The RRMSE values in spruce volume estimations of Kuopio varied between 96–109 % and in spruce proportion estimations between 49–64 %. In Äänekoski, the mean

volume and proportion of spruce was also quite high at 63.95 $m^3\ ha^{-1}$ and 37 %. The RRMSE values reached in spruce volume estimations of Äänekoski varied between 146–160 % and in spruce proportion estimations between 63–78 %. In rest of the study locations mean spruce volume and proportion were lower and varied between 14.15–32.82 $m^3\ ha^{-1}$ and 13–22 %. Also, the RRMSE values reached in spruce volume and proportion estimations were worse. The RRMSE values reached in spruce volume estimations varied between 173–207 % and in spruce proportion estimations between 113–155 %.

The mean broadleaved volumes and proportions in the field data were more similar between the study locations. The mean volumes varied between 13.93–37.67 $m^3\ ha^{-1}$ and the mean proportions varied between 20–31 % in all locations. Variability within this range did not explain the differences in the relative estimation errors reached for broadleaved volume and proportion estimations. The RRMSE values reached in broadleaved volume estimations varied between 116–202 % and the RRMSE values reached in broadleaved proportion estimations varied between 66–120 % in all study locations.

The differences in pine, spruce and broadleaved volumes and proportions within the study locations did not explain the differences in R^2 values reached in the estimations. Differences in species volumes and proportions between the study locations did not seem to explain the differences in reached R^2 values either. In all study locations, R^2 values reached in pine volume estimations varied between 0.07–0.47 and values reached in pine proportion estimations varied between 0.33–0.65. In spruce volume estimations R^2 values varied between 0.05–0.62 and in spruce proportion estimation values varied between 0.17–0.64. In broadleaved volume estimations R^2 values varied between 0.12–0.41 and values in broadleaved proportion estimations varied between 0.23–0.72.

The field data values of mean total volumes varied between 69.76–168.24 $m^3\ ha^{-1}$ in the study locations. Although the differences were quite large, they did not explain the differences in RRMSE and R^2 values reached in total volume estimations between the study locations. Opposed to the species-specific estimations, in total stem volume estimations the smallest relative estimation errors at 52–54 % were reached with the smallest mean total volume in the field data at Sodankylä. The largest relative estimation errors at 81–93 % were reached with the largest mean total volume in the field data at Äänekoski. Perho and Kuopio, though, reached smaller relative estimation errors with higher mean field data values of total volume when compared to Virrat.

The amount of field plots in the study locations varied between 98–287. The amount was the highest in Sodankylä and the lowest in Kuopio. Perho, Äänekoski and Virrat had intermediate values with 197–218 field plots. In total volume and broadleaved volume estimations the best RRMSE values were reached in Sodankylä, where also the amount of field plots was the highest. The best RRMSE values in broadleaved proportion estimations were reached in Kuopio and Sodankylä where the amount of field plots was the lowest and highest. In Kuopio, though, the lower amount of field plots could have been compensated with a slightly larger mean proportion of broadleaved species at 31 % in the field data, when compared to the other locations with mean proportions between 20–25 %. In rest of the locations, the differences in RRMSE values were not explained by the differences in the amount of field plots. Estimations related to pine and spruce were not compared, as these were affected by the differences in the species dominance.

Distribution of the best R^2 values, on the other hand, was quite even between study locations with different amount of field plots. Sodankylä had the best values in total stem volume estimations, Virrat and Sodankylä in broadleaved volume estimations and Perho and Kuopio in broadleaved proportion estimations.

Based on the results, the accuracy of total stem volume estimations was more consistent and less dependent on the characteristics of the study locations. The accuracy of species-specific estimations was affected by the differences in species dominance between the study locations, as larger species dominance was related to better estimation accuracies. The variation in the amount of field plots between 98–218 did not affect the estimation accuracy between the locations. However, the highest amount of field plots at 287 in Sodankylä might have resulted into slightly better estimation accuracies.

5.3 Suggestions for improvements and following research

When sufficient accuracy of estimations is reached, developed machine learning models can be applied to a wider area. For example, the developed SVR and RF models could be used to estimate the total stem volumes, species-specific stem volumes and species proportions in the whole extent of the applied hyperspectral satellite images.

Especially the estimation accuracy of total stem volume was consistent in different study locations, which supports its applicability in further use. However, the performed species-specific estimations seemed to benefit from a larger species dominance. This degrades the applicability of estimations performed for less dominant species and for locations where distribution of species is unknown. To overcome this limitation, instead of estimating stem

volumes and proportions related to single species, it could be beneficial to estimate for example the Shannon's index, that can be applied to describe the overall species richness in boreal forests (Pitkänen 1998; Reich et al. 2001).

It seemed that the highest amount of field plots at 287 in Sodankylä could have positively impacted the estimation accuracies. The amount of field plots applied in the studies by Hyyppä et al. (2000), Halme et al. (2019) and Tuominen et al. (2017) was even larger and varied between 483–2469. These findings suggest that better estimation accuracies might be reached by including a larger amount of field plots within each study location.

Currently, in addition to EnMAP mission also PRISMA mission offers hyperspectral satellite data with an easy and open access data portal (ASI 2024). The hyperspectral data collected by the PRISMA mission should be well applicable for similar estimations as performed in this study. Also, the planned CHIME and SBG missions should provide global hyperspectral satellite data applicable for similar estimations in the future (ESA 2021; NASA 2018). When the availability of hyperspectral satellite data grows, the accuracy of estimations can be investigated on a larger scale. This will enable a deeper understanding for the applicability of the estimations in assessment of boreal forest biodiversity.

6 Conclusions

The overall goal of this work was to gain information on estimation of different boreal forest attributes from hyperspectral satellite data. In general, the best estimation accuracies with RRMSE values between 52–93 % were reached in the estimations of total stem volume. In the species-specific estimations, better estimation accuracies were reached for the species proportions. The RRMSE values reached in species proportion estimations varied between 44–155 % with both approaches. The RRMSE values reached in species volume estimations varied between 82–207 %.

Interest was also in factors that might affect the estimation accuracies of different boreal forest attributes and boreal forest locations. It seemed that the slightly higher amount of field plots at 287 in Sodankylä, might have increased the accuracy of all estimations when compared to the other study locations with field plots varying between 98–218. The estimation accuracy of species-specific attributes was affected by the species dominance, both within and between the study locations. The best estimation accuracies were reached for the most dominant species of each location. The RRMSE values of the most dominant species of each location varied between 82–143 % in the stem volume estimations and between 44–73 % in the species proportion estimations. The best estimation accuracy for spruce proportion and volume were reached in Kuopio, where spruce was the most dominant species. Similarly, better estimation accuracies for pine proportion and volume were reached in rest of the study locations, where pine was the most dominant species. Distribution of broadleaved species was more even between the study locations, and the smaller differences did not explain the variability of RRMSE values reached in broadleaved volume and proportion estimations. Also, the estimation accuracy of total stem volume was more consistent between the study locations.

The R^2 values applied to describe the fit of developed SVR and RF models varied between -0.12–0.62 in the species-specific stem volume estimations, between 0.17–0.72 in the species proportion estimations and between 0.22–0.61 in the total stem volume estimations. Systematic factors behind the variability of R^2 values between the estimated attributes and study locations were not identified.

References

ASI. (2024). *PRISMA portal.* [Accessed 22 November 2024]. Available: https://prisma.asi.it/

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N. & Chanussot, J. (2013). *Hyperspectral remote sensing data analysis and future challenges.* IEEE Geoscience and remote sensing magazine, 1(2), 6-36, https://doi.org/10.1109/MGRS.2013.2244672

Breiman, L. (2001). *Random forests.* Machine learning, 45, 5-32, https://doi.org/10.1023/A:1010933404324

Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S. & Malo, J. (2022). *Remote sensing image processing,* Switzerland: Springer Cham, https://doi.org/10.1007/978-3-031-02247-0

Chabrillat, S., Guanter, L., Kaufmann, H., Foerster, S., Beamish, A., Brosinsky, A., Wulf, H., Asadzadeh, S., Bochow, M., Bohn N., Boesche, N., Bracher, A., Brell, M., Buddenbaum, Cerra, D. H., Fischer, S., Hank, T., Heiden, U., Heim, B., Heldens, W., Hill, J., Hollstein, A., Hostert, P., Krasemann, H., LaPorta, L., Leitão, P.J., van der Linden, S., Mauser, W., Milewski, R., Mottus, M., Okujeni, A., Oppelt, N., Pinnel, N., Roessner, S., Röttgers, R., Schneiderhan, T., Schickling, A., Soppa, M., Staenz, K. & Segl, K. (2022). *EnMAP Science Plan.* EnMAP Technical Report, GFZ Data Services, http://doi.org/10.48440/enmap.2022.001

Chabrillat, S., Foerster, S., Segl, K., Beamish, A., Brell, M., Asadzadeh, S., Milewski, R., Ward, K. J., Brosinsky, A., Koch, K., Scheffler, D., Guillaso, S., Kokhanovsky, A., Roessner S., Guanter, L., Kaufmann, H., Pinnel, N., Carmona, E., Stroch, T., Hank, T., Berger, K., Wocher, M., Hostert, P., Van Der Linden, S., Okujeni, A., Janz, A., Jakimow, B., Bracher, A., Soppa, M. A., Alvarado, L. M. A., Buddenbaum, H., Heim, B., Heiden, U., Moreno, J., Ong, C., Bohn, N., Green, R. O., Bachmann, M., Kokaly, R., Schodlok, M., Painter, T. H., Gascon, F., Buongiorno, F., Mottus, M., Brando, V. E., Feilhauer, H., Betz, M., Baur, S., Feckl, R., Schickling, A., Krieger, V., Bock, M., La Porta L. & Fischer, S. (2024). *The EnMAP spaceborne imaging spectroscopy mission: Initial scientific results two years after launch.* Remote Sensing of Environment, 114379, https://doi.org/10.1016/j.rse.2024.114379

Corona, P., Chirici, G., McRoberts, R. E., Winter, S. & Barbati, A. (2011). *Contribution of large-scale forest inventories to biodiversity assessment and monitoring.* Forest ecology and management, 262(11), pp. 2061-2069, https://doi.org/10.1016/j.foreco.2011.08.044

CRESDA. (2022a). *Satellite Resources: ZY-1 02D 30m Hyperspectral image*. [Accessed 3 December 2024]. Available: https://www.cresda.com/zgzywxyyzxeng/20220726141812992629/sid/article/2022081209364976591148 2.html

CRESDA. (2022b). *Satellite Resources: GF-5B 30m Hyperspectral image*. [Accessed 3 December 2024]. Available: https://www.cresda.com/zgzywxyyzxeng/20220726141812992629/sid/article/2022081209351584835919 7.html

DLR. (2019): *DESIS - Hyperspectral Images - Global*. [Accessed 19 December 2024]. Available: https://doi.org/10.15489/hxom21uqeo90

DLR. (2024a). *EnMAP: Data Access Portal*. [Accessed 22 November 2024]. Available: https://planning.enmap.org/

DLR. (2024b). *EnMAP HSI Level 1 / Level 2 Product Specification Document*. [Accessed 29 November 2024]. Available: https://www.enmap.org/data/doc/EN-PCV-ICD-2009-2_HSI_Product_Specification_Level1_Level2.pdf

Duren, R., Cusworth, D., Ayasse, A., Herner, J., Thorpe, A., Falk, M., Heckler, J., Guido, J., Giuliano, P., Chapman, J., Green, R. & Asner, G. (2021). *Carbon Mapper: on-orbit performance predictions and airborne prototyping,* In AGU Fall Meeting Abstracts (Vol. 2021, pp. A53F-05). [Accessed 14 January 2025]. Available: https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/915730

ESA. (2021*). Going hyperspectral for CHIME*. [Accessed 30.8.2024]. Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Going_hyperspectral_for_CHIME

ESA. (2023). *EoPortal: EO-1 (Earth Observing-1)*. [Accessed 19 November 2024]. Available: https://www.eoportal.org/satellite-missions/eo-1

ESA. (2024). *The Sentinel missions*. [Accessed 19 November 2024]. Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/The_Sentinel_missions

European Commission, Directorate-General for Environment. (2021). *EU biodiversity strategy for 2030: bringing nature back into our lives*. Publications Office of the European Union, https://data.europa.eu/doi/10.2779/677548

FAO. (2020). *Global Forest Resources Assessment 2020: Main report,* Rome, Italy: FAO, https://doi.org/10.4060/ca9825en

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C. & Ghosh, A. (2016*). Review of studies on tree species classification from remotely sensed data.* Remote sensing of environment, 186, 64-87, https://doi.org/10.1016/j.rse.2016.08.013

Feilhauer, H., Somers, B. & van der Linden, S. (2017). *Optical trait indicators for remote sensing of plant species composition: Predictive power and seasonal variability*. Ecological Indicators, 73, 825-833, https://doi.org/10.1016/j.ecolind.2016.11.003

Folkman, M. A., Pearlman, J., Liao, L. B. & Jarecke, P. J. (2001). *EO-1/Hyperion hyperspectral imager design, development, characterization, and calibration*. Hyperspectral Remote Sensing of the Land and Atmosphere, 4151, 40-51, https://doi.org/10.1117/12.417022

Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A. Z. & Schepaschenko, D. G. (2015). *Boreal forest health and global change*. Science, 349(6250), pp. 819-822, https://doi.org/10.1126/science.aaa9092

GHGSat. (2024). *GHGSat*. [Accessed 19 November 2024]. Available: https://www.ghgsat.com/en/

Goetz, A. F. (2009). *Three decades of hyperspectral remote sensing of the Earth: A personal view*. Remote sensing of environment, 113, pp. S5-S16, https://doi.org/10.1016/j.rse.2007.12.014

Green, R. O., Mahowald, N., Ung, C., Thompson, D. R., Bator, L., Bennet, M., Bernas, M., Blackway, N., Bradley, C., Cha, J., Clark, P., Clark, R., Cloud, D., Diaz, E., Ben Dor, E., Duren, R., Eastwood, M., Ehlmann, B. L., Fuentes, L., Ginoux, P., Gross, J., He, Y., Kalashnikova, O., Kert, W., Keymeulen, D., Klimesh, M., Ku, D., Kwong-Fu, H., Liggett, E., Li, L., Lundeen, S., Makowski, M. D., Mazer, A., Miller, R., Mouroulis, P., Oaida, B., Okin, G. S., Ortega, A., Oyake, A., Nguyen, H., Pace, T., Painter, T. H., Pempejian, J., Garcia-Pando, C. P., Pham, T., Phillips, B., Pollock, R., Purcell, R., Realmuto, V., Schoolcraft, J., Sen, A., Shin, S., Shaw, L., Soriano, M., Swayze, G., Thingvold, E., Vaid, A. & Zan, J. (2020). *The Earth surface mineral dust source investigation: An Earth science imaging spectroscopy mission*. In 2020 IEEE aerospace conference (pp. 1-15). IEEE, https://doi.org/10.1109/AERO47225.2020.9172731

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitão, P., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K. & Sang, B. (2015). *The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation*. Remote Sens. 2015, 7(7), 8830-8857, https://doi.org/10.3390/rs70708830

Guarini, R., Loizzo, R., Facchinetti, C., Longo, F., Ponticelli, B., Faraci, M., Dami, M., Cosi, M., Amuruso, L., De Pasquale, V., Taggio, N., Santoro, F., Colandrea, P., Miotti, E. & Di Nicolantonio, W. (2018). *PRISMA hyperspectral mission products*. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 179-182, https://doi.org/10.1109/IGARSS.2018.8517785

Halme, E., Pellikka, P. & Mottus, M. (2019). *Utility of hyperspectral compared to multispectral remote sensing data in estimating forest biomass and structure variables in Finnish boreal forest*. International Journal of Applied Earth Observation and Geoinformation, 83, 101942, https://doi.org/10.1016/j.jag.2019.101942

Haykin, S. (1999). *Neural networks—a comprehensive foundation,* 2nd edn. Prentice Hall, Upper Saddle.

Huang, Z., Turner, B. J, Dury, S. J., Wallis, I. R. & Foley, W. J. (2004). *Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis*. Remote Sens Environ 93(1):18–29, https://doi.org/10.1016/j.rse.2004.06.008

Hunter, M. L. (1999). *Biological diversity*. In: Hunter, M. L. (ed.). *Maintaining biodiversity in forest ecosystems*, Cambridge, UK: Cambridge university press, pp. 3-21.

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S. & Zhu, Y. H. (2000). *Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes*. Forest Ecology and Management, 128(1-2), pp. 109-120, https://doi.org/10.1016/S0378-1127(99)00278-9

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: springer.

Juola, J., Hovi, A. & Rautiainen, M. (2024). *Comparison of contemporaneous Sentinel-2 and EnMAP data for vegetation index-based estimation of leaf area index and canopy closure of a boreal forest*. European Journal of Remote Sensing, 57(1), https://doi.org/10.1080/22797254.2024.2432975

Kaufmann, H., Sang, B., Storch, T., Segl, K., Foerster, S., Guanter, L., Erhard, M., Heider, B., Hofer, S., Honold, H.-P., Penné, B., Bachmann, M., Habermeyer, M., Müller, A., Müller, R., Rast, M., Staenz, K., Straif, C. & Chlebek, C. (2016). *Environmental Mapping and Analysis Program – A German Hyperspectral Mission*. In: Qian, S. E. (ed.), *Optical Payloads for Space Missions,* Chichester: Wiley, p. 161-182, https://doi.org/10.1002/9781118945179.ch7

Kneeshaw, D. & Gauthier, S. (2003*). Old growth in the boreal forest: A dynamic perspective at the stand and landscape level*. Environmental Reviews, 11(S1), pp. 99–S114, http://www.jstor.org/stable/envirevi.11.s1.s99

Kuva Space. (2023). *Kuva Space*. [Accessed 25 September 2024]. Available: https://kuvaspace.com/

Lee, K. S., Cohen, W. B., Kennedy, R. E., Maiersperger, T. K. & Gower, S. T. (2004). *Hyperspectral versus multispectral data for estimating leaf area index in four different biomes*. Remote Sensing of Environment, 91(3-4), 508-520, https://doi.org/10.1016/j.rse.2004.04.010

Loizzo, R., Guarini, R., Longo, F., Scopa, T., Formaro, R., Facchinetti, C. & Varacalli, G. (2018). *PRISMA: The Italian hyperspectral mission*. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 175-178, https://doi.org/10.1109/IGARSS.2018.8518512

Metsäkeskus. (2024a). *Metsävaratiedot*. [Accessed 9 August 2024]. Available: https://www.metsakeskus.fi/fi/avoin-metsa-ja-luontotieto/metsa-tietoaineistot/metsavaratiedot

Metsäkeskus. (2024b). *Muut arvokkaat elinympäristöt ja luontokohteet*. [Accessed 20 August 2024]. Available: https://www.metsakeskus.fi/fi/metsan-kaytto-ja-omistus/metsien-suojelu-ja-elinymparistojen-hoito/muut-arvokkaat-elinymparistot-ja-luontokohteet

Metsäkeskus. (2024c). *Paikkatietoaineistot: Inventointikoealat*. [Accessed 29 November 2024]. Available: https://www.metsakeskus.fi/fi/avoin-metsa-ja-luontotieto/aineistot-paikkatieto-ohjelmille/paikkatietoaineistot

Metsäkeskus. (2024d). *Tietotuotekuvaus: Inventointikoealat*. [Accessed 29 November 2024]. Available: https://www.metsakeskus.fi/sites/default/files/document/tietotuotekuvaus-inventointikoealat.pdf

NASA. (2018*). Welcome to Surface Biology and Geology study*. [Accessed 19 November 2024]. Available: https://sbg.jpl.nasa.gov/

NASA. (2024). *Landsat science: Satellites*. [Accessed 19 November 2024]. Available: https://landsat.gsfc.nasa.gov/satellites/

Noss, R. F. (1990). *Indicators for monitoring biodiversity: a hierarchical approach*. Conservation biology, 4(4), pp. 355-364, https://doi.org/10.1111/j.1523-1739.1990.tb00309.x

Pearlman, J. S., Barry, P. S., Segal, C. C., Shepanski, J., Beiso, D. & Carman, S. L. (2003). *Hyperion, a space-based imaging spectrometer*. IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 6, pp. 1160-1173, June 2003, https://doi.org/10.1109/TGRS.2003.815018

Pitkänen, S. (1998). *The use of diversity indices to assess the diversity of vegetation in managed boreal forests*, Forest ecology and management, 112(1-2), 121-137, https://doi.org/10.1016/S0378-1127(98)00319-3

Planet. (2024). *TANAGER*. [Accessed 25 September 2024]. Available: https://www.planet.com/products/hyperspectral/

Rasmussen, C. E. & Williams, C. K. I. (2006*). Gaussian processes for machine learning*. The MIT Press, New York.

Reich, P. B., Bakken, P., Carlson, D., Frelich, L. E., Friedman, S. K. & Grigal, D. F. (2001). *Influence of logging, fire, and forest type on biodiversity and productivity in southern boreal forests.* Ecology, 82(10), 2731-2748, https://doi.org/10.1890/0012-9658(2001)082[2731:IOLFAF]2.0.CO;2

Richards, J. A. (2022). *Remote sensing digital image analysis,* 6th edn, Berlin/Heidelberg, Germany: Springer, https://doi.org/10.1007/978-3-030-82327-6

Räty, M. & Vaahtera, E. (2023). *Forest resources.* In: Kulju, I., Niinistö, T., Peltola, A., Räty, M., Sauvula-Seppälä, T., Torvelainen, J., Uotila, E. & Vaahtera, E. (eds.), *Finnish Statistical Yearbook of Forestry 2022, Finland, Helsinki:* Natural Resources Institute Finland, pp. 17 – 40, http://urn.fi/URN:ISBN:978-952-380-584-2

Schaepman, M. E., Ustin, S. L., Plaza, A. J., Painter, T. H., Verrelst, J. & Liang, S. (2009). *Earth system science related imaging spectroscopy—An assessment*. Remote Sens. Environ. 113, S123–S137, https://doi.org/10.1016/j.rse.2009.03.001

Scikit-learn. (2024a). *Scikit-learn: Machine learning in Python*. [Accessed 29 November 2024]. Available: https://scikit-learn.org/stable/index.html

Scikit-learn. (2024b). *Scikit-learn: SVR*. [Accessed 29 November 2024]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html

Scikit-learn. (2024c). *Scikit-learn: RandomForestRegressor*. [Accessed 29 November 2024]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

Scikit-learn. (2024d). *Scikit-learn: GridSearchCV*. [Accessed 29 November 2024]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

Scikit-learn. (2024e). *Scikit-learn: RBF SVM parameters*. [Accessed 29 November 2024]. Available: https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html

Shafri, H. Z. (2017). *Machine learning in hyperspectral and multispectral remote sensing data analysis*. In Artificial Intelligence Science and Technology: Proceedings of the 2016 International Conference (AIST2016) (pp. 3-9), https://doi.org/10.1142/9789813206823_0001

Smola, A. J. & Schölkopf, B. (2004). *A tutorial on support vector regression*. Statistics and Computing 14, 199–222, https://doi.org/10.1023/B:STCO.0000035301.49549.88

Storch,T., Honold, H. P., Chabrillat, S., Habermeyer, M., Tucker, P., Brell, M., Ohndorf, A., Wirth, K., Betz, M., Kuchler, M., Mühle, H., Carmona, E., Baur, S., Mücke, M., Löw, S., Schulze, D., Zimmermann, S., Lenzen, C., Wiesner, S., Aida, S., Kahle, R., Willburger, P., Hartung, S., Dietrich, D., Plesia, N., Tegler, M., Schork, K., Alonso, K., Marshall, D., Gerasch, B., Schwind, P., Pato, M., Schneider, M., De Los Reyes, R., Langheinrich, M., Wenzel, J., Bachmann, M., Holzwarth, S., Pinnel, N., Guanter, L., Segl, K., Scheffler, D., Foerster, S., Bohn, N., Bracher, A., Soppa, M. A., Gascon, F., Green, R., Kokaly, R., Moreno, J., Ong, C., Sornig, M., Wernitz, R., Bagschik, K., Reintsema, D., La Porta, L., Schickling, A. & Fischer, S. (2023). *The EnMAP imaging spectroscopy mission towards operations*. Remote

Sensing of Environment, 294, 11363,
https://doi.org/10.1016/j.rse.2023.113632

Tikka, T., Makynen, J. & Shimoni, M. (2023). *Hyperfield-Hyperspectral small satellites for improving life on Earth.* In 2023 IEEE Aerospace Conference (pp. 1-8). IEEE,
https://doi.org/10.1109/AERO55745.2023.10115806

Tuia, D., Volpi, M., Verrelst, J. & Camps-Valls, G. (2018). *Advances in Kernel Machines for Image Classification and Biophysical Parameter Retrieval.* In: Moser, G. & Zerubia, J. (eds.) *Mathematical Models for Remote Sensing Image Processing. Signals and Communication Technology.* Springer, Cham, https://doi.org/10.1007/978-3-319-66330-2_10

Tuominen, S., Pitkänen, T., Balazs, A. & Kangas, A. (2017). *Improving Finnish Multi-Source National Forest Inventory by 3D aerial imaging,* Silva Fennica vol. 51 no. 4 article id 7743, https://doi.org/10.14214/sf.7743

Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J. & Moreno, J. (2012). *Retrieval of vegetation biophysical parameters using Gaussian process techniques,* IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 5, pp. 1832-1843, May 2012,
https://doi.org/10.1109/TGRS.2011.2168962

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J. P., Lewis, P., North, P. & Moreno, J. (2019). *Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods.* Surveys in Geophysics, 40, pp. 589-629,
https://doi.org/10.1007/s10712-018-9478-y

Whittaker, R. H. (1972). *Evolution and measurement of species diversity.* Taxon, 21(2-3), pp. 213-251.