DISSERTATIONS IN
FORESTRY AND
NATURAL SCIENCES

Paras Pant

Optimizing Spectral Bands of Airborne Imager for Tree Species Classification

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences No 177

UNIVERSITY OF
EASTERN FINLAND

PARAS PANT

Optimizing Spectral Bands of Airborne Imager for Tree Species Classification

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences
No 177

Academic Dissertation
To be presented by permission of the Faculty of Science and Forestry for public examination in the Auditorium AU100 in Aurora Building at the University of Eastern Finland, Joensuu, on June, 11, 2015,
at 12 o'clock noon.

School of Computing

Kopio Niini Oy
Helsinki, 2015
Editors: Prof. Pertti Pasanen, Prof. Kai Peiponen, Prof. Pekka Kilpeläinen
and Prof. Matti Vornanen

Distribution:
University of Eastern Finland Library / Sales of publications
P.O. Box 107, FI-80101 Joensuu, Finland
tel. +358-50-3058396
http://www.uef.fi/kirjasto

ISBN: 978-952-61-1782-9 (printed)
ISSNL: 1798-5668
ISSN: 1798-5668
ISBN: 978-952-61-1783-6 (pdf)
ISSNL: 1798-5668
ISSN: 1798-5676

Author's address:
University of Eastern Finland
School of Computing
P.O.Box 111
FI-80101 Joensuu, FINLAND
email: paras.pant@uef.fi

Supervisors:
Ville Heikkinen, Ph.D.
University of Eastern Finland
School of Computing
P.O.Box 111
80101 Joensuu, FINLAND
email: ville.hekkinen@uef.fi

Professor Timo Tokola, Ph.D.
University of Eastern Finland
School of Forest Sciences
P.O.Box 111
80101 Joensuu, FINLAND
email: timo.tokola@uef.fi

Professor Markku Hauta-Kasari, Ph.D.
University of Eastern Finland
School of Computing
P.O.Box 111
80101 Joensuu, FINLAND
email: markku.hauta-kasari@uef.fi

Reviewers:
Adjunct Professor Matti Mõttus, Ph.D.
University of Helsinki
Department of Geosciences and Geography
P.O.Box 64, Gustaf Hällströminkatu 2
00014 Helsinki, FINLAND
email: matti.mottus@helsinki.fi

Professor (Associate) Alamin Mansouri, Ph.D.
Université de Bourgogne
LABORATOIRE Le2i
BP 16, Route des Plaines de l'Yonne
89010 AUXERRE Cedex, FRANCE
email: alamin.mansouri@u-bourgogne.fr

Opponent:
Professor Pekka Neittaanmäki, Ph.D.
University of Jyväskylä
Department of Mathematical Information Technology
P.O.Box 35 (Agora)
40014 Jyväskylä, FINLAND
email: pekka.neittaanmaki@jyu.fi

ABSTRACT

A hyperspectral sensor provides the possibility of recording a large number of spectral bands. However, considering a specific application some bands will be redundant and complicate data processing and transmission. To avoid these problems, feature selection methods have been used to select a subset of hyperspectral bands in a post-processing phase. Furthermore, current hyperspectral sensors capture lower spatial resolution imagery when compared with images captured using multispectral sensors. Efficient high-altitude, high spatial resolution data acquisitions are only feasible with the use of multispectral sensors. Here, it was assumed that in the future some hyperspectral sensors will be designed in such a way that imaging band position could be defined in the pre-flight setup. This approach can be used to remove redundant bands before the actual data collection phase. Alternatively, multispectral sensors could be designed with the optimized narrow / broadband spectral sensitivities needed for specific applications. In these contexts, using training hyperspectral data a suitable subset of bands can be found by using feature selection methods. There is, however, a need to investigate whether these selected bands or optimized spectral sensitivities provide a reasonable classification performance.

In this dissertation, a sparse regression-based feature selection method was used in the selection of hyperspectral bands in the 400–1000 nm wavelength range. In band selection, the effects of the spatial scale and balance in the training samples were evaluated. The band selection results showed that the use of a balanced plot-level scale dataset and sparse logistic regression with the Bayesian regularization feature selection method provided the minimum eight selected bands. The selected hyperspectral band positions were related to the Leica Airborne Digital Sensor (Leica Geosystems) sensitivity position and optimized 4 and 5-band multispectral sensor systems were proposed. Using the selected hyperspectral bands and simulated responses (standard multispectral sensor sensitivities and optimized sensitivities) the classification of the Scots pine

(*Pinus sylvestris L.*), Norway spruce (*Picea abies (L.) H. Karst.*) and deciduous birch (*Betula pubescens Ehrh. and Betula pendula Roth*) tree species was investigated using plot- and pixel-level scale datasets. In simulations hyperspectral radiance data were used and in the band selection estimated reflectance data were used. The classification performance of selected bands was investigated by either matching and changing the view-illumination geometry condition of the datasets used in the band selection and the classification. In addition, these results were compared with the results obtained using 64 AisaEAGLE II hyperspectral bands in the 400–1000 nm wavelength range.

The tree species classification results suggested that the classification performance of the optimized systems was remarkably improved (4–13%) when compared with the results obtained for simulated responses via standard multispectral sensor systems. For the plot-level scale dataset, the proposed 4 and 5-band multispectral sensor systems provided a similar tree species classification performance as the use of all 64 hyperspectral bands.

For the pixel-level scale dataset, the simulated response of the 4 and 5-band optimized multispectral sensor system classification accuracies was lower than with results obtained using all 64 hyperspectral bands. The eight selected bands provided, however, similar (difference 1–2%) or improved classification performance when compared with the results obtained using all 64 hyperspectral bands throughout the view-illumination geometry conditions. The obtained tree species classification results support the approach of designing optimized or tunable spectral imaging systems.

Universal Decimal Classification: 004.83, 004.93, 535.33, 582.091
Library of Congress Subject Headings: Remote sensing; Spectral imaging; Multispectral imaging; Wavelengths; Image analysis; Classification; Trees; Scots pine; Norway spruce; Birch; Support vector machines; Discriminant analysis; Regression analysis; Aerial surveys in forestry
Yleinen suomalainen asiasanasto: kaukokartoitus; ilmakuvaus; spektrikuvaus; kuva-analyysi; luokitus; puulajit; mänty; kuusi; koivu; regressioanalyysi; koneoppiminen

Preface

First, I would like to thank the University of Eastern Finland, School of Computing, for accepting me as a Ph.D. student and providing the facilities to conduct this research. This research was conducted during 2011–2014. The study was fully supported by the University of Eastern Finland Project No. 931043 (Multi-scale Geospatial Analysis). I am grateful to the project leader, Prof. Matti Maltamo, School of Forest Sciences, for this support.

I want to express my thanks to Prof. Jussi Parkkinen and Prof. Markku Hauta-Kasari, who offered me the opportunity to work in their research group. I sincerely thank my supervisors, Dr. Ville Heikkinen and Prof. Markku Hauta-Kasari, for supervising me in Computer Science and Prof. Timo Tokola for supervising and guiding me in Forestry. I am grateful to Ville and Timo for the endless discussions and valuable guidance in scientific writing. I am thankful to my co-authors, Dr. Ilkka Korpela and Aarne Hovi, for their collaboration.

I would like to thank Tuure Takala, Department of Geosciences and Geography, University of Helsinki, for providing the preprocessed image data, measurement details and answering my multiple queries regarding measurement details. Likewise, I would like to thank Dr. Lauri Mehtätalo School of Computing, UEF for his suggestion in thesis writing process. Furthermore, I would like to thank present and past colleagues in the research group with whom I have worked, shared an office and chatted with: Jussi Kinnunen, Jukka Antikainen, Juha Lehtonen, Tuija Jetsu, Pauli Fält, Oili Kohonen, Pesal Koirala, Jouni Hiltunen, Tapani Hirvonen, Niko Penttinen, Piotr Bartczak, Arash Mirhashemi, Zhengzhe Wu, Ana Gebejes, and Joji Sakamoto. The working environment and trips I shared with you all are highly appreciated.

I also greatly appreciate my friends, Manisha Singh, Manash Shah, Anup Nepal and all Nepalese in Joensuu, for the wonder-

ful memories. I am also very proud of my parents (Prof. Prahlad Raj Pant and Mrs. Shanta Pant) and my brother (Pabin Pant), who constantly encouraged and supported me throughout my studies.

Joensuu, May 5, 2015,
Paras Pant

LIST OF PUBLICATIONS

This thesis is based on the following publications:

I Pant P., Heikkinen V., Hovi A., Korpela I., Hauta-Kasari M., Tokola T., Nov 2013. "Evaluation of simulated bands in airborne optical sensors for tree species identification," *Remote Sensing of Environment* **138**, 27 – 37.

II Pant P., Heikkinen V., Korpela I., Hauta-Kasari M., Tokola T., Sept 2014. "Logistic Regression-Based Spectral Band Selection for Tree Species Classification: Effects of Spatial Scale and Balance in Training Samples," *IEEE Geoscience and Remote Sensing Letters* Vol. 11, No. 9, 1604–1608.

III Pant P., Heikkinen V., Hauta-Kasari M., Tokola T., 2014. "Assessment of Hyperspectral Bands for Tree Species Classification Under Changing View–Illumination Geometry," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **(Submitted)**.

Throughout the thesis, these publications are referred to as [*P1*], [*P2*] and [*P3*]. [*P1*] and [*P2*] are peer-reviewed journal articles, and [*P3*] is a submitted peer-reviewed journal article. The publications are included at the end of the thesis with the permissions of their copyright holders.

In addition, the author has participated in the preparation of other peer-reviewed publications [1, 2] as the lead or co-author in the study period.

AUTHOR CONTRIBUTIONS

The publications included in this dissertation are original research papers, and the contributions of the authors are summarized as follows.

The idea for [*P1*], *P2*] and [*P3*] originated from collaboration between the lead author and co-authors. The image data were acquired and processed to at-sensor radiance data in collaboration with the University of Helsinki. The Leica ADS40 sensitivity information was provided by Ulrich Beisl (Leica Geosystems) and the Vexcal Ultacam–D was provided by Susanne Scholz (Microsoft Photogrammetry). The co-authors Ilkka Korpela and Aarne Hovi provided the ground-based information on the tree species plots. In [*P1*], using the simulated responses of standard and proposed optimized multispectral sensor sensitivities, tree species classification performance was studied.

In [*P2*], tree species classification was studied using selected hyperspectral bands and the simulated responses of standard and proposed optimized multispectral sensor sensitivity. Optimized multispectral sensor sensitivities were proposed using the knowledge obtained from the selected hyperspectral band positions. A sparse logistic regression-based feature selection algorithm was used in band selection. In [*P3*], the use of the selected bands for the tree species classification performance was evaluated so that the view-illumination geometry conditions of the datasets used for band selection and classification either matched or deviated.

In all the publications, the lead author conducted the numerical computations, data selection and algorithmic implementations. In [*P1*], co-author Ville Heikkinen performed some numerical computations using a C-SVM classifier. The lead author drafted all the publications for this dissertation. The lead author and co–authors closely collaborated on the written portion of the manuscript. Among the co-authors cooperation with Ville Heikkinen has been particularly important.

LIST OF ABBREVIATION

ACTOR	Atmospheric and Topographic Correction
ADS	Airborne Digital Sensor
ATREM	Atmospheric Removal Code
BRDF	Bidirectional Reflectance Distribution Function
CCD	Charge-Coupled Device
DA	Discriminant Analysis
DMC	Intergraph-Z/I Digital mapping Camera
FLAASH	Fast Line-of-Sight Atmospheric Analysis of Hypercubes
FWHM	Full-Width-Half Maximum
LASSO	Least Absolute Selection and Shrinking Operator
LDA	Linear Discriminant Analysis
LOO	Leave-one-out
LS-SVM	Least Squares Support Vector Machine
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
PCA	Principal Component Analysis
QDA	Quadratic Discriminant Analysis
SVM	Support Vector Machine
UCD	Vexcel UltraCam-D
VNIR	Visible to Near Infrared

Contents

1 INTRODUCTION **1**

1.1 Research Problem 2

1.1.1 Simulated Multispectral Sensor Responses . . 4

1.1.2 Hyperspectral Band Selection 4

1.1.3 Assessment of Selected Bands 5

2 PASSIVE AIRBORNE IMAGING **7**

2.1 Optical Radiation Model 7

2.2 Panchromatic and Multispectral Imaging Sensor . . . 9

2.3 Hyperspectral Imaging Sensor 10

2.4 Atmospheric Correction 13

2.4.1 Absolute Correction Method 14

2.4.2 Relative Correction Method 15

3 HYPERSPECTRAL IMAGING CAMPAIGN **17**

3.1 Remote Sensing Data 17

3.2 Field Data 19

3.3 Data Preparation for Experiment 22

3.4 Noise Removal 24

3.5 Reduction of View-Illumination Geometry Condition Effect 24

4 HYPERSPECTRAL BAND SELECTION **27**

4.1 Sparse Linear Regression 29

4.2 Sparse Logistic Regression 31

4.3 Sparse Logistic Regression with Bayesian Regularization 33

5 CLASSIFIERS **37**

5.1 Discriminant Analysis 37

5.2 Support Vector Machine 39

5.2.1 C–SVM 39

5.2.2 Least Squares Support Vector Machine (LS–SVM) 40

6 EXPERIMENTS **43**
6.1 Hyperspectral Band Selection 43
6.2 Optimized Multispectral Sensor Sensitivities 44
6.3 Simulation of Sensor Responses 46
6.4 Plot- and Pixel-Level Tree Species Classification . . . 48
6.4.1 Simulated Sensor Responses 48
6.4.2 Selected Hyperspectral Bands 49
6.4.3 Assessment of Selected Bands under Changing View-Illumination Geometry Conditions . 50

7 DISCUSSION AND CONCLUSIONS **53**

BIBLIOGRAPHY **57**

APPENDICES: ORIGINAL PUBLICATIONS **72**

1 Introduction

Remote sensing of forest is currently possible with airborne optical sensors that include active (airborne laser scanning) or passive spectral imaging technology. The data obtained with an active sensor are efficient for probing target object shape [3, 4], vegetation density and forest parameters related to tree height [5, 6]. Passive imaging sensor data are useful for target identification and classification [6–8].

Currently, widely used passive multispectral sensors have 3–4 spectral bands and a panchromatic band. The role of multispectral sensors is mainly in tree species identification, and detailed tree species classification is important in forest inventories for technical, ecological, and economic reasons [9]. However, the spectral sensitivities of multispectral sensors have not been optimized for forestry applications but mainly for surveying and mapping purposes. Consequently, there is greater interest in using airborne hyperspectral sensor data in research and applications. A hyperspectral sensor can capture informative data on tens to hundreds of bands, and several studies show that the use of hyperspectral data yields a reasonably accurate vegetation classification [10–16].

In Finland, 87% of the land is classified as forest land [17], and national forest inventories have been conducted since 1921 [18]. The commercially important tree species are Scots pine, Norway spruce, and broadleaf species (mainly birch), which constitute 97% of the total stand volume [17]. In Finland, forest management plans are based on attributes compiled by field work and multi-source (LiDAR and multispectral sensors) data. Several airborne multispectral sensors data have been used in tree species classification [19–21]. Few studies have examined hyperspectral data measured at ground-level in tree species classification [15, 16, 22] and airborne hyperspectral data for estimating forest stand attributes [23, 24]. Airborne hyperspectral data have also been used to investigate tree

species classification in boreal forest [12]. There is limited research work of using airborne hyperspectral data from Finland to investigate supervised tree species (pine, spruce and birch) classification. Previously, seventeen band airborne AISA imaging spectrometer data has been use for classification of vegetation and soil areas [25, 26]. Furthermore, airborne measured hyperspectral data have been used for timber volume estimation [27], classify peatland biotopes [28] and mapping forest land fertility [29].

There are several factors which disturb forest remote sensing data collected by a sensor. There is often a presence of gases, particles, and clouds between the forest and sensor system. The forest objects on the Earth's surface interact with the transmitted and scattered sunlight by absorbing or reflecting the light differently at different wavelengths. Because objects reflect light differently, they can be differentiated on the basis of their spectral signatures. However, in forest remote sensing foliage optical properties, canopy structure, the properties of the underlying ground and view-illumination geometry condition affect how vegetation reflects light [30, 31].

1.1 RESEARCH PROBLEM

Airborne hyperspectral data have been used in investigating tree species classification [10, 10, 11, 32]. However, a larger number of bands may result in high processing costs and a delay in online data transmission and communication. Likewise, a problem often noted in the study of classification using hyperspectral data [33–35] was the large number of features (bands) in hyperspectral data and the small set of training data; thus difficult to obtain reliable classification results. This phenomenon is called the Hughes effect [36]. Furthermore, current hyperspectral sensors capture lower spatial resolution images compared with those captured using multispectral sensors. In Finland, there is scarcity of proper (clear, cloud-free sky) weather conditions. Likewise, national regulations for collecting aerial images for mapping recommend that the solar elevation during the imaging campaign be 33° above the horizon [37]. This

limits the effective hours of flight campaign per day; in other words, for efficient high-altitude, high spatial resolution data acquisitions.

Efficient high-altitude, high spatial resolution data acquisition is only feasible with the use of multispectral sensors. This is an important property of multispectral devices in reducing the costs of flight campaigns. However, the available multispectral sensors are general-purpose sensors, and their few discretely located spectral band sensitivities are not optimized for tree species classification. Therefore, to improve data classification performance there is a need for the application specific optimized bands.

The research done in this work is aimed to support development of sensor that allow efficient imaging and high accurate tree species classification. Previously, the development of a programmable imaging spectrometer was discussed to change sensor spectral characteristics and the signal-to-noise ratio (SNR) to fit specific application requirements [38]. Therefore, Dell'Endice et al. [38], presented software to generate a spectral binning pattern to optimize an imaging spectrometer spectral characteristic. Similarly, it can be assumed that in the future hyperspectral sensors will be designed to tuned (define) the imaging band position in advance (a pre-flight setup) depending on the application need. Alternatively, multispectral sensors could be designed with optimized narrow and broadband sensitivities.

This thesis supports the development of efficient sensors by identifying several narrow and broadband multispectral characteristics that could be suitable for accurate tree species classification. Identification of these system was based on using several computational techniques and hyperspectral modeling data. The research problem regarding the use of airborne hyperspectral data to define optimized bands for the classification of the Scots pine (*Pinus sylvestris L.*), Norway spruce (*Picea abies(L.) H. Karst.*), and deciduous birch (*Betula pubescens Ehrh.* and B*etula pendula Roth*) tree species is addressed as follows.

1.1.1 Simulated Multispectral Sensor Responses

The evaluation of tree species classification performance using various airborne multispectral sensor data is expensive due to the cost of imaging. Using accurate spectral sensitivity information from airborne multispectral sensors and airborne hyperspectral data, multispectral sensor responses can be simulated and evaluated in classification. This approach also allows to evaluate arbitrary spectral response systems. In this thesis, using imaged airborne radiance hyperspectral data, standard (existing) and proposed optimized 4 and 5-band multispectral sensor systems, sensor responses were simulated and tree species classification performance evaluated.

1.1.2 Hyperspectral Band Selection

Feature selection methods have been used to select a subset of hyperspectral bands in data post-processing phase to reduce the hyperspectral data dimensionality. Previously, Pal [39] evaluated the band selection performance of three sparse logistic regression-based feature selection methods and Support Vector Machines Recursive Feature Elimination (SVM-REF) and suggested that the sparse logistic regression method [40] offered the best band selection results and the selected bands provided better classification results than the use of all hyperspectral bands. In this thesis, sparse regression-based feature selection methods were chosen for band selection. To our knowledge these methods have not been used for band selection for tree species classification. In the application of these methods, each regression coefficient corresponded to a hyperspectral band. Due to the property of sparseness, the regression coefficient of several bands had a value of zero and bands with a zero regression coefficient value were discarded. The remaining bands with a non-zero regression coefficient were selected. Here, band selection was performed using pixel- and plot-level datasets and balance in training samples.

Previous research has not addressed the question of whether the

selected hyperspectral bands can be realized as physical bands with multispectral sensitivities or whether the selected band positions are related to the sensitivity positions of the existing multispectral sensor sensitivities. Here, the selected hyperspectral band positions were related to the sensitivity positions of the existing multispectral sensor system and used as information to define an optimized 4 and 5-band multispectral sensor system.

1.1.3 Assessment of Selected Bands

During the flight campaign, forest atmospheric and imaging view-illumination geometry conditions constantly change and affect imaged data and pose difficulties in band selection. Likewise, it is difficult to obtain reliable ground information that is similar to all the imaging view-illumination geometry conditions. The collected ground information may only provide information on specific imaging view-illumination geometry conditions. In the context of defining optimized bands, the selected bands (obtained from specific imaging view-illumination geometry conditions data) used in tree species classification performance have to be evaluated in order to provide reasonably accurate classification results, even though the selected bands are suboptimal with respect to the view-illumination geometry conditions of the training and test datasets used in classification. In this thesis, band selection was performed using the plot-level scale (plot size 10.5 m $\times$ 10.5 m) hyperspectral reflectance data collected from the images acquired in the morning. Using the selected bands, pixel-level scale (pixel size 0.3 m $\times$ 0.3 m, and 0.5 m $\times$ 0.5 m) tree species classifications were investigated for the data imaged in the morning and afternoon.

The addressed research problems and results have been presented in scientific publications [*P1*], [*P2*] and [*P3*].

2 Passive Airborne Imaging

In passive airborne imaging, the reflected information collected by the imaging sensor in the visible to shortwave infrared wavelength range originates from the sun. Longwave infrared imaging relies on the thermal emission of the object in a scene rather than on sunlight to create an image [41]. Some of the radiometric quantities associated with a light beam used in this dissertation are irradiance, radiance, and reflectance, and we define these following [41–43].

Irradiance refers to the incident light energy per unit time per unit area on the surface, and its unit is the watt per meter square (Wm^{-2}). The irradiance per wavelength of the light is termed as spectral irradiance and a unit in nanometer is given as $Wm^{-2}nm^{-1}$. Radiance is the irradiance per solid angle of the observation or the direction of the propagation of the light. The measuring unit for the solid angle is the steradian (sr), defined as the area of the radial projection of a surface element to the surface of the sphere with radius 'r'. The unit for spectral radiance is given as $Wm^{-2}nm^{-1}sr^{-1}$. Radiance from the object surface does not distinguish between the light illuminating or the light reflected from the surface [41].

Reflectance is a quantity which characterizes the fraction of incident light reflected from an object [41]. Surface reflectance information can be used to characterize properties of an object and is useful in many spectral-based pattern recognition applications. For example, in remote sensing the atmospheric and illumination conditions affect the collected radiance data, and the reflectance information can be used in comparing images taken from different flight campaigns.

2.1 OPTICAL RADIATION MODEL

Solar irradiance that reaches the top of the atmosphere is also called exo-atmospheric solar irradiance. Some of this exo-atmospheric so-

lar irradiance transmitted through the Earth's atmosphere reaches the surface, some scatters and some is absorbed. The transmittance is governed by the Earth's atmosphere, and is a function of wavelength. The transmitted and scattered irradiance by the atmosphere interacts with the object surface and an imaging sensor senses the reflected radiance traveling through the atmosphere. Generally, the reflected radiance information sensed by the imaging sensor in solar reflective remote sensing has three significant radiation components (Fig. 2.1), the un-scattered surface reflected radiance, down-scattered surface reflected (the effect of skylight) and up-scattered path radiance [7].

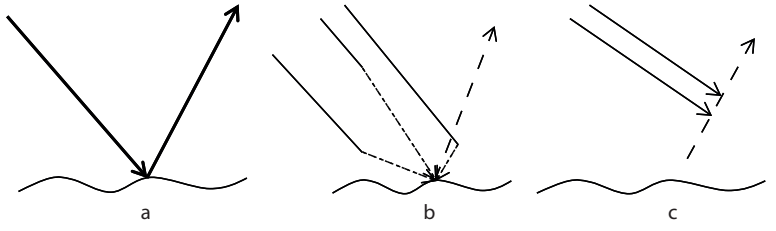


Figure 2.1: General surface reflected radiance component seen by sensor in solar reflective remote sensing a) Un-scattered b) Down-scattered and c) path-scattered. Figure adapted from [7].

When considering the Lambertian surface (perfectly diffuse reflecting surface) model, the total radiance component in the visible to shortwave infrared range sensed by the airborne imaging sensor can be presented as (2.1) [7],

$$R(\lambda) = r(\lambda)\frac{l_o(\lambda)\tau_s(\lambda)\tau_v(\lambda)}{\pi}cos(\Theta) + r(\lambda)\frac{l(\lambda)\tau_v(\lambda)}{\pi} + R^s(\lambda), \quad (2.1)$$

where $l_o(\lambda)$ is the exo-atmospheric solar irradiance, $l(\lambda)$ the irradiance at the surface due to skylight, $\tau_s(\lambda)$ the atmospheric transmittance along the solar path, $\tau_v(\lambda)$ the atmospheric transmittance along the sensor view path, $r(\lambda)$ the spectral reflectance (Lambertian) of the object, Θ the angle between the surface normal and the solar incident angle and $R^s(\lambda)$ is the path-scattered radiance

at-sensor component. Furthermore, the dependence on the spatial location is not explicitly written in the model (2.1). All objects on the Earth's surface might not have a Lambertian surface, and for a non-Lambertian surface, the term $r(\lambda)/\pi$ in (2.1) is replaced by the bi-directional reflectance distribution function of the incident and view angles [7].

With the development of optical sensing technology, different airborne optical sensors have been developed to sense the total reflected radiance component. These optical sensors capture reflected radiance in one to hundreds of spectral bands. Assuming a fixed geometry the interaction of reflected radiance $R(\lambda)$ (2.1) with a n-band sensor system can be modeled as,

$$X_i = \int_{\Lambda} R(\lambda)\tau_c(\lambda)s_i(\lambda)d\lambda,\ i = 1,\ldots,n, \tag{2.2}$$

where Λ is the wavelength range, λ the wavelength variable, X_i the spectral response of the i^{th} band, n the number of bands, $\tau_c(\lambda)$ transmittance of camera optics (lens, filter), $R(\lambda)$ the reflected spectral radiance from object surface and $s_i(\lambda)$ is the i^{th} spectral sensitivity function. Spectral sensitivity functions are positioned continuously or discretely in a given wavelength range.

2.2 PANCHROMATIC AND MULTISPECTRAL IMAGING SENSOR

The difference between a panchromatic and multispectral imaging sensor depends on the number of wavelength bands sensed in a given wavelength range. A panchromatic sensor has only one band for a given wavelength range. A multispectral sensor has a few narrow and discretely located bands in a given wavelength range. Some of the widely used multispectral sensors in forest remote sensing include the Vexcel Ultracam-D (UCD) [44], Intergraph-Z/I Digital Mapping Camera (DMC) [45] and Leica Airborne Digital Sensor (ADS) [46]. These multispectral sensors have four spectral bands and a panchromatic band. Properties of these multispectral

sensors are presented in Table 2.1. The normalized sensor sensitivities (maximum peak value 1) of these multispectral sensors are presented Fig. 2.2. The potential of these multispectral sensors in tree species classification have been investigated [3, 19–21], and 75–85% overall accuracies have been reported for Scots pine, Norway spruce and deciduous birch when using imaged data in summer. Furthermore, Holmgren, et al. [3] presented accuracy around 91% using multispectral sensor data imaged in autumn.

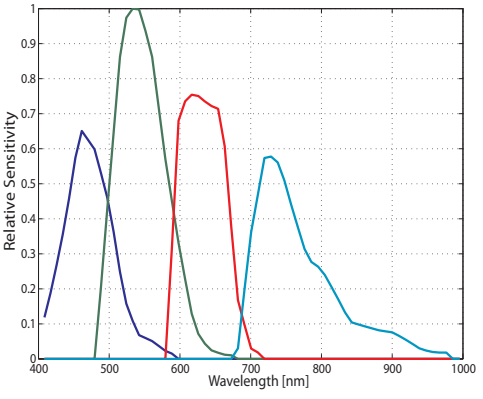
2.3 HYPERSPECTRAL IMAGING SENSOR

The hyperspectral imaging sensor sensed informative data on tens or hundreds of narrow bands. Airborne hyperspectral imaging sensors such as HyMap [49], Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [50], Compact Airborne Spectrographic Imager (CASI) [51] and the Airborne Imaging Spectrometer for Applications (AISA) [52] have been used in remote sensing and tree species classification [10, 11, 13]. However, current hyperspectral imaging sensors archived a lower spatial resolution when compared with the multispectral imaging sensor data obtained at the same altitude.

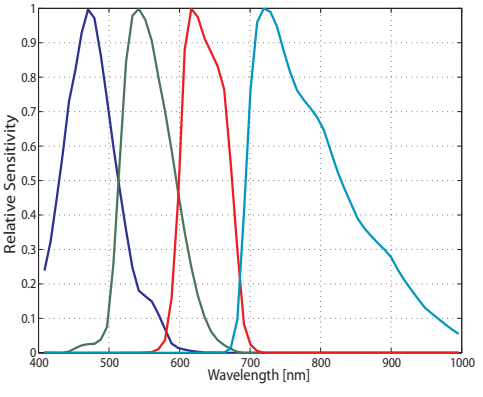
In this dissertation, the AisaEAGLE II hyperspectral sensor [53] was used for airborne measurements. AisaEAGLE II is an airborne sensor based on the pushbroom principle; it is manufactured by Specim, Spectral Imaging Ltd., Finland [53]. The sensor operates in the visible to near-infrared (VNIR) spectral range (400–1000 nm) with a 1024-pixel swath width and a 12 μm pixel size. The sensor electronics outputs using 12 bits. The minimum width of a spectral channel of the sensor is 1.2 nm, and the optimal spectral resolution of the sensor is 3.3 nm. The sensor has 516 channels with a 30 Hz sampling rate [54]. These channels can be combined to 258 (2x binning), 129 (4x) and 64 (8x) channels to obtain higher sampling rates [54]. The example of a hyperspectral cube of a forest plot (forest area 10.5 m $\times$ 10.5 m of single-tree species) is shown in Fig. 2.3, and corresponding forest plot mean radiance spectrum and estimated reflectance spectrum are presented in Fig. 2.4.

Table 2.1: Comparison of Vexcel Ultracam-D, Digital Mapping Camera and Leica ADS40 sensor/image characteristics [44–48]

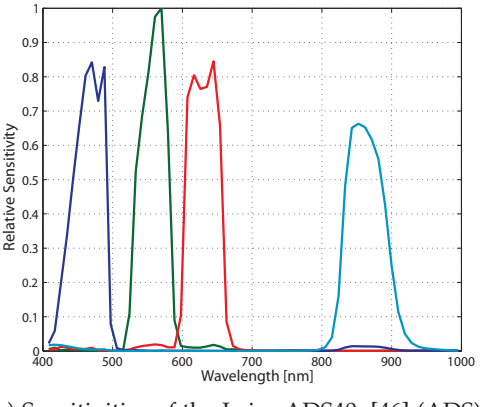
Sensor	**ULTRACAM**	**DMC**	**ADS40**
Scanning Principle	Frame grabbing	Frame grabbing	Pushbroom
Image Capture	Multi-head	Multi-head	Single-head
Smallest Ground Sample at	2.7 cm panchromatic at 1,000 feet above ground	3 cm panchromatic 1,000 feet above ground	5 cm panchromatic and multispectral at 1,500 feet above ground
Panchromatic sensor spectral resolution	380–720 nm	400–950 nm	465–680 nm
Multispectral sensor spectral resolution			
Blue	380–580 nm	400–580 nm	428–492 nm
Green	480–640 nm	500–650 nm	533–587 nm
Red	580–700 nm	590–675 nm	608–662 nm
Infrared	680–940 nm	675–850 nm	833–887 nm
Radiometric Resolution	12+ bit, 14 bit ADC, 16 bit storage	12 bit	12 bit (16 bit ADC storage)
Array Size	11,500 X 7,500 pixels (after pan/MS fusing)	13824 pixel X 7680 pixel (after pan/MS fusing)	12 lines x 12,000 pixels across track



(a) Sensitivities of the Vexcel UltraCam-D (UCD) [44].



(b) Sensitivities of the Z/I Digital Mapping camera (DMC) [45].



(c) Sensitivities of the Leica ADS40 [46] (ADS).

Figure 2.2: Spectral sensitivities of the three multispectral systems. Sensitivities are normalized to a maximum peak value of 1.

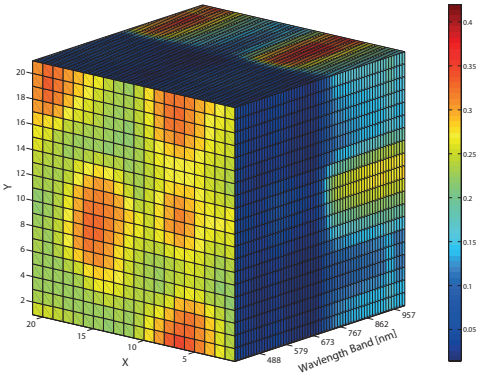


Figure 2.3: Spectral cube of a forest plot.

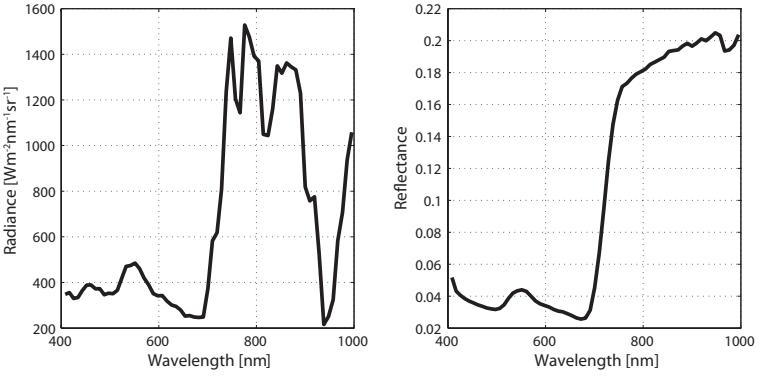


Figure 2.4: Mean at-sensor radiance and estimated reflectance spectrum.

2.4 ATMOSPHERIC CORRECTION

In airborne imaging, the components of the Earth's atmosphere (gases, particles, and clouds) scatter and absorb the light from the sun. This affect the reflected radiance spectra collected at the sensor. These effects must be corrected to obtain band selection and classifier data in the same [55].

Using the calibrated at-sensor radiance data, atmospheric and illumination effects can be reduced so that at-sensor radiance data

are transformed into reflectance data on the Earth's surface. To reduce the atmospheric and illumination effects and estimate the reflectance image in several studies different relative and absolute correction atmospheric methods have been used [7,54,56–61]. Relative and absolute correction atmospheric methods assume a reduction of the atmospheric and illumination effects that are independent of the viewing direction.

2.4.1 Absolute Correction Method

In absolute correction methods, radiative transfer codes, physically based correction methods are used to reduce the atmospheric effect and estimate the reflectance of the imaged data. Moderate Resolution Atmospheric Transmission (MODTRAN) or Second Simulation of a Satellite Signal in the Solar Spectrum (6S) [7,62,63] are two examples.

Software packages are available for absolute atmospheric correction, for example, Atmospheric Removal Program (ATREM) [64], Atmospheric and Topographic Correction (ATCOR, from ReSe Applications Schläpfer, Langeggweg 3, Switzerland), Atmospheric Correction Now (ACORN, from Analytical Imaging and Geophysics LLC, CO, USA) and Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) [65]. These programs reduce the effects of atmospheric attenuation, topographic conditions and other characteristics in an image. These correction programs utilize a physical method (e.g., MODTRAN) to model the atmospheric gas absorption, scattering effects required for data correction. The ATCOR-4 program has been previously used in atmospheric correction for AisaEAGLE II hyperspectral data [54]. The program takes radiance data and a physical parameter as the input and returns the corrected data. Similarly, the atmospheric correction performance of the three programs (ATREM, ACORN, and FLAASH) has been evaluated using the AVIRIS hyperspectral sensor data described by Kruse [66]. The author suggested that the three methods produce comparable atmospheric correction results and are quite similar in

their basics and operation [66].

2.4.2 Relative Correction Method

In addition to the absolute correction method to reduce the atmospheric effect, different relative correction methods have been used [7, 56, 57]. Applying relative correction methods for atmospheric correction are computationally less expensive than absolute correction methods. Relative correction methods are sometimes referred to as normalization techniques [7]. Some of the methods used in remote sensing studies are the following:

Internal Average Relative Reflectance (IARR): In IARR, the correction is performed so that first average spectrum of the entire image is calculated. Next, each pixel in an image is divided by the calculated average spectrum to obtain the reflectance of the image relative to the average spectrum. However, this method is not suitable for the correction of vegetation areas because the averaged spectrum may include spectral features that are related to the vegetation rather than just the effects of atmospheric and solar irradiance [57].

Flat Field Correction: In the flat field approach, the reflectance spectra are estimated so that a spectrum from each pixel in a scene is divided wavelength-wise by the mean spectrum of a known target area within the scene. The target area is assumed to be a spatially homogeneous, spectrally uniform, high reflectance area in the scene [7, 56]. The drawback of the method is that it is strongly scene-dependent [67]. Furthermore, in applying this method effect from solar irradiance and a solar path atmospheric transmittance are assumed to decrease, but the effect of view path radiance and topographic conditions still exist in the corrected data [7].

Empirical Line Method: This method assumes that there is one or more specially made calibration targets or a natural homogenous area within the image. The reflectance spectra of these targets are measured on the ground. Similarly, the radiance spectra of the targets recorded by sensors are extracted from the images. Then the radiance data over the surface targets are linearly regressed against

the ground-measured reflectance spectra in order to calculate the gain (slope) and offset (intercept) values for each band. These derived values are then applied to an image to estimate the surface reflectance [7, 58]. On applying this method, solar irradiance, solar path atmospheric transmittance and view path radiance are assumed to decrease, but the topographic effect is still present in the corrected data [7].

3 Hyperspectral Imaging Campaign

The AisaEAGLE II hyperspectral sensor [53] was used in airborne measurements over the Hyytiälä forest area in southern Finland (61.50′ N, 24.20′ E) on July 22nd, 2011, between 9:44 and 10:38 (morning) and 13:10 and 13:22 (afternoon) local time. The camera field of view at the time of measurement was 35.8°. The measurements were performed using an 8x binning mode [68], resulting in a 64 discrete channel in VNIR (400–1000 nm) (Table 3.1) with a full-width-at-half-maximum (FWHM) of approximately 9.3 nm. The sensor electronics work with 12 bits and the imaged data were stored as 16 bit unsigned integers.

3.1 REMOTE SENSING DATA

During the morning flight campaign, nine imaging strips (B1, B2, B3, B4, B5, B6a, B6b, B7, and B8) were imaged at an altitude of approximately 1000 m; these are collectively called B-Line strips (Fig. 3.1a). In addition, the B-Line strips were imaged in two flight directions. Five strips (B1, B2, B4, B6a, and B7) were imaged from southeast (SE) to northwest (NW), and four strips (B3, B5, B6b, and B8) were imaged from NW to SE. Likewise, in the afternoon, three strips (D1, D2, and D3) were imaged at an altitude of approximately 650 m; these are collectively called D-Line strips (Fig. 3.1b). This change in altitude was done to maximize the spatial resolution. Each pixel in an imaged strip from the B-Line and D-Line measured approximately 0.5 m $\times$ 0.5 m and approximately 0.3 m $\times$ 0.3 on the ground, respectively. In the D-Line, the D1 strip was imaged over a south to north flight direction; the D2 strip was imaged from northwest to southeast, and the D3 strip was imaged from north-

east to southwest. The D3 strip was partly affected by the presence of clouds. The image data acquisition details of the B- and D-Line strips are presented in Table 3.2. In the B4 and D1 strips, a 50% reflective 5 m $\times$ 5 m diffuse reference target [69] was placed on the ground.

Considering the position of the sun (Fig. 3.2a) and normal of the plane, when imaging B-Line strips and the D1 strip (D-Line), for the nadir view sensor, the solar plane is in a horizontal across-track direction and forest on either side of nadir view are equally illuminated. For D2 and D3 strips (D-Line), the solar plane is along (parallel) the across-track direction. One side of the nadir view can be highly illuminated compared with the other; this increases the within-species spectral variation.

Table 3.1: AisaEAGLE II [53] hyperspectral bands and corresponding peak wavelength (WL) value in nanometers with a full-width-at-half-maximum (FWHM) of approximately 9.3 nm.

Band	WL	Band	WL	Band	WL	Band	WL	Band	WL
1	408.39	**14**	524.20	**27**	644.58	**40**	766.61	**53**	890.52
2	417.03	**15**	533.20	**28**	653.92	**41**	776.14	**54**	900.04
3	425.67	**16**	542.20	**29**	663.26	**42**	785.68	**55**	909.57
4	434.33	**17**	551.37	**30**	672.60	**43**	795.22	**56**	919.11
5	443.24	**18**	560.69	**31**	681.95	**44**	804.76	**57**	928.67
6	452.24	**19**	570.01	**32**	691.29	**45**	814.30	**58**	938.22
7	461.23	**20**	579.33	**33**	700.65	**46**	823.84	**59**	947.78
8	470.23	**21**	588.65	**34**	710.04	**47**	833.37	**60**	957.33
9	479.23	**22**	597.97	**35**	719.42	**48**	842.89	**61**	966.89
10	488.22	**23**	607.29	**36**	728.81	**49**	852.42	**62**	976.44
11	497.22	**24**	616.61	**37**	738.19	**50**	861.94	**63**	986.00
12	506.21	**25**	625.93	**38**	747.58	**51**	871.47	**64**	995.55
13	515.21	**26**	635.25	**39**	757.07	**52**	880.99		

The digital number of the acquired images were first radiometrically corrected to the radiance using calibration coefficients provided by the manufacturer and the CaliGeo software [68] by SPECIM. Each pixel in a corrected image was further geometrically rectified into the WGS84 UTM zone 35 coordinate system using

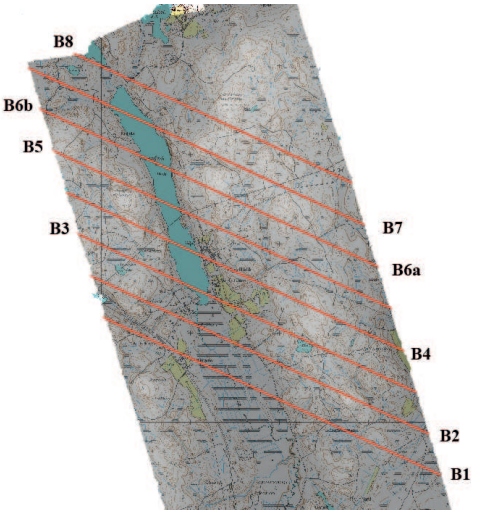
Table 3.2: Image data acquisition and field data of tree plots corresponding to the B– and D–Line image strips. F. H = Flight heading, F. Dir = Flight direction, F. A = Flight altitude, F. S = Flight speed, Az = Solar azimuth, Ev = Solar elevation, No. P. = Number of plots and GSD=Ground sampling distance.

Strip	Time UTC +3	F H [°]	F. Dir	F. A [m]	F. S [kn]	Az [°]	Ev [°]	Total No.P	No.P Pine	No.P Spruce	No.P Birch	GSD [m]
B1	9:44	297	SE–NW	952	105	111	33.5	65	46	17	2	0.5
B2	9:48	290	SE–NW	965	125	112	33.9	54	19	11	24	0.5
B3	9:59	244	NW–SE	972	105	115	35.2	49	15	12	22	0.5
B4	10:05	291	SE–NW	983	125	116	35.8	65	26	33	6	0.5
B5	10:10	245	NW–SE	949	105	117	36.3	72	31	27	14	0.5
B6a	10:21	290	SE–NW	975	120	120	37.5	54	14	14	26	0.5
B6b	10:25	244	NW–SE	975	110	121	37.9	58	18	14	26	0.5
B7	10:32	290	SE–NW	967	130	123	38.5	64	29	23	12	0.5
B8	10:38	245	NW–SE	967	105	125	39.1	96	56	26	14	0.5
D1	13:10	359	S–N	661	125	173	48.3	23	12	5	6	0.3
D2	13:15	242	NW–SE	656	115	175	48.4	13	7	3	3	0.3
D3	13:17	232	NE–SW	662	125	176	48.4	15	2	12	1	0.3

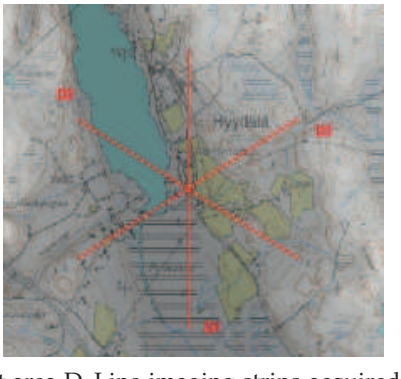
PARGE [70] software from the ReSe Company. A one-meter grid-sized digital elevation model (DEM) [71] and navigation data were used in the geometrical rectification.

3.2 FIELD DATA

The plots (i.e., the forest area) containing the trees of interest in B- and D-Line strips were identified by a photo interpretation expert who combined a visual inspection with additional ground information. The photo interpretation was based on the Vexcel UltraCamXp RGB images (pixel size approximately 15 cm) acquired on June 28th, 2010, at 16.00 local time from a flight altitude of 2.5 km. The identified forest plots contained only single-tree species (e.g., Fig. 3.2b). In the plot identification process, the mean, maximum and 95th percentile of the height distribution was estimated using LiDAR data imaged in the same forest area in 2010 and 2011. The

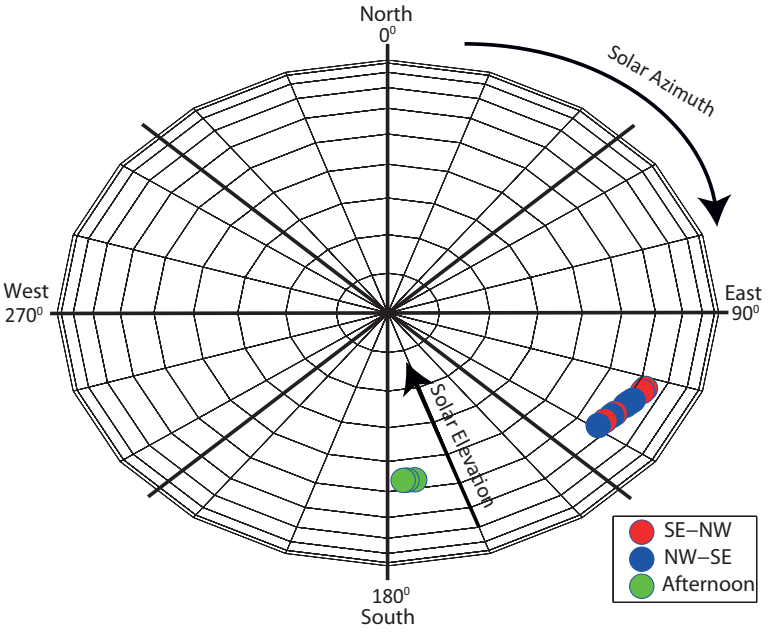


(a) Hyytiälä forest area B–Line imaging strips acquired in the morning (nine imaging strips: B1, B2, B3, B4, B5, B6a, B6b, B7 and B8). Five strips (B1, B2, B4, B6a and B7) were measured in the flight direction southeast (SE) to northwest (NW). Similarly, four strips (B3, B5, B6b and B8) were measured in the flight direction NW to SE.

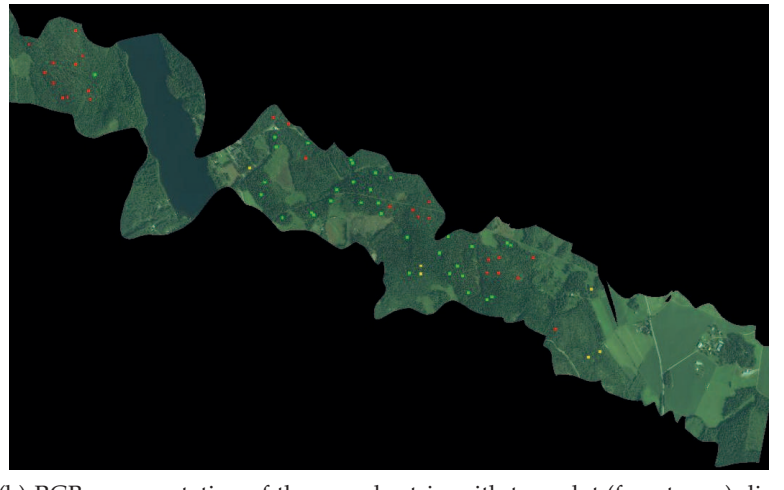


(b) Hyytiälä forest area D–Line imaging strips acquired in the afternoon. D–Line strips were imaged in three different flight directions: D1 south to north, D2 northwest to southeast and D3 northeast to southwest.

Figure 3.1: Hyytiälä forest area and imaged strip in the morning and afternoon.



(a) Polar plot of solar elevation and solar azimuth position for data imaged in the morning (Southeast (SE) to Northwest (NW) and vice versa) and afternoon.



(b) RGB representation of the sample strip with tree plot (forest area) distribution. Identified tree plots in the strip are marked with the colored boxes: Pine (red), spruce (green) and birch (yellow).

Figure 3.2: Polar plot of a solar elevation, solar azimuth and RGB representation of a sample imaged strip.

difference in the mean and maximum heights of the LiDAR points (2010 and 2011) was used to remove the identified plots where harvesting operations began after 2010. Finally, each plot was checked to determine in which hyperspectral imaged strips the plot was visible. In detail, the plot identification process is presented in [*P1* and *P3*]. Altogether, 577 plots (254 pine plots, 177 spruce plots and 146 birch plots) were identified from the B-Line and 51 plots (21 pine plots, 20 spruce plots and 10 birch plots) were identified from the D-Line imaged strips. From the identified plots, tree information was collected by drawing a 21 pixel $\times$ 21 pixel window around the plot center and tree species spectra (441 pixels) inside the window were extracted. This procedure produced an area of 10.5 m $\times$ 10.5 m (2–10 trees) for each identified forest plot in the B–line and 6.3 m $\times$ 6.3 m (2–5 trees) for those identified in the D-Line strips. Furthermore, the forest structures of the tree species plots extracted from the strips ranged from young to mature stands, where the LiDAR mean tree height varied between 2.3–24.5 m in the B-Line and 4.4–21.5 m in the D-Line strips. In addition, these extracted pixels come from sunlit and shaded regions of the vegetated and non-vegetated pixels.

3.3 DATA PREPARATION FOR EXPERIMENT

The B-Line and D-Line strips were measured at different times of the day, and the imaging and view-illumination geometry condition varied among the acquired images (see Fig. 3.2a). The extracted forest plot datasets were hyperspectral radiance data that are influenced by solar irradiance and atmospheric effects. These influences were corrected here using a 50% reflecting white reference surface placed on one strip in each B- and D-Line case (B4 strip in B-Line, D1 strip in D-Line) and the flat field correction method. We assumed that the two spectral radiance vector features from the reference targets represented the atmospheric and illumination conditions in other strips in the sets (B- and D-Line) because imaging was performed at the same flight altitude over a small geographical

area within a one hour time window.

In further processing, pixel- and plot-level scale datasets were prepared for use in the band selection and tree species classification. To prepare the plot-level dataset, the mean spectrum of the plot was chosen as the classification feature because the species information in the plot was homogeneous and was assumed to represent the plot-level spectral characteristics. The plot-level dataset was prepared using 577 mean plot spectra collected from the B–Line strips. Furthermore, this dataset was prepared with and without extracting the vegetation pixels in a plot. Vegetation pixels in a plot were extracted using Normalized Difference Vegetation Index (NDVI) [72] thresholding. In the NDVI calculation, the bands with peak wavelength value 814 nm and 691 nm were used. All pixels with an NDVI value greater than 0.7 were considered vegetation. The plot-level dataset was prepared using the hyperspectral radiance and reflectance data.

To prepare the pixel-level scale dataset, the estimated reflectance data for the identified tree species plots were first denoised as discussed in section 3.4. From the denoised dataset, vegetation pixels were extracted by applying NDVI thresholding. A dataset (*BL*) was prepared from the extracted vegetation pixels identified in tree species plots in the nine B-Line strips (Table 3.3). Subsequently, two pixel-level datasets were prepared using the extracted vegetation pixels from the tree species plots identified in the B-Line strips (Table 3.3). This step was performed to include varying view-illumination geometries conditions. In the first dataset, all vegetation pixels from the identified tree plots in the first two strips (B1 and B2) imaged from southeast to northwest, were combined and called the *BL1* dataset. The second dataset contained all vegetation pixels from the last two strips (B7, B8), measured from northwest to southeast and vice versa and was called the *BL2* dataset. Similarly, a dataset was assembled from the vegetation pixels extracted from all the identified tree species plots in the D-Line strips (Table 3.3) and called the *DL* dataset. The plot and pixel-level scale datasets used in the band selection and tree species classification are summarized

in Table 3.4.

3.4 NOISE REMOVAL

Noise is present in the airborne hyperspectral data from different sources. Different methods have been presented to denoise hyperspectral data [73–76]. In the thesis, noise in the reflectance data was estimated using the hyperspectral signal identification by minimum error (HySime) method presented in [74]. This noise estimation was based on the correlation between bands and the multiple regression theory [77]. For the algorithm to work noise type needs to be assumed. Here, assuming additive noise (zero-mean gaussian independent and identically distributed), a noise signal was estimated. Denoising was performed by subtracting the estimated noise from the data. The same denoising technique has been used in [78] before applying different band selection methods. From the observed result, Latorre-Carmona, et al. [78] suggested that this denoising reduces regression error and improves band selection performance in choosing a band selection method where the selection criterion uses the information given by the whole dataset rather than a distance criterion between sample neighbors. In this thesis, band selection and classification results based on this denoised dataset are presented in [*P2*] and [*P3*].

3.5 REDUCTION OF VIEW-ILLUMINATION GEOMETRY CONDITION EFFECT

The datasets (*BL1*, *BL2*, and *DL*) used in this thesis belong to different imaging view-illumination geometry conditions. Tree crown reflectance is dependent on the view-illumination conditions, and the method used (flat field) for atmospheric correction assumes a reduction in atmospheric and solar irradiance effects that is independent of the viewing direction. Generally, a surface can be characterized with the Bidirectional Reflectance Distribution Function (BRDF) model [79]. Accurate modeling and correcting the BRDF

effect on the forest canopy surface to a known accuracy is difficult [80]. Previous studies used a normalization process to reduce illumination effects so that each pixel (spectral vector) was divided by its L_1-norm for field-measured hyperspectral reflectance [81] and airborne-measured hyperspectral radiance data [12] to obtain a unit length vector. Furthermore, to reduce view-illumination geometry condition effect on airborne multispectral reflectance data Heikkinen, et al. [20] divided each pixel by its L_2-norm. In this thesis, it is assumed that imaging from different view-illumination geometry conditions caused scale differences, and the estimated reflectance is normalized (the spectral vector was divided by its L_2 norm) to obtain a unit length vector. The use of reflectance and normalized reflectance data for tree species classification was studied in [*P3*].

Table 3.3: Number of tree species pixels in the datasets, *BL*, *BL1*, *BL2* and *DL* the LiDAR plot mean tree height range.

Dataset	**Mean Height [m]**	**Pine**	**Spruce**	**Birch**
BL	2.3–24.5	80,959	54,945	64,272
BL1	4.1–24.5	16,064	8,298	11,441
BL2	6.8–19.7	28,321	15,588	11,435
DL	4.4–21.5	5,853	7,942	4,407

Table 3.4: Summary of dataset used in classification and band selection.

Study	Scale	Dataset	Remarks	Publication
Tree species Classification	Plot	577 Plot mean spectra	Radiance data with / without vegetation pixel extraction	*P1,P2*
	Pixel	*BL*, *BL1*, *BL2* and *DL*	Reflectance data with vegetation pixel extraction	*P2, P3*
Band selection	Plot	577 Plot mean spectra	Reflectance data with vegetation pixel extraction	*P2, P3*
	Pixel	*BL*	Reflectance data with vegetation pixel extraction	*P2*

4 Hyperspectral Band Selection

A hyperspectral sensor captures information via tens or hundreds of spectral bands. However, previous studies [33,77] have suggested that it is difficult to obtain reliable classification results when there are a large number of available features (bands) and a small set of training data. This phenomenon is termed the Hughes effect [36]. Thus, when using hyperspectral data in classification, a reduction in dimensionality must be considered. To reduce hyperspectral data dimensionality, feature extraction [16, 74, 77, 82, 83] and feature selection [11, 12, 35, 39, 78] methods have been used. In feature extraction, hyperspectral data are mapped to a lower dimensional space to compute new features [15, 16, 77].

Likewise, in the feature selection approach, a subset of original features is identified which is useful for separating the classes (objects) and reducing data dimensionality [35, 84]. In hyperspectral band selection, feature selection is used to select a subset of bands. However, in studies using the subset of hyperspectral bands for data classification, there has been no discussion on whether the selected hyperspectral bands could be realized as physical multispectral bands or if the selected band positions had any relation to the band positions of the existing multispectral sensitivity systems. In this thesis, it is assumed that discretely selected band position could be used as a tool to define optimized multispectral sensor sensitivity or considered as an optimized band.

The feature selection methods have been categorized as filter, wrapper and embedded methods based on whether the method used a classification algorithm to evaluate a generated subset of features [35, 84, 85]. The filter approach was used as a post-processing step to select a feature, and the method did not use a classification algorithm to evaluate selected features. This method maxi-

mizes an evaluation function and uses a search criterion to choose a subset of features. The wrapper method utilizes the feature selection algorithm as a black box. This approach finds the score for possible subsets of features according to their data discriminative power [86] and the best performing subset of features is output. In the embedded method, feature selection is a part of the classification process, and feature selection and classification cannot be separated [35,84,85].

In hyperspectral remote sensing to select a subset of bands different filter [10,35,78,87], wrapper [35,39,78] and embedded [35,39] feature selection methods can be used. Similarly, for band selection, the analysis of second-derivative spectra [87,88], interclass distances [10, 78], correlation coefficients [35, 78, 87], information theory measures [78, 87], regression-based methods [39, 78] and pattern classifier methods [35] have been used. Previously, Pal [39] investigated the band selection performance of three sparse logistic regression-based methods and indicated that the sparse logistic regression-based method of Cawley and Talbot [40] gives the best band selection results. Furthermore, this showed that the selected bands provided improved classification results than by using all bands and bands selected by other methods. To our knowledge, this method has not been previously used for the band selection in tree species classification. In this dissertation, the sparse logistic regression method [40] and two sparse regression-based feature selection methods were used in band selection. The sparse regression approach allowed us to obtain a sparse representation of the regression model via an L_1-penalty term.

For a given training data of size m as the input, $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots (\mathbf{x}_m, y_m)\} \subset \mathbb{R}^p \times \mathbb{R}$, with $\mathbf{x}_i \in \mathbb{R}^p$ and y_i is the response of the i^{th} observation, the following minimization problem is solved:

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}}\{\sum_{i=1}^{m} \mathcal{L}(y_i, \mathbf{x}_i^T \beta) + \gamma \parallel \beta \parallel_1\}, \tag{4.1}$$

where

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

and

$$\| \beta \|_1 = \sum_{j=1}^{p} |\beta_j|, \tag{4.2}$$

$\beta_j = (\beta_1, \beta_2, \ldots, \beta_p)$ represents the regression coefficient to be estimated. In model (4.1) the first term $\mathcal{L}(y_i, f(\mathbf{x}_i))$ is a loss function which measures the fit of the function to the given training data. The second term $\| \beta \|_1$ is the L_1-norm penalty (4.1) of the regression coefficient vector, and the term $\gamma \| \beta \|_1$ is called the regularization term, in which γ is the regularization parameter that controls the strength of the L_1-norm penalty. This penalization shrinks some coefficients to a value of zero, resulting in the sparse representation of a regression model.

We related the regression coefficient to a hyperspectral band. Due to the properties of the sparseness, several band regression coefficients with a value of zero were discarded. The remaining bands with non-zero regression coefficients were considered as selected features or bands.

4.1 SPARSE LINEAR REGRESSION

In general, the linear regression model for a given training set (S) of size m, where $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots (\mathbf{x}_m, y_m)\} \subset \mathbb{R}^p \times \mathbb{R}$ with $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \{0, 1, 2\}$ a response of the i^{th} sample is given as

$$y_i = \mathbf{x}_i^T \beta + \varepsilon \text{ where } E[\varepsilon] = 0 \tag{4.3}$$

The regression coefficients are often estimated using the least squares, in which the regression coefficients are selected to minimize the squared error loss,

$$SE = \sum_{i=1}^{m} (y_i - \mathbf{x}_i^T \beta)^2. \tag{4.4}$$

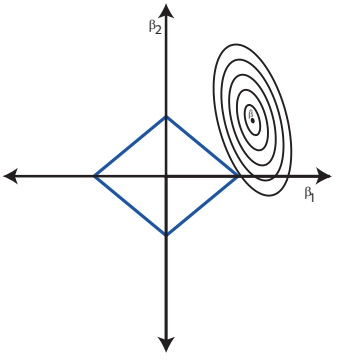


Figure 4.1: Estimation picture of lasso contour for error and constrain function. The area inside the diamond is the constraint regions $|\beta_1| + |\beta_2| \leq t$, and the ellipse are the contours of the least squares error function [90].

When solving for sparse linear regression, the loss function $\mathcal{L}(y_i, \mathbf{x}_i^T\beta)$ in (4.1) is replaced by the squared error loss (4.4), and the minimization problem (4.1) is given as,

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}}\{\sum_{i=1}^{m}(y_i - \mathbf{x}_i^T\beta)^2 + \gamma \parallel \beta \parallel_1\} \tag{4.5}$$

and $\widehat{\beta_j} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is the estimated regression coefficient vector.

This model formulation (4.5) is called LASSO (Least Absolute Selection and Shrinking Operator) [89, 90], where the L_1-norm penalty is added to the linear regression problem. The estimated regression coefficients are constrained ($\sum_{j=1}^{p} |\beta_j| \leq t$) [90] so that the coefficient vectors of the L_1-norm penalty lies in a specific geometric shape centered on the origin (see Fig. 4.1). Due to constraint solving (4.5) some of the estimated regression coefficients become zero, resulting in a sparse solution.

4.2 SPARSE LOGISTIC REGRESSION

In a two-class (C_1, C_2) problem, $\mathbf{x}_i \in \mathbb{R}^p$ is a vector of measurements and $y_i \in \{0, 1\}$ is associated with the binary class level of the i^{th} sample. The logistic regression model attempts to estimate the posterior probability of the class membership based on a linear combination of the input features and is given as,

$$p(y = 1|\mathbf{x}_i; \beta) = \frac{1}{1 + \exp\{-\mathbf{x}_i^T \beta\}}, \tag{4.6}$$

where $p(y = 1|\mathbf{x}; \beta)$ is the conditional probability that input sample $\mathbf{x}_i$ is in class C_1. The conditional probability that input sample $\mathbf{x}_i$ belongs to class C_2 is given as,

$$\begin{aligned} p(y = 0|\mathbf{x}_i; \beta) &= 1 - \frac{1}{1 + \exp\{-\mathbf{x}_i^T \beta\}} \\ &= \frac{\exp\{-\mathbf{x}_i^T \beta\}}{1 + \exp\{-\mathbf{x}_i^T \beta\}} \\ &= \frac{1}{1 + \exp\{\mathbf{x}_i^T \beta\}}. \end{aligned} \tag{4.7}$$

In a logistic regression, the regression coefficient vector β is estimated using the likelihood function [91–94]. Let us suppose a training set (S) of size m, where $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots (\mathbf{x}_m, y_m)\} \subset \mathbb{R}^p \times \{0, 1\}$ where $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \{0, 1\}$ is a response of the i^{th} observation. If the training set S represents an independent and identically distributed (i.i.d.) sample of a Bernoulli distribution, the likelihood function is given as,

$$\mathcal{L}(y_i|x_i; \beta) = \prod_{i=1}^{m} \pi(x_i)^{y_i} \{1 - \pi(x_i)\}^{1-y_i}, \tag{4.8}$$

where $y_i = (y_1, \ldots, y_m)^T$ and $\pi(x_i) = p(y = 1|x_i; \beta)$.

The regression coefficient vector β can be estimated by maximizing the log-likelihood of the training sample or its equivalent by minimizing the negative log-likelihood [40, 92, 93]. In this study,

the negative log-likelihood was minimized and is given as,

$$\begin{aligned}\ell &= -\log\, \mathcal{L}(y_i|x_i;\beta) \\ &= -\sum_{i=1}^{m}\{y_i \log\, \pi(x_i) + (1-y_i)\log\,(1-\pi(x_i)\}; \end{aligned} \tag{4.9}$$

this can be transformed into

$$\ell = \sum_{i=1}^{m} \log\{1 + exp(-y_i' \mathbf{x}_i^T \beta)\}, \tag{4.10}$$

where

$$y_i' = \begin{cases} 1 & \text{if } y_i = 1 \\ -1 & \text{if } y_i = 0 \end{cases}$$

In the logistic regression, due to the nonlinearity of the logistic sigmoid function, there is no closed-form solution [94]. A Newton method can be used to compute the regression coefficient vector beta [94]. In the Newton method, the first and second derivatives of an objective function are computed. Thus, after minimizing the negative log-likelihood (4.10) with respect to individual regression parameters, its first and second derivatives are computed to find the solution.

This logistic regression model solution is dense, and the regression coefficient β is never exactly zero. Thus, selection of the most informative regression coefficient is difficult. To solve this problem, a sparse logistic regression model [90, 95] can be formulated by adding the regularization term to the negative log-likelihood function (4.10) and the modified model is presented as follows,

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}}(\ell + \gamma \parallel \beta \parallel_1). \tag{4.11}$$

The addition of a regularization term (4.11) to the negative log-likelihood (4.10) can also be seen as the placement of a Laplacian prior [96] over the regression coefficient vector β. To solve (4.11), different methods have been proposed [93, 95, 97–99].

In (4.5) and (4.11), the regularization parameter γ controls the strength of the penalty. For $\gamma = 0$, $\widehat{\beta}$ is the same as solving for

the linear regression least square fit in (4.5) and the negative log-likelihood in the case of logistic regression (4.11). Similarly, when $\gamma = \infty$, then $\widehat{\beta} = 0$. Thus, when solving (4.5) and (4.11) for $\widehat{\beta}$ (regression coefficient) the regularization parameter γ is selected between these two extremes.

The addition of L_1-norm penalization raises two issues concerning how the objective function can be minimized and how to choose the free regularization parameter γ, because the selection of a different γ provides a different model solution [90, 95]. To solve the problem of model selection, a cross-validation routine can be applied to select a model solution with a γ value that provides a minimum error. In this thesis, the solution to the objective function (4.5 and 4.11) was found using the coordinate descent approach presented in [97, 98]. From the solutions, such a model solution was selected where the γ value provides the minimum error.

4.3 SPARSE LOGISTIC REGRESSION WITH BAYESIAN REGULARIZATION

In (4.5) and (4.11), the regularization parameter γ controls the results and needs to find the best γ value, which minimizes error. Thus, there is a need for a model selection stage to find the best γ. Cawley and Talbot [40] discussed the elimination of the model selection stage in the sparse logistic regression and introduced a Bayesian approach in which the need for a regularization parameter γ in (4.11) was eliminated. They stated that minimization of (4.11) provided a Bayesian interpretation, and the posterior distribution of the logistic regression coefficient (β) was given as,

$$p(\beta|S,\gamma) \propto p(S|\beta)p(\beta|\gamma), \tag{4.12}$$

where S is the training sample. The prior over the logistic regression coefficient (β) is given by the separable Laplace distribution as,

$$p(\beta|\gamma) = (\frac{\gamma}{2})^N \exp\{-\gamma \parallel \beta \parallel_1\}, \tag{4.13}$$

where N is the number of the non-zero logistic regression coefficient. According to Cawley and Talbot [40] the regularization parameter was eliminated using an analytical approach [96] in which the prior distribution over the logistic regression coefficient is given by marginalization over γ

$$p(\beta) = \int p(\beta|\gamma)p(\gamma)d\gamma. \tag{4.14}$$

Because γ is a scale parameter, Cawley and Talbot [40] stated that the improper Jeffrey's prior [100] $p(\gamma) \propto 1/\gamma$ provides the appropriate ignorance prior, corresponding to the uniform prior over the $log\gamma$. Substituting (4.13) in (4.14) and noting that γ is strictly positive we obtain,

$$p(\beta) = (\frac{1}{2})^N \int_0^\infty \gamma^{N-1} \exp\{-\gamma \parallel \beta \parallel_1\} d\gamma. \tag{4.15}$$

Using the Gamma integral $\int_0^\infty x^{\nu-1}e^{-\mu x}dx = \frac{\Gamma(\nu)}{\mu^\nu}$ [101] Cawley and Talbot [40] showed

$$p(\beta) = \frac{1}{2^N}\frac{\Gamma(N)}{(\parallel \beta \parallel_1)^N} \Rightarrow -logp(\beta) \propto Nlog \parallel \beta \parallel_1 . \tag{4.16}$$

This calculation gives the revised optimization function for the sparse logistic regression (4.11) with Bayesian regularization

$$\widehat{\beta_r} = \underset{\beta}{\operatorname{argmin}}\{\ell + Nlog \parallel \beta \parallel_1\}. \tag{4.17}$$

Furthermore, they [40] state that the differentiation of (4.11) and (4.17) gives

$$\nabla\widehat{\beta} = \nabla\ell + \gamma\nabla \parallel \beta \parallel_1 \text{ and } \nabla\widehat{\beta}_r = \nabla\ell + \tilde{\gamma}\nabla \parallel \beta \parallel_1, \tag{4.18}$$

where

$$\tilde{\gamma} = \frac{1}{1/N\sum_{j=1}^N |\beta_j|}. \tag{4.19}$$

From the gradient descent viewpoint, the minimization of (4.17) effectively becomes equivalent to the minimization of (4.11) so that

the regularization parameter in (4.11) is constantly modified as given in (4.19). This procedure eliminates the need for a user-defined regularization parameter. The details of the method can be found in [40]. To solve the modified sparse logistic regression with Bayesian regularization, Cawley and Talbot [40] introduced minor modifications into the sparse logistic regression solution described by Shevade et al. [95], who used the Gauss-Seidel method.

5 Classifiers

In the classification of airborne multispectral and hyperspectral data, several classifiers have been used [3,9–11,20,21,34,35,102]. For example, discriminant analysis (DA), maximum likelihood (ML), spectral angle mapper (SAM), support vector machine (SVM), Radial basis function (RBF) neural networks, K-nn classifier. Among the classifiers, Melgani and Bruzzone, and Camps-Valls and Bruzzone [34, 102] suggested that SVM is a promising classifier for remote sensing application as it is robust in regard to noise and high data dimension and less sensitive to the Hughes phenomenon. Furthermore, in several studies [10,34,102–104] improved classification results with SVM classifiers have been presented when compared with the classification results obtained from other classifiers.

In this dissertation, tree species classification was studied using reflectance data corresponding to the selected hyperspectral bands, simulated responses of standard and optimized multispectral sensors and full spectral reflectance features corresponding to the 64 AisaEAGLE II hyperspectral bands. Tree species classification was evaluated using the DA and SVM classifiers. The DA classifier was chosen for the experiments due to its simplicity in construction.

5.1 DISCRIMINANT ANALYSIS

Discriminant analysis is a supervised classifier that has been widely used to solve the machine learning problem for object classification. Linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) are widely used discriminant analysis methods.

Let us assume we have a multi-class problem and for class label k, where $k = 1, 2, \ldots, k$, the Bayes rule minimizes the total error of the classification by assigning each object to class k, which has a height conditional probability [91, 94, 105]. To estimate this, we can

follow the idea of discriminant analysis. If we have many classes and an input training data of size m, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_m\} \subset \mathbb{R}^p$ with $\mathbf{x}_i \in \mathbb{R}^p$ and assuming that the data within each class follow a normal distribution, then the conditional probability is given as

$$p(\mathbf{x}|k) = \frac{1}{(2\pi|\Sigma_k|)^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}-\mu_k)^T \Sigma_k^{-1}(\mathbf{x}-\mu_k)\}, \qquad (5.1)$$

where μ_k and Σ_k are the mean vector and covariance matrix, respectively, of class k. Taking the logarithm of (5.1) for differing prior class probabilities π_k, the optimal discriminant analysis classification rule is given as

$$f_x(k) = \max_{k=1,2..K} \{(\mathbf{x}-\mu_k)^T \Sigma_k^{-1}(\mathbf{x}-\mu_k) - 2log(\pi_k) + log(|\Sigma_k|)\}, \quad (5.2)$$

where $\mu_k = \frac{1}{m_k}\sum_{i=1}^{m_k} \mathbf{x}_k$ and $\pi_k = \frac{m_k}{m}$, m_k is the number of training data in each class and m the total number of training data. The covariance matrix of m-samples in each class k is defined as

$$\Sigma_k = \frac{1}{m_k} \sum (\mathbf{x}_k - \mu_k)(\mathbf{x}_k - \mu_k)^T. \qquad (5.3)$$

The basic difference between the two classifiers LDA and QDA is how the covariance information is obtained. In LDA $\Sigma_k = \Sigma \ \forall \ k$, Σ is covariance information computed combining the training samples in all classes and the linear discriminant function is defined as

$$\delta_k(x) = \mathbf{x}^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + log(\pi_k). \qquad (5.4)$$

In QDA, the covariance matrix Σ_k is calculated using training samples for each class k.

One of the problems in using these classifiers (LDA and QDA) for hyperspectral data is that when estimating the covariance matrices, it is possible that the covariance matrix is ill-conditioned due to the high-dimensionality of the data and, therefore, difficult to use in (5.4).

5.2 SUPPORT VECTOR MACHINE

The support vector machine (SVM) is also a supervised classifier and has been used to classify remote sensing data [10, 16, 20, 33–35]. The SVM detects the separating hyperplane by maximizing the margin between the classes [106, 107].

Let us assume we have a binary classification problem and a training data of size m, $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_m, y_m)\} \subset \mathbb{R}^p \times \mathbb{R}$ with $\mathbf{x}_i \in \mathbb{R}^p$ and response variable $y_i \in \{0, 1\}$). In the SVM framework the data are mapped to some feature space $\mathcal{F}$ with a feature map

$$\Phi : \mathbb{R}^p \to \mathcal{F} \tag{5.5}$$

and the decision function is written as

$$f(x) = \text{sign}(\mathbf{w}^{\mathrm{T}}\Phi(\mathbf{x}) + b), \tag{5.6}$$

where b is a bias term and $(\mathbf{w}^T\Phi(\mathbf{x}) + b) = 0$ defines the hyperplane in the feature space. Then $f(\mathbf{x}_i) \geq 1$ when $y_i = 1$ and $f(\mathbf{x}_i) \leq 0$ when $y_i = 0$.

Assuming that the two classes are not separable in the feature space, in the SVM framework the classification model is derived as the solution to the minimization problem. As a minimization problem, we used the C–SVM and least-squares support vector machine (LS–SVM) formulation of SVM.

5.2.1 C–SVM

C-SVM solves the nonlinear classification problem using quadratic programming. In C-SVM, the minimization problem is given as,

$$\begin{cases} \min\limits_{w,b,\zeta} J(w, \zeta) = \dfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum\limits_{i=1}^{m}\zeta_i & \\ \text{s.t. } y_i[\mathbf{w}^T\Phi(\mathbf{x}_i) + b] = 1 - \zeta_i & i = 1, \ldots, m \\ \text{and } \zeta_i \geq 0 & i = 1, \ldots, m \end{cases} \tag{5.7}$$

where the term $\mathbf{w}^T\mathbf{w}/2$ corresponds to the margin parameter and ζ_i^m represents the slack variable. This indicates the misclassification

of the sample $\mathbf{x}_i$ when $\zeta_i > 1$ [16]. The parameter C controls the tradeoff between the margin maximization and the tolerable classification errors. The solution is obtained by solving the quadratic programming problem using the dual space of Lagrange multipliers and the property

$$\mathcal{K}(\mathbf{x}, \mathbf{z}) = \Phi(\mathbf{x})^T \Phi(\mathbf{z}), \tag{5.8}$$

where kernel $\mathcal{K}$ defines the mapping $\Phi : \mathbb{R}^p \rightarrow \mathcal{F}$ of input samples $\mathbf{x}$ and $\mathbf{z} \in \mathbb{R}^p$ to the feature space $\mathcal{F}$ [106, 107]. The decision function for the C-SVM becomes

$$f(\mathbf{x}) = \text{sign}\left[\sum_{i=1}^{N_s} \alpha_i y_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + b\right], \tag{5.9}$$

where N_s is the number of the support vector. The training data corresponding to the non-zero α_i is called the support vector [107], α_i represents the calculated Lagrange multipliers and $\mathcal{K}$ is the selected positive definite kernel function [106, 107].

5.2.2 Least Squares Support Vector Machine (LS–SVM)

In simplifying the SVM formulation LS–SVM as a least-squares cost function has been proposed [107, 108]. The formulation of LS–SVM is given as a minimization problem [107, 108] and is defined as follows:

$$\begin{cases} \min\limits_{w,b,e} J(w, e) = \dfrac{1}{2}\mathbf{w}^T\mathbf{w} + \gamma\dfrac{1}{2}\displaystyle\sum_{i=1}^{m} e_i^2 \\ \text{s.t. } y_i[\mathbf{w}^T\Phi(\mathbf{x}_i) + b] = 1 - e_i, \qquad i = 1, \ldots, m \end{cases} \tag{5.10}$$

where $\mathbf{w}^T\mathbf{w}/2$ corresponds to the margin between classes, (w b) is the parameter of the linear approximation, $\gamma > 0$ a regularization parameter similar to parameter C in (5.7) and e_i is the error of the i^{th} sample. This formulation (5.10) consists of equality rather than inequality constraints in (5.7) and takes into account a squared error with a regularization term similar to the ridge regression [107].

To solve the optimization problem in (5.10), a Lagrangian function can be constructed [107].

The solution to the formulation (5.10) is obtained by solving the linear set of equations using the dual space of Lagrange multipliers and the property $\mathcal{K}(\mathbf{x}, \mathbf{z}) = \Phi(\mathbf{x})^T \Phi(\mathbf{z})$, where kernel $\mathcal{K}$ defines the mapping $\Phi : \mathbb{R}^p \rightarrow \mathcal{F}$ of input samples $\mathbf{x}$ and $\mathbf{z} \in \mathbb{R}^p$ to the feature space $\mathcal{F}$. The LS–SVM classifier is then constructed as follows:

$$f(\mathbf{x}) = \text{sign}\left[\sum_{i=1}^{N} \alpha_i y_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + b\right], \tag{5.11}$$

where N is the number of the support vector, α_i represents the calculated Lagrange multipliers and $\mathcal{K}$ is the selected positive definite kernel function [107].

LS–SVM has drawbacks when compared with C–SVM. C–SVM leads to a sparse representation, i.e., the support vectors are a subset of the original training sample. In the case of LS–SVM every training data point is a support vector because none of the α_i values are exactly zero [107].

In this thesis, LS–SVM and C–SVM were applied via a feature map a radial basis function (RBF) kernel defined as $\mathcal{K}(\mathbf{x}, \mathbf{z}) = \exp(- \parallel \mathbf{x} - \mathbf{z} \parallel_2^2 / \sigma^2)$ for $\sigma > 0$. The hyperparameter C or γ and the kernel parameter σ were estimated with the selected training dataset and a 10-fold cross validation. The three tree species (pine, spruce, birch) classification was conducted using the multiclass one-against-one method [109].

6 Experiments

In this thesis, Scots pine, Norway spruce and deciduous birch tree species classifications were evaluated using simulated responses and selected hyperspectral bands. Sensor responses were simulated using measured hyperspectral data, the spectral sensitivity of three existing multispectral sensors and two optimized multispectral sensor system. In band selection estimated reflectance data were used assuming they represent the spectral signature of the tree species since single-tree species forest plots were extracted. These classification results were compared with the results obtained by using all 64 band AisaEAGLE II hyperspectral radiance and reflectance data, depending on the use of the simulated response or selected bands. The tree species classification was investigated at pixel- and plot-level scales. In the plot-level scale Leave-one-out (LOO) classification was evaluated together with a case in which the training and test data were selected from different view direction datasets. Using the selected bands obtained from specific view-illumination geometry condition datasets, the pixel-level scale tree species classification was studied in cases where the view-illumination geometry conditions of the datasets used in band selection and classification either closely match or deviate.

6.1 HYPERSPECTRAL BAND SELECTION

Three different sparse regression-based feature selection methods (sparse linear regression (SLinR), Sparse Logistic Regression (SLogR) and Sparse Logistic Regression with Bayesian Regularization (SLogBR)) were used to select subset of bands. The band selection results for the use of these three methods are presented in [*P3*] using the balanced plot mean dataset (300 i.e. 100×3 plot mean spectra).

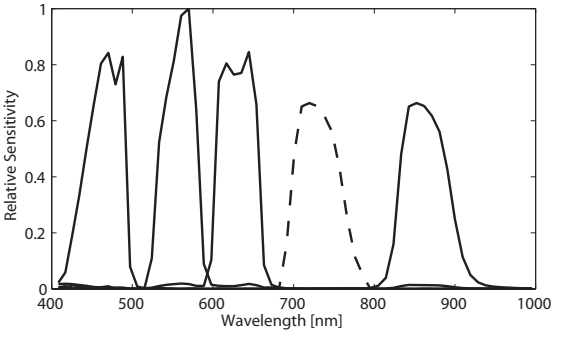
Using SLinR (4.5) 39 spectral bands were selected. Similarly,

using SLogR (4.11) 17 spectral bands were selected. When using SLogBR, the minimum number of bands was selected. An additional advantage of the method is that it avoids the model selection stage. The band selection performance of SLogBR was further evaluated using pixel- and plot-level datasets in the case of an imbalanced and balanced dataset. Using SLogBR with plot- and pixel-level datasets in balanced and imbalanced cases resulted in the selection of 8–11 narrow bands. The detailed band selection results for SLogBR are presented in [*P2*].

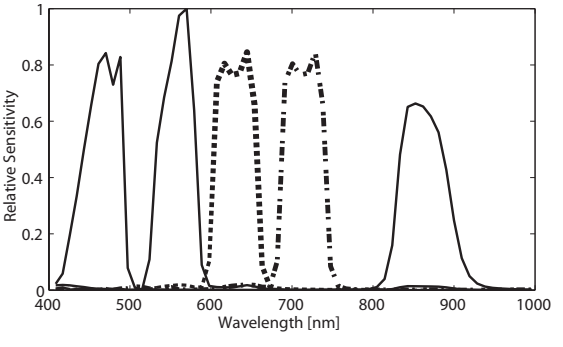
6.2 OPTIMIZED MULTISPECTRAL SENSOR SENSITIVITIES

Existing airborne multispectral sensors are general purpose remote sensing sensors and discretely located sensor sensitivities are not optimized for tree species classification. In this thesis, optimized 4 and 5-band multispectral sensor sensitivities were proposed either by repositioning the existing sensitivity or adding an extra sensitivity function in the standard 4-band Leica ADS system. The Leica ADS system was selected for the band modification because it lacks sensitivity in the wavelength range of 690–800 nm (red-edge), where the NIR band in UCD and DMC is extended to encompass the red-edge and NIR range (see Fig. 2.2 on page 12). Although the Leica ADS system (Fig. 2.2c on page 12) lacks sensitivity in the red-edge range, it has been reported that certain Leica systems exhibit an additional band in the wavelength range of 705–755 nm [46]; however, that sensitivity information was unavailable. In optimization, the NIR sensitivity function in the Leica ADS sensor was first replicated and relocated to achieve sensitivity in the 691–785 nm range so that there was no overlap with other existing Leica ADS bands (see Fig. 6.1a). This new sensitivity has the same FWHM as the NIR sensitivity in the Leica ADS system and a peak value at 719 nm. This proposed 5-band multispectral sensor system is referred to as ADS*. The motivation for this approach was based on simulation work presented in [16], which was based on the hyperspectral data measure on the ground and idealized Leica ADS sensitivities,

where the importance of the red-edge band for the classification of tree species has been presented. Similarly, the importance of the red-edge band has been emphasized in previous band selection studies on utilizing hyperspectral data for tree species classification [6, 10, 11].



(a) The five band system *ADS** proposed in [*P1*]. The four solid lines are the Leica ADS40 bands and the dashed line is the added fifth band.



(b) The dotted curve is the ADS40 red band and the dashed dot curve is the repositioned or added band of the proposed 4 and 5-band multispectral system *ADS-S* and *ADS-S**, respectively, proposed in [*P2*].

Figure 6.1: Repositioned/added band in the Leica ADS40 system.

Furthermore, the selected band position obtained using a sparse logistic regression with the Bayesian regularization method [40] was used to modify one Leica ADS multispectral band with broadband characteristics. To achieve this goal, the positions of the selected narrow bands were related to the Leica multispectral sensitivity positions. The selected red-edge bands were positioned in locations

where the Leica ADS lacked sensitivity (see figures [*P2*]). Based on this finding, it was assumed that the positions of the selected bands could be useful in defining new sensitivity or repositioning the existing sensitivity in the Leica ADS. Using the position information of the selected narrow bands, the Leica ADS sensitivity was modified in two ways to define an optimized multispectral sensor system. First, the sensitivity function of the red band of Leica ADS was repositioned for effective sensitivity starting from the position of 681 nm (see Fig. 6.1b, dashed dot curve). This repositioning was based on the position of the selected narrow bands in the red-edge range of 681–776 nm. This proposed optimized 4-band multispectral sensor system was referred to as ADS-S. Second, extra sensitivity was added to the Leica ADS system rather than redefining the location of the existing sensitivity. To achieve this goal, a 5-band multispectral sensor system was defined where the four original Leica ADS sensitivity functions and an additional sensitivity function was defined (see Fig. 6.1b). This additional sensitivity function represented the same repositioned sensitivity function as in ADS-S. The proposed 5-band multispectral sensor system was referred to as ADS-S*.

6.3 SIMULATION OF SENSOR RESPONSES

The evaluation of tree species classification performance with different airborne multispectral sensor data is expensive due to the cost of imaging. Using airborne radiance hyperspectral data and accurate multispectral sensor sensitivity information, multispectral sensor responses can be simulated and tree species classification performance can be evaluated. Previously, using simulated responses tree species classification has been studied using hyperspectral data (tree species reflectance and midday condition solar irradiance) measured on the ground, with idealized Leica ADS80 sensitivity [16].

For the purposes of this study, the multispectral sensor responses were simulated using airborne measured hyperspectral

(AisaEAGLE II) radiance data and accurate sensor sensitivity information for three airborne multispectral sensors, namely, Vexcel Ultracam-D (UCD) [44], the Intergraph-Z/I Digital Mapping Camera (DMC) [45], the Leica Airborne Digital Sensor (ADS) [46] (see Fig. 2.2) and the proposed optimized 4 and 5-band multispectral sensor sensitivity information by applying the weighted integration model (2.2).

For the simulation of responses (2.2), the discrete representation $\hat{R}$ was used for the hyperspectral radiance R, and the representation $\hat{s}_i$ was used for the discretely located spectral sensitivity s_i. The discrete representation of radiance R corresponds to the peak locations of the hyperspectral bands as defined in Table 3.1. The discretely located sensor sensitivity functions for the ADS (Fig. 2.2c) and the UCD (Fig. 2.2a) were obtained from the sensor manufacturer. The DMC (Fig. 2.2b) sensor sensitivities were manually digitized from the sensitivity information presented in [45]. In all the cases, the sensor sensitivity functions are a product of the lens properties, the filters and the sensitivity of the CCD [45, 46]. In the simulations, each sensor system sensitivity was scaled to have a maximum peak value of 1 (see Fig. 2.2).

Furthermore, the available discretely located sensor sensitivities were linearly interpolated to have values in wavelength locations that correspond to the peak locations of the hyperspectral bands (Table 3.1). The integration (2.2) was then approximated by using a discrete sum,

$$X_i = \sum_{j=1}^{64} \hat{R}(\lambda_j)\hat{s}_i(\lambda_j), \tag{6.1}$$

with sampling locations $\lambda_1, \ldots, \lambda_{64}$ corresponding to the band peaks (Table 3.1). Using this approach, the responses for all of the pixel locations in the tree plots were simulated.

6.4 PLOT- AND PIXEL-LEVEL TREE SPECIES CLASSIFICATION

The Scots pine, Norway spruce and deciduous birch tree species classifications were evaluated for pixel- and plot-level scale datasets. With these datasets, the classification results were investigated using simulated responses of the standard and proposed optimized multispectral sensor, the selected hyperspectral bands, and all 64 hyperspectral band features. The classification accuracy and kappa value [110] were calculated to evaluate the classification performance.

6.4.1 Simulated Sensor Responses

In [*P1*], using the plot-level scale dataset, tree species classification was studied using the DA and C-SVM classifiers and the simulated response of the widely used 4-band multispectral sensors (ADS, UCD and DMC) and the proposed 5-band multispectral sensor ADS* (see Fig. 6.1a). The QDA, LOO accuracy for the simulated 4-band responses of ADS, UCD and DMC was similar (the total LOO accuracy had a difference of approximately 1%). For the simulated 4-band responses, the LOO accuracy of the SVM was 2–5% higher than the accuracy obtained using QDA. However, using hyperspectral data, the QDA classifiers were ill-conditioned and in many cases led to poor tree species classification performance due to the smaller number of available plots (training data size) and the larger hyperspectral data dimensionality. In such cases, the SVM results were not affected by the size of the training data. In cases where QDA failed, the LDA classification results were calculated and presented.

The simulated responses of the proposed 5-band multispectral sensor ADS* yielded similar (a difference of approximately 1%) LOO results for the QDA and SVM classifiers. In all of the experiments, the simulated 5-band responses produced a 5–13% higher classification accuracy than the 4-band responses of ADS, UCD and DMC. Furthermore, the simulated responses of the proposed

5-band sensor system produced classification accuracy similar to those obtained using 64 band AisaEAGLE II hyperspectral radiance data.

The use of training and test data from the different view direction showed that the accuracy of the simulated 5-band response (ADS*) was high and stable, similar to the results obtained using all 64 hyperspectral bands. However, the accuracy of the simulated standard 4-band responses was lower and varied among the sensors.

Similarly, in [*P2*] the simulated responses of the proposed 4-band (ADS-S) and 5-band (ADS-S*) tree species classification were studied. When using the plot-level scale dataset, the simulated response of the proposed 4 and 5-band multispectral sensor provided a tree species classification performance that was similar to those obtained using all 64 hyperspectral bands.

Furthermore, when using the simulated response of the proposed 5-band (ADS*), the pixel-level classification accuracy was improved by approximately 2% versus the simulated responses standard 4-band ADS. Similarly, the simulated responses of the proposed 4-band (ADS-S) and 5-band (ADS-S*) provide approximately 4% improved tree species classification compared with the results obtained by using the simulated responses of the standard 4-band ADS. These classification results were lower (approximately 7%) than the results obtained by using all 64 hyperspectral bands.

6.4.2 Selected Hyperspectral Bands

For band selection hyperspectral reflectance data were used. Using different spatial scales and balance conditions for the training samples (pixel- and plot-level scales), 8–11 hyperspectral bands were selected using sparse logistic regression with the Bayesian regularization method [40]. The results are presented in [*P2*]. A minimum of eight bands was selected using a balanced plot-level scale dataset. With the selected bands, the pixel-level scale tree species classification was evaluated by selecting approximately 1% of the total data

in the training while the remaining data were test set. Because the training dataset was selected randomly, each classification experiment was repeated 10 times, and the average classification result is presented. Despite the differences in the selected band combinations, similar tree species classification accuracies were obtained at the pixel-level. The classification results obtained with the 8–11 selected bands in the pixel- and plot-level scale datasets were similar to those obtained using all 64 hyperspectral bands. To assess the similarity in the classification results obtained between selected bands and using all 64 hyperspectral bands, a non-inferior and difference in the results were computed using the method described in [111]. Furthermore, in [*P2*] the use of the first five selected narrow bands for the balanced dataset improved (approximately 3% higher) the pixel-level tree species classification compared with the results obtained using the simulated responses of the proposed 4 and 5-band multispectral sensor system.

6.4.3 Assessment of Selected Bands under Changing View-Illumination Geometry Conditions

Forest canopy reflectance varies with changes in the imaging view-illumination geometry condition. Furthermore, with a change in view direction different parts of the crowns are observed and affect the reflectance signal. In the context of defining an optimized band, it must be determined whether the selected band obtained with specific view-illumination geometry conditions and spatial scale in the dataset provide a reasonably accurate tree species classification despite deviations in the spatial and view-illumination geometry conditions of the datasets used for the band selection and classification. The classification results obtained using eight selected bands (obtained with plot-level scale reflectance data and morning view-illumination geometry conditions) and all 64 band hyperspectral reflectance were evaluated. Furthermore, the results were evaluated using pixel-level scale reflectance and a normalized reflectance dataset.

In the experiments described in [*P3*], when the tree species classification was performed using eight the selected bands and morning (*BL1*, *BL2*) datasets, i.e., classifier training and test sets, the view-illumination geometry conditions matched those of the data for band selection, and accuracy and kappa above 94% and 0.89, respectively, was obtained. When the afternoon (*DL*) dataset was used in the classification, the view-illumination geometry conditions of the band selection and classification dataset differed; we obtained accuracy and kappa, 93% and 0.89, respectively. Furthermore, tree species classification results using selected bands (39, 17 and 8) were evaluated and compared. We obtained the similar results (accuracy difference $< 1\%$) on using 39, 17 and 8 selected bands. These results showed that suboptimal band selection (with respect to view-illumination geometry conditions) still provides reasonable (accuracy approximately 93% and kappa 0.90) classification results. Furthermore, when comparing the classification results for the reflectance and normalized reflectance dataset, a similar (difference 1–2%) classification result was obtained. In addition, results in [*P3*] indicated that there was no significant difference between the results obtained with selected bands (39,17,8) and using all 64 hyperspectral bands, when classifier training and test dataset view-illumination geometry condition match.

When there was a small difference in the solar azimuth ($< 15^\circ$), the solar elevation ($< 6^\circ$) and the imaging view-illumination geometry conditions between the classifier training and test dataset (when using *BL1* and *BL2* dataset for training / test) an accuracy of approximately 90% was observed. However, a significant decrease in the accuracy was observed (approximately 26%) for the reflectance dataset when there were larger differences in the solar azimuth ($> 48^\circ$), solar elevation ($> 9^\circ$) and viewing-illumination geometry conditions between the classifier training and test dataset (i.e., using the data imaged in the morning and in the afternoon datasets as training and test). Moreover, when a larger difference in the view-illumination geometry conditions between classifier training and test dataset occurred, the normalization of the

reflectance vectors to unit vectors improved the accuracy (approximately 13%) compared with results obtained using reflectance data. We assumed that changes in view-illumination geometry conditions caused scale changes in the spectral reflectance, and that normalization reduced the scale; therefore, the classification performance was improved.

Furthermore, the classification result with the eight bands provide similar (difference $<1\%$) or 4% (on average) improved overall accuracy than the classification results obtained with all 64 hyperspectral bands. In addition, when using all 64 hyperspectral bands deviation in the classification results was higher than in results obtained using selected bands. These results suggest that some of the hyperspectral bands are problematic with respect to view-illumination geometry condition changes in the dataset and contribute to lowering classification accuracy.

7 Discussion and Conclusions

The aim of this thesis was to define optimized spectral bands in the 400–1000 nm wavelength range that accurately classify Scots pine, Norway spruce and deciduous birch tree species. The tree species classification performance was evaluated using simulated multispectral responses (via existing multispectral sensor sensitivities and optimized sensitivities) and selected hyperspectral bands. The classification results were compared with the results obtained using all 64 AisaEAGLE II hyperspectral bands in the 400–1000 nm wavelength range. Classifications were performed using support vector machines and discriminant analysis classifiers with pixel- and plot-level scale data.

The results presented in this dissertation are based on the airborne measured line-array imaging AisaEAGLE II hyperspectral sensor data on the wavelength ranging from 400–1000 nm. The hyperspectral data were collected from the forest in Hyytiälä, Finland, which has been widely used in developing methodology for aerial remote sensing [9, 19, 20, 29, 71, 112–115]. Single tree species forest plots were extracted from the acquired images.

The focus of [*P1*] was to compare the tree species classification performance of the simulated responses of three widely used multispectral sensors (Leica ADS40, Vexcel UltraCam-D (UCD), and Intergraph-Z/I Digital Mapping Camera (DMC)). The simulations with Leica ADS sensitivities were most similar to the data acquired using real Leica sensors because the Leica ADS and the hyperspectral sensor rely on the same line-scanning principle. Simulations with the UCD and DMC sensitivities correspond to a hypothetical line-scanning sensor with these spectral sensitivity shapes (Fig. 2.2). The simulations apply here to the level of spectral sensitivities and, therefore, do not give information on the performance of complete

sensor systems. All 4-band standard multispectral sensors provided a similar classification performance for plot-level scale data.

Based on the previous recommendation in [16] (based on ground level hyperspectral measurements), a 5-band optimized multispectral sensitivity was proposed with four original Leica ADS40 sensitivity and one additional red-edge sensitivity function. Using plot-level scale data, the tree species classification performance of this proposed 5-band sensor system was significantly improved (5–13% points) compared with simulated responses of standard 4-band multispectral sensor systems. In addition, the classification performance of the simulated response of the proposed 5-band sensor system was similar (approximately 99%) to the performance of 64 bands AisaEAGLE II hyperspectral system. However, for the pixel-level scale dataset, classification accuracy for the proposed 5-band was lower than that obtained using all 64 hyperspectral bands. This finding was expected because the pixel-level classification is more complex than the plot-level classification due to increasing variability (effects of shadows and background).

We considered methods for selecting hyperspectral bands and the use of selected band in tree specie classification was evaluated. In [*P2*], band selection results using the sparse logistic regression with the Bayesian regularization method [40] was presented for pixel- and plot-level scale data in balanced conditions. Minimum eight bands was selected using the balanced (balancing the class sizes in the training dataset) plot-level dataset. Furthermore, the band selection performance of other regression-based methods, sparse linear regression [98] and sparse logistic regression [95], were evaluated using the balanced plot-level scale dataset and presented in [*P3*]. These methods resulted in 39 and 17 selected bands. In previous logistic regression-based classification, the imbalance in number of training samples in each class affects logistic regression performance [116]. It was also validated in our experiments ([*P2*]) that balancing the number of training samples in each class lead to a smaller number of selected bands than when compared to the bands selection with imbalanced training data.

Previously, hyperspectral band selection had been performed [10,11,35]. Whether the selected hyperspectral bands could be considered physical multispectral bands or whether the selected band positions had any relation to the existing multispectral sensor sensitivity position had not been addressed. In [*P2*] the band selection outcome obtained using the method in [40] was related to the 4-band Leica ADS40 multispectral sensor sensitivity positions, and modified 4 and 5-band broadband multispectral sensor systems were proposed. Only one Leica ADS band position was modified because tuning all Leica bands for tree species classification is unreasonable. Using the simulated response of the proposed 4 and 5-band multispectral sensor systems, tree species classification was improved when compared with results obtained via the standard 4-band sensor system. Furthermore, in the experiments [*P2*], it became clear that the first five hyperspectral bands that were selected with the balanced data set provided a better classification result than the proposed 4 and 5-band system. This result was expected since the proposed system was based on the Leica ADS, and we only modified one Leica ADS sensitivity position.

When defining an optimized band the selected bands should provide a reasonably accurate tree species classification performance despite differences in the view-illumination geometry conditions of the datasets used for the classification and band selection. In all the experiments presented in [*P3*] we demonstrate that the selected eight bands (optimized bands) resulted in similar (1–2% difference) or improved tree species discrimination than those obtained using all 64 hyperspectral bands and the 39 and 17 selected bands throughout the view-illumination geometry conditions.

In this thesis, hyperspectral data were acquired at the peak of the growing season, where differences between species are likely to be difficult to identify. Therefore, the classification results presented in the thesis could be improved if the data were acquired at different stages of phenology. For example, in previous studies, imaging in autumn [3] and using multi-temporal imaging [117] suggest improved tree species classification performance.

The optimized bands are sensitive to forest types, structure, and temporal spectral changes in vegetation. The results presented in this thesis using the selected hyperspectral bands and the simulated responses of the proposed 4 and 5-band multispectral systems suggest that optimized or tunable spectral imaging systems with a small number of bands will enable the extraction of adequate information for tree species classification. Therefore, it is possible to find suitable application-specific band combinations for certain geographic areas.

Bibliography

[1] P. Pant, J. Mutanen, A. Mankki, M. Hauta-Kasari, and J. Parkkinen, "Spectral analysis and classification of dirt particles in pulp," *Nordic Pulp and Paper Research Journal* **27,** 577–589 (2012).

[2] P. Koirala, P. Pant, M. Hauta-Kasari, and J. Parkkinen, "Highlight detection and removal from spectral image," *Journal of the Optical Society of America A* **28,** 2284–2291 (2011).

[3] J. Holmgren, A. Persson, and U. Söderman, "Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images," *International Journal of Remote Sensing* **29,** 1537–1552 (2008).

[4] U. Valerie and T. Livia, "Airborne lidar: advances in discrete return technology for 3D vegetation mapping," *Remote Sensing* **3,** 416–434 (2011).

[5] J. Hyyppä, H. Hyyppä, D. Leckie, F. Gougeon, X. Yu, and M. Maltamo, "Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests," *International Journal of Remote Sensing* **29,** 1339–1366 (2008).

[6] M. Dalponte, L. Bruzzone, and D. Gianelle, "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data," *Remote Sensing of Environment* **123,** 258 – 270 (2012).

[7] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 3rd ed. (Amsterdam, The Netherlands: Elsevier, 2007).

[8] Y. Xie, Z. Sha, and M. Yu, "Remote sensing imagery in vegetation mapping: a review," *Journal of Plant Ecology* **1,** 9–23 (2008).

[9] I. Korpela, *Individual Tree Measurements by Means of Digital Aerial Photogrammetry* (Finnish Society of Forest Science, 2004).

[10] M. Dalponte, L. Bruzzone, L. Vescovo, and D. Gianelle, "The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas," *Remote Sensing of Environment* **113,** 2345 – 2355 (2009).

[11] M. L. Clark, D. A. Roberts, and D. B. Clark, "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales," *Remote Sensing of Environment* **96,** 375 – 398 (2005).

[12] M. Dalponte, H. O. Ørka, T. Gobakken, D. Gianelle, and E. Næsset, "Tree Species Classification in Boreal Forests With Hyperspectral Data," *Geoscience and Remote Sensing, IEEE Transactions on* **51,** 2632–2645 (2013).

[13] P. Bunting and R. Lucas, "The delineation of tree crowns in Australian mixed species forests using hyperspectral Compact Airborne Spectrographic Imager (CASI) data," *Remote Sensing of Environment* **101,** 230 – 248 (2006).

[14] B. L. Becker, D. P. Lusch, and J. Qi, "A classification-based assessment of the optimal spectral and spatial resolutions for Great Lakes coastal wetland imagery," *Remote Sensing of Environment* **108,** 111 – 120 (2007).

[15] T. Jääskelainen, R. Silvennoinen, J. Hiltunen, and J. Parkkinen, "Classification of the reflectance spectra of pine, spruce, and birch," *Applied Optics* **33,** 2356–2362 (1994).

[16] V. Heikkinen, T. Tokola, J. Parkkinen, I. Korpela, and T. Jääskelainen, "Simulated Multispectral Imagery for Tree

Species Classification Using Support Vector Machines," *Geoscience and Remote Sensing, IEEE Transactions on* **48,** 1355–1364 (2010).

[17] E. Ylitalo, *Finnish Statistical Yearbook of Forestry* (Finnish Forest Research Institute, 2010).

[18] E. Tomppo and J. Heikkinen, "National Forest Inventory of Finland: Past, Present and Future," in *Statistics, Registries, and Science: Experiences* (Statistics Finland, Helsinki, Finland, 1999), pp. 89–108.

[19] I. Korpela, V. Heikkinen, E. Honkavaara, F. Rohrbach, and T. Tokola, "Variation and directional anisotropy of reflectance at the crown scale – Implications for tree species classification in digital aerial images," *Remote Sensing of Environment* **115,** 2062 – 2074 (2011).

[20] V. Heikkinen, I. Korpela, T. Tokola, E. Honkavaara, and J. Parkkinen, "An SVM Classification of Tree Species Radiometric Signatures Based on the Leica ADS40 Sensor," *Geoscience and Remote Sensing, IEEE Transactions on* **49,** 4539–4551 (2011).

[21] P. Packalén, A. Suvanto, and M. Maltamo, "A Two Stage Method to Estimate Species–specific Growing Stock," *Photogrammetric Engineering & Remote Sensing* **75,** 1451–1460 (2009).

[22] E. Puttonen, A. Jaakkola, P. Litkey, and J. Hyyppä, "Tree classification with fused mobile laser scanning and hyperspectral data," *Sensors* **11,** 5158–5182 (2011).

[23] Juha Hyyppä and Hannu Hyyppä and Mikko Inkinen and Marcus Engdahl and Susan Linko and Yi-Hong Zhu, "Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes," *Forest Ecology and Management* **128,** 109 – 120 (2000).

[24] K. Makisara and E. Tomppo, "Airborne imaging spectrometry in National Forest Inventory [Finland]," in *Geoscience and Remote Sensing Symposium, IGARSS '96. 'Remote Sensing for a Sustainable Future.', International*, Vol. 2 (1996), pp. 1010–1013.

[25] J. Lumme, "Classification of vegetation and soil using imaging spectrometer data," in *Geoscience and remote sensing symposium* (2004), p. 83.

[26] J. Lumme, "Forest Classification Using High Spectral and Spatial Resolution Data," in *Proceedings of 4th EARSeL Workshop on Imaging Spectroscopy. New quality in environmental studies, EARSeL and Warsaw University, Warsaw* (2005), pp. 611–616.

[27] A. Pekkarinen, "Image segment-based spectral features in the estimation of timber volume," *Remote Sensing of Environment* **82,** 349 – 359 (2002).

[28] M. Middleton and P. N "Ordination and hyperspectral remote sensing approach to classify peatland biotopes along soil moisture and fertility gradients," *Remote Sensing of Environment* **124,** 596 – 609 (2012).

[29] M. M "A forestry GIS-based study on evaluating the potential of imaging spectroscopy in mapping forest land fertility," *International Journal of Applied Earth Observation and Geoinformation* **33,** 302 – 311 (2014).

[30] Benjamin Kötz and Michael Schaepman and Felix Morsdorf and Paul Bowyer and Klaus Itten and Britta Allgöwer, "Radiative transfer modeling within a heterogeneous canopy for estimation of forest fire fuel properties," *Remote Sensing of Environment* **92,** 332 – 344 (2004).

[31] J. M. Chen, X. Li, T. Nilson, and A. Strahler, "Recent advances in geometrical optical modelling and its applications," *Remote Sensing Reviews* **18,** 227–262 (2000).

[32] Q. Xiao, S. L. Ustin, and E. G. McPherson, "Using AVIRIS data and multiple-masking techniques to map urban forest tree species," *International Journal of Remote Sensing* **25,** 1451–1460 (2004).

[33] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment* **113, Supplement 1,** S110 – S122 (2009), Imaging Spectroscopy Special Issue.

[34] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *Geoscience and Remote Sensing, IEEE Transactions on* **42,** 1778–1790 (2004).

[35] M. Pal and G. Foody, "Feature Selection for Classification of Hyperspectral Data by SVM," *Geoscience and Remote Sensing, IEEE Transactions on* **48,** 2297 –2307 (2010).

[36] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *Information Theory, IEEE Transactions on* **14,** 55–63 (1968).

[37] Publications of The Finnish Society of Photogrammetry and Remote Sensing, "Recommendations for aerial photogrammetry for mapping in Finland," (1995), `http://foto.hut.fi/seura/julkaisut/erillisjulkaisu1_1995/tekst i.html`, (Accessed: 21st May, 2012).

[38] F. Dell'Endice, J. Nieke, B. Koetz, M. E. Schaepman, and K. Itten, "Improving radiometry of imaging spectrometers by using programmable spectral regions of interest," {*ISPRS*} *Journal of Photogrammetry and Remote Sensing* **64,** 632 – 639 (2009).

[39] M. Pal, "Multinomial logistic regression-based feature selection for hyperspectral data," *International Journal of Applied Earth Observation and Geoinformation* **14,** 214 – 220 (2012).

[40] G. Cawley and N. Talbot, "Gene selection in cancer classification using sparse logistic regression with Bayesian regularisation," *Bioinformatics* **22,** 2348–2355 (2006).

[41] G. A. Shaw and H. hua K. Burke, "Spectral Imaging for Remote Sensing," *Lincoln Laboratory Journal* **14,** 3–28 (2003).

[42] G. Schaepman-Strub, M. Schaepman, T. Painter, S. Dangel, and J. Martonchik, "Reflectance quantities in optical remote sensingdefinitions and case studies," *Remote Sensing of Environment* **103,** 27 – 42 (2006).

[43] F. Nicodemus, J. Richmond, J. Hsia, I. Ginsberg, and T. Limperis, "Geometric Considerations And Nomenclature For Reflectance," *Monograph 161, National Bureau of Standards (US)* (1977).

[44] M. Kropfl and M. Gruber, "UltraCam Calibration Geometry and Radiometry Analysis," (2006), `http://www.gtbi.net/export/sites/default/GTBiWeb/productos/descargas/VexcelUltracamD-Calibracion-en.pdf` (Accessed: 22nd May, 2012).

[45] R. Rayn and M. Pagnutti, "Enhanced Absolute and Relative Radiometric Calibration for Digital Aerial Cameras," in *Proceedings of the Photogrammetric Week* (2009), pp. 81–90.

[46] U. Beisl, "Absolute Spectroradiometric Calibration of the ADS40 Sensor," in *Proc. Congrés ISPRS Commission Technique I, Symposium* (2006), pp. 14 – 18.

[47] E. Honkavaara and L. Markelin, "Radiometric Performance of Digital Image Data Collection - A Comparison of ADS40/DMC/UltraCam and EmergeDSS," in *Photogrammetric Week 2007* (2007), pp. 117–129.

[48] L. Markelin, E. Honkavaara, J. Peltoniemi, E. Ahokas, R. Kuittinen, J. Hyyppä, J. Suomalainen, and A. Kukko, "Radiometric calibration and characterization of large-format digital

photogrammetric sensors in a test field," *Photogrammetric Engineering & Remote Sensing* **74,** 1487–1500 (2008).

[49] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, "The HYMAP airborne hyperspectral sensor: the system, calibration and performance," in *First EARSeL Workshop on Imaging Spectroscopy (Remote Sensing Laboratories, University of Zurich, Zurich, Switzerland)* (1998), pp. 37–42.

[50] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, "Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sensing of Environment* **65,** 227 – 248 (1998).

[51] C. D. Anger, S. Mah, and S. Babey, "Technological enhancements to the Compact Airborne Spectrographic Imager (CASI)," in *First International Airborne Remote Sensing Conference and Exhibition, Strasbourg, France* (1994), pp. 205–213.

[52] K. Makisara, M. Meinander, M. Rantasuo, J. Okkonen, M. Aikio, and K. Sipola, "Airborne imaging spectrometer for applications (AISA)," in *Geoscience and Remote Sensing Symposium, 1993. IGARSS '93. Better Understanding of Earth Environment., International*, Vol. 2 (1993), pp. 479 –481.

[53] SPECIM, "AISAEAGLE HYPERSPECTRAL SYSTEM," (2012), `http://www.specim.fi/index.php/products/airborne/aisaeagle/` (Accessed: 5th September, 2012).

[54] L. Markelin, E. Honkavaara, T. Takala, D. Schläpfer, J. Suomalainen, and P. Pellikka, "A novel approach for the radiometric correction of airborne hyperspectral image data," in *XXII Congress of the International Society for Photogrammetry and Remote Sensing ISPRS* (2012).

[55] B.-C. Gao, C. Davis, and A. F. H. Goetz, "A Review of Atmospheric Correction Techniques for Hyperspectral Remote

Sensing of Land Surfaces and Ocean Color," in *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on* (2006), pp. 1979–1981.

[56] M. Rast, S. J. Hook, R. E. Alley, and C. D. Elvidge, "An Evaluation of Techniques for the Extraction of Mineral Absorption Features from High Spectral Resolution Remote Sensing Data," *American Society for Photogrammetry and Remote Sensing* **57,** 1303 – 1309 (1991).

[57] F. A. Kruse, "Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in the northern Grapevine Mountains, Nevada, and California," *Remote Sensing of Environment* **24,** 31 – 51 (1988).

[58] G. M. Smith and E. J. Milton, "The use of the empirical line method to calibrate remotely sensed data to reflectance," *International Journal of Remote Sensing* **20,** 2653–2662 (1999).

[59] I. Norjamaki and T. Tokola, "Comparison of Atmospheric Correction Methods in Mapping Timber Volume with Multitemporal Landsat Images in Kainuu, Finland," *American Society for Photogrammetry and Remote Sensing* **73,** 155–164 (2007).

[60] B.-C. Gao, K. B. Heidebrecht, and A. F. Goetz, "Derivation of scaled surface reflectances from AVIRIS data," *Remote Sensing of Environment* **44,** 165–178 (1993), Airbone Imaging Spectrometry.

[61] A. Berk, G. P. Anderson, L. S. Bernstein, P. K. Acharya, H. Dothe, M. W. Matthew, S. M. Adler-Golden, J. H. Chetwynd, Jr., S. C. Richtsmeier, B. Pukall, C. L. Allred, L. S. Jeong, and M. L. Hoke, "MODTRAN4 radiative transfer modeling for atmospheric correction," in *Proc. SPIE, Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III*, Vol. 3756 (1999), pp. 348–353.

[62] S. Y. Kotchenova, E. F. Vermote, R. Matarrese, and F. J. Klemm, "Validation of a vector version of the 6S radiative

transfer code for atmospheric correction of satellite data. Part I: Path radiance," *Applied Optics* **45,** 6762–6774 (2006).

[63] A. Berk, G. P. Anderson, P. K. Acharya, L. S. Bernstein, L. Muratov, J. Lee, M. Fox, S. M. Adler-Golden, J. H. Chetwynd, M. L. Hoke, R. B. Lockwood, J. A. Gardner, T. W. Cooley, C. C. Borel, and P. E. Lewis, "MODTRAN 5: A Reformulated Atmospheric Band Model with Auxiliary Species and Practical Multiple Scattering Options: Update," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, Vol. 5806 (2005), pp. 662–667.

[64] B.-C. Gao, M. J. Montes, C. O. Davis, and A. F. Goetz, "Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean," *Remote Sensing of Environment* **113, Supplement 1,** S17 – S24 (2009), Imaging Spectroscopy Special Issue.

[65] G. P. Anderson, G. W. Felde, M. L. Hoke, A. J. Ratkowski, T. W. Cooley, J. H. Chetwynd Jr, J. Gardner, S. M. Adler-Golden, M. W. Matthew, and A. Berk, "MODTRAN4-based atmospheric correction algorithm: FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes)," in *AeroSense 2002* (International Society for Optics and Photonics, 2002), pp. 65–71.

[66] F. A. Kruse, "Comparison of ATREM, ACORN, and FLAASH Atmospheric Corrections using Low-Altitude AVIRIS Data of Boulder, Colorado," in *Proceedings 13th JPL Airborne Geoscience Workshop* (Jet Propulsion Laboratory, Pasadena, CA, USA, 2004).

[67] E. Ben-Dor, B. Kindel, and A. Goetz, "Quality assessment of several methods to recover surface reflectance using synthetic imaging spectroscopy data," *Remote Sensing of Environment* **90,** 389 – 404 (2004).

[68] SPECIM, "SPECIM CaliGeo Brochures," (2012), `http://www.spectralcameras.com/files/Brochures/CaliGeo-ver5-08.pdf` (Accessed: 5th September, 2012).

[69] L. Markelin, E. Honkavaara, J. Peltoniemi, E. Ahokas, R. Kuittinen, J. Hyyppä, J. Suomalainen, and A. Kukko, "Radiometric calibration and characterization of large-format digital photogrammetric sensors in a test field," *Photogrammetric Engineering & Remote Sensing* **74,** 1487–1500 (2008).

[70] D. Schläpfer, M. E. Schaepman, and K. I. Itten, "PARGE: Parametric Geocoding Based on GCP-Calibrated Auxiliary Data," in *Descour, M.R. and Shen, S.S., Imaging Spectrometry IV, SPIE* (1998), pp. 3438–334.

[71] I. Korpela, H. O. Ørka, M. Maltamo, T. Tokola, and J. Hyyppä, "Tree Species Classification Using Airborne LiDAR – Effects of Stand and Tree Parameters, Downsizing of Training Set, Intensity Normalization, and Sensor Type," *Silva Fennica* **44,** 319–339 (2010).

[72] R. Crippen, "Calculating the vegetation index faster," *Remote Sensing of Environment* **34,** 71–73 (1990).

[73] H. Zhang, "Hyperpsectral Image Denoising with Cubic Total Variation Model," in *22th International Society for Photogrammetry and Remote Sensing Congress* (2012).

[74] J. Bioucas-Dias and J. Nascimento, "Hyperspectral Subspace Identification," *Geoscience and Remote Sensing, IEEE Transactions on* **46,** 2435 –2445 (2008).

[75] A. Green, M. Berman, P. Switzer, and M. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *Geoscience and Remote Sensing, IEEE Transactions on* **26,** 65–74 (1988).

[76] J. Lee, A. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted

principal components transform," *Geoscience and Remote Sensing, IEEE Transactions on* **28,** 295–304 (1990).

[77] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Plenum Publishing Co., 2003).

[78] P. Latorre-Carmona, J. Martinez Sotoca, F. Pla, J. Bioucas-Dias, and C. Ferre, "Effect of Denoising in Band Selection for Regression Tasks in Hyperspectral Datasets," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **6,** 473–481 (2013).

[79] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, "Geometric Considerations and Nomenclature for Reflectance," *National Bureau of Standards* (1977).

[80] J. D. Shepherd and J. R. Dymond, "BRDF Correction of Vegetation in AVHRR Imagery," *Remote Sensing of Environment* **74,** 397 – 408 (2000).

[81] B. Yu, M. Ostland, P. Gong, and R. Pu, "Penalized discriminant analysis of in situ hyperspectral data for conifer species recognition," *Geoscience and Remote Sensing, IEEE Transactions on* **37,** 2569–2577 (1999).

[82] F. Tsai and W. Philpot, "A derivative-aided hyperspectral image analysis system for land-cover classification," *Geoscience and Remote Sensing, IEEE Transactions on* **40,** 416 –425 (2002).

[83] F. Tsai, E.-K. Lin, and K. Yoshino, "Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species," *International Journal of Remote Sensing* **28,** 1023–1039 (2007).

[84] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research* **3,** 1157–1182 (2003).

[85] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence* **97,** 245–271 (1997).

[86] R. Kohavi and G. H. John, "Wrappers for Feature Subset Selection," *Artificial Intelligence* **97,** 273–324 (1997).

[87] P. Bajcsy and P. Groves, "Methodology for Hyperspectral Band Selection," *Photogrammetric Engineering & Remote Sensing* **70,** 793–802 (2004).

[88] B. L. Becker, D. P. Lusch, and J. Qi, "Identifying optimal spectral bands from in situ measurements of Great Lakes coastal wetlands using second-derivative analysis," *Remote Sensing of Environment* **97,** 238–248 (2005).

[89] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B* **58,** 267–288 (1994).

[90] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2 ed. (Springer, 2009).

[91] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2 ed. (Wiley Interscience, 2000).

[92] M. Collins, R. Schapire, and Y. Singer, "Logistic Regression, AdaBoost and Bregman Distances," *Machine Learning* **48,** 253–285 (2002).

[93] J. Liu, J. Chen, and J. Ye, "Large-scale Sparse Logistic Regression," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09 (2009), pp. 547–556.

[94] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006).

[95] S. K. Shevade and S. S. Keerthi, "A simple and efficient algorithm for gene selection using sparse logistic regression," *Bioinformatics* **19,** 2246–2253 (2003).

[96] P. Williams, "Bayesian regularisation and pruning using a Laplace prior," *Neural Computation* **7,** 117143 (1995).

[97] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *The Annals of Applied Statistics* **1,** 302–332 (2007).

[98] J. H. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software* **33,** 1–22 (2010).

[99] M. Schmidt, G. Fung, and R. Rosales, "Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches," in *Proceedings of the 18th European Conference on Machine Learning*, ECML '07 (2007), pp. 286–297.

[100] H. Jeffreys, "An Invariant Form for the Prior Probability in Estimation Problems," *Royal Society of London Proceedings Series A* **186,** 453–461 (1946).

[101] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Seventh ed. (Elsevier/Academic Press, Amsterdam, 2007), Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, with one CD-ROM (Windows, Macintosh and UNIX).

[102] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on* **43,** 1351–1362 (2005).

[103] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas," *Geoscience and Remote Sensing, IEEE Transactions on* **46,** 1416 –1427 (2008).

[104] G. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *Geoscience and Remote Sensing, IEEE Transactions on* **42,** 1335–1343 (2004).

[105] D. Michie, D. J. Spiegelhalter, C. C. Taylor, and J. Campbell, eds., *Machine Learning, Neural and Statistical Classification* (Ellis Horwood, Upper Saddle River, NJ, USA, 1994).

[106] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, USA, 2001).

[107] J. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, *Least Squares Support Vector Machines* (World Scientific, Singapore, 2002).

[108] T. V. Gestel, J. Suykens, S. V. B. Baesens, J. Vanthienen, G. Dedene, B. D. Moor, and J. Vandewalle, "Benchmarking Least Squares Support Vector Machine Classifiers," *Machine Learning* **54,** 5–32 (2004).

[109] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on* **13,** 415–425 (2002).

[110] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* **20,** 37–46 (1960).

[111] G. M. Foody, "Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority," *Remote Sensing of Environment* **113,** 1658 – 1663 (2009).

[112] I. Korpela, H. O. Ørka, J. Hyyppä, V. Heikkinen, and T. Tokola, "Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies," *ISPRS Journal of Photogrammetry and Remote Sensing* **65,** 369 – 379 (2010).

[113] J. Pitkänen, "Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods," *Canadian Journal of Forest Research* **31,** 832–844 (2001).

[114] J. Vauhkonen, I. Korpela, M. Maltamo, and T. Tokola, "Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics," *Remote Sensing of Environment* **114,** 1263 – 1276 (2010).

[115] L. Korhonen, I. Korpela, J. Heiskanen, and M. Maltamo, "Airborne discrete-return {LIDAR} data in the estimation of vertical canopy cover, angular canopy closure and leaf area index," *Remote Sensing of Environment* **115,** 1065 – 1080 (2011).

[116] D. Olson, "Data Set Balancing," in *Data Mining and Knowledge Management*, Vol. 3327, Y. Shi, W. Xu, and Z. Chen, eds. (Springer Berlin, Heidelberg, 2005), pp. 71-80.

[117] R. A. Hill, A. Wilson, M. George, and S. Hinsley, "Mapping tree species in temperate deciduous woodland using time-series multi-spectral data," *Applied Vegetation Science* **13,** 86–99 (2010).

Paras Pant

Optimizing Spectral Bands of Airborne Imager for Tree Species Classification

This thesis concentrated on optimizing and selecting spectral bands of airborne imagers for pine, spruce and birch tree species classification. Band optimizations and selections were performed in 400-1000 nm wavelength range using simulations based on airborne measured hyperspectral image data. Classification results were presented using simulated responses of proposed 4 and 5-band multispectral systems and selected hyperspectral bands via sparse regression-based feature selection methods. Results suggest that 4-8 multi or hyperspectral bands can be used to achieve accurate classification of the tree species.

UNIVERSITY OF
EASTERN FINLAND

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences

ISBN 978-952-61-1783-6