Hyperspectral close-range and remote sensing of soils and related plant associations
Spectroscopic applications in the boreal environment

Maarit Middleton
HYPERSONTRAL CLOSE-RANGE AND REMOTE SENSING OF SOILS AND RELATED PLANT ASSOCIATIONS
Spectroscopic applications in the boreal environment

by
Maarit Middleton
Geological Survey of Finland
P.O. Box 77
FI-96101 Rovaniemi, Finland

ACADEMIC DISSERTATION

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Engineering for public examination and debate in Auditorium M1 at the Aalto University School of Engineering (Otakaari 1, Espoo, Finland) on the 26th of September 2014 at 12 noon.

Unless otherwise indicated, the figures have been prepared by the author of the publication.

Geological Survey of Finland
Espoo 2014
Instructor:
Doc. Raimo Sutinen
Geological Survey of Finland
Rovaniemi, Finland
Department of Geosciences
University of Oulu
Oulu, Finland

Supervisor:
Prof. Henrik Haggrén
Department of Real Estate, Planning and Geoinformatics
Aalto University
Espoo, Finland

Reviewers:
Dr. Angela Harris
School of Environment, Education and Development
University of Manchester
Manchester, United Kingdom

Prof. Alexander McBratney
Department of Environmental Sciences
University of Sydney
Sydney, Australia

Opponents:
Assoc. Prof. John Carranza
School of Earth and Environmental Sciences
James Cook University
Townsville, Australia

Prof. Nicholas Coops
Department of Forest Resources Management
University of British Columbia
Vancouver, Canada

Front cover: Peatland site type classification at southern boreal study site in Keminmaa, Finland (paper IV). Mapping was based on fuzzy support vector machine classification of airborne HyMap data. Peatland class-specific probabilities are presented with RGB composition. Class ‘bog’ is represented as red, ‘sedge fen’ as green, and ‘eutrophic fen’ as blue. Probability values are scaled with histogrammic equalization to emphasize spatial details. Pure RGB colors depict high probabilities. Normalized differential vegetation index (NDVI) was calculated for background. High NDVI values are illustrated as light shades of grey.

Maarit Middleton
Geological Survey of Finland, P.O. Box 77, FI-96101 Rovaniemi, Finland

E-mail: maarit.middleton@gtk.fi


Layout: Elvi Turtiainen Oy
Printing house: Juvenes Print – Suomen Yliopistopaino Oy

– One of the primary resources upon which man directly depends for his existence is soil moisture. [...] This resource, among others, man now seeks to monitor from space, and towards this end he is applying some of the most sophisticated technology and management expertise. – Idso et al. (1975)

To Kent, Sofia and Benjamin
Abstract

Hyperspectral close-range and remote sensing techniques have been available to the research community since the 1980’s but applications have focused on forestry and land use. The objective of the study was to explore relevant applications of visible and short wavelength infrared spectroscopy (350–2500 nm) for detection of physical and chemical properties of glacial till soils and plant species communities related to the soil properties in the boreal environment of northern Finland.

Empirical single and multivariate regression techniques (MVR) were applied for predicting glacial till soil dielectric permittivity (ε, i.e. soil moisture) and till elemental concentrations from close-range spectrometry. Predictive kernel and neural network based fuzzy classification approaches were applied for classification of data acquired with AISA and HyMap™ airborne imaging spectrometers. Ordination techniques were used for revealing plant community structures and optimizing the thematic class hierarchal level.

The till soil ε was well predicted from VSWIR spectra with the exponential single-spectral variate but also with MVR techniques. The most accurate results were gained with relevance vector machines. Prediction of till soil chemical element concentrations of Al, Ba, Co, Cr, Cu, Fe, Mg, Mn, Ni, V, and Zn was also statistically valid. Soil moisture based site suitability for Scots pine (Pinus sylvestris) from imaging spectroscopic data was moderately successful as the highest area under the receiver operating characteristics curve (AUC) value was 0.741. Site type mapping of aapa peatlands with support vector machines was highly successful with AUC values 0.946–0.999 for bog, sedge fen, and eutrophic fen.

Understanding the ε-reflectance relationship would be evident when artificial regeneration to Scots pine, intolerant of wet soils, is considered on clear-cuts with high soil moisture variability. The site suitability on site prepared forest compartments could be predicted using exposed soil pixels in high spatial resolution imagery but also with indirect imaging of soil moisture through understory species patterns. The high success of the peatland site type mapping was attributed to optimization of class hierarchal levels with a constrained ordination based approach which was used to test the spectral and ecological class separability prior to classification. These novel applications of imaging spectroscopic data can readily be applied in practice once cost-effective satellite based data is available. Further research is required to make the close-range spectroscopy operational for quantification of element concentrations to serve forest soil research and mineral potential mapping.

Keywords (GeoRef Thesaurus, AGI): remote sensing, airborne methods, hyperspectral analysis, spectroscopy, dielectric properties, chemical elements, multivariate analysis, till, biotopes, vegetation, forest soils, peatlands, classification, Finland

Empiiristen yksi- ja monimuuttujaregressiomenetelmien avulla laboratoriossa spektroradiometrillä mitatuista spektreistä mallinettiin moreenin dielektrisyyttä (ε, vesipitoisuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsinen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.

Ymmärrys ε:n ja heijastuskertoimen suhteesta tulee ilmeiseksi silloin, kun arvioidaan männyn uudistamisvalotus tai -tarpeen (e, vesipitoisuuutta) ja alkukaideen pitoisuuksia. Tuki- ja neuroverkkomenetelmillä tehtiin useanlogiikan mallin oikeellisuutta ja lihitöitä luotettavasti (ROC-käyrä) alapuolelle jäävä alue (AUC) oli männylle soveltuvalle luokalle 0,741. Hyperspektsisen kaukokartoitusaineisto soveltui hyvin myös aapasoiden suotyyppiluokitteluun, koska AUC-arvot vaihtelivat 0,946 ja 0,999:n välillä suolistoille rajahkeena, neva ja letto.
# CONTENTS

List of original publications ................................................................. 8  
Author's contribution ........................................................................ 8  
List of abbreviations ........................................................................... 9  
List of symbols ................................................................................... 9  

1 Introduction ...................................................................................... 10  
1.1 Background and research environment .......................................... 10  
1.2 Objectives and scope .................................................................... 12  
1.3 Research realization and dissertation structure ............................... 13  

2 Theoretical foundation ...................................................................... 14  
2.1 Spectra of soil .............................................................................. 14  
2.2 Spectra of plant species in shrub, herb and bryophyte layers .......... 17  
2.3 Spectroradiometry ....................................................................... 18  
2.4 Prediction of soil water content from VSWIR spectra .................. 19  
2.5 Prediction of till soil chemical elements from VSWIR spectra ....... 20  
2.6 Mapping of site suitability for artificial regeneration of Scots pine from  
  hyperspectral imaging of soil water based species patterns ............. 21  
2.7 Optimization of peatland class hierarchal level from hyperspectral remote sensing data ......................................................... 22  
2.8 Chapter summary ......................................................................... 23  

3 Materials and methods .................................................................... 23  
3.1 Study sites .................................................................................... 24  
3.2 Sampling and application of numerical techniques on research data ................................................................. 25  
3.3 Hyperspectral instrumentation and data ........................................ 28  
  3.3.1 Close-range spectroscopy ......................................................... 28  
  3.3.2 Airborne hyperspectral imaging spectrometry ......................... 28  
3.4 Reference field data ....................................................................... 29  
  3.4.1 Geophysical measurements ..................................................... 29  
  3.4.2 Geochemical, particle size and mineralogical analysis .......... 29  
  3.4.3 Vegetation analysis ................................................................. 29  
3.5 Ordination techniques ................................................................... 30  
  3.5.1 Non-metric multidimensional scaling ....................................... 30  
  3.5.2 Canonical correspondence analysis ......................................... 30  
  3.5.3 Partitioning around medoids .................................................... 30  
3.6 Classification and hyperspectral feature extraction ...................... 31  
  3.6.1 Feature extraction .................................................................. 31  
  3.6.2 Supervised non-parametric classification .............................. 31  
3.7 Chemometric techniques .............................................................. 31  
  3.7.1 Spectral transformations ......................................................... 31  
  3.7.2 Partial least squares regression ............................................. 32
3.7.3 Bagging regression trees ................................................................. 32
3.7.4 Multivariate adaptive regression splines ........................................ 32
3.7.5 Random forests .................................................................................. 33
3.7.6 Relevance vector machines ................................................................. 33
3.7.7 Mixed-effects modeling .................................................................... 33
3.8 Model validation .................................................................................... 33

4 Results ........................................................................................................ 34
4.1 Predicting till soil dielectric permittivity from reflectance spectra ........ 34
4.2 Predicting till soil element concentrations from reflectance spectra .... 38
4.3 Mapping site suitability for artificial regeneration to Scots pine based on
hyperspectral imaging data ........................................................................... 38
4.4 Peatland site type mapping from hyperspectral imaging data .......... 42
4.5 Chapter summary .................................................................................... 44

5 Discussion and conclusions ...................................................................... 44
5.1 Theoretical implications ......................................................................... 45
  5.1.1 Close-range detection of till soil water content and chemical element
       concentrations .................................................................................. 45
  5.1.2 Remote sensing of boreal ground cover plant associations .............. 46
5.2 Practical implications .............................................................................. 49
  5.2.1 Close-range detection of till soil water content and chemical element
       concentrations .................................................................................. 49
  5.2.2 Remote sensing of boreal ground cover plant associations .............. 51
5.3 Reliability and validity ........................................................................... 52
5.4 Recommendations for future research ................................................ 55
5.5 Chapter summary .................................................................................... 56
Acknowledgements ....................................................................................... 57
References ..................................................................................................... 58
Original publications
LIST OF ORIGINAL PUBLICATIONS

This thesis consists of an overview and of the following publications, which are referred to by Roman numerals in the text.


The papers are PEER reviewed journal articles. Copyright of the published articles belongs to Elsevier and Taylor & Francis.

AUTHOR’S CONTRIBUTION

Maarit Middleton wrote the papers, and was the corresponding author of the papers being responsible for all contents and errors. The coauthors have contributed to the text of the manuscripts by correcting the contents and the language. In addition to the original papers, the dissertation includes previously published data which are further processed by a plan made by M. Middleton. The plan was executed by Paavo Närhi who conducted the analyses, made Figures 3 and 4, and calculated the statistics for Table 5 presented in this synopsis.

I Raimo Sutinen and M. Middleton chose the research topic and made the sampling plan. M. Middleton conducted the measurements of dielectric permittivity and spectral reflectance in the laboratory, and made Figures 1 and 3 and the tables. M. Middleton and Ari Teirilä designed the statistical analysis. A. Teirilä was responsible for conducting correlation analysis, mixed effects modeling, and making Figures 2, 4, and 5. M. Middleton interpreted the results and discussed them.

II The research topic and sampling were designed by R. Sutinen. M. Middleton was solely responsible for spectroscopic measurement design and conducting the measurements in the laboratory. She also made Figures 1, 2, and 3 and Table 2. P. Närhi made Figures 4, 5, and 6. Tables 1 and 3 were made by M. Middleton and P. Närhi together. M. Middleton and P. Närhi planned the statistical and chemometric analysis and P. Närhi executed them. M. Middleton and Viljo Kuosmanen interpreted the results and M. Middleton discussed them.

III M. Middleton and R. Sutinen formulated the research scope. R. Sutinen designed the hyperspectral data collection and M. Middleton and P. Närhi the field data sampling design and field data collection. They also acquired the dielectric and vegetation data in the field. M. Middleton was responsible for designing the data analysis methods. P. Närhi conducted the ordination analysis, accuracy assessment with receiver operating characteristics curves, and made Figures 2, 3, and 5.
M. Middleton conducted the spatial analysis with neural networks, and made the tables and Figures 1 and 4. The results were interpreted by M. Middleton and P. Närhi and conclusions were drawn by all authors.

IV The research was initiated by Hilkka Arkimaa, Eija Hyvönen, M. Middleton, and R. Sutinen. The HyMap data collection was organized by V. Kuosmanen and H. Arkimaa. She and E. Hyvönen participated in conducting the geometric rectification and H. Arkimaa and V. Kuosmanen in the atmospheric correction and bidirectional reflectance correction. M. Middleton, H. Arkimaa, E. Hyvönen, P. Närhi, and R. Sutinen designed the sampling scheme and field measurements and conducted the field work. Data analysis was designed by M. Middleton, P. Närhi, and Paul Treitz. P. Närhi conducted the ordination and ordination related statistical including Tables 1–3 and Figures 2–4. M. Middleton conducted the spatial classification with support vector machines (Fig. 1) and its accuracy assessment (Table 4, Fig. 1 and 6). The results were interpreted and discussed by M. Middleton (Fig. 4 and 5).

LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AISA</td>
<td>Airborne Imaging Spectrometer for Applications</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Networks</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the receiver operating characteristics Curve</td>
</tr>
<tr>
<td>BRT</td>
<td>Bagging Regression Trees</td>
</tr>
<tr>
<td>ex situ</td>
<td>out of original place</td>
</tr>
<tr>
<td>HyMap™</td>
<td>Hyperspectral Mapper</td>
</tr>
<tr>
<td>in situ</td>
<td>in place</td>
</tr>
<tr>
<td>LWIR</td>
<td>long wavelength infrared, 8–15 μm</td>
</tr>
<tr>
<td>MANOVA</td>
<td>permutational multivariate analysis of variance</td>
</tr>
<tr>
<td>MARS</td>
<td>Multivariate Adaptive Regression Splines</td>
</tr>
<tr>
<td>MVC</td>
<td>Multivariate Calibration Techniques</td>
</tr>
<tr>
<td>MWIR</td>
<td>mid wavelength infrared, 3–5 μm</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>NIR</td>
<td>near-infrared, 700–1700 nm</td>
</tr>
<tr>
<td>OMC</td>
<td>Organic Matter Content</td>
</tr>
<tr>
<td>PAM</td>
<td>Partitioning Around Medoids</td>
</tr>
<tr>
<td>PLSR</td>
<td>Partial Least Squares Regression</td>
</tr>
<tr>
<td>PNN</td>
<td>Probability Neural Network</td>
</tr>
<tr>
<td>RBFLN</td>
<td>Radial Basis Functional Link Network</td>
</tr>
<tr>
<td>RF</td>
<td>Random Forests</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristics</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machines</td>
</tr>
<tr>
<td>SWC</td>
<td>Soil Water Content</td>
</tr>
<tr>
<td>SWIR</td>
<td>short wavelength infrared, 1700–3000 nm</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet, 10–350 nm</td>
</tr>
<tr>
<td>VIS</td>
<td>visible, 350–700 nm</td>
</tr>
<tr>
<td>VSWIR</td>
<td>visible and short wavelength infrared, 350–2500 nm</td>
</tr>
<tr>
<td>XRD</td>
<td>X-ray diffractometry</td>
</tr>
</tbody>
</table>

LIST OF SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ε</td>
<td>dielectric permittivity</td>
</tr>
<tr>
<td>σ</td>
<td>electrical conductivity</td>
</tr>
<tr>
<td>λ</td>
<td>wavelength</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer</td>
</tr>
<tr>
<td>n</td>
<td>number of observations</td>
</tr>
<tr>
<td>μm</td>
<td>micrometer</td>
</tr>
<tr>
<td>p</td>
<td>statistical significance</td>
</tr>
<tr>
<td>r_s</td>
<td>Spearman correlation coefficient</td>
</tr>
<tr>
<td>R²</td>
<td>coefficient of multiple determination</td>
</tr>
<tr>
<td>R_adj</td>
<td>residual degrees of freedom based adjusted coefficient of multiple determination</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 Background and research environment

Until recently soil properties have gained only minor research attention as abiotic factors affecting the spatial distribution, abundance, and diversity of species in the boreal previously glaciated landscapes in Lapland, northern Finland (Cajander 1926, Kujala 1929, Sutinen et al. 2002a, Sutinen et al. 2002b, Sutinen et al. 2007a, Mäkitalo 2009). Climatic factors and effects of climate change have gained much more interest as factors contributing to biogeography of tree species (e.g., Holopainen et al. 1996, Veijola 1998, Kultti et al. 2006). Bedrock mineralogy and texture has a significant impact on the spatial heterogeneity of glacial till soil hydrology and nutrient potential (Koljonen 1992, Sutinen et al. 2007a). The extremes of the till soil gradients are coarse textured tills derived from felsic rocks and fine-grained tills derived from mafic rocks (Lintinen 1995). The finer the till texture is, the higher the ionic exchange capacity and nutrient potential will be but also soil water content (Timoney et al. 1993). The edaphic properties such as moisture and nutrient potential are driving factors for spatial distribution, diversity and abundance of tree species (Sutinen et al. 2002a, Sutinen et al. 2007a, Mäkitalo 2009). Till has a significant role in the species distribution in regional scales as glacial tills are the most common sediment type on Fennoscandian forest floors with 75% aerial coverage. Moreover, tills commonly exhibit high spatial variability, which emphasizes the spatial complexity of soils as substrate for boreal vegetation (Hänninen 1997, Penttinen 2000).

The glaciogenic sediments in Lapland act as substrate for the biosphere of the Scandinavian and Russian taiga ecoregion characterized by open canopy coniferous forests with Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies* L. Karst), downy birch (*Betula pubescens* Ehrh.), and their understorey dominated by shrubs, mosses and lichens, and peatlands, i.e., organic wetlands or mires. The dominant tree species thriving in dry nutrient poor tills are Scots pine and silver birch (*Betula pendula*), whereas Norway spruce and downy birch dominate in the wet and mesic nutrient-rich tills (Sutinen et al. 2007a). Similarly, understory species follow edaphic properties (Salmela et al. 2001, Närhi et al. 2011), as well as species on organic soils which follow patterns soil of moisture, nutrients, and also pH (Tahvanainen et al. 2002, Närhi et al. 2010).

Soils are a fundamental factor for the biosphere and survival in the North. From the local economic perspective, forestry and reindeer husbandry benefit from soils in northern Finland followed by agriculture, peat production, infrastructure development, and mineral industry. Mapping the properties of soils and plant assemblages should be facilitated to benefit the local industries and government. More accurate and cost efficient ways are needed since the most common practices require extensive field work, sampling, and expensive laboratory analysis. In this study, radiation sensed at close range or remotely in the visible and short wavelength infrared (VSWIR, 400–2500 nm) range of the electromagnetic range is explored. The study focuses on hyperspectral sensing which is considered as a suite of techniques recording the intensity of emitted energy reflected from a surface in many narrow bands within the optical wavelength range. The term close-range soil sensing, in this study, is considered as *ex situ* stationary laboratory spectroscopy. The ‘closeness’ (distance of centimeters or meters from a target) of the material and the non-destructive rapid nature of the measurement makes VSWIR spectroscopy a non-invasive and relatively inexpensive active close-range soil sensing tool. In addition, potentially several soil properties can be measured at one instance with minimal sample preparation and without hazardous chemicals.

The idea behind hyperspectral remote sensing in soil studies is to measure reflectance spectra in a spatially continuous way for the purpose of mapping the patterns of soil classes and properties cost-efficiently and consistently. To circumvent the vegetation canopy obscuration on exposed surfaces the published geologic and pedologic hyperspectral applications have focused on arid deserts, steppes, and savannas where mineral top soils are exposed (Ben-Dor et al. 2009, Mulder et al. 2011, Kruse 2012, Van der Meer et al. 2012). However, in densely vegetated areas, such as the boreal zone, the use of hyperspectral remote sensing is mostly limited to vegetation, more specifically for inventorying forest properties or land cover (Treitz et al. 2010). The notion of spatial interpolation of soil
properties between soil observation or samples with the aid of remote sensing has been more successful on sparsely vegetated areas than on densely vegetated areas. On densely vegetated areas, the use of optical spectroscopy relies on indirect retrieval using soil indicators such as plant functional groups (Schaepman et al. 2007), ‘optical types’ (Ustin & Gamon 2010), biogeographical gradients (Schmidt & Hewitt 2004, Mahecha & Schmidtlein 2008), geobotany (Ustin et al. 1999), indicator species (Mücher et al. 2009), productivity changes (Dobos 2006, Hansen et al. 2009), and Ellenberg indicator values (Schmidtlein et al. 2007).

The idea in using vegetation as a proxy of soil assumes that the biogeographical and geobotanical information could reveal underlying patterns of soil properties. This would be very helpful, e.g., in spatial assessment of selecting an appropriate tree species for artificial regeneration. The success rate of artificial regeneration for dry soil tolerant Scots pine could be significantly improved by including a soil water based assessment of site suitability in the reforestation planning process especially in Lapland where forest compartment sizes can be large. Previously it has been reported that 96.5% of the mature pine stands in Finnish Lapland occur on dry sites with dielectric permittivity (Ɛ) less than 15 (water content 0.27 cm³/cm³, Sutinen et al. 2002a, Sutinen et al. 2002b). The indirect link of soil-vegetation-remote sensing is a complex but fascinating research topic, and also rarely researched. Although there lies the potential in hyperspectral sensing for detecting properties of soils and other vegetation layers besides upper forest canopy (Salmea et al. 2001) the topic is poorly explored, especially in the boreal.

The limited exposure of soils in Lapland constrains the applications of direct hyperspectral sensing of soils merely to a close-range *ex situ* approach conducted in laboratory. The significance of lab spectrometry lays in cost reduction (see Nduwamungu et al. 2009) and selection of samples for more expensive soil measurements. The main idea behind spectroscopy and chemometrics, i.e., sciences of extracting information from spectrochemical systems by data-driven means, is that the spectra are measured of all samples and they are empirically calibrated with a conventionally determined subsample. The estimated quantities of the remaining samples are then obtained through application of the calibrated model on the rest of the samples. VSWIR spectroscopy has been shown to be a very promising tool for rapid estimation of multiple soil chemical, physical and biological properties both in field (Stevens et al. 2008, Bendor et al. 2009) and in laboratory (Sorensen & Dalsgaard 2005, Cécillon et al. 2009).

Currently, VSWIR and mid wavelength infrared (MWIR, 3–5 μm) spectroscopy and chemometrics are highly active research fields, particularly in precision agriculture (e.g., Muller & Décamps 2001, The Second Global Workshop on Proximal Remote Sensing 2011). In recent years, the chemometric methods have advanced greatly including also the non-linear multivariate regression (MVR, e.g., Artificial Neural Networks, ANN; Support Vector Machines, SVM, Viscarra Rossel & Behrens 2010), besides linear MVRs (e.g., partial least squares regression, PLSR, Geladi & Kowalski 1986, Martens & Naes 1992, Wold et al. 2001) and single variate regression (e.g., Dematté et al. 2006), for empirical calibration of the spectra with subsample of laboratory measured soil properties. Spectroscopy has been shown to have high potential for predicting soil mineralogy, soil texture, organic carbon (soil organic matter), soil moisture, nitrogen content, and clay content especially for agricultural soils (see reviews in Ben-Dor et al. 2009, Stenberg et al. 2010, Mulder et al. 2011, Sarkhot et al. 2011). The various forms of C are probably the most accurately predicted of all soil elements (e.g., Viscarra Rossel et al. 2006, Ludwig et al. 2008, Vasques et al. 2009, Sarkhot et al. 2011). On the other hand, variability in the success predicting, e.g., cation exchange capacity, base saturation, pH, exchangeable bases, extractable phosphorus, and hydraulic properties is high (Santra et al. 2009). The weak nature of the absorption features, overlapping of them, influence of constituents on the absorption features of other constituents, too wide or narrow variance of concentrations, geological variability, coating of the bigger particles by a variety of compounds and grains, and variability in the quality of the predicted property are factors contributing to highly variable success.

In northern Finland, close-range spectroscopy would most prominently benefit mineral exploration. Overall, the lack of information on suitability of chemometric predictions for mineral soil properties prevents the use of VSWIR spectroscopy for the purpose of mineral exploration and forest soil research. Moreover, soil spectroscopy has been shown to hold high potential as a mapping tool for a variety of soil properties of agricultural soils but
the application of soil spectroscopy in geochemical exploration and pedogeochemical research, however, is very rare. For mineral exploration purposes till samples are usually taken from relatively unweathered low-organic-matter podzolic C-horizon. As a matter of fact, the low organic matter content (OMC) would be beneficial for mineralogical and elemental quantification of soils with spectroscopy as organic matter has a tendency to mask absorption features by other soil constituents (Galvão & Vitorello 1998). The fine mineral powder fraction (<0.06 mm) of glacial tills is commonly used in geochemical exploration in glacial terrains because compared to coarse fraction the concentrations of trace elements are higher in the fine fraction (see Tarvainen 1995). Eventually, the purpose is to evaluate if reduction in the number of laboratory analyzed samples would be possible with the aid of spectroscopy especially in terms of the critical elements for mineral exploration, and thus cost-efficiency could be improved.

From a forestry perspective, the soil spectral reflectance could aid in tree species selection at sites formerly dominated by Norway spruce. Due to the high spatial variability of soil moisture, only subareas of these sites are usually suitable for pine. Dry soil tolerant Scots pine is preferred in artificial re-forestation over the other predominant species due to the higher timber value and productivity of pine over spruce and birch. A tendency to regenerate pine on former spruce-birch stands, which naturally also occur on mesic and wet soil water regimes (Sutinen et al. 2002a), has led to heterogeneous stocking success (Hansson & Karlman 1997, Witzell & Karlman 2000). Bare soils are routinely exposed by soil mounding or trenching in northern Finland. Information on soil moisture derived from the bare soil pixels would provide a means for evaluating the site suitability for Scots pine by using high spatial resolution optical remotely sensed data (Siira et al. 2000b). In situ ε would be practical for field referencing of such remotely sensed data, as the ε measurements can be efficiently conducted and are immediately usable for model calibration. However, the relationship between soil ε and spectral reflectance has not yet been reported.

Besides upland mineral soil properties, hyperspectral data are also expected to be applicable for peatland studies. Peatland soils have global importance as a large carbon stock (Yu et al. 2011). Emissions of carbon dioxide, methane and nitrous oxide are controlled by peatland ground water and nutrient levels, and dominant species associations (Moore & Knowles 1989, Silvola et al. 1996). Understanding them, as a function of increasing trophic level (nutrient-richness: oligotrophic<mesotrophic<eutrophic) and surface moisture, would be important as global change is predicted to have impacts on peatland hydrology and the resulting changes may act as a major triggering mechanism for greenhouse gas release. Spatial representations of species associations would, therefore, be beneficial as inputs into models of greenhouse gas emissions and into regional conservation plans. Furthermore, Finland is missing a national wetland inventory of the ecological surface properties although they cover over 30% of the land area.

Hyperspectral remote sensing could potentially provide data for such mapping as hyperspectral data should be available from satellites in the near future (Buckingham & Staenz 2008). Hyperspectral data are spectrally superior to other optical datasets. Therefore, a high number of peatland site types can be expected to be classified. The optimal class hierarchy for peatland site types, determined with the traditional separability analysis, could be a laborious task. This would be especially true with boreal peatlands which commonly exhibit gradiental nature. Thus purely data driven methods which consider the aspects of nutrient-potential and moisture for determining optimal class hierarchical levels are needed in order to process hyperspectral data for peatland classifications effectively.

1.2 Objectives and scope

The main motive behind the research is the lack of understanding of hyperspectral close-range and remote sensing in applications that are relevant to research of soil properties. Compared to other optical datasets, hyperspectral data have superior spectral resolution, and thus has significantly higher potential of identifying mineral materials and plant species. These methods of optical spectroscopy, utilizing the wavelength range 350–2500 nm, have shown potential in a variety of soil related applications across biomes on the Earth. However, this study focuses on applications
Hyperspectral close-range and remote sensing of soils and related plant associations. Spectroscopic applications in the boreal environment

relevant in the boreal environment of Finnish Lapland. The aim of the study is on practical applications on site specific and target scales rather than on regional or global scales. In this study strictly soil related abiotic factors, i.e., non-living chemical and physical factors in the environment, which affect ecosystems, were considered. Resource gradients (i.e., light, nutrient contents, carbon dioxide, and oxygen), direct gradients (i.e., temperature) or indirect gradients (i.e., distance to peatland margin, see Franklin, 1995, for explanation) are not considered in this study. The main interest is in soil properties such as soil moisture and nutrient potential (measured as bulk soil electrical conductivity, \( \sigma \), and elemental concentrations) of both mineral and organic soils. The hyperspectral close-range soil sensing is used for bare mineral soil spectra. Spatial patterns of species associations and indirect detection of soil properties are imaged with airborne remote sensing.

The main research problem addressed in this dissertation is stated as follows:

*How can hyperspectral spectroscopy be used for quantifying and classifying soil properties and plant associations dependent on the soil properties in the boreal environment?*

The problem is studied from the perspective of four potential practical applications of hyperspectral data based on which four research questions (RQ) were formed (Table 1). Each RQ is answered with a journal article (Table 2). The contributions of the articles are combined in the dissertation discussion.

### Table 1. Research questions.

<table>
<thead>
<tr>
<th>RQ #</th>
<th>Research question</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ 1</td>
<td>What is the most appropriate empirical modeling technique to predict bulk soil dielectric permittivity of glacial till soils from VSWIR spectral reflectance measurements?</td>
</tr>
<tr>
<td>RQ 2</td>
<td>Which element concentrations of glacial till soils can be predicted with VSWIR spectral reflectance?</td>
</tr>
<tr>
<td>RQ 3</td>
<td>Can soil moisture based suitability for artificial regeneration to Scots pine be mapped spatially on clear cuts using hyperspectral imaging of ground cover plant species patterns?</td>
</tr>
<tr>
<td>RQ 4</td>
<td>How to determine an optimal peatland class hierarchical level for medium resolution hyperspectral imaging spectroscopic data?</td>
</tr>
</tbody>
</table>

### Table 2. Overview of research papers.

<table>
<thead>
<tr>
<th>Paper</th>
<th>RQ #</th>
<th>Title</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>RQ 1</td>
<td>Response of reflectance to dielectric properties of bare tills</td>
<td>International Journal of Remote Sensing</td>
</tr>
<tr>
<td>II</td>
<td>RQ 2</td>
<td>Quantification of glacial till chemical composition by reflectance spectroscopy</td>
<td>Applied Geochemistry</td>
</tr>
<tr>
<td>III</td>
<td>RQ 3</td>
<td>Imaging spectroscopy in soil-water based site suitability assessment for artificial regeneration to Scots pine</td>
<td>ISPRS Journal of Photogrammetry and Remote Sensing</td>
</tr>
<tr>
<td>IV</td>
<td>RQ 4</td>
<td>Ordination and hyperspectral remote sensing approach to classify peatland biotopes along soil moisture and fertility gradients</td>
<td>Remote Sensing of Environment</td>
</tr>
</tbody>
</table>

### 1.3 Research realization and dissertation structure

This research was initiated after years of working with hyperspectral data in projects which examined the applicability of hyperspectral data in geoenvironmental studies. During the work, it was realized that summarizing and reporting the hyperspectral applications suitable to boreal environments would be necessary to increase awareness of spectrometry and its applications which have a geological or ecological background or significance. Close-range hyperspectral data were...
acquired in laboratory conditions from soil samples gathered from the field. Mineralogical and chemical analyses were conducted for referencing the hyperspectral data. Two airborne campaigns for acquiring the remotely sensed hyperspectral data were organized. In situ field referencing was done with geophysical measurements and vegetation analysis. Data were interpreted with statistical, data mining, chemometric, and image processing methods. Finally, conclusions were drawn by reviewing the results against existing literature. The research process typical to all four papers follows the sequence of: literature review, data gathering, data analysis with quantitative methods, and comparison of the results to existing literature.

This compilation dissertation consists of four individual papers and the summary of their results. This summary is organized as follows: Chapter 2 presents the theoretical foundation of VSWIR spectroscopy and each RQ. Chapter 3 shortly reviews the research methods and materials. The most important results answering each RQ are summarized in chapter 4. Finally, chapter 5 discusses the results from theoretical and practical points of view, reviews the reliability and validity of study, and suggests some future research.

2 THEORETICAL FOUNDATION

Soil and plant spectra are a sum of inherent scattering and absorption properties of their constituents. The soil matrix is composed of mineral and organic matter, and the air and water filling their pore spaces. Soil spectral response is primarily a product of mineral particles as minerals approximately account for half of the soil volume (Hillel 1998). Grain size, water, organic matter content, and soil precipitates also play a significant role in the overall soil spectra. Similarly, plant spectra are composed of a variety of plant properties and constituents: pigments, proteins, cellulose, lignin, water, tissue structure, and canopy structure. Therefore, any measured spectrum is a mix of spectrally active components.

The following two chapters 2.1 and 2.2 explain the theoretical framework of soil and plant spectra. Then brief summaries of hyperspectral close-range and imaging spectroscopy are given in chapter 2.3. Finally, the four RQs are explained with respect to previous studies. As background to the RQ 1, the relationship between soil water and spectral reflectance is explained, in addition the strengths and weaknesses of different regression approaches are discussed in chapter 2.4. To answer RQ 2 the mineralogy of tills is discussed in chapter 2.5 as the prediction of chemical elements concentrations is based on indirect information from the spectrally active mineralogical components in till. Chapter 2.6 reviews the environmental drivers for distribution of understory plant species and principles of soil hydrology relevant to understanding why remotely sensed spectra can be indirectly used for estimation of soil moisture in order to answer RQ 3. Steps towards analyzing the usability of satellite based hyperspectral data and developing data classification methods are needed for the purpose of nationwide peatland mapping. To answer the RQ 4 particularly, chapter 2.7 explains the needs for a data processing chain for resolving the optimal level of class detail of the peatland site types. This would be extremely useful for the future goal of an ecological national wetland inventory.

2.1 Spectra of soil

Soil VSWIR spectra is generally rather flat and characterized by a few broad weak absorption features (Fig. 1A). They are caused by electronic processes, including crystal field transitions, electron transfer between transition metal ions and charge transfer between transition metal ions and anions, and molecular vibrational processes of the material (Hunt 1977). The overtones and combinations of molecular fundamental vibrations in the MWIR are the main sources of absorption features in the VSWIR. In the visible (VIS, 350–700 nm) range, on the contrary, the primary process causing absorption is electronic excitation. Extensive reviews of the soil absorption features caused by the main overtones of OH, SO₄, and CO₃, and combinations of H₂O and CO₂ are presented by Clark (1999) and Stenberg et al. (2010).

The soil particles are larger than the VSWIR
wavelengths, which causes soils to exhibit diffuse reflectance. When radiation interacts with dry soil a portion of it reflects from particle surfaces and the rest penetrates into the particles, where it is either absorbed or scattered. The scattered spectral reflectance is thus a sum of the specular reflectance and internal volume reflectance. Soils exhibit bidirectional reflectance properties (Cierniewski & Courault 1993) thus observation conditions including illumination and observation direction impact soil reflectance. Surface roughness, surface aggregation, and the way the components are arranged to form the structure of the soil, have a significant effect on the overall reflectance (e.g., Wu et al. 2009).

The till spectra as a combination of their mineral bearing constituents is not fully understood. In till fine fraction, which was the interest of this research, the rock forming primary minerals such as quartz, alkali feldspars, plagioclase, and amphibole dominate. Secondary minerals such as kaolinite, mixed-layer clay minerals such as vermiculite, chlorite, illite and swelling-lattice vermiculite may also be present (Pulkkinen 2004, paper II). Quartz is featureless in VSWIR having overall high reflectance (Fig. 1B, Hunt & Salisbury 1970). Feldspars are also highly reflective (e.g., albite in Fig. 1B) but plagioclase, in the presence of Fe(II), can have a broad absorption around 1100–1300 nm (Crow & Pieters 1987). The OH-bearing minerals such as amphiboles (e.g., termolite in Fig. 1B), chlorite, talc, and illite are characterized by a number of absorption features between 2000–2500 nm (Fig. 1B). They are caused by combination vibrations of hydroxyl stretch and metal-OH bend of AOH at 2200 nm, MgOH at 2300 nm (and FeOH at 2290 nm) for till bearing layered silicates such as chlorite, talc and illite (Farmer 1974, in Clark 1999). When present, carbonates strongly absorb at 2335 nm and 2500 nm, and weakly at 1870 nm, 1990 nm and 2160 nm (Clark 1999). Structural and electrically bound molecular water may partially mask the OH-absorption (Hunt 1977).

The Fe related absorption features in soils dominate the ultraviolet (UV, 10–350 nm) and VIS part of the spectrum. Features by Fe(III)- and Fe(II)-bearing constituents are situated at approximately 670 nm and 950 nm (Hunt et al. 1971). Iron(III) produces crystal field transition at 450 nm, 550–650 nm and 750–950 nm (e.g., ferrihydrite in Fig. 1A) whereas absorption bands caused by Fe(II) crystal field transitions can be found at 450 nm, 510 nm, 550 nm, and 900–1100 nm (e.g., goethite in Fig. 1A, Hunt & Ashley 1979). The Fe ions can also be present in layered minerals such as biotite, muscovite, sericite, and mixed-layer clay minerals such as mica-vermiculite and AOH-interlayer-vermiculite, which could be present in till soil horizons (see Righi et al. 1997, Peuraniemi et al. 1997). These secondary minerals may be products of podzolisation, which results in crystalline Fe and amorphous Fe precipitates. They may coat mineral grains and therefore mask the absorbance features of the primary minerals. In soils, Al and Mg, which have an overall high VSWIR reflectance, can be present in secondary precipitates thus affecting the entire spectra. Electronic excitations also by other transition metals ions such as Cr, Co, and Ni can cause absorption in the VIS range (Clark 1999).

Particle size has an effect on the depth of the absorption features and the level of the overall reflectance. Larger mineral grains absorb more light than smaller grains causing decreased reflectance. Absorption depth increases with an increase mineral grain size as with larger grains greater internal path causes more light to be absorbed. With the smaller grains surface reflection dominates over internal scattering, which makes the absorption features shallower compared to coarse grains. (Clark 1999, Van der Meer et al. 2012)

In the case of moist soils, water molecules are in the lattices of minerals and electrically bound on the surfaces of the particles in the matrix. Solid particles can be surrounded by several layers of molecules until free water appears into the pore spaces. This is observed as decreased reflectance throughout the VSWIR spectrum (Fig. 1A, Bowers & Hanks 1965, Whiting et al. 2004), and better known as visual darkening (Ångström 1925, Planet 1970). Darkening is caused by multiple mechanisms some of them might still be unknown (Twomey et al. 1986, Lekner & Dorf 1988, Zhang & Voss 2006). Soils darken with increasing soil water content (SWC) only until a certain moisture level is reached. Beyond a so-called cut-off thickness the SWC-reflectance relationship reverts, i.e., the reflectance starts to increase with surface moisture. The cut-off thickness is defined as a thickness of soil mixture that transmits only 5% of the incident light. This point is close to SWC at hygrosopic limit where water film covers all the particles (Neema et al. 1987), and also SWC at the field holding capacity, which typically lies within a
soil moisture range of 0.15–0.40 g/cm³ (Zhu 1984, in Weidong et al. 2002). The dynamics of SWC-reflectance relationship seem to be closely related to soil type and used wavelengths. At the shorter wavelengths, where scattering processes dominate, reflectance change rate is very rapid. At the longer wavelengths (λ>1900 nm), where water absorption have a greater impact, variation is more progressive (Weidong et al. 2002). This supports the use of short wavelength infrared (SWIR, 1000–3000 nm) for soil moisture prediction especially at low moisture levels. Shorter wavelengths would be more appropriate for moisture estimation at high moisture levels.

In addition to the general darkening of the soils, the hydroxyl bands reveal the presence of water in soils. They are seen as very strong specific overtone and combination absorption bands around at 1400 nm and 1920–1935 nm, and weaker bands 450 nm, 574 nm, 760 nm, 986 nm (942–1135 nm), 1200 nm, 1379–1450 nm, 1672 nm, 1800 nm, 2189 nm, and 2400 nm (Fig. 1A, Hunt 1977, Bishop et al. 1994, Ben-Dor et al. 1999, Weidong et al. 2002, Demattê et al. 2006). The presence of the weaker bands is dependent on the mineralogy of the soils, i.e., the presence of mineral-OH bonds. The effect of increasing water is also seen as widening and deepening of the absorption features especially at 1400 nm but also less intensively at 1900 nm (Demattê et al. 2006). In contradiction, hydroxyl absorption at 2200 nm has shown to exhibit a less evident behavior with increasing water content (Ben-Dor et al. 1999, Demattê et al. 2006). Another fundamental water absorption band is centered at 2800 nm and an OH combination band at 3100 nm. They are located beyond the SWIR range considered in this paper but they have a wing effect on the spectra at λ<2500 nm (Bishop et al. 1994).

Similarly to water, soil organic compounds have their fundamental absorption features in MWIR, and their overtones and combinations appear as numerous weak absorption features throughout the VSWIR (Fig. 1A, Stenberg et al. 2010). However, overall decrease in reflectance due to OMC is generally observed in the VIS range as organic matter has a broad absorption in those wavelengths. Several bands have been recognized to be important for OMC in SWIR but especially 1000 nm, 1600 nm, 1799 nm, 1800 nm, 2000 nm, and 2200–2400 nm are related to organic molecules and proteins with O-H, C-C, C-H, C=O, and N-H (Coleman & Montgomery 1987, Henderson et al. 1992, Ben-Dor & Banin 1995, Clark 1999, Viscarra Rossel & Behrens 2010, Stenberg 2010).
2.2 Spectra of plant species in shrub, herb and bryophyte layers

Plant spectra are characterized by low reflectance (<5%) in the VIS, and especially around 680 nm, caused by absorption of plant pigments (Fig. 1C). Radiative transfer modeling results indicate that chlorophyll a and b contents causes about 60% of the canopy reflectance variation in VIS wavelengths (Jacquemoud et al. 2009). Spectroscopy has been used for quantification of vegetation biochemical constituents such as plant pigments: chlorophyll a and b, carotenoids, anthocyanides, and nitrogen at the leaf and canopy levels successfully (see summary Ustin et al. 2009). Pigments have absorbance features in the VIS but nitrogen is found to occur in chlorophyll molecules thus enabling its detection through cross-correlation with chlorophyll absorption (Daughtry et al. 2000). The most unique property of plant spectral reflectance occurs at red-edge, between 680 nm and 730 nm, which is characterized as the most extreme reflectance gradient in naturally occurring materials (Fig. 1C). Broadening and deepening of chlorophyll absorption feature between 660–759 nm as a result of the increase in chlorophyll content is related to a shift of the red edge towards the longer wavelengths (Vogelmann & Moss 1993). The high spectral reflectance (<50%) in near-infrared (NIR, 700–1700 nm) is associated with scattering from the internal tissue structure, i.e., air-water interfaces, and consequently lower NIR reflectance. On the canopy level, leaf angle distribution and leaf area index (effect of dry matter) both have been estimated to contribute equally up to 40% of the NIR reflectance (Jacquemoud et al. 2009) causing the correlation between normalized difference vegetation index (NDVI) and amount of biomass (Lillesand et al. 2008). At the same time, quantification of other leaf constituents such as water, protein, cellulose, and lignin has been successful using NIR wavelengths (see summary Kokaly et al. 2009). In the SWIR, water on average contributes 50% to the reflectance (Jacquemoud et al. 2009). Similarly to soils, vegetation has the primary specific absorption features at 1450 nm, 1940 nm, and 2700 nm, and secondary features throughout the spectra at 480 nm, 680 nm, 960 nm, 1120 nm, 1540 nm, 1670 nm, and 2200 nm (Wessman 1990, Carter 1991). On the canopy level, the effect of leaf area index on SWIR can be up to 60% around 1450 nm and 1950–2500 nm where the absorption coefficient of water is high (Jacquemoud et al. 2009).

Despite the ability of hyperspectral remote sensing to discriminate between landscape components, difficulties arise when separating vegetation species. Within species reflectance varies depending on factors such as season, illumination, nutrient regime, and moisture availability. These factors contribute to the changing of single species reflectance in space and time. The main contributors to the failure to separate and identify plant species is the spectral similarity of plant species, and the inadequacy of hyperspectral sensors in spatial, spectral, and temporal resolutions (Treitz et al. 2010).

Besides the green vegetation components, several other ground cover components contribute to the reflectance in the boreal zone. Some example spectra are presented in Figure 1C. The samples were gathered at the study site of paper III and were measured in laboratory with a FieldSpec spectroradiometer (see chapter 3.3.1). Moss spectra resembles green vegetation spectra but are highly variable between species, within species, and even within an individual according to genotype, light, nutrient, and water status (Fig. 1C, Harris et al. 2006b, Van Gaalen et al. 2007, Arkiama et al. 2009). Lichens, on the other hand, have negligible red-edge as NIR reflectance is lowered compared to green vegetation (Fig. 1C, Rees et al. 2004, Peltoniemi et al. 2005). Plant litter, similarly to organic matter (Coûteaux et al. 2003), is characterized by several broad and narrow absorption features in the NIR and SWIR caused by cellulose, hemicellulose, lignin, starch, pectins, waxes or other structural compounds (Fig. 1A, 1C, Elvidge 1990). Strong absorption features, associated to cellulose and lignin, are at 2100 nm and 2300 nm (Elvidge 1990, Nagler et al. 2000). Another diagnostic feature of the dry plant material is an absorption wing in the VIS portion of the spectrum produced by the intensive absorption of lignin and humic acid in the blue and UV ranges. Collapse of cell structure, caused by desiccated water, induces reduced NIR reflectance compared to green leaves. Burnt litter, organic soil, and wood with charcoal top gives a distinct ‘dark’ visual effect (Fig. 1C, Lang et al. 2002).
2.3 Spectroradiometry

Spectroradiometry is the technique of measuring the spectrum of radiation emitted by a source. The use of the term ‘hyperspectral’ is more vague and undefined in the field of spectroscopy indicating a strong need for a proper internationally acknowledged terminology definition. ‘Hyperspectral’ is a ‘catch phrase’ which is established for scientific use. In the strictest meaning, it is associated with high number spectrally narrow (i.e., low full-width-half-maximum) spectral bands which are contiguously spaced in the optical wavelength range. In the broad sense, the number of bands is commonly more than ten and the data are not necessarily contiguous in the spectral domain. Most recently, with the developing sensor technology, data at the MWIR and long wavelength infrared (LWIR, 8–14 μm) and ranges are also considered hyperspectral if bands are spectrally narrow and numerous. In this study, hyperspectral data are considered in this broad sense, including all spectral data in the VSWIR domain with a higher number of bands and narrow bandwidths compared to multispectral data.

Remote sensing or remote detection is considered as a collection of information about an object without making physical contact with it (Rees 2001). Remotely sensed hyperspectral data including the spatial domain can be acquired with imaging spectroscopy, hyperspectral remote sensing, ultraspectral imaging or hyperspectral imaging as proposed by Goetz et al. (1985). They define imaging spectrometry as “acquisition of images in hundreds of contiguous, registered, spectral bands such that for each pixel a radiance spectrum can be derived”. Whereas Schaepman (2007) suggests the following definition for imaging spectroscopy: “simultaneous acquisition of spatially co-registered images, in many narrow, spectrally continuous bands, measured in calibrated radiance units, from a remotely operated platform”. Often imaging sensors are programmable to record a selected number of bands, on selected spectral ranges with variable bandwidths. In this work, spectrometry conducted in tacked to a measurement target is considered as close-range sensing.

In analytical chemistry hyperspectral close-range spectroscopy is better known as near-infrared analysis (NIRS). In chemical spectroscopy the NIR region is considered to cover the wavelength range 750–2500 nm whereas in remote sensing terminology NIR region covers only the range 700–1700 nm which is followed by SWIR at 1700–3000 nm. NIRS was first developed for analysis of agricultural commodities in the 1960’s (Norris et al. 1976). The history of hyperspectral close-range detection of soil properties also dates back to 1960’s (Obukhov & Orlov 1964, Bowers & Hanks 1965, Huete & Escadafal 1991), and hyperspectral remote sensing of mineral Earth materials to the 1970’s and 1980’s (Abrams et al. 1977, Goetz et al. 1985). Development of imaging spectrometers closely relate to geology as they were primarily developed for the purpose of mineral mapping of soils and outcrops in the 1970’s and 1980’s (Goetz 2009). Nowadays, hyperspectral sensing is used extensively for routine analysis of raw materials, in-process materials, and quality control, e.g., in the food, polymer, pharmaceutical, textile, and petroleum industries (Davies 1998).

Several advantages lie in spectroscopy. The high spectral resolution of hyperspectral data offers the potential to discriminate between materials which have subtle spectral signature differences. Therefore, more information is potentially available from high-resolution spectra with a broad coverage of wavelengths. Compared to multispectral data, hyperspectral data give a greater likelihood of making an optimal choice of wavebands for discriminating between materials or relating the optical properties to the property of the surface (Rees et al. 2004). Hyperspectral data are acquired non-invasively and non-destructively allowing monitoring of several surface properties at once. Although currently quite expensive and difficult to obtain, the availability of hyperspectral remote sensing systems is improving steadily. On the contrary, the major disadvantage of hyperspectral information is shallow penetration. The technology is still developing thus the most affordable data are often characterized by low signal-to-noise (S-N) ratio. Even though capable of deriving several surface properties at once the properties also may influence and obstruct detection of other surface properties. Hyperspectral data are currently lower in spatial resolution compared to high resolution multispectral data. Thus the field-of-view is often a mixture of several features of interest, which requires special image processing methods for ‘unmixing’ the data to acquire the needed information. Similarly to other optical methods, spectral
reflectance is dependent on the soil-target-sensor geometry making it difficult to compare the data from different platforms, configurations, illumination settings, and timing of acquisition.

### 2.4 Prediction of soil water content from VSWIR spectra

Estimating SWC from a single spectral reflectance variable by regression analysis has been shown to be statistically feasible using a variety of reflectance factors (e.g., Haubrock et al. 2008). The overtone absorption feature around 1900 nm has proven to be especially powerful for estimating SWC in the laboratory (Weidong et al. 2002). Several studies reveal the dependence between reflectance and SWC being non-linear, often exponential, regardless of moisture being expressed as volumetric water content (Lobell & Asner 2002, Weidong et al. 2002), gravimetric water content (Whiting et al. 2004), or degree of saturation (Lobell & Asner 2002). However, linear relationships are often reported but on a narrow range of SWCs (<35%, Muller & Décamps 2001). A problem using this relationship for practical purposes arises when SWC at the cut-off thickness is unknown. Different models should be constructed for predicting soil moisture on both sides of this critical point but it would be practical only for SWC monitoring. When mapping SMC the spatial variation in soil co-variates (e.g., texture, OMC, mineralogy) also means spatial variation in the position of the cut-off thickness. In other words, the inflection point of the SWC-reflectance relationship varies. Therefore, using single variable models for mapping spatial distribution of SWC might not be effective. The use of single reflectance values as well as overall albedo, when estimating moisture for a set of field samples, is unreliable since co-variates can affect soil brightness significantly, especially in the VIS-NIR region.

Utilizing water absorption depth in soil moisture estimations is only possible for monitoring soil moisture of a specific sample in laboratory conditions and when other co-variates of soil stay constant. Several workarounds for the dilemma are presented. Continuum removal could be performed prior to modeling in order to get relative absorption band depths (Clark & Roush 1984). Dematté et al. (2006) utilized the area of an absorption feature. They found a linear relationship between SWC and the area of the water absorption features around 1400 nm, 1900 nm, and 2200 nm. Removing the effect of soil co-variates could also be done by measuring the dry spectra of each soil sample (Weidong et al. 2002). However, this requires prior information regarding the soil type or dehydration of the soil sample for a dry spectrum measurement limiting the application to laboratory use. One of the most promising approaches is presented by Whiting et al. (2004). Their approach relies on the water absorption feature at 2800 nm which decreases the reflectance over the entire range of NIR and SWIR (Bishop et al. 1994, Whiting et al. 2004, Haubrock et al. 2008). A Gaussian curve is fit though the convex hulls and the slope of the curve is regressed to SWC successfully.

The idea behind using multiband approaches in chemometric regressions is that incorporation of multiple bands generally makes the regression models more robust compared to using only single band reflectance. Multiband approaches also normalize the effect of overall reflectance variation due to soil co-variates (Haubrock et al. 2008). The SWC has been modeled from multiband reflectance measurements in several different ways. Soil moisture indexes have been successfully used for modeling the SWC (e.g., Nocita et al. 2011). Haubrock et al. (2008) review the common soil moisture indices and other quantification methods applying two or more bands which are mostly used for laboratory data. The advantage of the MVR techniques is that the entire spectra can be input to model SWC. It is argued that different wavelengths can be more valuable for predicting low SWCs than the high SWCs. The longer wavelengths are better suited for estimating volumetric moisture contents at >20% as reflectance saturates at much lower moisture levels at VIS-NIR wavelengths than on SWIR wavelengths (Lobell & Asner 2002, Haubrock et al. 2008). This trend is valid for several soil classes but with varying intensities. Weidong et al. (2002) present the successful use of longer SWIR wavelengths (>1900 nm) for soil moisture prediction, especially at low moisture levels (<0.3 g/cm³), similarly to Neema et al. (1987). Shorter wavelengths would be more appropriate for moisture estimation at high moisture levels. The fact that the most optimal wavelength ranges are
different for predicting soil moisture at low and high water contents supports the use of the entire measured spectral range for SWC predictions. An additional advantage of all multiband techniques is that water absorption bands, which are influenced by the atmospheric water vapor, can be excluded. This enables the use of the prediction techniques in field conditions. A few studies yet show the superiority of MVR and, especially the nonlinear data mining techniques, for predictions of soil properties (Mouazen et al. 2010, Viscarra Rossel & Behrens 2010).

The classical partial least squares regression (PLSR) has been used in chemometrics to predict concentrations of constituents for a couple of decades now (Geladi & Kowalski 1986, Martens & Naes 1992, Wold et al. 2001). The idea is to input an appropriate spectral range, which can include an unlimited number of spectral variables. The currently most advanced chemometric MVR techniques such as boosted regression trees (BRT) are shown to be effective for soil predictions, specifically for large datasets and wide range of values (Brown et al. 2006, Viscarra Rossel & Behrens 2010). Although the number of input variables can be high in spectroscopic data, the number of observations might still remain relatively low. Therefore, a MVR technique which could also produce stable results with a range of sample sizes would be optimal for predicting SWC spatially. It would also be important to consider if till-specific models are needed or if a global model would be accurate enough. The value of finding a robust prediction method and understanding the \( \varepsilon \)-reflectance relationship would be evident in Lapland, e.g., when artificial regeneration to dry soil tolerant Scots pine is being considered on clear-cuts with high soil moisture variability. The SWC of mechanically exposed soil pixels in high spatial resolution optical data could be modeled.

### 2.5 Prediction of till soil chemical elements from VSWIR spectra

Although reflectance spectroscopy has proven to be useful in quantifying many properties of mineral and organic soils (Ben-Dor et al. 2009, Stenberg 2010) the number of studies dealing with chemical element concentrations of mineral soils is few. The overtones and combinations in the NIR are much weaker than the fundamental MWIR bands, and can be difficult to distinguish. Their location might also have shifted from the theoretical location because molecules do not exhibit harmonic vibration. Therefore, interpretations of VSWIR spectra can be difficult, particularly when the target species is present in small amounts or when the soil property being studied does not exhibit absorption features but it can only be detected through correlation to the target species. The low cost of spectroscopy, however, is the driving force for applying it to new research problems.

In contradiction to the soil SWC predictions, chemical elements rarely produce absorption features. Absorption features of the chemical elements appear through molecular bonding, e.g., mineralogy. Several rock forming minerals have spectral absorption features that allow for their quantification. Mainly, it is the features associated with transition metals, OH, and water and combinations of and, \( \text{CO}_3 \), \( \text{PO}_4 \), \( \text{BO}_3 \), and \( \text{SO}_4 \) (Hunt 1977, Hunt & Ashley 1979, Crowley et al. 2003) that allow for the development of the calibration models. The fine fraction (<0.06 mm) of glacial tills in northern Finland is mainly composed of rock-forming minerals such as quartz, plagioclase, amphibole, microcline and low amounts of secondary minerals such as chlorite and illite, vermiculite, chlorite, and swelling-lattice vermiculite (Pulkkinen 2004). As discussed in chapter 2.1, silicates (quartz, plagioclase, microcline) are characterized by high reflectance and by lack of absorption features in the VSWIR. Amphiboles and secondary minerals can be distinguished by their absorption features. Thus predictions of till chemical element concentration merely rely on the spectrally active components. The OH-bearing secondary minerals have a number of absorption features in 2000–2500 nm range. However, one of the few chemical elements that has a pronounced absorption features at UV and VIS ranges is Fe. Electronic transfers of Fe ions cause absorption especially at 440 nm, 670 nm, and 900–1100 nm (Crowley et al. 2003, Richter et al. 2009). Concave spectral shape around 850 nm is related also to crystalline form of iron but amorphous iron does not exhibit this particular feature (Demattê et al. 2006).

Even in the absence of elemental absorption features, concentrations of several soil chemical elements have been successfully predicted using
VSWIR. Stenberg et al. (2010) review the literature on soil nutrient predictions for agricultural purposes and conclude that determination of soil chemical elements can be successful but not consistent. Concentrations of organic C (see Cozzolino & Morón 2003, Terhoeven-Urselmans et al. 2008), total and organic N (see Cozzolino & Morón 2003) are particularly well estimated across a variety of soil types. Success has been reported to be highly variable for mineral N, available K, exchangeable K, Ca, Fe, Na, Mg, P, and some micronutrients (Cozzolino & Morón 2003, Stenberg et al. 2010). Prediction of contaminant metals has also been quite successful (Kemper & Sommer 2002, Wu et al. 2007, Bray et al. 2009, Stenberg et al. 2010). The occasionally successful calibrations are attributed to locally present co-variation to spectrally active constituents. This phenomena is called cross-correlation. Such co-variations are naturally sample set specific.

2.6 Mapping of site suitability for artificial regeneration of Scots pine from hyperspectral imaging of soil water based species patterns

The role of understory in forestry related hyperspectral remote sensing studies has, thus far, been insignificant. Research has focused on inventory of the upper canopy, whereas the focus regarding understory has been on removing its effect on forest canopy estimations (e.g., Pisek et al. 2010). With the arrival of high spatial and spectral resolution VSWIR data the interest in understory related applications has grown. Remote sensing based understory inventory has been suggested as a means to improve estimates of, e.g., site productivity (Lieffers et al. 1996). Remotely sensed data have also been demonstrated to improve mapping of invasive understory species (Joshi et al. 2006) and species vital to conserving the habitat of endangered mammals (Linderman et al. 2004). In Finnish Lapland, the quality of reindeer pastures has been evaluated with remotely sensed optical data (Kumpula 2006).

VSWIR detection of the understory has previously been used as an indirect method to map soil conditions. McMichael et al. (1999) suggest that spectrally described vegetation is an indirect indicator of soil environmental CO₂ conditions in the undisturbed ecosystem of Alaska. Schmidtlein & Sassin (2004) and Schmidtlein (2005) modeled the Ellenberg indicator values from hyperspectral imagery in estimation of underlying soil quality in alpine meadow habitats. However, reports on soil moisture modeling from remotely sensed patterns of understory are rare. Ritari & Saukkola’s (1985) study, conducted in southern Lapland, aims towards this goal. They showed the spectral disparity between dry nutrient-poor sites and moist nutrient-rich sites.

The occurrence, diversity and abundance of plant species in the boreal zone is controlled by climatic (precipitation, temperature, wind, snow cover, photosynthetically active radiation), edaphic (coarse fraction content, aeration, soil temperature, soil nutrients, and pH), biotic (competition), and/or orographic (slope, aspect) variables, as well as the state of the forest succession after natural and/or anthropogenic disturbances (Archibold 1994, Wieser & Tausz 2007). As demonstrated on an uniform climatic area of central Lapland, Ca concentration and Ca:Al ratio are the main edaphic variables determining forest vegetation composition and diversity (Närhi et al. 2011). However, a couple of examples in the Pacific Northwestern mountains of the USA (Lookingbill et al. 2004), and in subboreal (Wang 2000) and subarctic northwestern Canada (Timoney et al. 1993) suggests surveying of understory species as a means of revealing relative spatial differences in SWC. According to Salmela et al. (2001) soil water can explain 42% of a species coverage and 60% of a species occurrence in central Lapland. This may open up a possibility to map the spatial distribution of the understory species associations with remotely sensed methods to reveal the underlying patterns of soil moisture. Soil moisture mapping could be turned into thematic information supporting species selection in artificial regeneration of formerly Norway spruce-downy birch dominated sites which are spatially heterogeneous in soil moisture.

The basis for the artificial regeneration site suitability analysis for Scots pine is the temporal persistence, i.e., time stability of the SWC patterns (Sutinen et al. 2007b). Time stability of soil moisture means that well drained soils with high hydraulic conductivities are characterized by low SWCs throughout the year, whereas poorly drained soils with low hydraulic conductivity
have higher SWCs regardless of the fact that till soils in Lapland experience a seasonal hydrological cycle with spring snowmelt, thawing, and saturation, summer drying and fall wetting (Sutinen et al. 1997). Snowmelt saturation (ε > 30) may last 2–8 weeks in fine-grained tills of Norway spruce stands, but in the case of sandy soils of Scots pine stands snowmelt saturation is absent and seasonal SWC is low (Hänninen 1997, Sutinen et al. 1997). Heavy autumn rainfalls may also instigate high SWC (Sutinen et al. 2007b). However, the time stability of soil implies that the magnitude difference of SWCs between sites tend to follow a consistent spatial pattern from year to year regardless of changes in amount of precipitation. Spatial soil water regimes of tills are controlled by physical properties of tills such as soil texture and structure. Particularly at the study site for paper III, macroporosity caused by the pebble and boulder fractions (>6 mm) was found to be a vital contributor to the spatial pattern of redistribution and infiltration of soil water of the fine grained tills (Sutinen et al. 2007b). The spatially stable behavior of the soil water patterns is recognized also in different climates (e.g., Cosh et al. 2008, Hu et al. 2009) aside from the humid climate of central Lapland with low summer evaporation (50 mm lower than precipitation, Solantie 1987).

In paper I it is suggested that soil surfaces exposed by mechanical site-preparation could be exploited in site suitability mapping for Scots pine. Presently, site-preparation by plowing is merely replaced with less destructive soil mounding (Hunt & McMinn 1988, Hagner 1999). As mounding exposes mineral soil surfaces to a lesser extent, an assessment approach which does not require soil exposure, would be more desirable for pine site suitability. In ideal circumstances, detecting the understory species patterns rather than exposed soil mounds would also provide a more continuous representation of the site suitability. As species differentiation with optical remote sensing can be cumbersome (Treitz et al. 2010), Airborne Imaging Spectrometer for Applications (AISA) data with a high spectral resolution and with the highest possible spatial resolution available at the time were chosen for the task. If the site suitability methodology was successful it would be an asset for artificial regeneration of Scots pine, in all cool and humid regions throughout Fennoscandia and Eurasia.

2.7 Optimization of peatland class hierarchal level from hyperspectral remote sensing data

Spatially continuous peatland inventory encompassing the heterogeneity of plant communities and environmental drivers, including information on the trophic level and hydrology, would provide a suitable data source for conservation planning and better means for quantifying greenhouse gas emissions. National wetland inventories of the wetland surface properties are conducted for these purposes using merely remotely sensed optical or radar data, at least in Sweden (Boresjö Bronge et al. 2006) and Canada (Fournier et al. 2007). It is hoped that hyperspectral data will be available from satellites during the 2010’s (Buckingham & Staenz 2008) to facilitate such nationwide mapping also in Finland. Steps towards analyzing the usability of satellite based hyperspectral data and developing data classification methods are needed for the purpose of nationwide mapping. In particular, resolving the optimal level of class detail of the peatland site types would be extremely useful. Water table depth (and consequently surface soil moisture), and its magnitude of seasonal fluctuation are the primary factors for peatland species composition (Laitinen et al. 2007, Närhi et al. 2010). Moreover, soil pH (Sjörs & Gunnarsson 2002, Tahvanainen et al. 2002) and nutrients, primarily N, P, Ca, and Mg (Wells 1996, Økland et al. 2001, Bragazza & Ger dol 2002, Sjörs & Gunnarsson 2002) determine the trophic level of peatlands. Soil fertility and hydrology (Malmer 1986, Bridgham et al. 1996) are included as the main environmental drivers for peatland site types in the Finnish floristic peatland classification systems (Cajander 1926, Eurola et al. 1984, Laine & Vasander 2005) providing a good basis also for national mapping. In those classification systems, peatlands are divided into swamps, spruce peatlands, pine peatlands, fens, and eutrophic fens on the basis of species associations at the top of the hierarchy, but at a lower hierarchical level into 25–37 site type classes. The process of combining them into higher level ecological classes with the traditional spectral separability approach would be a tedious task. In addition, a large variation in the
classification accuracies between high level hierarchical peatland classes (Huang & Sheng 2005, Töyrä & Pietroniro 2005, Haapanen & Tokola 2007) and low level classes (Dechka et al. 2002, Grenier et al. 2007, Korpela et al. 2009) leads one to think that each dataset, based on their spectral, spatial, and temporal resolution, has an optimal class level (Middleton et al. 2009). Therefore, it would be beneficial to develop a methodology to resolve the issue of the optimal class hierarchical level appropriate to each dataset.

In order to make the classes meaningful ecologically, community structure and spectral separability of the plant associations could be solved by applying ordination as vegetation cover fractions, environmental variables, and remotely sensed spectra as inputs. Since a prior understanding of environmental drivers in northern boreal peatland environments is available (Tahvanainen et al. 2002, Närhi et al. 2010) constrained ordination could be applied. Firstly, unsupervised classification and multivariate analysis of variance was applied in the ordination space in order to separate peatland clusters which could also be distinguished by their spectral reflectance. Secondly, a fuzzy clustering approach was employed on hyperspectral data to produce continuous representations of the peatland plant associations and map their extent spatially.

2.8 Chapter summary

Based on the literature review of spectral properties of soils it obvious that the water molecule has several absorption features in the VSWIR but out of the chemical elements only Fe clearly and a few others exhibit identified absorption features. Thus single variate non-linear regression techniques might be possible for predictions of water content. The signature of chemical elements is 'hidden' in the spectra, as absorption features are caused by molecular and mineral bonds. This phenomenon is called cross-correlation. Thus sample-set-specific multiband regression techniques are required. In addition, SWC predictions may also be improved by MVR as the variation by the effects of soil covariates, soil-specific cut-off thicknesses, and optimal wavelength ranges can be incorporated into the models.

Close-range soil spectroscopy could be applied for site suitability assessment of Scots pine but a remote sensing based method through recognition of understory species patterns would provide a more continuous representation to aid artificial regeneration. Literature of indirect mapping of soil water through remote sensing of vegetation patterns is rare and several environmental factors contribute to the understory species patterns. However, time stability of soil moisture patterns and relative high impact of soil moisture on species occurrence and abundance in the cool and humid boreal zone it may provide a relatively successful guideline. Similarly to mineral soils, species patterns on organic peatlands are also dependent on a range of environmental factors. Utilizing statistical techniques in the constrained ordination space could aid in revealing the vegetation community structure, spectrally separable peatland site types, and spectral feature selection for ecologically meaningful and optimized peatland site type mapping. This methodology could aid a possible national scale peatland site type mapping.

3 MATERIALS AND METHODS

In chapter 3.1 the ecological, geological, and climatic conditions in central and southern Lapland are shortly described. The data for the four original research papers I, II, and III were mainly collected in central Lapland but the study area for paper IV was located in southern Lapland (Fig. 2). Secondly, the sampling schemes, measurement techniques, and application of the numerical methods to answer each research question are explained in chapter 3.2. Thirdly, a short description of the data collected with close-range and airborne hyperspectral instrumentation (chapter 3.3), and reference data collected with geophysical, geochemical, particle size, and mineralogical techniques of soil are given followed by a description of vegetation analysis (chapter 3.4). Fourthly, the numerical data processing techniques are described including the ordination techniques (chapter 3.5), classification
and hyperspectral feature extraction (chapter 3.6), and chemometric regression (chapter 3.7). Lastly, the validation measures calculated for efficiency of prediction and classification are reviewed in chapter 3.8. Table 3 summarizes the devices used for data collection and Table 4 lists the statistical and data mining techniques applied on the data to answer each RQ. For background and more detailed descriptions of applying each method the reader is advised to refer to the original papers.

3.1 Study sites

The study sites in central Lapland belong to the northern boreal zone. Yet the study site of paper IV belongs to the central boreal zone following the compilation of harmonized vegetation regions by the Nordic countries (Arbetsgrupper för naturvårdsfrågor 1977, Fig. 2). The glacial landscapes in

Fig. 2. Map of northern Finland study area. Study sites and areas specific to each original research paper are given in symbols and as black area for paper IV. On background light grey resembles peatlands, dark grey lakes and main road network. Names of main towns are given in capital letters. In lower right corner map colours represent vegetation regions (Arbetsgrupper för naturvårdsfrågor 1977). Dark green corresponds to northern boreal zone, beige to central boreal zone, light green to hemiboreal zone, and blue to southern boreal zone. Base maps from the National Land Survey of Finland Topographic Database 03/2013 © NLS and HALTIK.
both areas are gently undulating with a few-meter-thick drift cover superimposing the bedrock. The central Lapland study sites are characterized by basal tills because of its unique location in the so-called ice divide zone of the Fennoscandia ice sheet during the last Pleistocene glaciation (Tanner 1915). Tills are local in origin and strongly reflect the lithology of the underlying crystalline and weathered bedrock (see Lestinen 1980). The southern Lapland area is also dominated by tills, yet it is also characterized by active ice moraines such as drumlins, and also fluvial and glaciofluvial sorted sediments (Nenonen et al. 2004, Nenonen 2004). The Proterozoic bedrock types in the central Lapland sites vary from mafic greenstones of the Lapland Greenstone belt to felsic granitic intrusions (Lehtonen et al. 1998). The southern Lapland sites are located on the Peräpohja Schist Belt whose main rock type is phyllite (Perttunen 1991). The southern Lapland site is characterized by an abundance of aapa mires which is a minerotrophic peatland complex with concave surface topography receiving supplementary nutrients from surrounding uplands. The boreal climate is typically cool and humid with low summer evaporation that is below precipitation by about 50 mm (Solantie 1987). The annual mean temperature is +1 °C in southern Lapland and -1–-2 °C in central Lapland. The precipitation sum is higher (550–600 mm) in southern Lapland than in central Lapland (500–550 mm, Vajda & Venäläinen 2003) and roughly 40–50% of annual precipitation falls as snow (Heino & Hellsten 1983).

### 3.2 Sampling and application of numerical techniques on research data

For paper I four undisturbed till soil samples were taken (Fig. 2) from sites with varying underlying lithology. After wetting the samples to near full saturation $\varepsilon$ of each sample was measured ex situ in laboratory to acquire a longitudinal data-set measured daily up to 3 months or until $\varepsilon=3$–$4$ ($n=396$, Table 5). Dielectric permittivity was measured with a Percometer using a non-destructive flat probe and spectral reflectance with a FieldSpec FR spectroradiometer (Table 3). This resulted in a total dataset of 95 for further processing of paper I data as two to six daily measurements per till were averaged and some measurements were removed due to small S-N ratio (Table 5). Particle size distribution was determined with wet sieving and sedimentation, and OMC colorimetrically (Metrohm spectrocolourimeter E 1009). Spectral transformations with continuum removal, derivatives, and wavelet transformation were applied (Table 4). To predict $\varepsilon$ from spectra, a single variate regression of a selected spectral bands was performed with mixed effects modelling for paper I. In further processing of the data MVR models with BRT, multivariate adaptive regression splines (MARS), PLSR, random forests (RF), and relevance vector machine (RVM) were constructed. The efficiency, tested with the calibration dataset, was reported as coefficients of multiple determination ($R^2$) and root mean square errors (RMSE, Table 4).

For paper II a complete set of 300 till soil samples at depth of 0–20 cm were taken from two sites (150 samples at each site, Fig. 2) 60 km apart but with similar metavolcanic lithology. Spectral reflectance was measured with the FieldSpec FR spectroradiometer and chemical element concentrations were assayed with inductively coupled plasma atomic emission spectrometry (ICP-AES, Table 3) from digestion of hot (90 °C) aqua regia. The final population was narrowed to 217 when heterogeneous samples with standard deviation <1% (in reflectance) throughout the spectral range 350–2500 nm were removed. To predict element concentrations from till spectra PLSR was applied on continuum removed and untransformed spectra, and model efficiencies were compared to a separate validation set reporting $R^2$ and RMSE (Table 4). To get an overview of the mineralogy X-ray diffractometry (Philips X’Pert MPD) was applied on a subset of the data ($n=16$, Table 3).

For paper III, the AISA instrument was flown in a twin-engine cargo plane owned by the Helsinki University of Technology on the 14th of September, 1999, at 9.00 a.m. UTC (11.00 local time, Table 3). At the 46.9 ha study site (see Fig. 2), $\varepsilon$ measurements were made with a Percometer (Table 3) on 573 micro sites, which were every 25 m on parallel lines spaced 50 m apart. Measurements were made between June 27th and 29th, 2000. In the summer 2007 a random selection of 51 these sites was chosen for vegetation inventory (Table 3). Neural network models were built with a radial basis functional link network (RBFLN) and probabilistic
neural network (PNN) functions (Table 4). Their sensitivity to a variety of validation set properties and their performance were compared with a separate calibration dataset of field ε-measurements. Receiver operating characteristics (ROC, Park et al. 2004) analysis was used in model validation (Table 4). Non-metric multidimensional scaling (NMDS) ordination and correlation analysis were used for revealing the understory community structure and their dependence on soil variables (Table 4).

For paper IV the Hyperspectral Mapper (HyMap) sensor was flown in a Dornier 288 aircraft operated by the German aerospace center in the middle of the growing season on the 29th of July, 2000, between 11.35 a.m.–2.30 p.m. local time (8.35 a.m. UTC, Table 3). The calibration data were gathered at 61 sites representing a variety of peatland site types in July 2007. The sampling scheme was based on unsupervised classification (Middleton et al. 2009) and segmentation of the HyMap imagery into spatially homogenous regions and choosing large enough segments for placing the field plots. This was done to ensure that the field observations would match the remotely sensed data as an up to 30 m spatial deviation was observed for the HyMap data at maximum. Data for validation were gathered in August-September 2005 at 115 sites using a stratified sampling limited to locations reasonably accessible by road (Närhi et al. 2010). In these datasets ε was measured with the Percometer, σ with a conductivity fork, and pH with a IQ160 pH meter (Table 3). Plant species coverages and site types were also determined (Table 3). For spatial classification of HyMap with SVMs and the validation of the SVM result with ROC analysis (Table 4) the sample plots were spatially extended based on visual interpretation of the HyMap data.
<table>
<thead>
<tr>
<th>Data processing technique</th>
<th>Software</th>
<th>Organization</th>
<th>Place</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st derivative</td>
<td>KernSmooth soil.spec ver. 1.4 in R</td>
<td>Terhoeven-Urselmans 2012, Wand &amp; Ripley 2012</td>
<td>University of Zürich, Switzerland</td>
<td>Richter 2003</td>
</tr>
<tr>
<td>2nd derivative</td>
<td>KernSmooth soil.spec ver. 1.4 in R</td>
<td>Terhoeven-Urselmans 2012, Wand &amp; Ripley 2012</td>
<td>University of Zürich, Switzerland</td>
<td>Richter 2003</td>
</tr>
<tr>
<td>Atmospheric correction</td>
<td>2012, Peters &amp; Hotton 2012</td>
<td>Terhoeven-Urselmans 2012, Wand &amp; Ripley 2012</td>
<td>University of Zürich, Switzerland</td>
<td>Richter 2003</td>
</tr>
<tr>
<td>GCA</td>
<td>1994a, 1994b</td>
<td>Mäkisara et al. 1994a, Mäkisara et al. 1994b</td>
<td>VTT, Finland</td>
<td>Zürich, Switzerland</td>
</tr>
<tr>
<td>Geometric correction</td>
<td>2001</td>
<td>Schläpfer 2001</td>
<td>Zürich, Switzerland</td>
<td>Geomat ReSe Applications, University of Zurich</td>
</tr>
<tr>
<td>Mixed-effects modelling</td>
<td>S-Plus 4.5 Professional</td>
<td>Therneau 2012, Peters &amp; Hotton 2012</td>
<td>University of Zürich, Switzerland</td>
<td>Richter 2003</td>
</tr>
<tr>
<td>MNDS</td>
<td>vegan ver. 1.15-4 in R</td>
<td>Oksanen et al. 2009</td>
<td>ESRI</td>
<td>x</td>
</tr>
<tr>
<td>Ordinary kriging</td>
<td>vegspec ver. 1.4-3 in R</td>
<td>Oksanen et al. 2009</td>
<td>R Development Core Team</td>
<td>x</td>
</tr>
<tr>
<td>PAM</td>
<td>vegspec ver. 1.4-3 in R</td>
<td>Oksanen et al. 2009</td>
<td>R Development Core Team</td>
<td>x</td>
</tr>
<tr>
<td>RF</td>
<td>vegspec ver. 1.4-3 in R</td>
<td>Oksanen et al. 2009</td>
<td>R Development Core Team</td>
<td>x</td>
</tr>
<tr>
<td>ROC</td>
<td>vegspec ver. 1.4-3 in R</td>
<td>Oksanen et al. 2009</td>
<td>R Development Core Team</td>
<td>x</td>
</tr>
<tr>
<td>RMSE</td>
<td>vegspec ver. 1.4-3 in R</td>
<td>Oksanen et al. 2009</td>
<td>R Development Core Team</td>
<td>x</td>
</tr>
<tr>
<td>ROC curve comparison</td>
<td>vegspec ver. 1.4-3 in R</td>
<td>Oksanen et al. 2009</td>
<td>R Development Core Team</td>
<td>x</td>
</tr>
<tr>
<td>SVM</td>
<td>vegspec ver. 1.4-3 in R</td>
<td>Oksanen et al. 2009</td>
<td>R Development Core Team</td>
<td>x</td>
</tr>
<tr>
<td>Wavelet</td>
<td>vegspec ver. 1.4-3 in R</td>
<td>Oksanen et al. 2009</td>
<td>R Development Core Team</td>
<td>x</td>
</tr>
</tbody>
</table>
Table 5. Descriptive statistics of dielectric permittivity ($\varepsilon$) and spectral data used in thesis synopsis to answer research question I. Number (n) of measurements made of each till sample and the final population used for regression models are given along with descriptive statistics of till soil $\varepsilon$. Sample sizes and statistics of $\varepsilon$ are also reported for calibration and validation datasets combined of four tills (Haipankuusikko, Kuorajoki, Vaalolehto and Vuotsonkangas).

<table>
<thead>
<tr>
<th></th>
<th>Vuotsonkangas till 1</th>
<th>Vaalolehto till 2</th>
<th>Kuorajoki till 3</th>
<th>Haipankuusikko till 4</th>
<th>All calibration</th>
<th>All validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>n of measurements</td>
<td>25</td>
<td>152</td>
<td>83</td>
<td>136</td>
<td>396</td>
<td></td>
</tr>
<tr>
<td>final n</td>
<td>12</td>
<td>28</td>
<td>29</td>
<td>26</td>
<td>95</td>
<td>47</td>
</tr>
<tr>
<td>mean $\varepsilon$</td>
<td>13.1</td>
<td>13.3</td>
<td>16.9</td>
<td>16.8</td>
<td>15.3</td>
<td>15.4</td>
</tr>
<tr>
<td>median $\varepsilon$</td>
<td>15.0</td>
<td>12.8</td>
<td>18.0</td>
<td>18.1</td>
<td>15.5</td>
<td>15.2</td>
</tr>
<tr>
<td>standard deviation $\varepsilon$</td>
<td>5.9</td>
<td>6.9</td>
<td>6.1</td>
<td>5.2</td>
<td>6.3</td>
<td>6.3</td>
</tr>
<tr>
<td>minimum $\varepsilon$</td>
<td>3.6</td>
<td>3.9</td>
<td>3.7</td>
<td>4.9</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>maximum $\varepsilon$</td>
<td>21.4</td>
<td>26.3</td>
<td>25.5</td>
<td>25.8</td>
<td>26.3</td>
<td>26.3</td>
</tr>
</tbody>
</table>

3.3 Hyperspectral instrumentation and data

3.3.1 Close-range spectroscopy

The soil spectral reflectance between 350–2500 nm was measured with a FieldSpec FR spectroradiometer (see Table 3 for instrument details). The samples were artificially illuminated from nadir with a tungsten lamp configuration (Makita Inc., U.S.A.) with an approximately 25° viewing angle of an optical fiber. The fiber optic cable had a 25° full angle cone of acceptance field-of-view. The distance between the head of the fiber optic cable and the sample was 7.1 cm on average. With this configuration the measured surface area was estimated to be 5 cm². The spectroradiometer sampling resolution was 1.4 nm but the data were internally resampled to 1-nm-resolution. The instrument was calibrated to a white reference spectrum (i.e., Spectralon plate, Labsphere, NH, U.S.A.) between each sample, and the reflectance readings for each wavelength were expressed in relative reflectance readings. For improved S-N ratio the instrument was allowed to warm up for 20 min prior to data collection. Figure 1 in this thesis synopsis, Figure 1 in paper I, and Figure 3 in paper II illustrate the FieldSpec FR measured spectra of tills with varying moisture contents and varying concentrations of elements.

3.3.2 Airborne hyperspectral imaging spectrometry

Two airborne datasets with different imaging spectrometers and spectral settings were acquired for this work. The Hyperspectral Mapper imaging spectrometer (HyMap™, Cocks et al. 1998, Table 3) was programmed to record a true hyperspectral dataset with 126 10-to-22-nm-wide (full-width-half-maximum) bands across the 437–2485 nm spectral range. The Airborne Imaging Spectrometer for Applications (AISA, Mäkisara et al. 1993, Table 3) instrument was programmed to record the first nine spectral bands in the VIS (448.99–652.27 nm) and eight bands in the NIR (667.46–874.04 nm) 7–8 nm in width. For the HyMap data a flight altitude of 2000 m, a ground speed of 278 km/h, a 60° field of view, and 2.5 mrad along- and 2.0 mrad across-track instantaneous fields of view resulted in a spatial resolution of 5 m for a 2 km ground swath. For the AISA data an average flight altitude of 1300 m and speed of 200 km/h resulted in 1.1 m spatial resolution.

The AISA and HyMap datasets were preprocessed with radiometric and geometric corrections, geocoding, and mosaicking of the flight stripes. For the raw radiance HyMap data, supplied by imaging service provider Hyvista Corp. (Sydney, Australia), geometric rectification was performed with PARametric GEocoding (Table 4). The HyMap data were not optimal due to the unavailability of the inertial movement unit (IMU) data and lack of clearly identifiable ground control points. Georectification was performed using elevation, position, and one tie point for each 100 pixels along a scan line (points/flight swath >50) as inputs resulting in a mean root mean square error (RMSE) of 15 m and 30 m maximum deviation.
The AISA data were first geometrically corrected as GPS navigation data and data gathered with an IMU as inputs, and geocoded based on 1:20000 digital topographic map and 25 m pixel size digital elevation model by the National Land Survey of Finland with GEOVIS software (Table 4).

For the AISA data, the spectrometer recorded digital numbers were radiometrically normalized to correspond to 9.00 a.m. UTC assuming a Lambertian surface model and converted to raw radiance values using instrument and band-specific gain and offset values. The HyMap data were also subjected to atmospheric correction, for which reflectance was measured at five pseudo invariant features with a spectroradiometer during the flight campaign. Pseudo invariant features were supplied as inputs into ATCOR4 (Table 4), which was also used for bidirectional reflectance correction with nadir normalization procedure. The atmospheric correction was done assuming a flat terrain model since the terrain was gently undulating (6–102 m a.s.l., 99% of peatlands 13–44 m a.s.l.). Finally, continuous coverages of both datasets were created by mosaicking flight swaths. HyMap spectra of different peatland site types are illustrated in Figure 5 of paper IV. AISA spectra of some landscape components on the study area of paper III can be viewed in Figure 4 of Siira et al. (2000a) and close-range spectra in Figure 1C of this thesis synopsis.

3.4 Reference field data

3.4.1 Geophysical measurements

The soil root zone bulk ε (as a measure of volumetric water content, Roth et al. 1992) was acquired with an electrical capacitance probe. The device measures the change in the electrical capacity of the electrode due to the influence of the soil the probe is inserted (Plakk 2008, Table 3). The device operates at a frequency of 30–50 MHz. The measurement volume of the tube probe, used in situ, was estimated to 500 cm³ and approximately 100 cm³ for the 5-cm-diameter flat probe, which was used ex situ. The σ (as a rough estimate of the soil solute content, Rhoades & Oster 1986) was measured with a conductivity fork with a galvanic four electrode Wenner configuration having 15 cm steel rods and 16 cm spacing (Puranen et al. 1999, Table 3). With this configuration the soil σ measurements represent approximately the top 30 cm of the soil. Soil water pH was measured with an IQ160 pH meter (Table 3). Three to five measurements were made at each sampling site and the mean values were used as inputs into the statistical analysis to avoid bias caused by pebbles and boulders in the soil.

3.4.2 Geochemical, particle size and mineralogical analysis

The partial leaching mineral soil samples were digested with hot (90 °C) aqua regia. Element concentrations from soil solution 0.2 g/3 mL (mineral sample/HCl + HNO₃ dilution) were analyzed with ICP-AES (Table 3). The particle size distribution the standard wet sieving followed by sedimentation with the areometric flotation method was applied (Gee & Bauder 1986). The OMC was determined spectrocolorimetrically (Table 3). The mineralogy of the till samples was determined with X-ray diffractometry (XRD, Table 3). Specimens were prepared by grinding the mineral powder to remove clumps, and were then fixed on a glass slide with acetone. For clay mineralogy, oriented samples were prepared by mixing with water and allowing the mixture to settle before submerging a glass slide into the suspension.

3.4.3 Vegetation analysis

The in situ species inventory was done by estimating species coverages in terms of cover-% within a 1m-by-1m sampling frame (Table 3). Species cover fractions were estimated in herb and bryophyte (i.e., bottom and field) layers separately. Alternatively, the coverages of all species were also estimated as one layer, total coverage being 100%. The coverages of grass and moss litter were also estimated. On peatlands the classification system described by Eurola et al. (1984) and Eurola et al. (1995) was followed.
3.5 Ordination techniques

The purpose of ordination techniques, commonly applied in ecology, is to reveal the community structure and its relation to the environmental variables such as soil properties and, in this work, also the spectral variables. Ordination techniques are visual in nature representing multivariable vegetation and environmental data in a few dimensions, so that distances in the ordination reflect similarities in vegetation composition between sample sites. Values of the coordinates are not essential since they only achieve the desired spacing. The variables are given as lines in the ordination graphics, each line indicating the direction of a gradient, and length indicating strength of correlation between the variable and ordination. (Jongman et al. 1995)

3.5.1 Non-metric multidimensional scaling

Non-metric multidimensional scaling (NMDS, Kruskal 1964) using Bray-Curtis dissimilarities with Wisconsin double standardization (Bray & Curtis 1957, Table 4) was applied to vegetation data in this study to analyze soil variables contributing to vegetation composition. The NMDS is an unconstrained method based on vegetation dissimilarities only, no attempt to maximize correlations with environmental variables is made. The Kruskal’s NMDS is constructed by first calculating a distance matrix, then randomly selecting the initial configuration of samples in n-dimensions, and finally calculating the ‘stress’ and moving the samples in the direction of the stress. Stress is considered as a mismatch between the rank order of distances in the data and the rank order of distances in the ordination. The stress is iteratively recalculated until the stress appears to reach minimum. The final configuration of points may be rotated. NMDS is demonstrated superior to other unconstrained ordination methods when applied to ecological data (Kenkel & Orlóci 1986, Minchin 1987). The significance of the correlations between vegetation and environmental variables were tested with several thousand permutations.

3.5.2 Canonical correspondence analysis

Direct gradient analysis can be chosen when there is *a priori* understanding of the environmental variables that constrain the vegetation composition. The constrained canonical correspondence analysis (CCA, Ter Braak 1986, Ter Braak 1987, Table 4) was used in this study. As a constrained ordination, the CCA does not try to display all variation in the data, yet only the part that can be explained by the utilized constraints. The CCA is calculated using reciprocal averaging from the correspondence analysis such that at each cycle of the averaging process, a multiple regression is performed on the sample scores of the environmental variables. New site scores are calculated based on this regression, and then the process is repeated, continuing until the scores stabilize (Jongman et al. 1995). Compared to another common constrained ordination technique, redundancy analysis, the CCA has been demonstrated to be more robust in extracting variation in species cover fractions in relation to the environmental variables, specifically in cases, where relations among species variables and environmental attributes are non-linear (Ter Braak 1986, Närhi et al. 2010).

3.5.3 Partitioning around medoids

It is suggested that classification in ordination space should be applied (Feoli & Orlóci 1991) to reveal natural community clusters reliably. Thus unsupervised classification method partitioning around medoids (PAM, Van der Laana et al. 2003, Table 4) was applied in this study. PAM is a $k$-medoids clustering algorithm technique. In comparison to the commonly used $k$-means it is considered more resistant to noise and outliers. The reason for it is that it minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances and uses the most centrally located objects in clusters, i.e., medoids, instead of the cluster means. (Van der Laana et al. 2003, Kaufman & Rousseeuw 2008)
3.6 Classification and hyperspectral feature extraction

3.6.1 Feature extraction

Minimum noise fraction (MNF) is a feature extraction or selection algorithm performed to reduce dimensionality of hyperspectral data (Green et al. 1988, Table 4). Noise can be effectively removed from multispectral data by transforming the original hyperspectral data to the MNF space, and then rejecting the most noisy components by selecting, e.g., MNF bands with eigenvalues greater than five (similarly to, e.g., Harris et al. 2006a). The benefits of MNF are not only in feature extraction but also saving in computational costs. Because of the effective noise removal MNF also improves vegetation classification (Yang et al. 2009, Zhang & Xie 2012).

3.6.2 Supervised non-parametric classification

Both supervised non-parametric classification methods used in this study, artificial neural networks (ANN) and support vector machines (SVM), determine the association between calibration points and datasets, and provide a fuzzy membership value between 0 and 1 as an output. The advantage of non-linear classifiers is that a single output class can be comprised of a series of spectral clusters in the spectral feature space. For a successful result, the SVMs and ANNs can be trained with a lower number of calibration points compared to, e.g., statistical classifiers (e.g., Mas & Flores 2008). Furthermore, it is possible to use the non-linear classifiers on data without atmospheric correction, especially if a study area was small (paper III, 46.9 ha in size) and atmospheric properties influencing the spectra were, therefore, fairly constant.

The ANN architecture is inspired by the brain structure. The main processing unit is a neuron, composed of synapses that connect the input signals to a synaptic weight, a component that sums the weights, and a non-linear activation function that combines the summed weights to the output signals. The neurons are organized as several interconnected layers, some of them hidden, where the signals can propagate forward or backward (Kanellopoulos et al. 1997). Two supervised ANN algorithms were applied for hyperspectral image classification in this study: RBFLN and PNN. The RBFLN is a three-layer feed-forward network where full propagation is used in calibration (Looney 2002, Table 4). The PNN is a multilayer feed-forward network that uses sums of Gaussian distributions to estimate the probability density function for a calibration dataset (Specht 1990, Table 4).

SVMs, on the other hand, deal with class separation by attempting to find optimal hyperplanes to differentiate discrete classes in the multidimensional feature space (Burges 1998). Generalization power of SVMs relies on maximizing the margins of the hyperplane by determining support vectors, i.e., a subset of calibration points which define the position of the hyperplane margins. The SVMs deal with linearly non-separable classes by mapping the data into a higher dimensional feature space by a kernel function. The kernel approach enables fitting of a linear hyperplane between classes in the new variable space. The SVMs outperform traditional statistical classifiers because of their ability to generalize based on a limited number of calibration pixels, higher robustness to noise, higher learning efficiency, and in general, less requirement for prior knowledge (see Bishop 2006 and Du et al. 2012 for summary). The method has proven successful in processing of remotely-sensed data (Mountrakis et al. 2011). In this study the image-SVM software, which utilized the LIBSVM for the calibration of the SVMs, was used (Table 4).

3.7 Chemometric techniques

3.7.1 Spectral transformations

The purpose of spectral pretreatments is to correct for non-linearity, measurement and sample variations, and noisy spectra (Stenberg et al. 2010). Three different spectral transformations were applied in further processing of the data in paper I to enhance the absorbance peaks, reduce baseline shifts and overall curvature of samples, and to ‘normalize’ the spectral reflectance. The derivatives, continuum removal and wavelet transformation were applied using routines in an R package ‘soil.spec’ (Table 4). The ‘soil.spec’ package uses respective function ‘locpoly’ for derivatives and ‘chull’ for continuum removal by convex hull method from a package ‘KernSmooth’. To smooth and compress
the spectra also wavelet coefficients were applied (Percival & Walden 2000). The ‘dwt’ function in ‘soil.spec’, used for the wavelet transformation, also refers to an R package ‘wavelets’ (Table 4).

3.7.2 Partial least squares regression

Partial least squares regression (PLSR, Wold et al. 2001) is the most popular technique in chemometrics for quantitative analysis of reflectance spectra. It is used to construct predictive models when there are many predictor variables (spectral variables at a wavelength range) that are highly collinear. The PLSR algorithm integrates data compression and regression steps and it selects successive orthogonal factors that maximize the co-variance between predictor and response variables. The number of factors to use in the models is selected by cross-validation. By fitting a PLSR model, one hopes to find a few PLSR factors that explain most of the variation in both predictors and responses (Martens & Naes 1992). In the current study, an R package ‘pls’ was used (Table 4). Overview of the chemometrics in spectroscopy, its history and main concepts are provided by Geladi (2003). The more recent MVR methods, presented hereafter, have emerged in chemometrics from the fields of data mining and machine learning.

3.7.3 Bagging regression trees

In recursive partitioning the feature space of predictor variables is recursively partitioned into a set of decision rules, i.e., smaller groups with binary splits based on a single predictor variable. Splits of all of the predictors are examined by an exhaustive search procedure and the split that maximizes the homogeneity of the two resulting groups with respect to response variable is chosen. The output is a tree diagram with the branches determined by the splitting rules and a series of terminal nodes that contain the mean nodes. Initially, maximal trees are grown and then cross-validations are applied to prune the tree to avoid overfitting (Breiman et al. 1999).

Bootstrap aggregation, or bagging, proposed by Breiman (1996) is used with regression trees to reduce the variance associated with prediction, and thereby improve the prediction process. In bagging regression trees (BRT) many bootstrap samples are drawn from the available data. A prediction method is applied to each bootstrap sample and the results are then combined by averaging to obtain an overall prediction with reduced variance (Sutton 2005). BRT has been reported to be very effective in reducing variance and error in high dimensional datasets. In comparison to one-tree regression tree approaches, better estimates of errors are provided by bootstrapping, and the bias component of the error is marginally better than for single regression trees. Difficulty comes from interpretation of the results as a large number of trees are averaged for the final result (Prasad et al. 2006). In this study, recursive partitioning algorithm ‘rpart’ was bagged in an R package ‘ipred’ (Table 4).

3.7.4 Multivariate adaptive regression splines

Multivariate Adaptive Regression Splines (MARS TM, Friedman 1991, Table 4) produce a non-parametric regression model. MARS is a generalization of recursive partitioning regression approaches, which generate piecewise linear models instead of piecewise constant models. The adaptive piecewise linear regressions are called basis functions. Changes in slope of the basis functions occur at points called knots. The regression can bend at the knots, which mark the end of one region of data and the beginning of another with different behavior of the function. Knots are established in a forward-backward stepwise way. A model which clearly overfits the data is produced first. In subsequent steps, knots that contribute least to the efficiency of the model are discarded by backwards pruning steps. The best model is selected via cross-validation, a process that applies a penalty to each knot added to the model to keep low complexity values (Muñoz & Felicísimo 2004). By applying linear basis functions between the splits of the partitioned space, MARS overcomes the restriction of the regression trees of returning discontinuous response surfaces. Hence, prediction accuracy can be expected to be higher, output smoother and less coarse grained with MARS than with regular tree based methods. The downsides of MARS are the local nature of the model, which can lead to unstable predictions, and cumbersome selection of the input parameters (Friedman 1991, Prasad et al. 2006).
3.7.5 Random forests

Random forests (RF, Breiman 2001, Table 4) are similar to BRTs in that bootstrap samples are drawn to construct multiple trees but the difference is that each tree is grown with a randomized subset of predictors. A large number of trees are grown, i.e., forest of trees. Each tree is grown to maximum size without pruning and these multiple predictions are aggregated by averaging the individual tree outputs for the final regression prediction. Out-of-bag samples can be used to calculate an unbiased error rate and variable importance eliminating the need of a calibration set or cross-validation. Diaz-Uriarte & Alvarez (2006) conclude that the most important advantages of RFs are suitability for datasets with many predictors and few samples, robustness to noise and irrelevant features, they are unlikely to overfit if trees are carefully sized, random selection keeps bias low, and almost no fine-tuning of the parameters is necessary to produce good predictions. Because of the additional complexity compared to BRT, it is usually considered even more of a ‘black-box’ in interpretability (Prasad et al. 2006).

3.7.6 Relevance vector machines

The relevance vector machine (RVM, Tipping 2001) is another sparse kernel based learning machine that has the same functional form as the SVM. It uses Bayesian inference to obtain parsimonious solutions for regression and classification but it is used only for chemometric regression in this study. An RVM is formed as a linear combination of data-centered basis functions, called relevance vectors, which can be non-linear. RVM provides truly probabilistic predictions through Bayesian inference whereas SVM can provide pseudoprobabilistic outputs only through postprocessing. Its decision function is more sparse (i.e., depends on fewer input data) than a comparable SVM, because SVM minimizes the calibration error under the constraint of maximum smoothness, requiring more decision points (Bishop & Tipping 2000). Moreover, the number of relevance vectors in an RVM is not proportional to the number of samples. The benefit of a sparser classifier is decreased overfitting and thus RVM models are more generalized than SVM predictions. Therefore, the RVM has been shown to provide equivalent and often superior results to the SVM both in generalization ability and sparseness of the model (Gómez-Chova et al. 2009).

3.7.7 Mixed-effects modeling

Mixed-effects models make it possible to take into account for the dependencies in data by including more than one source of random variation, i.e., more than one random effect. Non-linear mixed-effects models are mixed-effects modes in which some, or all, of the fixed and random effects occur non-linearly in the model functions. Non-linear mixed effects models provide a tool for analyzing repeated-measurements data in which the relationship between the explanatory variable and the response variable can be modeled as a single function, allowing the parameters to differ between individuals (Zuur et al. 2009). The model used in this study was formulated by (Lindstrom & Bates 1990). In paper I, the mixed-effects models were created to compare regressions of $\varepsilon$-reflectance for different tills (treatments).

3.8 Model validation

The root mean square error (RMSE) was reported as a prediction error estimate for the estimations of soil $\varepsilon$ and element concentrations from spectra. RMSE is the square root of the sum of the squared differences between predicted $(y)$ and measured $(x)$ values divided by the number of fitted spectra $(n)$. The following formula was applied:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y-x)^2}$$

The coefficients of multiple determination ($R^2$), i.e., square of the correlation between response values and the predicted response values, were also calculated for regression goodness of fit. Due to a small sample size, in paper I and further processing of the data, the calibration $R^2$ and RMSE values were compared for the till-specific models, meaning that the same sample set was used for model construction and goodness of fit and error estimations. Validation goodness of fit and error estimations could be calculated for the universal fits and results in paper II as the population sizes were large enough to be split into calibration and validation sets. Calibration $R^2$ and RMSE were also
recorded for the generic fit so that the values could be comparable to the till-specific fits. For paper I, residual degrees of freedom based adjusted $R^2$ ($R_{adj}^2$) was also calculated as it is suitable for comparing nested models. Although commonly used for goodness of fit, $R^2$ does not directly characterize accuracy as the values measure the strength of the relationship between spectral and the soil variables. For the purposes of characterizing uncertainty, the RMSE may thus be preferable (Woo’dock 2006).

The receiver operating characteristics (ROC, Park et al. 2004) curves and their area under the curve (AUC) values were used to assess the prediction efficiency of classifiers which produced a real number class probability as an output as recommended by Manel et al. (2001). An ROC curve is a plot of test sensitivity (i.e., true positive rate, TPR) on the Y axis and 1-specificity (i.e., false positive rate, FPR) on the X axis generated using a set of varying cut-off points. The TPR equals to the fraction of true positives in the validation dataset out of all positives in the classification result. Similarly, the FPR equals to the fraction of false positives out of all the negatives in the classification result. The best possible model would yield a point in the upper left corner of the ROC space, representing TPR of 1 (no false negatives) and FPR of 0 (no false positives). Such a model would produce a perfect classification with a AUC values of 1. A model which is no better than a random guess (AUC=0.5) would yield a point along a change diagonal that is a line from the left bottom to the top right corners. In case of two models with equal AUCs but different shapes, the better model reaches high TPR values at low FPR values.

In paper IV, the class-specific prediction probabilities were also turned into discrete classification results such that each pixel was assigned into a class with the highest output probability. Thus accuracy could also be assessed with measures determined from an error matrix. Besides overall accuracy and kappa values, for the individual classes both user’s and producer’s accuracies were calculated (Congalton 1991). All classification accuracies were calculated using independent validation sets. For comparison of the spatial patterns of Scots pine suitability created in paper III a probability surface was also created from in situ $\varepsilon$ measurements with ordinary kriking in ArcMap (Cressie 1988, Table 4).

### 4 RESULTS

In this chapter the key results to answer each research question are presented. Chapter 4.1 summarizes the results related to prediction of glacial till soil $\varepsilon$ from close-range VSWIR spectra and comparison of the prediction efficiency between the different regression techniques. In chapter 4.2 the mineralogical description of the studied tills is first given and then the differences in the PLSR prediction efficiencies between the two studied tills in relation to variation and correlation of chemical element concentrations are explained. In chapter 4.3, the selection and utilization of the best performing ANN model for Scots pine site suitability assessment and the sensitivity of the ANN model performance to parameterization and calibration set properties are described. In addition, the results of the NMDS are interpreted to explain the dependence of the ground cover components on soil $\varepsilon$. Finally, results related to the peatland site type mapping are given in chapter 4.4. The logic in conducting the peatland thematic class level optimization with CCA is described and the SVM classification results of the peatland are given. For a more detailed description of the results the reader is advised to refer to the original research papers I–IV.

#### 4.1 Predicting till soil dielectric permittivity from reflectance spectra

Paper I and further data processing for the paper I data answer the RQ 1. The key finding of paper I is that a three parameter exponential function, $y = ae^{bx} + c$, where $x = \varepsilon$ and $y =$ reflectance, explains the relationship between the $\varepsilon$ and VSWIR reflectance statistically significantly. It also better fits the dry end of the gradient ($\varepsilon = 3–15$) compared to other non-linear functions. The shape of the function is such that reflectance sharply decreases as a function of the $\varepsilon$ from $\varepsilon = 3$ to $\varepsilon = 15$. The
function becomes linear when $\varepsilon=15–40$ leveling off at reflectance values between 0.04–0.15 (660–2320 nm) with a slope of a function close to 0. The results of the mixed effects modeling showed that a statistically significant generic exponential model, $y = 0.564e^{-0.205x} + 0.072$, explains the relationship between $\varepsilon$ and reflectance, e.g., when mean of reflectance between 1444-1464 nm was used as a reflectance variable ($R_{adj}=0.847$).

The data used in paper I was also further processed with MVR methods such as PLSR, MARS, BRT, RF, and RVM and compared against the single-variate exponential function. Results are illustrated in Figure 3. The further processing of the data with MVR techniques demonstrated that the till soil $\varepsilon$ could be predicted well not only with the exponential single variate model (mean of reflectance between 1444–1464 nm, validation $R^2=0.77$, RMSE=2.98, Fig. 3B, D) but also with MVR techniques: PLSR, MARS, BRT, RF and RVM (validation $R^2=0.81–0.94$, RMSE=1.59–2.77, Fig. 3B, D). The highest goodness of fit ($R^2=0.89–0.94$) and the lowest error (RMSE=0.99–1.23) were calculated for RVM predicted models, indicating that the RVM outperformed the other methods. When the till-specific calibration sets with smaller sample size were also used, PLSR and RF produced equally good predictions of $\varepsilon$ compared to RVM (Fig. 3A, C). It was also seen from the Figure 3 (B, D) that the validation goodness of fit and error values usually produced more optimistic results than the validation with an independent sample.

The mixed effects modeling results in paper I showed that the generic exponential model was not statistically valid for the tills with OMC>1.7% (dag/kg) and especially when a reflectance variable from the VIS range was chosen (mean reflectance at 640–660 nm, $R_{adj}=0.585$). Although the generic models estimate $\varepsilon$ statistically significantly, the till-specific models are usually more appropriate (Fig. 3). When calibration stage goodness of fit and errors are compared between the generic (Fig. 3B, D) and till-specific models (Fig. 3A, C) the PLSR, RF, and RVM produced better predictions for till-specific calibrations. However, with MARS and BRT the $\varepsilon$ could be better predicted for all tills (generic fit) than for each till specifically. The spectral transformations did not have a significant impact on the results (Fig. 3). Only the RVM validation results were improved by the continuum removal and wavelet transformations. In conclusion, individual multivariate calibrations are not necessary for the studied tills (clay fraction content 2.4–5.5%, fine fraction content 23.5–47.1%), which are typical for northern Finland. According to the results of paper 1 for single variate predictions of $\varepsilon$ the reflectance variable should definitely be chosen outside of VIS ($\lambda>700$ nm) and rather at SWIR ($\lambda>1740$ nm) especially when dealing with tills with high OMC >1.7% (dag/kg).
Fig. 3. Coefficients of multiple determination ($R^2$, 3A, 3B) and root mean squared error (RMSE, 3C, 3D) values for models of till dielectric permittivity predicted from reflectance spectra (350–2500 nm) with exponential (EXP), partial least squares regression (PLSR), multivariate adaptive regression splines (MARS), bagging regression trees (BRT), random forests (RF), and relevance vector machines (RVM). Till-specific models (3A, 3B) were constructed on Vuotsonkangas till (till 1), Haipankuusikko till (till 2), Vaalolehto till (till 3) and Kuorajoki till (till 4), and generic models (3B, 3D) for four tills combined. For generic models (3B, 3D) both calibration and validation $R^2$ and RMSE values are given. Original reflectance spectra (ref) was pre-treated with transformations such as first derivative (1.der), second derivative (2.der), continuum removal (cr), and wavelet transformation (wt). Figures edited by Paavo Närhi, GTK.
4.2 Predicting till soil element concentrations from reflectance spectra

The results of paper II answer the RQ 2. According to the XRD measurements, the till 1 was composed of plagioclase feldspar (XRD sample n=9, mean 48.3%), quartz (41.3%), amphibole (9%) and clinichlorite (1.6%). Till 2 (XRD samples n=7) was very similar in composition but had slightly higher quartz content (44.9%), lower amphibole content (5.4%), and had less than 1% of talc and illite. Based on the literature it is known that a number of other soil chemical/mineralogical constituents that were not analyzed, could be present in these soil samples (Righi et al. 1997, Peuraniemi et al. 1997). For example, soil precipitates and clay mineralogy which have absorption features were not quantified nor identified.

When the PLSR models were built for a combined dataset of till 1 and till 2, concentration of several elements were predicted: Al, Ba, Be, Co, Cr, Cu, Fe, K, La, Mg, Mn, Ni, P, S, Sc, Sr, Ti, V, Y, and Zn (Fig. 4A). The statistically significantly (validation R²≥0.5) predicted elements were bolded in the Figure 4. For till 1, only concentration of Fe could be predicted with spectra (Fig. 4B), whereas concentrations of Al, Ba, Co, Cr, Cu, Fe, K, La, Mg, Mn, Ni, S, Sc, Sr, Ti, V, Y, and Zn could be predicted for till 2 (Fig. 4C). The most significantly (validation R²≥0.8) predicted element for till 1 was Fe but for till 2 Ba, Co, Cu, Mg, and V could be predicted with the highest significance (validation R² ≥ 0.8). For the combined till data, elements predicted with R²≥0.8 were Al, Co, Cr, Cu, Fe, Mg, Mn, and Zn. Predicting concentrations of the till 1 with the model created for the till 2, and vice versa, was not possible.

The Figure 4 demonstrates the dependence of the predictive accuracy (error expressed as RMSE) on standard deviation of the element concentrations. The predictive error of the PLSR model is higher for elements with larger variation in concentration. For example, the prediction error for Fe, which occurs in high concentrations and has large variation in concentration, is predicted with a large error. On the other end, e.g., Be (Till 2, Fig. 4C) with low standard deviation in concentration, has a small prediction error. There is a very slight difference in standard deviations between the till 1 and till 2, if concentrations of till 1, having smaller variation, is compared to till 2 (compare Fig. 4B and 4C). However, Spearman correlation analyses indicated statistically significant inter-correlation between concentrations of analytes. When absolute values of the Spearman correlation coefficients (n=217) between all analysed elements (n=25) were summed together a value of 114.1 was received. The corresponding value for till 1 was 115.3 and for till 2 it was 159.7. This means the element concentrations are more inter-correlated in till 2 compared to till 1 and the combined dataset (till 1 + 2).

4.3 Mapping site suitability for artificial regeneration to Scots pine based on hyperspectral imaging data

The RQ 3 is answered with the results of paper III. The success of the pine suitability assessment with ANN prediction from hyperspectral data could be considered moderate. The best performing model was an RBFLN model trained with 15 calibration points representing the ‘suitable’ and 15 points representing ‘unsuitable for pine’ class (0.64 points/ha). The soil ε-limit between the ‘suitable’ and ‘unsuitable for pine’ classes was set to ε=15 (later referred to as 15/15) and the resulting fuzzy membership band was then median re-sampled to 14 m pixel size. This particular model produced the highest area under the ROC curve value (AUC = 0.741) and model performance was better in the low false positive rate range than the five next best models (see Park et al. 2004).

This model was chosen to produce the final discrete map with the distribution of the ‘suitable’ and ‘unsuitable for pine’ classes (Fig. 5). If little risk regarding the anticipated pine survival rate is required, the class is delineated to low true and false positive rates. In case A (Fig. 5), 34.4% (true positive rate) of all suitable for pine (ε≤15) sites would be correctly identified as suitable for pine, and 17.9% (false positive rate) of all unsuitable for pine (soil ε>15) sites would be incorrectly identified as such. At higher risk rates, larger true (68.9%) and false (33.9%) positive rates are chosen to threshold the class (case B in Fig. 5). Therefore a larger area would be subjected to Scots pine regeneration. Compared to a prediction map created with ordinary kriking the ANN classification
Fig. 4. Element-specific root mean squared errors (RMSE) of partial least squares regression models for till soil element concentrations from reflectance spectra (350–2500 nm) plotted against the standard deviation of each element concentration. For till 1 (4B) concentration of only one element (Fe, bolded) could be predicted statistically significantly (validation $R^2 \geq 0.5$) whereas for till 2 (4C) concentrations of 17 elements. For the combined dataset of till 1 and till 2 (4A) concentrations of 16 elements could be predicted. Figures edited by Paavo Närhi, GTK.
results resemble very much the spatial patterns of probability for soil $\varepsilon \leq 15$ (Fig. 5).

The sensitivity of the classification on the calibration set $\varepsilon$ limits ('suitable' $\varepsilon \leq 13$ and 'unsuitable' $\varepsilon > 17$ or 15/15), calibration set size (0.64 points/ha, 2.43 points/ha, 6.06 points/ha), ANN model selection (PNN or RBFLN) and resampling of the output probability bands with different methods (mean or median) into different pixel sizes (4 m, 6 m, 8 m, 10 m, 12 m, 14 m) was tested statistically. The results show that these variables can have a profound impact on the efficiency of prediction as the variation of the AUCs was large (n=156 combinations of ANN classification variables, $\min_{\text{AUC}}=0.43$, $\max_{\text{AUC}}=0.74$, $\text{mean}_{\text{AUC}}=0.62$, $\text{std}_{\text{AUC}}=0.09$). The RBFLN models overall produced higher accuracies than the PNN models as tested with the Mann-Whitney test of the AUC values ($U=702$, $p<0.001$, $n_{\text{RBFLN}}=78$, $n_{\text{PNN}}=78$, $\text{mean}_{\text{RBFLN}}=0.61$, $\text{mean}_{\text{PNN}}=0.52$). A Spearman rank correlation coefficient of 0.66 ($p<0.001$, $n=78$) was obtained between the RBFLN and PNN. This indicated that, besides the ANN model, also the calibration set had a great impact on the accuracy.

In addition, the best AUCs were obtained with the same calibration set (15/15, 0.64 points/ha) both with PNN and RBFLN. It indicated that a very small number of calibration points could also be sufficient for successful calibration if the points were representative of the classes. The greatest potential to produce the most accurate RBFLN model was with the calibration set which was produced setting a threshold between the 'suitable'/‘unsuitable' classes to $\varepsilon > 15/\varepsilon \leq 15$ (compared to 13/17), and which had calibration point density of 2.43 points/ha (compared to 0.64 points/ha and 6.06 points/ha). The classification could be further improved by resampling the output ANN probability channel...
Hyperspectral close-range and remote sensing of soils and related plant associations. Spectroscopic applications in the boreal environment

from 2 m to 12 m or 14 m pixel size (compared to 4 m, 6 m, 8 m, 10 m, U=241, p<0.001, n_{4-6-8-10m}=24, n_{12-14m}=54). Statistical difference between mean and median resampling was not found (U=281, p=0.89, n_{MEAN, 12-14m}=24, n_{MEDIAN, 12-14m}=24, mean_{MEAN}=0.64, mean_{MEDIAN}=0.65).

The NMDS ordination analysis (Fig. 6) and correlation of the species cover percentages with $\varepsilon$ showed that the most common ground cover components are somewhat dependant on the soil $\varepsilon$. Positive correlation was found between $\varepsilon$ and cover-% of Betula pubescens (n=51, p<0.05, $r_s=0.51$), Vaccinium uliginosum ($r_s=0.35$), Luzula piloza ($r_s=0.34$), Deschampsia flexuosa ($r_s=0.34$), Salix phylicifolia ($r_s=0.34$), Equisetum sylvaticum ($r_s=0.41$), Empetrum nigrum ($r_s=0.32$), Carex globularis ($r_s=0.35$), and Polytrichum commune ($r_s=0.77$) indicating that these components were more abundant on mesic and wet sites than on dry sites. Moreover, Betula nana, Betula pubescens, Carex brunnescens, Polytrichum commune, Rhododendron tomentosum, and Vaccinium uliginosum were oriented in the direction of $\varepsilon$ in the NMDS ordination showing that they are associated with mesic and wet soils. The following ground cover components were more abundant on dry soils than on wet or mesic soils: Solidago virgaurea ($r_s=-0.36$), Epilobium angustifolium ($r_s=-0.40$), Polytrichum juniperinum ($r_s=-0.31$), rocks ($r_s=-0.55$), Pohlia nutans ($r_s=-0.35$), and Cladonia ($r_s=-0.55$) since their abundance was negatively correlated to $\varepsilon$. Bare soil, Cladina, Epilobium angustifolium, Hylocomium splendens, litter, Lycopodium clavatum, Nephroma arcticum, Pleurozium schreberi, and rocks were oriented to the opposite direction from $\varepsilon$ because they are most common on dry soils. Out of the most common species in the study site, only Vaccinium myrtillus can tolerate a wider range of SWCs as it is not correlated to soil $\varepsilon$.

Fig. 6. Non-metric multidimensional scaling ordination of micro sites (diamonds) at paper III study site in northern Finland based on field measured coverages of plant species and ground cover components. Indicative variables are displayed by lines including AISA radiance bands 2–17 (448.99–874.04 nm), normalized difference vegetation index (NDVI), electrical conductivity ($\sigma$), and soil dielectric permittivity ($\varepsilon$). Figure edited by Paavo Närhi and Viena Arvola, GTK.
Paper IV answers the RQ 4. The CCA ordination (Fig. 7) showed that soil ε and σ were significantly (p<0.01) correlated with the peatland vegetation composition similarly to soil pH (p<0.001) thus peat soil ε, σ and pH represent as important drivers for vegetation composition. The permutation test also showed the HyMap bands between 570–1324 nm and 1530–1793 nm were significantly correlated (p<0.001) with the vegetation CCA ordination showing that that entire HyMap spectral range contributes to the spectral separation of *aapa* peatland site types but the VIS and NIR plateaus provide most the information.

The CCA biplot revealed that the fifteen floristic *in situ* peatland site type classes (Eurola et al. 1984) could not be separated in the CCA ordination. The unsupervised PAM classification of the sites by their CCA scores showed that eight classes composed compact clusters. However, a pairwise comparison of the classes by permutational multivariate analysis of variance (MANOVA) showed that the classes did not separate spectrally. The process of merging the PAM clusters without a non-significant difference by their HyMap reflectance created four final clusters: swamp, eutrophic fen, sedge fen, and bog (Fig. 7, F=4.9–29, r²=0.19–0.53, p<0.05). According to the MANOVA these clusters also deviated by their species abundances (F=94.8–24.9, r²=0.29–0.45, p<0.001). When compared to the *in situ* subjectively determined generalized four classes the overall accuracy was 82.0% (kappa 0.737) indicating moderate success. Contrast between the *in situ* and CCA-PAM-MANOVA classifications existed especially in the transitions between site types.

The CCA ordination is very helpful in interpreting the four spectrally separable classes ecologically. In the CCA ordination (Fig. 7) the 'bog' class consisted of a distinct cluster primarily dominated by *Sphagnum fuscum*, *Empetrum nigrum*, *Rubus*...
chamaemorus, and Sphagnum angustifolium. The ‘bog’ class was located far from the soil pH, ε and σ indicating that bog was a dry nutrient-poor acidic site type. In the opposite ordination direction ‘sedge fen’ sites were clustered in the direction of high ε. They were associated with water on the surface and with low nutrient content (Eurola et al. 1984). ‘Sedge fens’ have the highest abundances of Sphagnum lindenbergii (mean 47.9%) and Eriophorum vaginatum (mean 4.6%). The nutrient-rich basic site types, i.e., ‘swamp’ and ‘eutrophic fen’, clustered together with high pH and σ. The CCA biplot (Fig. 7) and statistics (see Table 1 in paper IV) revealed that ‘eutrophic fens’ were dominated by Carex lasiocarpa and Trichophorum cespitosum, and swamps with Comarum palustre, Sphagnum majus, and Carex rostrata. Coverage of grass litter (mean 22.8% and 27.3% respectively) was high in both classes. Furthermore, ‘eutrophic fens’ were also characterized by high coverages of moss litter (mean 12.3%).

A permutation test run for the HyMap bands revealed that the entire measured spectral range excluding the water absorption feature around 1300 nm was significantly (p≤0.001–0.1) correlated to CCA ordination. Thus CCA contributed to feature selection of HyMap bands such that only the bands between 449–1793 nm and 2064–2358 nm were input into the SVM classifier. The bands at ranges 570–1324 nm and 1530–1793 nm were most significantly (p≤0.001) correlated to the CCA ordination indicating that those spectral ranges would be the most important for separation of the peatland site types.

The spatial distribution of the four classes was then predicted with a fuzzy multiclass SVM classification from the HyMap data (Fig. 8). The SVM results exhibited high efficiencies of prediction: the most accurate model returned by a ROC analysis gave AUC=0.946 for ‘bog’, AUC=0.951 for ‘sedge fen’, and AUC=0.999 for ‘eutrophic fen’. The overall accuracy varied between 78.5–87.8% (kappa 0.611–0.781) when each pixel was assigned into a class by the highest pixel probability. Efficiency of prediction was not available for the ‘swamp’ class due to the low number of calibration and valida-

Figure 8. Area under the curve (AUC) values derived from receiver operating characteristics curves (left axis) and overall accuracy for all three classes (right axis, grey diamonds) for accuracy assessment of support vector machine prediction for the peatland biotope classes. Accuracy of prediction was at its highest when only 10% of all calibration pixels were used for training. Figure edited by Paavo Närhi, GTK
tion areas in the study area. All accuracy measures indicated that the SVM provided good prediction accuracies with calibration sets of all sizes, yet the highest prediction success was gained with 10% of the complete calibration set (pixel \(n=2308\)): AUC=0.946 for ‘bog’, AUC=0.951 for ‘sedge fen’ and AUC=0.999 for ‘eutrophic fen’. The class-specific producer’s accuracy for: ‘bog’ was 89.1%; ‘sedge fen’ 84.8%; and ‘eutrophic fen’ 96.0%. Similarly user’s accuracy for: bog was 88.1%; ‘sedge fen’85.1% and; ‘eutrophic fen’ 96.0% when 10% of all calibration area pixels were used for training the SVMs.

### 4.5 Chapter summary

The results of this study showed that glacial till soil \(\varepsilon\) and concentrations of several chemical elements could be estimated statistically significantly with close-range VSWIR spectroscopy. The relationship between till soil \(\varepsilon\) and reflectance at a single narrow spectral range was well predicted with an exponential function, however, multivariate regression techniques, commonly used in chemometrics and machine learning, turned out to be most robust. Especially, the estimations with relevance vector machines performed slightly steadier compared to partial squared regression and random forests. The spectroscopic-chemometric predictions were only possible for tills with low organic matter content (<1.7 dag/kg) but could be improved by selecting spectral features only at SWIR wavelengths (>1740 mm) as inputs into the models. Prediction of chemical element concentrations in till soil the fine fraction with partial least squares regression was statistically significant for Al, Ba, Be, Co, Cr, Cu, Fe, K, La, Mg, Mn, Ni, P, S, Sc, Sr, Ti, V, Y, and Zn. However, the success of the predictions was highly sample set specific and most prosperous for sample sets with highly inter-correlated element concentrations.

The soil moisture based site suitability assessment for Scots pine was moderately successful. Applying fuzzy classification with neural networks on airborne AISA data and validation of it with receiver operating characteristics curves was found as an appropriate data processing solution because the element of risk could thus be incorporated into the suitability assessment. The neural network model performance was significantly impacted by optimization of training set size, representativeness of the training set, and NN model selection. Also post processing with filtering was found necessary in order to improve the classification accuracy for high spatial resolution AISA data (1.1 m) which spatial accuracy was relatively poor. Very high efficiencies of classification were gained for peatland site type mapping with fuzzy support vector machine classification and receiver operating characteristics curve validation. The key to success was optimization of the peatland class hierarchical content with canonical correspondence analysis, finding the natural clusters in the ordination space, merging the clusters based on multivariate analysis of variance, and naming the final clusters with ecologically significant nomenclature. As a visual technique the ordination approach is very helpful in interpreting the spectrally separable classes ecologically, but also as a feature selection method of spectral bands from the used medium resolution HyMap data.

### 5 DISCUSSION AND CONCLUSIONS

The purpose of the chapter 5.1 is to discuss and conclude the results of soil close-range sensing and imaging spectroscopy of vegetated surfaces from theoretical point. In chapter 5.2, concluding remarks concerning the practical implications of the results are made. Then the reliability and validity of hyperspectral data and data processing steps are explained in chapter 5.3. Finally, recommendations for future research are suggested in chapter 5.4.
5.1 Theoretical implications

5.1.1 Close-range detection of till soil water content and chemical element concentrations

Statistically significant predictions of $\varepsilon$ (3.6–26.3) were made with chemometric modeling of close-range VSWIR spectroscopic measurements in laboratory conditions (Fig. 3). The soil $\varepsilon$ could be predicted well with the exponential single variate model (mean of reflectance between 1444–1464 nm, validation $R^2=0.77$, RMSE=2.98) and also with several MVR techniques (validation $R^2=0.81–0.94$, RMSE=1.59–2.77). The MVR methods PLSR, MARS, BRT and RF, and RVM produced slightly higher accuracies than single variate exponential functions. Overall, RVM slightly outperformed the other methods. RVMs produced the highest validation goodness of fit and the smallest error values for the generic models including all four tills (Fig. 3B, D). Other MVR models did almost equally well (Fig. 3A, B, C, D). Especially, RF models had virtually equal RMSEs to RVM models (calibration RMSE<1.5, validation RMSE<2.5 $\varepsilon$ units, Fig. 3B, D). Although data mining techniques, and especially tree based methods, are designed for large sample sizes, RF and RVM also performed well predicting $\varepsilon$ also for datasets with small sample sizes (n=12–29, Fig. 3A, C). A regression method which could produce reliable methods with a wide range of sample sizes would be ideal in chemometry to avoid unnecessarily large sample sizes, and thus enabling the use of spectrometry in a variety of soil moisture monitoring and mapping applications. To answer the RQ 1, when predicting the till soil $\varepsilon$ from VSWIR it might be advisable to use the data mining techniques.

The relatively large variation between the model performances (Fig. 3A, 2C) might not be only a product of the selected regression technique but also the relatively small sample sizes in the till-specific predictions. With small sample sizes, such as in case of this dataset, the uncertainty in parameter estimation increases compared to larger sample sizes. Till 1 (Vuotsonkangas till, n=12) had the smallest population followed by till 4 (Haipankuusikko till, n=26), till 2 (Vaalolehto till, n=28), and till 3 (Kuorajoki till, n=29). This is not ideal for comparison of MVRs. For further evaluation of the models larger sample sizes are needed. Similarly, the results are valid only for the studied tills which have a relatively small spread of data: $\varepsilon$=3.6–26.3, clay fraction (<0.002 mm) 2.4–5.5%, fine fraction (<0.063 mm) 23.5–47.1%, and OMC 0.6–5.8% (dag/kg). For datasets with a larger spread the results might be different. For example, Viscarra Rossel & Behrens (2010) concluded that ANN, SVM, MLR, MARS, and PLSR outperformed RF and BRT when predicting $C$ (0.01–13.9%) and clay contents (2.8–79.20%) and pH (4.8–10).

The prediction errors of $\varepsilon$ (validation RMSE <2.5 $\varepsilon$ units) were somewhat in the same range compared to results of Liang et al. (2010) who also used MVR, specifically ANN, to predict soil moisture. They report RMSE-% as low as 1.04 (g/g, equals app. 1 $\varepsilon$ unit according to Saarenketo 2006) and maximum absolute errors of 2.66% (g/g, app. 1 $\varepsilon$ units) when using ANN regression with reflectance from three wavelengths (1314 nm, 1475 nm, 1979 nm, n=115) chosen by correlation analysis. The soil moisture Gaussian model by Whiting et al. (2004) produced an RMSE of 0.026 g/g at the low gravimetric water contents (range 0–0.32 g/g, n=2503), which is a magnitude lower error than gained with the MVR methods reported here. In this respect, the soil moisture Gaussian model might be even superior to the MVRs when predicting SWC from VSWIR spectra for soils at low water contents, i.e., below the critical water content (cut-off thickness). Even though SWC is better predicted with methods utilizing ranges of wavelengths in VSWIR, the results of paper I also show that monitoring of SWC in laboratory is also possible by utilizing the well known physically sound water absorption bands, especially the absorption feature at 1444–1464 nm. For field measured spectra, feature selection by inputting reflectance values from the ranges of 500–1300 nm, 1500–1850 nm, and 2100–2500 nm, to avoid the atmospheric water absorption features, should be done prior to regression modelling.

Spectral transformations had no significant effect on predicting till soil $\varepsilon$ whatsoever (Fig. 3). Only a slight increase in validation efficiency was produced with continuum removal and wavelet transformation, and also with second derivatives when RVM was applied (Fig. 3D). Some studies suggest that wavelets might produce a slight improvement to the prediction efficiency (Viscarra Rossel & Lark 2009, Viscarra Rossel & Behrens 2010). Mouazen et al. (2010) showed that feature selection by PLSR prior to ANN regression pro-

Hyperspectal close-range and remote sensing of soils and related plant associations.

Spectroscopic applications in the boreal environment
duced up to a 0.2 improvement in $R^2$ values when predicting soil element concentration. At the same time, Brown et al. (2005) express their concern about the use of derivatives as it can be unstable due to changing spectral contributions of various soil minerals. Shifts in absorption features can occur when other compounds are added or their concentration increases as demonstrated, e.g., in paper I for water at 1900 nm. Currently, there is no consensus on the use of transformations in preprocessing of the soil spectra. It would be necessary to systematically study all preprocessing procedures in order to draw further conclusions on the advantages of feature selection, noise removal, corrections or enhancement of spectral features for prediction of soil properties.

The masking effect of OMC>1.7% (dag/kg) coincides well with the finding by Galvão & Vitorello (1998) who conclude that OMC>1.7% (dag/kg) tends to obliterate the effect of chemical constituents in VIS. Organic matter has weak absorption features throughout the VSWIR spectra, and because of the relatively good spectral response OMC can be quantified quite well in many cases (see summary in Stenberg et al. 2010). However, the results of this study indicate that quite low quantities of organic matter may already impede accurate quantification of other soil constituents. In case of forest soils in Lapland, the organic matter content in illuvial and eluvial horizons is usually insignificant. The OMC of parent tills has been observed to be less than 1.5%, and 80% of data from plowing depth (0.3–0.5 m) indicate OMC to be less than 2% (Sutinen et al. 1993). Thus organic matter does not pose a major problem for applicability of VSWIR spectroscopy for the shallow northern boreal soils.

To answer the RQ 2, the chemical element concentrations of till fine fraction were predicted with PLSR statistically significantly ($R^2 \geq 0.5$) for Al, Ba, Be, Co, Cr, Cu, Fe, K, La, Mg, Mn Ni, P, S, Sc, Sr, Ti, V, Y, and Zn. The rock forming minerals in tills, i.e., quartz, feldspars, and amphibolites are spectrally merely featureless (Clark et al. 2007) but the spectral contribution for statistically significant detection of elemental concentrations are due to spectrally active crystalline and amorphous secondary minerals. Prediction of mineral soil chemistry is only possible in the presence of spectrally active mineral materials and cross-correlation of elements with the spectrally active or extraneous soil components. The results of paper III demonstrate this issue very well. In till 1 only the concentration of Fe could be predicted with spectra (Fig. 4B), whereas concentrations of a number of elements could be predicted for till 2. In paper II, it was argued that the low standard deviation of elemental concentrations would be the main reason for this difference. However, the Figures 4B and 4C reveal that the concentration differences are minor. Further processing of the paper II data revealed that the overall correlation between the concentrations of elements is much higher for till 2 than for till 1. The inter-correlation of elements is a determining factor for the number of elements which concentration can be simultaneously predicted with VSWIR spectroscopy. It is also the key issue in understanding why prediction of soil constituents can be highly variable in success (Stenberg et al. 2010).

For modeling any soil property generic, i.e., global empirical models would be more desirable than till-specific models. The results of predicting both soil $\varepsilon$ and chemical elements concentrations indicate that generic models can be statistically significant (Fig. 3). Overall till-specific models were, however, more accurate than generic models when predicting soil $\varepsilon$ (Fig. 3). In case of till chemical element concentrations, cross-modeling was not statistically feasible (Fig. 4). Modeling elemental concentrations of tills from one area with a model created with data from another area was not tangible. However, applying the generic model to site-specific modeling was not attempted. In conclusion, successful application of generic models is more likely for predicting SWC than concentrations of elements due to the presence of absorption features of water and weak features of minerals which are indirectly correlated to chemical element concentrations.

5.1.2 Remote sensing of boreal ground cover plant associations

The ordination results of the papers III and IV confirm the findings of some earlier studies (Salmela et al. 2001, Tahvanainen et al. 2002, Närhi et al. 2011) about the edaphic contribution of moisture ($\varepsilon$) and nutrient potential ($\sigma$) on distribution and abundance of several vascular species and lichens on mineral soil in central Lapland (Fig. 7). According to the NMDS and correlation analysis, 12 species were found to be abundant on wet and mesic sites, and nine species of dry sites. Some species, e.g., *Vaccinium myrtillus* can tolerate a wide range
of moisture conditions but is one of the most common species. In addition, pH was a driver of vegetation composition on organic soils of aapa peatlands in southern Lapland (Fig. 7). Based on CCA ordination analysis, 11 species were related to the constrained ordination, where five of them were associated with high pH and σ, and one with high ε (Fig. 7). These species are also the most abundantly occurring species on the peatlands. The contribution of the soil gradients on the species distribution of the most abundant species is the primary requirement of successful application of VSWIR remote sensing in detection of the underlying root zone soil properties. Moreover, several inorganic and senesced ground cover components such as a soil, rock, and litter attributed to the spectra and thus to the successful recognition of the Scots pine suitability classes and peatland site types. This is further demonstrated by correlation of the hyperspectral bands to the ordination space (Fig. 6, 7). The ordination approach might also provide means for feature extraction of the hyperspectral bands.

The results of the study demonstrated moderate success (AUC=0.741) of soil moisture based suitability assessment for artificial regeneration to Scots pine (Fig. 5). To answer the RQ 3, the clear-cut treated with prescribed burning and mechanical site-preparation could be mapped into ‘suitable’ and ‘unsuitable for pine’ subareas using hyperspectral imaging of ground cover plant species patterns. The results suggest that applying MNF transformation of spectral AISA radiance channels and careful calibration of the RBFLN function, instead of PNN, would produce consistently the most accurate results. The classification accuracies were not compared between AISA radiance and those transformed with MNF but several papers show that reduction of the spectral dimensionality with MNF improves classification by extracting noise from data (Harris et al. 2006a, Zhang & Xie 2012). The comparison of PNN and RBFLN showed that, although RBFLN is much slower than PNN, RBFLN is capable of generalizing more efficiently than PNN with heterogeneous spectral data, such as remotely sensed understory vegetation patterns.

Sensitivity testing of the classification variables showed that the RBFLN should be trained with >2.43 calibration points/ha and ε class division between ‘suitable’/‘unsuitable for pine’ classes should be set to ε≤15/ε<15 rather than ε≤13/ε≥17 for optimal performance. The final accuracy can be improved by median or mean resampling of the resulting fuzzy membership image to at least 12 m pixel size. These options are dataset specific but demonstrate that calibration set selection, size, resampling pixel size, and ANN algorithm can have a profound impact on the classification accuracy. These results agree with the earlier studies which suggest that 1) ANN algorithms train better if the calibration vectors cover the entire distribution of values (Hepner et al. 1990, Foody 2000), 2) a higher number of calibration points produces increased accuracies to a certain extent (Aurora & Foody 1997), and 3) optimum number of calibration points are very much algorithm, network complexity, and data dimension dependent (Kavzoglu & Mather 2003). This means reoptimizing of the parameters for each dataset and ANN classifier independently.

The question remains whether optical remote sensing of the soil moisture by spatial species patterns is applicable in practice. The applicability of the method remains untested in different habitats, e.g., on areas recently clear-cut but not yet disturbed by site-preparation nor burning. Acquiring remotely sensed data immediately after clear-cutting, but before other site treatments, would be the most logical from a practical perspective. The study site was a formerly Hylocomium-myrtillus type Norway spruce-downy birch mixed forest but site-prepared and treated with prescribed burning ten years prior to the hyperspectral data collection. The comparison of the species abundances in surrounding forest and the clear-cut demonstrated that completely different species dominate in the forest than on the clear-cut. Many of the most abundant ones, however, are indicators of underlying soil moisture (Salmela et al. 2001). Secondly, the true success of methodology can be verified on the study site by inventorying of the artificially regenerated Scots pine approximately 20 years after the seeding that took place in 1989 and 1992 (Sutinen et al. 2002b). Eventually, the costs of the approach should be subjected to economical calculations which would also consider method uncertainties.

The ability of the ordination methods to reveal the community structure was utilized in this study to cluster peatland site types which were spectrally separable. A constrained CCA ordination was first constructed and the field determined site types were plotted in the ordination (Fig. 7). Then natural clusters of the study sites were revealed with unsupervised PAM clustering with the ordination
scores, of which the spectral separability was tested with MANOVA. Based on the MANOVA the peatland site type classes were combined until four generalized spectrally separable biotopes were left. The field determined classes were also generalized to match the hierarchal level of the purely data driven CCA-PAM-MANOVA classes. The disagreement between the CCA-PAM-MANOVA and in situ subjectively determined generalized classes (overall accuracy 82.0%, kappa 0.737) was visualized, especially in the transition between classes. It indicates that at this scale, clusters of spectrally separable cohesive communities existed but they gradually changed from one plant association to another.

Three of the four classes were then mapped with SVMs to produce a continuous fuzzy representation of the class extents (Fig. 8). Moderate-to-good successes was calculated with the class-specific ROC analysis: 0.946 for ‘bog’, 0.951 for ‘sedge fen’, and 0.999 for ‘eutrophic fen’, and overall accuracy: 87.8%, and kappa: 0.781. These data processing steps combine the benefits of feature selection and optimization of class hierarchical level. Thus spectrally separable classes can be named in an ecological context. For these reasons the proposed approach can be considered an improvement in combining the ecological in situ and remotely sensed information. Compared to the traditional separability analysis of spectral classes with the ordination approach the ecological basis of combining the classes can be more readily considered and provides a substantial improvement to the signature separability analysis. To answer the RQ 4, the success of the SVM classification is a result of optimizing the class levels with the CCA-PAM-MANOVA approach, and can thus be recommended as a processing step in future remote sensing based classifications of peatlands. Ordination based methods are visual and thus the ecological interpretation of a combination of ecological-environmental-spectral data may be more readily accomplished with them than with other statistical separability analysis only.

The two common non-linear classifiers, ANNs and SVMs were applied to classify the vegetation in shrub, herb, and bryophyte layers on organic and mineral soils in this study. The aim was not to compare the classification methods but to select an approach that can generalize complex hyperspectral data. Our results support the general consideration that SVM is fairly reliable, even though SVMs have difficulties dealing with noisy data and data with extremely high dimensionalities (Muntekalis et al. 2011). Often SVMs have outperformed ANNs (Atkinson & Tatnall 1997, Mas & Flores 2008) but recent developments in ANN techniques have improved their capabilities to equal the SVMs. For example, adaptive learning vector quantization by Zhang & Xie (2012) and a supervised classifier based on the learning vector quantization by Kohonen (2001) have been demonstrated to perform equally well to SVMs (Trinder & Salah 2011). Advanced versions of SVMs, such as adaptive binary tree, which creates a binary tree automatically based on Jeffries-Matusita distances, are recently introduced (Du et al. 2012). Individual classifiers are shown to perform well in some situations and fail under other conditions (Dudoit et al. 2002). The identification of the best procedure is often rather difficult, especially when the performance of a number of candidates is evaluated for learning samples of limited size. In conclusion, the results of this study demonstrate that SVMs and ANNs can be considered a valuable tool for processing of hyperspectral data (e.g., Bruzzone & Melgani 2004, Van der Linden et al. 2007).

Because of the distinctly clustered appearance but continuous community variation of the peatland and clear-cut canopies, fuzzy classification can be considered the most suitable for mapping of the class extents. Fuzzy classification with ANN, SVM or other non-linear classifier could predict both spatially sharp boundaries and transitional zones between peatland site types. In contrast to discrete classifiers, fuzzy approach allows visualization of gradual transition of plant communities which can be viewed as low and/or equal class membership values, homogeneity of communities with abrupt transitions between stands, and within stand heterogeneity (see Schmidtlein et al. 2007 for discussion). Fuzzy membership bands can be visualized in a way that they reveal uncertainties as low class membership scores or as high spatial variation of membership scores, which could indicate possibly an unconsidered thematic class in the study area. Therefore, local variation in species composition and the degree of softness of the concrete boundaries can be visualized more readily with the fuzzy than the discrete approaches, and thus are preferred to sharp boundaries.
5.2 Practical implications

5.2.1 Close-range detection of till soil water content and chemical element concentrations

This research demonstrates the potential of VSWIR spectroscopy in predicting SWC in laboratory conditions with high spectral resolution data. In the paper the focus was on determining the statistical relationship between ε and reflectance. Thus most of the reflectance variables were selected on the basis of theoretical absorption features centered at 942–1135 nm, 1379–1450 nm, 1800 nm, and 1920–1935 nm (Curcio & Petty 1951, Hunt 1977) caused by absorption of water molecules. Because atmospheric water interferes with wave propagation in remotely sensed data especially at the strongest water absorption features at 1379–1450 nm and 1920–1935 nm it would not be ideal to use these wavebands when airborne or spaceborne data are applied. In addition, prediction efficiency is deteriorated in the VIS, which is caused by the masking effect of organic matter (Galvão & Vitorello 1998). Therefore, remote sensing of soil moisture in the SWIR bands approximately around 1100–1300 nm, 1600–1850 nm, and 2100–2500 nm would be most suitable for remote sensing of soil moisture.

The results of further processing of data in paper I promote the use of MVR techniques for predicting soil moisture. They are also easy to use, and require little knowledge about the spectra of the measured material. They can be applied quickly to produce a predictive model. But chemometric methods are also a ‘black box’ for an average user and may be prone to overfitting if appropriate care is not taken with regard to the independency and proper size of calibration and validation dataset (Prasad et al. 2006). From a practical point of view, it would be essential to establish rough guidelines for MVR technique selection. New improved MVR techniques emerge constantly in data mining. At the same time, applicability of the present MVRS to chemometric modeling is not evident based on theory. The applicability of each algorithm depends on the dataset, i.e., soil property to be predicted, ranges of these response variables and spectrally active variables, existence of the spectrally active variables, spectral range, pretreatment of samples, spectral instrument, and illumination setup. Hastie et al. (2009) conclude that nature of the target function has a strong influence on the performance of prediction approaches. Yet, only a few studies compare the data mining algorithms in chemometry (Viscarra Rossel & Behrens 2010). Overall best performance was by ANN and the worst by RF and BT for a large dataset (n=1104) which had a large spread of soil C, clay content, and pH. However, the differences between the models really were not very significant and also varied by spectral transformation (Viscarra Rossel & Behrens 2010). The MVR comparison should be intensively conducted by the research community to gain understanding of the MVR algorithms in chemometry (Nduwamungu et al. 2009) but the most important issues are quality of the spectral data and understanding the meaning of cross-correlation between spectrally active soil components and the predicted variable.

Empirical SWC modeling with VSWIR spectra over a wide SWC range in theory is restricted to monitoring a single sample. This is because the critical point, i.e., cut-off thickness, is soil dependent (Weidong et al. 2002). The critical point has been shown to differ according to field capacity and porosity. Thus SWC can be spatially estimated with spectra over a range of samples only up to the lowest critical point of all samples. These facts limit the practical use of spectroscopy only to applications which allow higher uncertainty or for initial investigations which are further going to be improved with another approach. Increase in till soil reflectance at high ε was not observed but more of leveling off the reflectance at ε>15. Once ε>15 the soil reflectance is almost constant and water induced information is lost within the variability in reflectance caused by other factors contributing to the soil reflectance, e.g., organic matter, mineralogy, and surface roughness. Reflectance increase at high ε was not observed because we did not reach full saturation of the samples (max. ε=42). Full saturation of surface soils is rare during the growing season in the boreal climate (Hänninen 1997, Sutinen et al. 1997). Thus at least two moisture classes: ε≤15 and ε>15 could be separated based on VSWIR data.

The practical application of the ε-reflectance relationship would be assessing subareas of clear-cut and mechanically site-prepared forest compartments which would be suitable for artificial regeneration to Scots pine. The exponential nature of reflectance decrease with increasing ε demonstrated...
in this study for till soils enables such a classification approach. A simple single band cut-off point at reflectance corresponding to $\varepsilon=15$ could be used to divide a dataset of extracted soil pixels from optical high spatial resolution data and field measured $\varepsilon$ to produce a spatial representation of site suitability for Scots pine. Another alternative to thresholding would be classification as demonstrated by Siira et al. (2000b). First pure soil pixels were extracted from digitized high resolution (0.8 m) false color aerial photographs. The maximum likelihood classification with field referencing of $\varepsilon$ measurements produced the overall highest user’s (up to 87%) and producer’s (up to 92%) accuracy measures for the suitable and unsuitable Scots pine regeneration areas. Two unsupervised classification were used and validated with in situ $\varepsilon$ measurements. The isodata clustering and k-means classifier provided almost identical results of 77% overall accuracy. Both approaches could serve as effective alternatives to traditional soil survey methods when determining soil suitability for Scots pine during forest regeneration on site-prepared clear-cuts on formerly Norway spruce-downy birch clear-cuts in Fennoscandia.

The key question is whether calibrations in field conditions would be accurate enough for quicker more cost-efficient data processing in practical applications such as geochemical work or forest soil studies. Although concentrations of several elements measured from partially leached solutes were predicted statistically significantly ($R^2>0.5$, Fig. 4A, B), the confidence of the predictions may still be too low for practical applications. The elements with a large range of concentrations were also predicted with higher RMSE errors (Fig. 4A, B). PLSR models with $R^2>0.9$, according to Reeves & Smith (2009), would provide sufficiently high accuracy. The usability can be estimated by interpreting the element-specific RMSE values, which are application specific. For example, in mineral exploration, metals which often occur in low quantities should be accurately predicted as they are most interesting from the viewpoint of geochemical exploration. False positives should be avoided as they initiate unnecessary target drilling and then resources are wasted. On the other hand, false negative anomalies would mean undiscovered mineral resources.

Applicability of global chemometric prediction models is currently unresolved. Universal geochemical calibrations have proven impossible with diffuse reflectance Fourier transform spectroscopy (Reeves & Smith 2009) and with diffuse reflectance spectroscopy (Cohen et al. 2007). On the other hand, results reported by Brown et al. (2006) suggest that there is great potential for worldwide calibrations of spectroscopic methods and exhaustive global spectral libraries. Combining large databases with many different calibrations (Sankey et al. 2008) might facilitate the use of global models in predicting elemental concentrations of tills. Global databases have been proposed as a solution (e.g., McCarty et al. 2002, Brown et al. 2006). Several studies, however, note that for large datasets with a large range in soil properties the prediction accuracy drops because of non-linearity in the relation between reflectance and the soil properties. This would limit the use of large databases in local scale studies. Thematic or spectral stratification of the large dataset into smaller subset have been shown to improve the estimates of several soil properties (Bartholomeus et al. 2011, Ge et al. 2011). The implementation of VSWIR in close-range soil sensing is limited because of the low robustness of the chemometric techniques, and consequently many local calibrations. Currently, most models perform better under controlled conditions and yield better results on a local scale.

Several other practical issues have to be solved before VSWIR spectroscopy can be applied in predicting mineral soil element concentrations. There is a great contradiction in sample preparation by wetting to improve the predictions of element concentrations and other soil properties. The best estimations are often gained with dried samples. For example, Bogrekci & Lee (2006) found that removal of water from soil samples prior to spectral measurements significantly improved the predictions of P by measuring the SWC and artificially removing the effect of water on the spectra. Similar results are also demonstrated by Sudduth & Hummel (1993) for organic C, and by Chang et al. (2005) for a variety of soil properties. At the same time, some mineral absorption features which can be indirect indicators of presence of elements are perhaps accentuated or may not appear until the soil is wetted. Dematte et al. (2006) shows that this phenomena is especially pronounced in the presence of hydroxide bearing minerals such as kaolinite and montmorillonite. Kusumo et al. (2008) demonstrate quite accurate predictions of total C and N despite moisture differences between samples. Viscarra Rossel et al. (2009) propose the
showed that the use of External Parameter Orthogonalisation (EPO-PLSR) was a more powerful tool for removing the effect of moisture when predicting soil C in field conditions. The EPO-PLSR quite successfully removes an effect of an external parameter which has an impact on the spectra.

The desirability of spectroscopy is primarily based on cost-savings. Cohen et al. (2007) estimate the cost of a VSWIR measurement to be less than 10% of the traditional wet soil chemical measurement costs. However, only a few papers deal with the optimal sample sizes for calibration and validation. The results of Brown et al. (2005) found that ~20–35% of the full sample size is needed for adequate calibration of PLSR modes. This would mean that the technique is not as cost efficient as intended. Moreover, greater care has to be taken for selecting the validation dataset as random selection can overestimate the predictive accuracy (Brown et al. 2005).

5.2.2 Remote sensing of boreal ground cover plant associations

To be accepted as a standard forest management practice, a SWC based Scots pine suitability assessment methodology should meet the data requirements of affordability, high accuracy, and large spatial extent. A few methods to map SWC for site suitability exist. Classification of low altitude airborne radiometric gamma data has been suggested for regional mapping of site suitability (Hyvönen et al. 2007). In central Lapland, on the basis of this low resolution mapping heterogeneity of forest compartments could be estimated but it cannot considered detailed enough for recognition of the spatial soil water patterns within forest compartments. Several dielectric ground methods have been applied to map site suitability at the compartment level (Hänninen 1997). To cut down the number of field measurements and costs (Siira et al. 2000b) proposed classification of exposed mineral surface SWC from inexpensive digital high resolution false colour aerial photographs. In comparison to aerial photography, the airborne hyperspectral data used in this study are currently too costly to be applied in practice. Resampling of the classification results into larger pixels sizes (original 1.1 m) demonstrated that medium resolution (≥12 m) data might be more appropriate than high resolution data. In addition, NDVI turned out to be a significant variable in the NMDS ordination space. Thus medium spatial and spectral resolution satellite data might be adequate for optical remote sensing based Scots pine suitability assessment already before hyperspectral satellite data will be available at affordable prices.

For soil moisture estimations several other potential remote sensing methods also exist. A remote sensing based method could be based on thermal inertia or microwave frequency backscatter. In physical principles they might be more appropriate than the use of VSWIR data, e.g., water strongly affects the bulk soil ε and thus the backscatter of microwave radiation (reviews by Stafford 1988 and Wang & Qu 2009). Due to longer electromagnetic wavelengths the penetration depth of microwave and LWIR is also better compared to VSWIR. However, the limitation of LWIR and microwave data is the poor spatial resolution. The data from cost-efficient satellite based sensors are currently too coarse for site-specific studies of soil moisture (see Kaleita et al. 2005, Mulder et al. 2011). For this reason VSWIR domain is the most operational as the data are most available and affordable (Muller & Décamp 2001).

The goal in paper IV was to study the applicability and develop processing methods of hyperspectral imaging data for peatland site type mapping keeping in mind the future prospective national wetland inventory. Such spatial representation of peatlands across Finland is missing in contrast to other peatland rich countries, such as Sweden (Boresjö Bronge et al. 2006) and Canada (Fournier et al. 2007), which are currently conducting the work based on satellite optical and microwave data. The high hierarchal class levels including fen, eutrophic fen, bog, and swamp matches well with the classes suggested for the Canadian wetland inventory including bog, fen, marsh, swamp, and shallow water. Therefore, four might be close to an optimal number of thematic classes. Thus four classes may be appropriate for the national level inventory when hyperspectral data from satellite platforms become available.

A widely applicable outcome of paper IV was the ordination based spectral signature separation procedure combined with fuzzy mapping. Previously, correspondence analyses combined with unsupervised clustering and indicator species methods, have been employed for a meaningful class division
of peatland environments (Thomas et al. 2003), on arctic tundra (Atkinson & Treitz 2012), and landscape-forest level (Malik & Husain 2006) with moderate success. A non-classificatory continuous mapping of ordination scores has been performed by predicting the ordination axis score values from spectral channels and then interpreting the patterns spatially as single axis scores (Ohmann & Gregory 2002, Schmidtlein & Sassin 2004, Thessler et al. 2005, Schmidtlein et al. 2007, Feilhauer et al. 2011). The advantage of the CCA-PAM-MANOVA approach compared to the previously suggested ones is that it combines community clustering and continuous nature of community transition in the spatial domain. The CCA-PAM-MANOVA approach is only applicable in situations where environmental variables are well known as constrained ordination was used. However, unconstrained ordination could also be applied in a variety of environments and applications.

In the case of peatlands, several remotely sensed datasets acquired during one summer season might improve the classification accuracy. Flooding period in the spring should be avoided as differences in surface moisture between site types have a great impact on the final success. Van der Meer et al. (2012) are also asking for more attention to process related studies with hyperspectral remote sensing to improve the efficiencies of classification. In northern Finland acquisition of temporal datasets, however, is difficult because of the low frequency of cloud free days during the short growing season. Therefore, incorporation of temporal datasets into vegetation classification may be operational only occasionally.

Although hyperspectral close-range and remote sensing of plant and soil media have been available since the 1970’s and became available for the wider scientific community since the 1990’s, their progress in becoming widely used research tools has not been as rapid and successful as was suggested. Issues such as complexity of the datasets, radiometric and atmospheric corrections and selection of appropriate data modeling and classification techniques, high cost of the data, shortage of people working in the hyperspectral science community, and educational opportunities limit the expansion of imaging spectroscopy for more common use. Many authors also agree that low S-N of the imaging spectrometers (Ben-Dor et al. 2009, Van der Meer et al. 2012) and low spatial resolution (Price 1997) hinders the separation of viewed materials. Advances in algorithm research are constantly published (e.g., Chanussot et al. 2010) but the time lag until the methods are available for a user might be significant. Ben-Dor et al. (2009) concludes some practical aspects of imaging spectrometry that require development in order for imaging spectroscopy to expand. These include breakthroughs in electro-optics technology to obtain near laboratory quality spectra with high S-N ratio, enlarging the spectral window to UV and LWIR, obtaining temporal coverage, training of specialists, encouraging commercial activity, educating the market, reducing prices, providing nearly real-time services, and developing new successful applications.

5.3 Reliability and validity

A study is reliable if the results are consistent over time and if the results accurately represent the total population under study (Golafshani 2003). In quantitative research, such as remote sensing, the concept of reliability is commonly considered as uncertainty. In statistical terminology, uncertainty consists of accuracy, bias, and precision (Atkinson & Foody 2002). In this study the uncertainty is reported as accuracy, which is considered a sum of unbiask and precision. Uncertainty of a model derived from remotely sensed spatial data consists of inaccuracies related to the measurement process and ones added by the processing steps (Schott 2006). The total uncertainty of a thematic spatial representation derived from remotely sensed data consists of uncertainties related to the input variables, used models, their parameterization, spatial support of the data, and position (Dungan 2002). In the following, the most significant error sources in these uncertainty categories are discussed.

The quality and usability of the spectroscopic signature is largely determined by the S-N performance, uniformity and stability of imaging systems, and also the atmospheric interferences. The measurement devices used in this study were spectrally and radiometrically calibrated prior to the surveys. The AISA instrument was calibrated by the Finnish Forest Research Institute (Mäkisara...
et al. 1993) and HyMap by Hyvista corp. (Cocks et al. 1998). A long data preprocessing chain applying a variety models is needed to correct for the hyperspectral remotely sensed data, which in turn causes error accumulation by assumptions related to models and their parameters.

Datcu et al. (1998) suggest that model selection might be one of the largest sources of uncertainty. In the preprocessing phase, the raw digital numbers from the instruments were first turned to on-board-radiance values according to the instrument-specific calibration functions provided by the instrument suppliers. Further, bidirectional correction with nadir normalization was conducted on both datasets to compensate for the viewing and solar illumination geometries. To compensate for the signal interference by atmosphere the HyMap data also went through an atmospheric correction. The atmosphere was assumed to be constant on the small AISA flight area and, therefore, atmospheric correction was skipped as it was not expected to improve the classification. Geometric correction and geocoding were done to correct for the distortions caused by the aircraft movements. The MNF transformation was applied to paper III data to remove as much of the noise in the data as possible. The affects of MNF transformation were not tested in this research, and MNF was not applied to HyMap data in paper IV.

Uncertainty was also introduced by the classification and regression models which were applied to produce the thematic end products. Often times model selection is not performed as it can be laborious. In this work, ANN models were compared in paper III and regression models in further processing of the paper I data. In paper III RBF LN was clearly found to be more efficient than PNN. In further processing of paper I data, the sample sizes were small and validation could not be performed with an independent dataset for till-specific models. For the generic $\varepsilon$-reflectance model proper validation was performed. The RVMs performed slightly better than other models but this is in contradiction to some other studies such as Viscarra Rossel & Behrens (2010) performed with larger sample size. In conclusion, the model selection can be a major source of uncertainty but very difficult to control as selection of models and their parameters are strongly dataset dependent.

Parametric uncertainty in this study was accomplished using calibration data in the supervised ANN and SVM approaches and the MVR methods. Emphasis was put in the sensitivity analysis of the model parameters in paper III. The conclusion was that with ANN algorithms the quality of the calibration set and the choice of the ANN algorithm had a greater impact on classification accuracy rather than optimizing the ANN calibration parameters (number of functions and number of iterations) to perfection. Although this does not apply until a sufficient number of calibration points is exceeded. The quantity of calibration might also be very low if the calibration is representative of the classes. The kernel based SVM used in paper IV is similar such that the optimal performance of the algorithm is determined in the calibration stage by using calibration areas. Over- and underfitting are avoided by optimizing the kernel width parameter and regularization parameter in a cross-validation process. Testing of sensitivity for the calibration set could not be performed in such detail as there was a lack of calibration sites of all four peatland site type classes. In the chemometric regression, models were also constructed based on calibration sets, which were chosen from the complete datasets randomly.

Spatial support in remotely sensed data refers to the area covered by an effective resolution element, which is a function of the instantaneous field of view, flight variables, and atmospheric effects (Dungan 2002). The cell size or ground instantaneous field of view are used as approximations of the spatial support. From the user perspective, the spatial support of airborne data is provided by the manufacturer and flight operator and they do not usually estimate the uncertainties involved in the spatial support. The spatial resolution is a determining factor for the model performance such that the thematic class hierarchical level is determined by cell size. In paper III optimization of the cell size was attempted for the two class classification and the best accuracies were gained with cell sizes $>10 \text{ m}$. In paper IV the number and thematic content of peatland classes were optimized to match the cell size with the CCA-PAM-MANOVA approach. Both approaches significantly improved the accuracies.

A major source of uncertainty for the HyMap data was created as the IMU data were missing due to a human error causing a 30 m discrepancy between true coordinate locations and the HyMap data. The lack of ground control points in sparsely or uninhabited study sites such as in papers III and IV is a problem. In case of the AISA data the lack
of ground control points caused an RMSE of 20 m. In addition, spatial distortion in the HyMap data was highly variable from place to place. The spatial inaccuracies are very common in remotely sensed data. When combining in situ data with remotely sensed data the positional uncertainty can be improved by averaging pixel values around expected field locations. In paper III, pixel values from a same size window were extracted and averaged to represent each 5 m by 5 m field site. To better compensate for the spatial discrepancy it might have been beneficial to choose a larger window. In paper IV, spatial discrepancy was better compensated by conducting a pixel based unsupervised classification and also a segmentation prior to planning of the sampling scheme.

More accurate predictions can be considered to be drawn from close-range spectral data than from remote sensed spatial data. In close-range sensing the uncertainties related to spatial support and position are non-existent or irrelevant. With remotely sensed data the correction steps taken to compensate for the atmospheric effects, platform movements, and geocoding cause increase in the total uncertainty of the model due to the uncertainties in the parameterization of each data processing step. Therefore, models constructed on close-range sensing data such as $\varepsilon$ and element concentrations may be less uncertain than the models build on remotely sensed data. Factors such as relatively low S-N ratio of remotely sensed data compared to laboratory data, spatial variability in surface roughness, viewing angles, the bidirectional reflectance property of soils (Cierniewski & Courault 1993) and illumination conditions, and atmospheric effects on wave propagation would increase the uncertainty of remote sensing of soil properties of exposed soils.

It is a common practice in spectroscopy to validate the information derived from hyperspectral data against observations acquired by other measurement techniques or by visual in situ observations. Uncertainty measures of those reference data should also be reported in order to truly monitor the uncertainty of the results. For example, in geochemical research uncertainty related to geochemical, sampling, and analytical variability are frequently monitored with measures of uncertainty calculated against field duplicates and reference samples (e.g., Ramsey 1998). Such quality control and assurance aspects of field or laboratory reference data were not considered in this study and are often ruled out by limited resources but, in fact, they should be conducted both for reference data and spectral data to produce a complete estimate of uncertainty.

Validity, on the other hand, is a more general term determining whether the research truly measures what it was intended to measure or how truthful the research results are (Golafshani 2003). Heavy assumptions of causality between vegetation reflectance and soil properties have strong implications on the validity of this study. However, it is demonstrated in papers III and IV with statistical methods and with accuracy assessments that a statistical co-variation between SWC, nutrients and pH, and vegetation compositions exists. Moreover, it was also shown that the spatial species patterns have distinct spectral responses and that these spatial patterns can be mapped with remotely sensed hyperspectral data. In paper III, moderate success was gained for the soil moisture site suitability for Scots pine assessment when the accuracy was assessed towards measurements of $\varepsilon$. However, an issue of validity remains. Although remotely sensed spatial patterns of the understory are adjusted to the time stable patterns of soil moisture variations, the field measured $\varepsilon$ represent soil moisture only in one point of time. Due to the time stability of soil $\varepsilon$ those measurements made late in June represent the magnitude difference of SWCs between sites during the growing season. Although soil $\varepsilon$ is used only as a proxy of pine survival it all comes down to the uncertainty of the prediction. The true uncertainty can be evaluated after the pine seedlings have exceeded the critical 20 years of a plantation to stabilize (Hansson & Karlman 1997). Such long term studies should be conducted in order to evaluate the true uncertainty of the approach. Although, the mortality of pine saplings has been related to excess SWC and incidences of concomitant diseases, Gremmeniella abietina in particular the causality is still unexplained (Witzell & Karlman 2000, Sutinen et al. 2002b, Mäkitalo 2009). Despite all of the uncertainties related to the site suitability approach, it still might serve as the most effective approach in assessing soil moisture based site suitability for Scots pine on the forest compartment level.

An issue of validity was also present in paper II where the till soil chemical element concentrations were predicted from spectra. The absorption properties of soils are merely caused by molecular
processes, i.e., mineralogy, although electronic processes also cause soil absorption features. Thus the elemental information extracted from spectra is largely indirect mineralogical information, and the success of quantification of elements is highly resultant from cross-correlation between elements due to their co-existence in minerals or of minerals. Therefore, it would have been very important to be able to quantify some spectrally active secondary minerals such biotite, muscovite, sericite and mixed-layer clay minerals in tills in order to draw further conclusion of the location of the detected elements in the till soil. However, the study was lacking such proper pretreatment of the till samples followed by XRD analyses.

5.4 Recommendations for future research

The close-range soil sensing with VSWIR and chemometry is a relatively young branch of science, and thus several research steps are needed to make conclusions concerning its usefulness. One of the major drawbacks of spectroscopy-chemometrics with respect to soil geochemistry is that the calibration and validation datasets should cover the entire range of concentrations as extrapolation would cause unreliable concentrations. In geochemical work the out of line samples are usually the most interesting and in mineral exploration may potentially be the ones indicating underlying mineralizations. Samples on the flanks of the distribution are usually few and thus models might not be able to predict those values accurately. An additional issue related to distribution is the geological range. Some studies state that development of reliable VSWIR prediction models for soil chemical properties should be limited to geologically homogenous areas (Chodak et al. 2002, Udelhoven et al. 2003). Others suggest that there is great potential for spectroscopic methods with worldwide calibrations and exhaustive worldwide spectral libraries would solve much of these problems (e.g., Brown et al. 2006). Future research in geochemical studies should involve using spectroscopy as a screening tool for more expensive laboratory analyses such as suggested by Chang & Laird (2002) when predicting soil C and N. Additional research is also needed to develop spectral similarity measures to detect spectral outliers (e.g., Ramirez–López et al. 2011) and use principal component analysis to plot the observations into a standardized principal component space to identify spectrally unique or dissimilar observations that might be difficult to model (Shepherd & Walsh 2002, Brown et al. 2005).

Another weak point of empirical models is the difficulty in extending predictive equations to other datasets. Ustin et al. (2009) concludes that compared to empirical models, physically based radiative transfer models have the potential to produce more accurate and consistent predictions of pigment interactions because the radiative transfer models are based on physics and the use of full spectrum. Therefore, calibrations would not be required for each dataset. Compared to vegetation, less modeling work has been conducted in the field of soil transfer models implying that the theoretical approach might bring new insight into modeling quantities of soil constituents. Radiative transfer modeling has been applied extensively to lunar surface research. Applying Hapke’s (2005) radiative transfer theory Li & Li (2011) were able to use inverse modeling to predict FeO, mineral abundances, and particle size with reasonable success. Liang (1997) proposes numerical radiative transfer modeling to understand the optical depth of soil but also notes that for large particle soils, e.g., sandy soils, geometric-optical models might be more appropriate. In addition, the microtopographical issues caused by macrotopography of agricultural soils (tilling) may have a significant influence on the reflectance and should be considered in modeling (Cierniewski et al. 2010). Overall, more research should be conducted on radiative transfer modeling of soils.

Nduwamungu et al. (2009) discuss the developments in soil spectroscopy practices which are needed to make it fully operational. According to them the key issues for better comparison of the model performance is standardization of sample preparation, measurement techniques, spectrum acquisition, spectral pretreatment techniques, calibration statistics, and reporting of the metadata. An interesting observation by Stenberg (2010) suggest that rewetting of a soil sample to 20-30% before spectral measurement would improve the prediction of especially organic C but also clay. Although the mechanisms are yet unknown, further studies are needed for exploring its usability. Overall, uncertainties in modeling for soil
properties might be too large with current VSWIR sensors but improvement of sensor technology at MWIR and LWIR ranges hold a high potential of improving the soil predictions (Linker 2011).

Besides the constant advances in non-linear classifiers, object based image analysis methods will also become more common in hyperspectral data processing. Currently SVMs are already implemented in the commercial object based image analysis software packages (e.g., eCognition, Trimble Geospatial Imaging, Munich, Germany). Incorporation of object based texture into vegetation classification is also available for wide user communities when processing hyperspectral data. It has already been demonstrated to be very powerful in, e.g., subtropical wetland studies (Zhang & Xie 2012). However, when working with transitional community structure the applicability of an object based approach has to be carefully evaluated. Hyperspectral classifiers are advanced to include contextual information. The spectral-spatial classifiers are reported as being successful for data with large homogenous regions and where spectral signatures of multiple classes overlap (Chanussot et al. 2010).

5.5 Chapter summary

The main conclusions of this thesis were:
1) Hyperspectral spectroscopy was feasible in novel geoenvironmental applications in the boreal region of Finnish Lapland. Quantification of glacial till soil dielectric permittivity and concentrations of several elements was statistically significant using single and multivariate regression techniques. Classifying ground cover plant associations for peatland site type mapping and assessment of site suitability for artificial regeneration of Scots pine from hyperspectral imaging spectroscopic data was also successful. These applications of imaging spectroscopy are progressed further towards being applied in practice but significantly more research is needed to solve the issues related to prediction of chemical element concentrations of glacial till soils.
2) Dielectric permittivity of glacial till soils can be slightly better predicted from close-range sensing of VSWIR spectra with RVM and RF compared to exponential single variate regression. Although this study was conducted on small samples set with relatively narrow spread of data (i.e., $\varepsilon$=3.6–26.3, clay fraction 2.4–5.5%, fine fraction 23.5–47.1%, OMC 0.6–5.8%). Thus further studies with datasets having large spread and variability are needed for more generalized conclusions.
3) The exponential soil $\varepsilon$-VSWIR reflectance relationship and leveling-off reflectance at $\varepsilon$~15 would enable classifying exposed soil pixels on mechanically prepared forest compartments in order to assess the soil moisture based suitability for artificial regeneration of dry soil tolerant ($\varepsilon$<15) Scots pine. High spatial resolution optical remote sensing data and ground $\varepsilon$ calibration are required for this task.
4) Prediction of concentrations of several elements was possible with SWIR spectra for fine fraction of glacial tills but only in the presence of spectrally active mineral material and cross-correlation of elements with the spectrally active or extraneous components.
5) The sample set specificity of the empirical calibrations, great prediction uncertainties related to global models, and lack of physical models currently restrict the applicability of VSWIR-chemometric modeling in estimation of till soil chemical element concentrations.
6) Hyperspectral imaging of soil moisture driven spatial plant community patterns for soil moisture based site suitability assessment for Scots pine was moderately successful on a previously Norway spruce-downy birch dominated clear-cut in central Lapland. The success was related to the soil moisture subjectibility of several abundant understory species.
7) Using medium spectral and spatial resolution optical remote sensing data for the suitability assessment could be a cost efficient way for forest managers to minimize risk related to Scots pine regeneration. Economic analyse are still needed to assess the final cost-benefit and applicability to a variety of site types.
8) A constrained ordination based statistical approach was developed for optimizing the class hierarchal level of $aapa$ peatland site types from medium resolution hyperspectral imaging data. The future intent would be to use this approach as a preprocessing step for mapping of peatlands, e.g. for national wetland inventories to support conservation planning and modeling of greenhouse gas emissions.
ACKNOWLEDGEMENTS

This thesis is a product of years of interacting, communicating and sharing ideas between people who have a passion for nature, science and sophisticated risk taking in order to push the boundaries of 'the traditional'. I remain in debt to my instructor and mentor Raimo Sutinen for introducing spectroscopy and interdisciplinary thinking when I entered work life as a young university student, and for the long term encouragement and support over the years. Thank you also to my unofficial instructor Markus Törmä for promptly giving feedback in the writing stages of this work. Although he mostly ‘sensed’ the work ‘remotely’ due to the significant geographic distance between us, this approach felt very appropriate for the research topic. Sincere thanks also go to my supervisor professor Henrik Haggrén for accepting me as his student without a background in engineering and for guiding me through the practical stages of my studies.

I want to express my sincere gratitude to the co-authors Hilkka Arkimaa, Eija Hyvönen, Viljo Kuosmanen, Paavo Närhi, Raimo Sutinen (Geological Survey of Finland, GTK), Ari Teirilä (Kajaani University of Applied Sciences), and Paul Treitz (Queen's university) for their contribution to the original research papers. Paavo Närhi and Ari Teirilä devoted their statistical skills for bringing this work up to the publishable standards. Hilkka Arkimaa, Eija Hyvönen, Viljo Kuosmanen, Raimo Sutinen, and Paul Treitz shared data, ideas and their precious time for guiding me. The expertise of GTK personnel Terttu Aaltonen, Viena Arvola, Timo Hirvasniemi, Pentti Hölttä, Jukka Laitinen, Päivi Kuikka-Niemi, Eira Kuosmanen, Tapio Muurinen, Markku Mäkilä, Vesa Nönkäinen, Marja Liisa Räisänen, Mia Tiljander, and Kimmo Virtanen has contributed to data processing, interpretation of the results and editorial work. Instructions also given by Gary Raines (University of Nevada, USGS), Carlos Roberto de Souza Filho (State University of Campinas), Kai Mäkisara (Finnish Forest Research Institute), and Lea Hämäläinen (Labtium Oy) contributed to data acquisition and interpretation. Special thanks go to GTK personal Paula Haavikko, Matti Piekkari, and Markku Virtanen and to Mari Jakonen and Mari Kuoppamäki (University of Oulu), for conducting field work, sample preparation, spectral measurements, and some data preprocessing. For the vegetation inventories I sincerely thank Paavo Närhi, Saara Salmela (WSP Finland Ltd.), and Oddvar Skre (Skre Nature and Environment). I have enjoyed working with all of you and all the ones I forget to mention at GTK and elsewhere. Kent Middleton revised the language of all the papers several times. Thank you also to Ed Cloutis (University of Winnipeg), Neal Scott (Queen's University) and several anonymous reviewers for constructive comments on the manuscripts of the research papers. My sincere gratitude also goes to the reviewers Angela Harris and Alexander McBratney for their thorough commenting of this thesis synopsis.

Over the years I have also been privileged to study remote sensing with the Canadian academia. Lectures and discussion with professors Myryka Hall-Beyer and Steven E. Franklin at the University of Calgary in 1998-1999 inspired me greatly, and reconfirmed my desire to study remote sensing. Professor Paul Treitz provided me with a peaceful place to retreat with proper facilities and a great research environment in the LARSEES laboratory, at the Queen's university Geography Department in 2010-2011 where I was able to concentrate on finalizing this work. Warm hearted wishes go to the LARSEES crowd including Karin van Ewijk, Greg McQuat, Adam Collingwood, and Fiona Gregory, and all the faculty, staff and students at the department for making my year memorable.

This thesis was financed by several sources. Emil Aaltosen säätiö, Seppo Säynäjäkankaan Tiedesäätiö and GTK international fund provided financial support for my 'writing sabbatical' to Queen's university. The AISA data were acquired as part of the HYDO (Hyperspectral remote detection and mapping of geological objects) – project between the GTK, National Technology Agency of Finland and mining industry in 1998-2001. The HyMap data were acquired as a part of the MINEO project (Monitoring and assessing the environmental Impact of mining in Europe using advanced Earth Observation Techniques, EU funded, IST–1999-10337, 2000-2003). Viljo Kuosmanen was the engine for both of these projects. The laboratory spectroscopic work was initiated in the Academy of Finland project Mosaic patterns of forest soils (1997-2000) and continued in the GTK funded project Global change and forest soils (2007-2012) managed by Raimo Sutinen. I want to express my gratitude to all of my present and past
managers at GTK for enabling me to conduct a majority of the work on salary.

The greatest thanks of all go to my husband Kent who’s been running our family of four for the past few years. Thanks to Sofia and Benjamin for keeping me busy also on times off. Thank you also to my Finnish and Canadian families, especially to my mother-in-law Rita Middleton and my sister Marjo Siira, and all my friends near and far for being there for me. I will now stop neglecting you with the excuse of the thesis. My loving thoughts go especially to my parents Pirkko and Kari Siira always promoting education to me. Thanks dad for letting me know already years back that once I will get my Ph.D. you will dig into your pocket one more time for the celebrations. Hey, it’s time to party!

REFERENCES


Hyperspectral close-range and remote sensing of soils and related plant associations.


Terhoeven-Urselmann, T., Schmidt, H., Georg Joergensen, R. & Ludwig, B. 2008. Usefulness of near-infrared spectroscopy to determine biological and chemical soil...


This Ph.D. thesis comprises of a synopsis and four original papers dealing with geoenvironmental applications of hyperspectral close-range and remote sensing relevant to northern Finland. Visible and short wavelength infrared (350–2500 nm) close-range spectral data was used for quantification of glacial till soil water content and chemical element concentrations. Remotely sensed imaging spectroscopic data was applied in mapping of site suitability for artificial regeneration to Scots pine (Pinus sylvestris) and peatland site types. The predictions were produced with statistical and machine learning techniques. The results demonstrate that glacial till soil element concentrations and moisture can be statistically significantly predicted to improve cost efficiency of soil analyses. Hyperspectral remotely sensed data is readily applicable for soil moisture based site suitability assessment as part of forest management practices. In the future, conservation planning or estimation of green house gas emissions could benefit from ecological peatland site type mapping.