JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE
284

Aleksi Räsänen

Developing and Comparing Methods for Mapping Habitat Types and Conservation Values Using Remote Sensing Data and GIS Methods

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 284

Aleksi Räsänen

Developing and Comparing Methods for Mapping Habitat Types and Conservation Values Using Remote Sensing Data and GIS Methods

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212
kesäkuun 18. päivänä 2014 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in building Seminarium, auditorium S212, on June 18, 2014 at 12 o’clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2014

Developing and Comparing Methods for Mapping Habitat Types and Conservation Values Using Remote Sensing Data and GIS Methods

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 284

Aleksi Räsänen

Developing and Comparing Methods for Mapping Habitat Types and Conservation Values Using Remote Sensing Data and GIS Methods

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2014

Editors
Anssi Lensu
Department of Biological and Environmental Science, University of Jyväskylä
Pekka Olsbo, Ville Korkiakangas
Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-5712-4
ISBN 978-951-39-5712-4 (PDF)

ISBN 978-951-39-5711-7 (nid.)
ISSN 1456-9701

Jyväskylä University Printing House, Jyväskylä 2014

ABSTRACT

Räsänen, Aleksi
Developing and comparing methods for mapping habitat types and conservation values using remote sensing data and GIS methods
Jyväskylä: University of Jyväskylä, 2014, 77 pp.
(Jyväskylä Studies in Biological and Environmental Science
ISSN 1456-9701; 284)
ISBN 978-951-39-5711-7 (nid.)
ISBN 978-951-39-5712-4 (PDF)
Yhteenveto: Elinympäristötyyppien ja suojeluarvojen kartoitus kaukokartoitus-aineistojen ja paikkatietomenetelmien avulla
Diss.

In this research, new methods for mapping habitat patches and types, conservation values, and species richness were developed in order to locate areas that are or are not valuable from a conservation perspective. The study was performed in two primarily forested southern boreal zone landscapes. The main datasets were airborne laser scanning data and high spatial resolution multispectral images. First, segmentation methods for delineating habitat patches were compared by means of supervised evaluation measures and visual interpretation. It was found that choosing the best segmentation is often arbitrary and that supervised measures give inconsistent results. Second, classification of the delineated segments was performed using a random forest classifier and several different features derived from remotely sensed datasets. For classification, several features were needed in order to map different habitat types and to get the highest classification accuracy. Third, different habitat types were given values, based on potential number of species, their rarity and naturalness. Conservation value maps were then drawn using different methodologies, taking habitat type connectivity and complementarity into account. The resulting maps were compared with maps of selected ecosystem services. Conservation value maps differed largely from each other, and thus great care should be used when mapping conservation values. Fourth, habitat type classification was used together with topographic, geodiversity, and other landscape features to explain and predict vascular plant species richness patterns. Most of the variance in species richness could be explained through mean altitude, which had a negative relationship with species richness, and landscape variability, which had a positive relationship. While the methods developed here can be utilized (e.g. for land-use planning), different uncertainties must be taken into account.

Keywords: Airborne laser scanning; conservation value; habitat type; object-based image analysis; segmentation; species richness; spectral images.

Aleksi Räsänen, University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Finland

Author's address Aleksi Räsänen
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
t.aleksi.rasanen@jyu.fi

Supervisors Anssi Lensu, PhD
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

Professor Markku Kuitunen
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

Reviewers Adjunct Professor Timo Kumpula
The Department of Geographical and Historical Studies
P.O. Box 111
FI-80101 Joensuu
Finland

Duccio Rocchini, PhD
Fondazione Edmund Mach
Research and Innovation Centre
Department of Biodiversity and Molecular Ecology
GIS and Remote Sensing Unit
Via Mach 1
IT-38010 San Michele all'Adige (Trento)
Italy

Opponent Professor Petri Pellikka
Department of Geosciences and Geography
P.O. Box 64
FI-00014 University of Helsinki
Finland

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following original papers, which will be referred to in the text by the Roman numerals I-IV.

I am the first and corresponding author in all the articles (I-IV). I planned the articles together with my co-authors and did the most of the data analyses. I wrote the first draft of the articles and completed them in co-operation with my co-authors.

I Räsänen, A., Rusanen, A., Kuitunen, M. & Lensu, A. 2013. What makes segmentation good? A case study in boreal forest habitat mapping. *International Journal of Remote Sensing* 34: 8603–8627.

II Räsänen, A., Kuitunen, M., Tomppo, E. & Lensu, A. Coupling high-resolution satellite imagery with ALS-based canopy height model and digital elevation model in object-based boreal forest habitat type classification. Accepted to *ISPRS Journal of Photogrammetry and Remote Sensing*.

III Räsänen, A., Lensu, A., Tomppo, E. & Kuitunen, M. Comparison of different remote sensing and GIS methods to evaluate conservation values and ecosystem services in Southern Finland. Submitted manuscript.

IV Räsänen, A., Kuitunen, M., Hjort, J., Vaso, A., Kuitunen T. & Lensu, A. The role of landscape, topography, and geodiversity in explaining and predicting vascular plant species richness in a fragmented landscape, Southern Finland. Submitted manuscript.

ABBREVIATIONS

ALS	airborne laser scanning
CHM	canopy height model
CLC	Corine Land Cover
DSM	digital surface model
DTM	digital terrain model
DTW	distance to water
FNEA	fractal net evolution approach
GAM	generalized additive model
GIS	geographic information system
GLCM	gray-level co-occurrence matrix
GLM	generalized linear model
HTC	habitat type classification
MRVBF	multiresolution index for valley bottom flatness
MS-NFI	multi-source National Forest Inventory
NDVI	normalized difference vegetation index
NLS	National Land Survey of Finland
OBIA	object-based image analysis
SDM	species distribution model
SH	spectral heterogeneity
SR	solar radiation
SWI	SAGA wetness index
TPI	topographic position index
TRI	terrain ruggedness index
TWI	topographic wetness index
WGS84	World Geodetic System 84
WV-2	WorldView-2 satellite imagery

1 INTRODUCTION

1.1 General introduction

The loss of biodiversity and the degradation of ecosystem functioning have been noted as among the most serious – if not the most serious – environmental problems on Earth. Changes in land use are a major driver in this loss, both globally (Rockström *et al.* 2009) and nationally in Finland (Rassi *et al.* 2010). Thus, land use and the planning of land use play a key role in addressing the loss of biodiversity. In many planning processes, such as land use planning or environmental impact assessment, evaluation of the natural environment has been used in order to prioritize areas for development or for conservation (Margules and Usher 1981, Smith and Theberge 1986, Spellerberg 1992). At least in Finland, however, it has been noted that ecological impact assessments performed in planning processes are often superficial. They are done too late in the process, they do not consider broader treatments of biodiversity and ecosystem services, and they are not integrated well enough into the process (Söderman 2012).

One of the approaches for early stage ecological impact assessment is habitat ranking (Rossi and Kuitunen 1996, Hilli and Kuitunen 2005). In habitat ranking, habitat types are given values on the basis of their potential number of species and species' conservation status. In other words, in this approach, habitat types that potentially include a large number of species overall and many threatened species are the most valuable. The mapping of habitat types is performed manually by experts by means of interpreting different maps. In this study, habitat ranking is automated, updated and evaluated critically. Automation means that geographic information systems (GIS) and geographic information analysis methods are used and, for the most part, data is remotely sensed.

Traditionally, the habitat ranking method took into account vascular plant species and vertebrates. In this study, however, only vascular plants are considered (III, IV). The reason why only plants are considered is that even though plants are just one taxon and partly depend on other organisms, they are the basis of the food chain (Whittaker *et al.* 2001).

The introductory chapter provides a background on the main topics of this research, and some of the key concepts are defined and discussed. The introduction starts with a discussion of conservation values and mapping methods for different conservation values. After this, an object-based image analysis (OBIA) methodology for mapping habitat types is presented, with a focus on segmentation evaluation (topic of I) and data selection (II). Then, different approaches and principles of spatial conservation prioritization (III) are discussed. The introduction ends with a short discussion about vascular plant species richness and modeling of species richness (IV).

1.2 Explaining conservation values

Conservation value (Fig. 1) can be considered as the value of nature from a conservation perspective. High conservation value areas are those that should primarily be conserved. Nonetheless, the concept of conservation value is not widely established. Generally speaking, the concept itself can be rather problematic, since in evaluations of (natural) areas the focus is not always on conservation. In terms of why something may have a high conservation value, several different criteria or principles have been suggested. Arguably, the most used principle is biodiversity, which is usually presented as consisting of three levels: genes, species and ecosystems. Usually, however, instead of considering all levels of biodiversity, often the focus is on species diversity (Orme *et al.* 2005, Wilson *et al.* 2006), and some species (e.g. threatened, rare, endemic, keystone) can be regarded as more valuable or targeted for conservation (Margules and Usher 1981, Smith and Theberge 1986, Groves *et al.* 2002, Duelli and Obrist 2003, Orme *et al.* 2005).

Biodiversity or even species diversity is tedious to measure as such; therefore, surrogates for biodiversity are usually used. Surrogates include, for example, habitat types or a subset of species (Margules and Pressey 2000, Pressey 2004). There are also many different indices for measuring species or other types of diversity (Tuomisto 2010). One set of indices measures diversity in a community or a location as α-diversity, differentiation of diversity between communities or locations as β-diversity, and diversity inside an area (i.e. a combination of α and β) as γ-diversity (Whittaker 1960). Another set of indices combines evaluations of richness and evenness of the species composition. These indices include, for example, the Shannon index and Simpson index (Tuomisto 2010).

Biodiversity is not the only suggested principle for conservation. Geodiversity can be regarded as important as such (Gray 2013) or as providing conditions for biodiversity (Parks and Mulligan 2010). Naturalness (i.e. lack of human influence) has been argued to be the basis for conservation and even for deciding how much diversity should exist (Angermeier 2000). Terms such as ecological or biological integrity can also be related to the natural state of a system (Karr 1990, Angermeier and Karr 1994). Ecological health (Karr 1990), however, refers to ecosystem functioning and can be defined metaphorically as an absence of distress (Callicott *et al.* 1999).

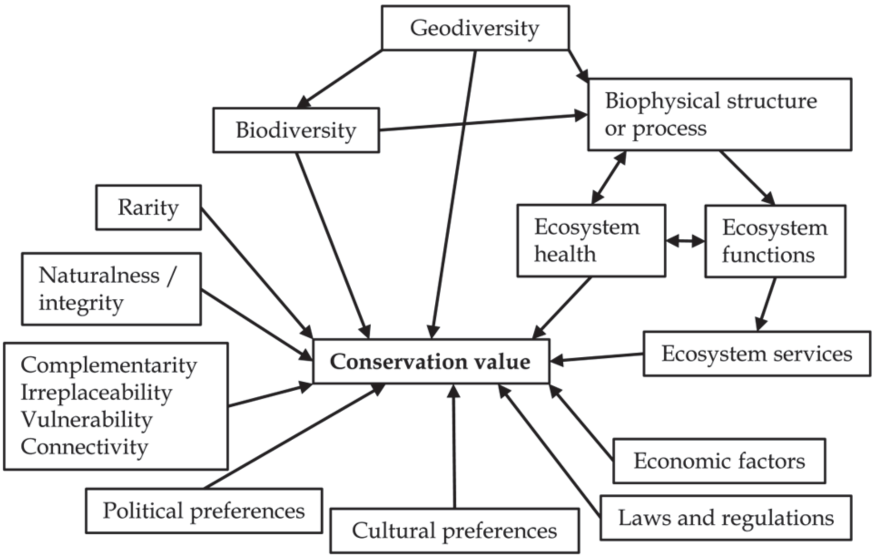

FIGURE 1 A simplified illustration of some of the factors that reduce or increase the conservation value of something. In the end, conservation value is an anthropogenic measurement set by humans. Some biophysical or geophysical factors may reduce or increase conservation value, but humans set the value of these factors.

Ecosystem services, which are used in conservation decisions more and more widely, can be defined in numerous ways. Roughly speaking, they are the contributions that natural ecosystems provide for human well-being. Ecosystem services can be divided into four main types: provisioning services, regulating services, habitat services and cultural services (De Groot *et al.* 2010). Ecosystem services can be understood as providing conditions of instrumental values of nature to humans, as opposed to intrinsic values of nature (Chan *et al.* 2012). Intrinsic value means that something has a value as an end in itself. Instrumental value is the value of something as a means to another's ends (Callicott 1997). Another way to define these terms is that if something has intrinsic value, it is valuable as such, independent of the valuer. Additionally, if something has instrumental value, it is regarded as valuable by the valuer (Justus *et al.* 2009). In other words, all of the value given to nature can be considered as instrumental value, since it needs to be given by someone. In this manner, the "intrinsic" value of biodiversity or nature, irrespective of the direct or indirect benefits it provides, can also be defined as an ecosystem service (Mace *et al.* 2012).

Through the application of different principles, two schools of conservation biology can be described. These schools are not antagonistic but complementary, and different principles can also be used together. For the compositionalist school, humans are considered to be separate from nature, and emphasized principles include integrity and diversity. The functionalist school regards humans as being a part of nature, and it emphasizes principles such as ecosystem services and ecological health (Callicott *et al.* 1999).

In addition to certain defining principles, a plethora of different criteria have been suggested for conservation. Such criteria include principles, but also other factors that are taken into account when conservation decisions are made. These criteria can be divided into two categories. Scientific criteria are measurable and can consist of biodiversity, rarity and naturalness. Political criteria are not based on biological, ecological or biogeographic concepts. These criteria include availability and amenity value (Margules and Usher 1981). It can be argued, however, that all conservation criteria are essentially political or at least normative (i.e. value laden) (Callicott *et al.* 1999, Kalamandeen and Gillson 2007).

When nature is valued, valuation can also be divided into subjective and objective components (Geneletti 2003, Villarroya and Puig 2012). It has been claimed, nonetheless, that all valuation is performed by humans (Justus *et al.* 2009). In this manner, subjectivity cannot be removed from planning decisions. Therefore, it is not possible to be objective; instead, one should be explicit in the valuation and selection of criteria (Smith and Theberge 1986, Margules and Pressey 2000, Pressey 2004). Overall, it is crucial that conservation criteria are selected with care, since they define the actions that are done in conservation or in land use planning (Angermeier 2000, Kalamandeen and Gillson 2007).

1.3 Spatial distribution of conservation values

For mapping different conservation values, several different solutions have been suggested. In this work, 'mapping' means visualizing or analyzing spatial distribution of something (e.g. conservation values) on maps. When biodiversity or species distributions have been mapped, both direct and indirect methods are suggested. Direct mapping means the identification of a single species or of species assemblages with remotely sensed information, such as satellite imagery. In indirect mapping, some environmental variables are used as surrogates (Nagendra 2001, Turner *et al.* 2003).

Different indirect methods include using spectral values or the spatial variation of spectral values as proxies for species diversity (Nagendra 2001, Turner *et al.* 2003, Rocchini *et al.* 2010, Nagendra *et al.* 2013). Other habitat requirements can also be related to the distribution of species or species richness (Nagendra 2001, Turner *et al.* 2003). The term 'habitat' is usually defined as the resources that are present in an area and needed by organisms. The term 'habitat type' is defined as a mappable land unit in which vegetation and environmental factors are fairly homogenous. However, the terms 'habitat' and 'habitat type' are often used interchangeably (Corsi *et al.* 2000).

Remote sensing based habitat type mapping is often based on land use / land cover classification. Land cover refers to biophysical surface characteristics of the Earth and land use to land utilization by humans (e.g. Kerr and Ostrovsky 2003, McDermid *et al.* 2005). Habitat types can be inferred from land cover only or from land cover and other environmental factors (McDermid *et al.* 2005). One methodological approach for mapping habitat types is explained in Section 1.4. Habitats, on the other hand, are usually mapped using species distribution models (for SDMs, see Section 1.6). In SDMs, environmental and spatial characteristics needed by different species are mapped instead of vegetation or habitat type (Elith and Leathwick 2009).

When the distribution of species or another level of biodiversity is known, spatial conservation prioritization can be used for mapping sites with the highest conservation value (Margules and Pressey 2000, Moilanen *et al.* 2005, Moilanen *et al.* 2009). Spatial conservation prioritization approaches are explained in more detail in Section 1.5.

Mapping approaches also exist for other conservation principles than biodiversity. For naturalness, different simple indices have been constructed. In these indices, the highest value is given to the most natural environment, and the state of naturalness is examined using remote sensing data interpretation and fieldwork (Machado 2004, Villarroya and Puig 2012). In more complex indices, several metrics of human influence can be combined in the assessment of integrity or naturalness (e.g. McGarigal *et al.* 2011). The state of naturalness or habitat condition can also be included in habitat type mapping methods (Vanden Borre *et al.* 2011, Spanhove *et al.* 2012, Nagendra *et al.* 2013). Remote sensing methods can be used to map invasive species (He *et al.* 2011, Nagendra *et al.* 2013) that lessen the state of naturalness and can reduce the instrumental values of nature.

Ecosystem services have been quantified by using direct observations, proxy data or process models (Egoh *et al.* 2012, Maes *et al.* 2012). The simplest method for assessing the potential supply of ecosystem services is a reference table or assessment matrix. In these matrices, regions are divided into different land use, land cover or habitat types, and the potential production of ecosystem services in them is valued either monetarily or otherwise (Costanza *et al.* 1997, Troy and Wilson 2006, Burkhard *et al.* 2012). There are also, however, more sophisticated efforts for mapping single or many ecosystem services. In these approaches as well, the primary GIS-data has been land cover or habitat type maps, but other GIS-datasets have also been used (Egoh *et al.* 2012, Maes *et al.* 2012, Crossman *et al.* 2013). The focus in ecosystem service mapping can be either in biophysically (or monetarily) accurate predictions (e.g. Troy and Wilson 2006, Willemen *et al.* 2008, Nelson *et al.* 2009) or in spatially explicit mapping (e.g. for making spatial comparisons or prioritizations of the studied areas) (Raudsepp-Hearne *et al.* 2010, Vihervaara *et al.* 2010, Egoh *et al.* 2011, Burkhard *et al.* 2012). Finally, approaches exist for mapping conservation value *per se* by means of a diverse set of criteria (e.g. Geneletti 2007, Samu *et al.* 2008). In many of the approaches for mapping conservation values, habitat type or land cover maps are used (e.g. Kerr and Ostrovsky 2003, Turner *et al.* 2003, Gillespie *et al.* 2008, Newton *et al.* 2009). In the next section, one methodological framework for mapping habitat types is introduced.

1.4 Object-based image analysis (OBIA)

1.4.1 Introduction to OBIA

OBIA, or Geographic object-based image analysis (GEOBIA), has been an emerging topic in remote sensing and geographic information science literature since the early 2000s, and it has even been noted to be a paradigm (Castilla and Hay 2008, Blaschke *et al.* 2014). OBIA differs from traditional pixel-based analyses. In pixel-based analyses, pixels are the main studied units. In OBIA, pixels are instead merged into objects using segmentation. In OBIA classification, segments are classified rather than pixels (Blaschke 2010, Blaschke *et al.* 2014). Ideally and in the widest terms, OBIA should mimic human observations by means of, among other methods, contextual and multi-scale analysis (Hay and Castilla 2008, Blaschke *et al.* 2014). In contextual analysis, objects are often not observed being in isolation; instead, their relationships to neighboring areas are surveyed. Multi-scale analysis means, for instance, that when observed from above (e.g. from an aerial or satellite image), a forest area consists of forest patches made up of individual trees and gaps in the canopy. Finally, OBIA is not only object-based; instead, it can also be object-oriented. More specifically, objects are often not only used in the first part of the analysis; they can also be altered during the analysis from primitive objects to objects of interest (Baatz *et al.* 2008). In this research, however, OBIA is used in a narrower scope of meaning through a typical segmentation–classification approach, using single scale and including no contextual information.

Compared to pixel-based analyses, the advantages of OBIA are also numerous in this narrower approach. Objects produced by OBIA are more meaningful entities and a more realistic representation of a landscape, especially when very high resolution (pixel size < 5 m) remote sensing data are used (Castilla and Hay 2008). Resulting maps do not suffer as much from salt-and-pepper effects (Blaschke 2010). In other words, resulting maps have more contiguous areas instead of pixels with varying values. Additionally, OBIA has produced better classification accuracies than pixel-based analyses (e.g. Bock *et al.* 2005, Díaz Varela *et al.* 2008, Whiteside *et al.* 2011, Yan *et al.* 2006). The availability of high resolution data, together with advances in computer software, can even be seen as two main reasons for doing OBIA analyses (Blaschke 2010). OBIA, however, is no panacea; pixel-based approaches may also give as good or better results. In some studies, OBIA has not given a statistically significantly better classification accuracy compared to pixel-based analysis (Dingle Robertson and King 2011, Duro *et al.* 2012). Moreover, although objects are more meaningful in OBIA, small rare classes can easily be merged into larger objects (Dingle

Robertson and King 2011). OBIA is also more labor-intensive than pixel-based analysis (Duro *et al.* 2012).

1.4.2 Segmentation and segmentation evaluation (I)

Segmentation is usually the first part of OBIA. In segmentation, images are partitioned into image objects (i.e. segments), which are prototypes of geographic objects or landscape objects, such as patches (Castilla and Hay 2008, Hay and Castilla 2008). In segmentation, the goal is to create meaningful objects (Castilla and Hay 2008). Meaningful objects are representations of geographic objects (Castilla and Hay 2008), mimic real-world objects (Zhang *et al.* 2008, Clinton *et al.* 2010), or minimize heterogeneity inside regions and maximize heterogeneity between regions (Zhang *et al.* 2008, Hou *et al.* 2013). Remote sensing segmentation methods are a special case of more general image segmentation methods, which have been also utilized in remote sensing for decades (Blaschke 2010). Segmentation methods can be divided into two types: similarity or region-based segmentation and discontinuity-based segmentation. In region-based segmentation, a similarity measure is used to find suitable regions. In discontinuity-based segmentation, discontinuities of the images, usually boundaries, are detected (Zhang 1997, Gonzales and Woods 2002). These groups are, however, only ideal types. For instance, watershed segmentation combines approaches from both groups. In watershed segmentation, boundaries between basin areas are sought by flooding the image (Gonzales and Woods 2002).

For deciding when segmentation is good (i.e. meaningful and appropriate), segmentation evaluation has generally been performed (Clinton *et al.* 2010, Marpu *et al.* 2010). One of the stated reasons for performing segmentation evaluation has been the argument that segmentation quality affects classification accuracy. In other words, it has been argued that better segmentations give higher classification accuracies (Kim *et al.* 2009, Clinton *et al.* 2010, Ke *et al.* 2010, Gao *et al.* 2011).

Segmentation evaluation has concentrated on developing new segmentation methods, comparing them, and optimizing parameters inside segmentation methods. In general, segmentation evaluation can be divided into two major categories: subjective (visual) evaluation and objective evaluation. Objective evaluation can be further divided into system-level evaluation, which evaluates the overall system in which segmentation is performed, and direct evaluation (Zhang *et al.* 2008). Segmentation quality has been assessed using final classification output as a system, for example (Smith 2010, Wang *et al.* 2010, Gao *et al.* 2011). Direct evaluation can be either analytical or empirical, with the former evaluating the method itself and the latter its results. Empirical methods consist of supervised and unsupervised methods (i.e. if ground truth is used as a reference or not) (Zhang *et al.* 2008).

In terms of remote sensing, analytical methods have been used in segmentation evaluation (Hay *et al.* 2003). However, most segmentation evaluation has been performed using unsupervised and supervised methods. In unsupervised segmentation evaluation, the most used approaches have arguably been based on heterogeneity and/or spatial autocorrelation. One approach seeks to minimize variance within segments and spatial autocorrelation between segments (Espindola *et al.* 2006, Gao *et al.* 2011, Yue *et al.* 2012). In another approach, optimal segmentations are hypothesized to have only low autocorrelation between segments (Kim *et al.* 2008, 2009). In supervised segmentation evaluation, either area-based or location-based measures have been used. Area-based measures evaluate if segmentation is either too coarse (under-segmentation) or too fine (over-segmentation). Over-segmentation and under-segmentation measures can also be combined. Location-based measures, on the other hand, are based on distances between segment centroids and reference polygon centroids or distances between boundary pixels (e.g. Clinton *et al.* 2010). Supervised measures of quality have been mostly applied when extracting features with clear borders, usually in urban and agricultural areas (Lucieer and Stein 2002, Möller *et al.* 2007, Zhan *et al.* 2005, Tian and Chen 2007, Weidner 2008, Clinton *et al.* 2010, Wang *et al.* 2010). In some studies, however, quality measures have been used in natural area segmentations (Carleer *et al.* 2005, Ke *et al.* 2010, Bar Massada *et al.* 2012). Another example of supervised segmentation evaluation is the comparison of thematic quality between segments and reference polygons (Pekkarinen 2002, Mustonen *et al.* 2008). Supervised segmentation evaluation can be performed using automated approaches with many reference polygons inside a large area (e.g. Clinton *et al.* 2010) or only a couple of distinct reference polygons and semi-automated approaches (e.g. Marpu *et al.* 2010).

1.4.3 Data and features in OBIA (II)

In OBIA studies, the most commonly used data types have been satellite or aerial imagery, often in very high spatial resolution (Blaschke 2010). It has been stated that one of the additional advantages of OBIA is its better chance of allowing textural analysis (Benz *et al.* 2004). Texture refers to changes in image brightness values (i.e. if neighboring pixel values are similar or different). Different textural features have also been used in pixel-based analyses (e.g. Coburn and Roberts 2004), but they are perhaps more meaningful in OBIA, since texture per image object can be quantified (Benz *et al.* 2004). In terms of different textural features, probably the most commonly used approach is the Gray-Level Co-occurrence Matrix (GLCM, Haralick *et al.* 1973, Haralick 1979) features, which are widely used in OBIA analyses (e.g. Yu *et al.* 2006, Johansen *et al.* 2007, Kim *et al.* 2009, 2011, Murray *et al.* 2010, Han *et al.* 2012, Sasaki *et al.* 2012). The GLCM quantifies how often different combinations of pixel brightness values occur, and several features can be calculated using the matrix (Haralick *et al.* 1973, Haralick 1979). There are also other approaches for quantifying texture and extracting features. These include semivariograms, filters and neural networks (Balaguer *et al.* 2010, Li *et al.* 2011), as well as wavelet features, which can quantify orderliness in imagery in larger neighborhoods than allowed by the GLCM (Arivazhagan and Ganesan 2003, Ruiz *et al.* 2004, Ouma *et al.* 2008, Su *et al.* 2012, Wang *et al.* 2012). It can also be argued that whereas the GLCM quantifies fine-scale texture, wavelets extract coarse-scale patterns (Morgan *et al.* 2010).

In addition to the multispectral images, also other data types have been used in OBIA studies. One of the most used data types is airborne laser scanning (ALS), a.k.a light detection and ranging (lidar) data. The usage of ALS has become increasingly popular in OBIA studies (Blaschke 2010). ALS is an active remote sensing technique, compared to passive techniques such as aerial or satellite imagery. In ALS, the sensor emits laser pulses that hit target surfaces and record their reflections; in passive sensors, only the reflectance of other electromagnetic radiation is recorded. The ALS reveals features that cannot be seen from above using passive sensors. Using ALS data, the three-dimensional structure of vegetation or buildings and a high-resolution digital terrain model (DTM) can be accurately quantified. The usefulness of ALS is evident, for instance, in forest habitats (Lefsky *et al.* 2002, Vierling *et al.* 2008). In the case of boreal forests, spectral reflectance seen from imagery does not give information about tree canopy height and soil moisture, for instance, which are vital characteristics in boreal forest habitat type mapping. In OBIA, vegetation structure has been quantified with ALS data (Antonorakis *et al.* 2008, Ke *et al.* 2010, Bar Massada *et al.* 2012, Sasaki *et al.* 2012), but ALS data have also been used to derive DTMs (Bar Massada *et al.* 2012, Ke *et al.* 2010). Furthermore, in terms of segmentation, it has been found that coupling imagery and ALS provides promising segmentation results and that the selection of input data has an effect on segmentation quality (Geerling *et al.* 2007, Mustonen *et al.* 2008, Ke *et al.* 2010, Hou *et al.* 2013).

Several different topographic features can be calculated from a DTM. These features include slope, aspect and curvature, which are widely used in OBIA studies (Yu *et al.* 2006, Thompson and Gergel 2008, Thompson *et al.* 2008, Ke *et al.* 2010, Morgan and Gergel 2010). Topographic features also include features that quantify local hydrologic or climatic conditions. Of the different hydrologic features, possibly the most widely used is the topographical wetness index (TWI, Beven and Kirkby 1979). The TWI models local soil moisture, taking local slope and upslope contributing area into account. The TWI has also been used in OBIA studies (Thompson and Gergel 2008, Thompson *et al.* 2008, Ke *et al.* 2010, Morgan and Gergel 2010). There are also other topographic features that are or can be incorporated into the OBIA workflow. Finally, also other types of data, such as soil or bedrock data, have been used in OBIA studies (e.g. Bock *et al.* 2005).

1.5 Spatial conservation prioritization (III)

Systematic conservation planning means that sites are prioritized in a systematic manner in order to conserve the most valuable areas (Margules and Pressey 2000). When prioritization is performed spatially using computational tools, the activity can be called ‘spatial conservation prioritization’ (Kukkala and Moilanen 2013). Usually spatial conservation prioritization is performed using species distribution data as input data (Margules and Pressey 2000, Kukkala and Moilanen 2013). Other types of data as well, such as ecosystem services (Chan *et* *al.* 2006) or habitat types (Lehtomäki *et al.* 2009), can be used. Furthermore, SDM results can be used. However, it has been found that prioritization results based on SDMs and species presence data may markedly differ (Altmoos and Henle 2007).

The most valuable areas are not necessarily the most species-rich areas. Instead, factors such as the vulnerability, irreplaceability, complementarity, connectivity and size of areas are taken into account (Margules and Pressey 2000, Wilson *et al.* 2009, Kukkala and Moilanen 2013). Vulnerability of the area refers to threats that the area will be converted to extractive uses. Irreplaceability means that the features that an area possesses cannot be replaced by conserving other areas. Complementarity means that conserved areas should complement each other (Margules and Pressey 2000). In other words, if all species in an area are to be conserved, a suitable minimum set of conserved locations is the minimum set of locations where all species exist. Connectivity can be defined as the degree to which a landscape facilitates or impedes movement of different organisms. Connectivity is divided into structural and functional connectivity. Structural connectivity refers to habitat contiguity and is analyzed only using landscape structure. With functional connectivity, on the other hand, responses of the organisms are considered (Tischendorf and Fahrig 2000). Structural connectivity, size and other spatial composition and configuration calculations that can be computed (e.g. with landscape metrics and indices) have also been used in planning and conservation value mapping outside the spatial conservation prioritization methodology (Botequilha Leitão and Ahern 2002, McGarigal *et al.* 2011, Uuemaa *et al.* 2013). Other factors that can be included in spatial conservation prioritization have also been included in other spatial analyses. In addition to ecological factors, political or economic factors can be included in spatial conservation prioritization approaches. It has even been claimed that the focus has been too much on economic factors, compared to other socio-political or ecological factors (Arponen *et al.* 2010).

1.6 Species richness and species distribution models (IV)

Species richness has been widely explained by means of area size and energy, using species-area (Connor and McCoy 1979) and species-energy relationships (Evans *et al.* 2005). That is, the larger the area or the more energy that is available there, the more species there are. Many of the factors that explain species richness can be reduced by means of different mechanisms to energy and area (Honkanen *et al.* 2010). Energy can be understood rather broadly and divided into solar and productive energy. Solar energy metrics, such as temperature, quantify the amount of available solar energy. Productive energy metrics, such as net primary productivity, measure the resources available for consumers (Evans *et al.* 2005, Honkanen *et al.* 2010). There are also many other factors used to explain species richness that cannot be reduced to area and energy. These factors can also be random; therefore, explaining why a particular area has a certain degree of species richness is far from straightforward. Other factors include historical factors related, for example, to dispersal and speciation, environmental stress, environmental stability, disturbance, and ecological interactions (Fraser and Currie 1996, Whittaker *et al.* 2001).

Recently, species richness has partly been explained in terms of geodiversity (Parks and Mulligan 2010, Anderson and Ferree 2010, Hjort *et al.* 2012, Ruddock *et al.* 2013), which can be defined as the variability of the Earth's surface materials, forms and physical processes (Gray 2013). Because geodiversity explains species richness and because it is more constant, as opposed to a perennially changing climate, it has been argued that areas to be conserved could be selected on the basis of geodiversity (Anderson and Ferree 2010, Beier and Brost 2010, Ruddock *et al.* 2013). Furthermore, it has been found that geodiversity and habitat diversity are positively connected (Jačková and Romportl 2008). Habitat diversity is, on the other hand, positively linked with species richness (Honkanen *et al.* 2010). Therefore, the factors that have an effect on species richness have also interactions.

SDMs use a range of different factors to map species richness (Guisan *et al.* 2006, Austin 2007, Elith and Leathwick 2009). SDMs are especially applicable on a mesoscale or landscape level. Although SDMs are usually constructed for a single species only, they have been widely used in richness mapping (Austin 2007, Elith and Leathwick 2009). When SDMs are used in species richness mapping, species distributions can be modeled one species at a time (Ferrier *et al.* 2002a, Thuiller *et al.* 2004, Raes *et al.* 2009), by assessing plant family species richness (Parviainen *et al.* 2009), at a community level (Ferrier *et al.* 2002b, Ferrier and Guisan 2006) or by looking at all species together (Heikkinen and Neuvonen 1997, Honnay *et al.* 2003, Hjort *et al.* 2012). There are few comparisons between different approaches. In one study, it was found that "one species at a time" and "all species together" approaches bring similar results in the current climate. In a changing climate, however, the "all species together" approach brings more realistic results, since in this approach high species richness areas were correlated with meaningful landscape structures (Guisan and Theurillat 2000).

Several different explanatory features have been used in SDMs. Perhaps the most often used have been bioclimatic features in large-scale studies and topographic features in mesoscale studies (Guisan and Zimmermann 2000). Other predictors include different remotely sensed variables. A normalized difference vegetation index (NDVI) is widely used as a proxy for productivity (e.g. Nagendra 2001, Turner *et al.* 2003, Parviainen *et al.* 2009), while spectral heterogeneity serves as a proxy for habitat heterogeneity (Rocchini *et al.* 2010, 2011a). Furthermore, remote sensing-based thematic land cover data has been used in SDMs that model species richness (Honnay *et al.* 2003, Thuiller *et al.* 2004). Finally, there are a few studies where different measures of geodiversity have been used (Heikkinen and Neuvonen 1997, Lobo *et al.* 2001, Pausas *et al.* 2003, Titeux *et al.* 2009, Hjort *et al.* 2012). Nevertheless, in SDMs relationships can be found between the explanatory features and the response feature, even if these relationships are not ecologically meaningful. Therefore, more effort should be put into finding the causal mechanisms (Araújo and Guisan 2006).

2 OBJECTIVES

The main objective of this study was to develop new methods for mapping habitat types and conservation values, as well as to critically evaluate those new methods. First, an object-based methodology for mapping habitat types was developed. Second, habitat type maps were converted into conservation value maps by taking species richness and rarity, naturalness, connectivity, complementarity, and selected ecosystem services into account. Third, the habitat type map was used to explain and predict vascular plant species richness. In the study, a critical approach was taken; in other words, methods, datasets, features, and approaches were compared. The study was divided into four consecutive phases, all of which had their own objectives and research questions.

1. Testing segmentation method, parameter value and data layer combinations in order to find a solution that works well in boreal forest habitat type mapping (I). In this framework, the following questions were assessed:
    a. Are supervised segmentation evaluation measures applicable in boreal forest habitat type mapping?
    b. Are the different evaluation measures sensitive to changes in reference polygons?
    c. Does the best segmentation lead to the best classification accuracy?
2. Developing a working habitat type classification (HTC) workflow that is applicable to boreal forest habitat type mapping (II). Two more specific research questions were formulated, as follows:
    a. What approaches, features and layers are needed to classify different habitat types?
    b. Is the examination of heterogeneity inside and between different habitat types useful in assessing problems in HTC?
3. Giving values to habitat types and patches from a conservation perspective and comparing different methods for mapping

conservation values (III). More specifically, the following questions were asked:

   a. To what extent do conservation value maps change if different habitat type mapping methods, valuation methods, or connectivity and complementarity mapping methods are used?
   b. To what extent do conservation value maps differ from maps of selected ecosystem services (timber production potential, carbon storage, landscape recreational value)?

4. Assessing the usefulness of the HTC method in vascular plant species richness mapping and explaining vascular plant species richness by means of landscape, topographic and geodiversity features (IV). Based on these general objectives, three more specific research questions were formulated:
   a. How does the developed HTC perform compared to two other landscape type classifications?
   b. Do explicit measures of geodiversity explain species richness in a fragmented landscape?
   c. Do results differ if only native species are taken into account?

3 MATERIALS AND METHODS

3.1 Study areas

This work included two study areas, both located in Southern Finland in the southern boreal vegetation zone (Fig. 2, Ahti *et al.* 1968). Both of the areas are predominantly rural and mostly covered by different forest types. Most of the forests are coniferous, but there are also some mixed and broadleaf forests. In both study areas, the main tree species are Scots pine (*Pinus sylvestris*), Norwegian spruce (*Picea abies*) and birches (*Betula pubescens* and *B. pendula*). Other main land cover types include peatlands, which are partly covered by forest, agricultural areas and lakes. Most of the forest areas are used for timber production with rotation-based forestry, including regeneration cutting with either artificial or natural regeneration. Most of the peatlands are drained for forestry. The relief of the areas varies between flat and moderately hilly.

The first study area (I, II) is southwest of the city of Jyväskylä. The geographic coordinates (WGS84) of the area are 62° 10´ 30´´ – 62° 13´ 30´´ N and 25° 29′–25° 38´ E. The size of the area is about 15 km^2. For the study, the area was divided into three separate blocks that all had different landscape characteristics. In the area, there are two nature conservation areas, which mostly consist of a semi-natural, over 100-year-old mesic forest. One of the protected zones is also part of a NATURA 2000 area.

The second study area (III, IV) is located southeast of the city of Tampere. The geographic coordinates (WGS84) of the area are 61° 16´ – 61° 30´ N and 24° 26´–24° 55´ E. The size of the area is about 390 km^2 and the central parts of the area are covered with one 38.9 km^2 lake. The flora in the second study area is a bit more diverse than in the first area. The area is located more to the south; hence, the annual effective temperature sum is larger. In the bedrock, there are some alkaline and limy patches in addition to granite. Additionally, there are many herb-rich forest patches, many of which are protected. In total, there are three NATURA 2000 areas and approximately 20 nature conservation areas.

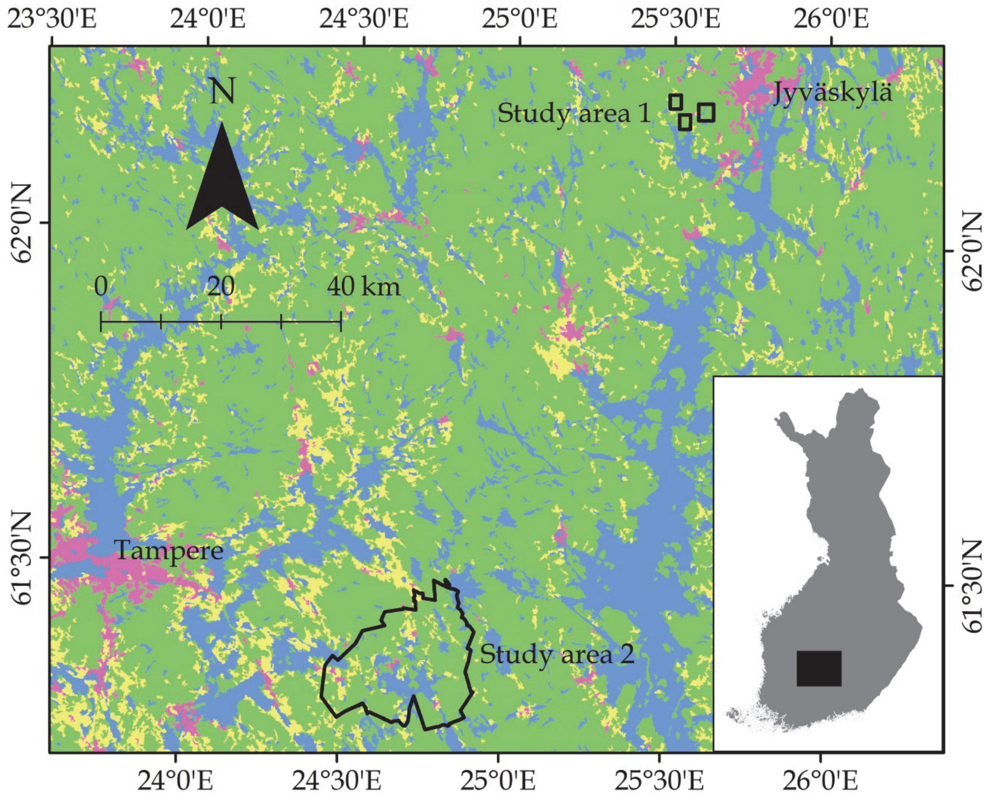

FIGURE 2 Location of the two study areas in Southern Finland. Study areas are outlined with a black border. Background maps: Corine Land Cover 2006 land use/land cover (generalized 25 ha) © SYKE, EEA; General map © NLS Finland 2010.

3.2 Datasets used

Numerous different remotely sensed and GIS datasets were used in the study (Table 1). The most important datasets were ALS data from the National Land Survey of Finland (NLS) and aerial and satellite imagery. Also included was 7 km^2 fieldwork data from the first study area, which was collected in 2010 and used as training and reference data in classifications (I and II) and as reference data in segmentation evaluation (I). None of the other datasets were collected for this study.

TABLE 1 Different datasets used in the study.

Data	Resolution / Scale	Articles	Year	Purpose	Producer / Source
WorldView-2 (WV-2) satellite imagery	2 m	I, II	2010	segmentation, classification	Digital Globe Inc.
ALS data	0.5 pts m^{-2}	I, II, III, IV	2008, 2010, 2012	segmentation, classification	The National Land Survey of Finland (NLS)
Aerial imagery	20 cm	I, II	2007	orthorectification of WV-2, assisting fieldwork	The City of Jyväskylä
Aerial imagery	40 cm	III, IV	2011	segmentation, classification, spectral heterogeneity (SH)	The Finnish Forest Centre Pirkanmaa / Terratec
Aerial imagery	50 cm	III, IV	2010–2012	segmentation, classification	NLS
Topographic database	1:10 000	II, III, IV	2010	classification, geodiversity	NLS
SLICES land use database	1:50 000	II, III, IV	2010	classification	NLS
Digital soil map	1:20 000	II, III, IV	1972–2007	classification, geodiversity	The Geological Survey of Finland
Digital bedrock map	1:200 000	III, IV	2009	classification, geodiversity	The Geological Survey of Finland
Multi-source national forest inventory	20 m	II, III	2009	classification, ecosystem services	The Finnish Forest Research Institute
Fieldwork data	polygons > 45 m^2	I, II	2011	segmentation reference, training and reference for classification	Antti Rusanen
Forestry planning data	polygons > 270 m^2	I	2005	segmentation reference	The City of Jyväskylä
Forestry planning data	polygons > 280 m^2	III, IV	1998–2011	training and reference for classification	The Finnish Forest Centre Pirkanmaa
Biotope classification data	polygons > 190 m^2	I	2006	segmentation reference	The Finnish Forest and Park Service
Vascular plant species inventory	1 km^2	III, IV	1983–2011	reference in species richness mapping, valuation of different habitat types	Tuomo Kuitunen
Habitat ranking data	not spatial	III	1993	valuation of different habitat types	Rossi and Kuitunen (1996)
Forest Act habitat polygons	polygons > 180 m^2	III, IV	1998–2011	classification	The Finnish Forest Centre Pirkanmaa
Corine Land Cover (CLC) 2006	25 m	IV	2010	land use / land cover classification	Finnish Environment Institute (SYKE)

3.3 Introduction to methodology

A general workflow of the approaches used is presented in Fig. 3. The approaches are divided into four phases. The first phase of the methodology was segmentation and segmentation evaluation (I). In this phase, the goal was to find a good and meaningful segmentation that could be used in HTC. Segmentation and segmentation evaluation form a loop, since segmentation evaluation results help in the parameter selection of new segmentations. In the second phase, classification workflow was developed and tested, and several features derived from used datasets were evaluated and compared (II). As part of this phase, the classification approach was modified to match the datasets available from study area 2 (III, IV). In the third phase, HTC was used in conservation value mapping (III). Conservation value mapping included different habitat type valuations that used potential species richness and rarity, naturalness, and habitat type connectivity and complementarity mapping. Conservation value maps were compared with ecosystem service maps. In the fourth phase, to explain species richness (IV), the HTC system was used together with two other landscape classification schemes (Corine Land Cover (CLC) and spectral heterogeneity (SH)), geodiversity and topographic features. SH was also based on segmentation.

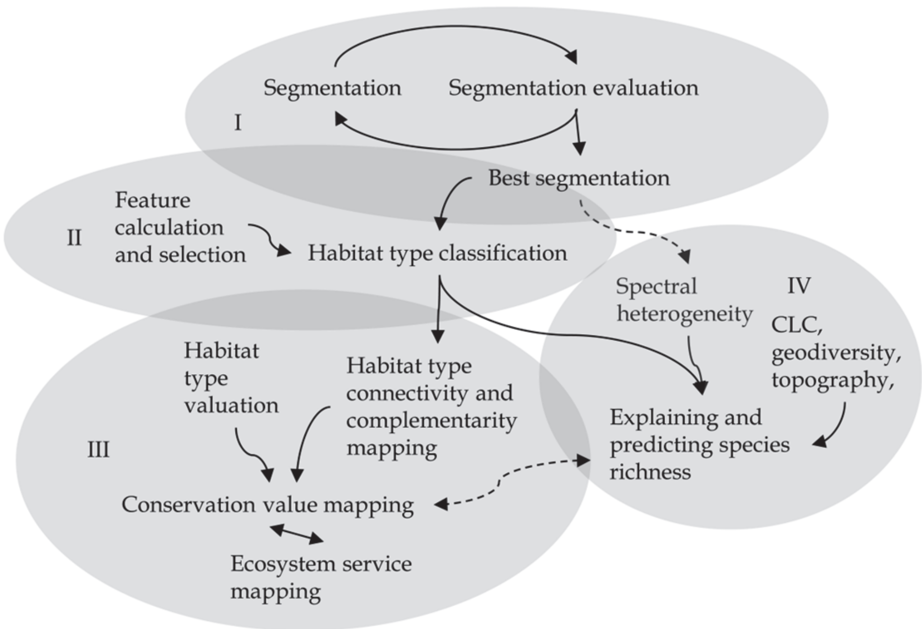

FIGURE 3 A workflow of the different approaches used in the study. CLC refers to Corine Land Cover.

3.4 Habitat type classification (HTC) system

The HTC system was based on habitat ranking (Rossi and Kuitunen 1996), in which the classification system was as detailed as possible, based on species' habitat type preferences given in species identification literature (e.g. Hämet-Ahti *et al.* 1986) and the Finnish Red List (Rassi *et al.* 1992, 2010). Therefore, the habitat types in the framework do not correlate to vegetation types or plant communities (as seen in Braun-Blanquet 1932, for example). Instead, habitat types are rather homogenous land units in which different species are likely found. Habitat types are partly based on vegetation and partly on abiotic factors. They are defined with the help of information given in species identification literature (e.g. Hämet-Ahti *et al.* 1986). In forests and peatlands, habitat types follow widely used Finnish classification systems (Cajander 1949, Eurola *et al.* 1995).

In the first study area (I, II), HTC was modified to match with remotely sensed imagery. It included habitat types that were mapped during fieldwork in June and July 2011. In the second study area (III, IV), habitat types consisted of the types included in the habitat ranking and that existed in the study area. The different habitat types are listed in Table 2.

3.5 Layers and features used

From the datasets, several features and layers were calculated. A set of features and layers was used for data segmentation and classification in study area 1 (Table 3, I, II). In study area 2, a slightly different set of features was used in segmentation and classification (Table 3, III, IV). For segmentation, derived raster layers were used as such; for classification, a number of features per each segment were calculated from the layers. Finally, to explain species richness, again different features were used.

Most of the features used in classification were derived from imagery and ALS data. The imagery layers were different spectral bands, together with a NDVI (II). The ALS layers were constructed from the DTM, a digital surface model (DSM) and an intensity layer. The canopy height model (CHM) was calculated by subtracting the DTM from the DSM. From the CHM and imagery layers, GLCM features (Haralick *et al.* 1973, Haralick 1979; II, III, IV) and Daubechies wavelet features (Daubechies 1992; II) were calculated. From the DTM, several different features were calculated using different parameter values. A SAGA wetness index (SWI) is a modification of the TWI. In the SWI, large TWI values are spread into their neighborhood if the area is almost flat. A distance to water (DTW) layer is the cost distance from each cell to the nearest stream, using slope as the cost surface (Murphy *et al.* 2007). A terrain ruggedness index (TRI) quantifies the amount of local altitudinal differences (Riley *et al.* 1999). A topographic position index (TPI) is a measure of the relative altitudinal position

TABLE 2 Different habitat type classes that were mapped in the study. I–IV refer to original publications of the thesis.

Habitat type	I	II	III	IV
xeric (pine dominated) forests	A	A	A	X
mesic (spruce dominated) forests	B	B	A	X
herb-rich (mixed/deciduous) forests	C	C	A	X
esker forests	-	-	A	X
rocky areas	X	X	-	-
non-calcareous rocky areas	-	-	X	X
calcareous rocks and quarries	-	-	X	X
pine mires	X	X	D	X
spruce mires	D	D	D	X
open mires	X	X	D	X
rich fen	-	-	X	X
water (lakes and streams)	X	X	-	-
oligotrophic lakes	-	-	X	X
eutrophic lakes	-	-	X	X
streams and rivers	-	-	X	X
small creeks	X	X	-	-
riparian habitats	-	-	X	X
flooded areas	-	-	X	X
beaches	-	-	X	X
springs	X	X	X	X
meadows	X	X	-	-
dry meadows	-	-	X	X
wet meadows	-	-	X	X
cultivated areas	X	X	X	X
parks and gardens	-	-	X	X
yards	X	X	-	-
industrial and urban areas	-	-	X	X
roads	X	X	-	-
sand pits	-	X	-	-

X: used in the publication
-: not used in the publication
A: 4 successional stages: open regeneration area (0), sapling stand (1), young (2), mature (3)
B: 5 successional stages: 0, 1, 2, 3, natural (4)
C: 4 successional stages: 1, 2, 3, 4
D: 2 drainage statuses: not-drained, drained (d)

TABLE 3 Different layers and features calculated for segmentation and habitat type classification. For segmentation, different layers were used. For classification, features calculated from layers were used. Features were calculated per segment or updated to the habitat type map as such. The MS-NFI data was only used for an alternative classification in II and III.

Data	Layer	Resolution	I and II segmentation	I classification	II classification	III and IV segmentation	III and IV classification
WV-2	Bands 1–8	2 m	X	mean	mean, sd, GLCM, wavelets	-	-
	Bands 1–8	10 m	-	-	mean, sd	-	-
	NDVI	2 m	-	-	mean, sd, GLCM, wavelets	-	-
	NDVI	10 m	-	-	mean, sd	-	-
Aerial imagery / Terratec	Bands 1–3	10 m	-	-	-	X	mean, sd, GLCM
Aerial imagery /NLS	Bands 1–4	10 m	-	-	-	X	mean, sd, GLCM
ALS	CHM	2 m	X	mean	mean, sd, range, GLCM, wavelets	-	-
	CHM	10 m	-	-	mean, sd, range	X	mean, sd, GLCM
	intensity	2 m	-	-	mean, sd	-	-
	intensity	10 m	-	-	mean, sd	-	-
	slope	2 m	-	-	mean, sd	-	-
	slope	10 m	-	-	mean, sd	-	-
	SWI	2 m	X	mean	mean, sd	-	-
	SWI	5 m	-	-	mean, sd	-	-
	SWI	10 m	-	-	mean, sd	X	mean, sd
	DTW	2 m	-	-	mean, sd	-	-

	DTW	10 m	-	-	mean, sd	-	mean, sd
	TRI[1]	2 m	-	-	mean, sd	-	-
	TRI[2]	10 m	-	-	-	-	mean, sd
	TPI[3]	2 m	-	-	mean, sd	-	-
	TPI[4]	10 m	-	-	-	-	mean, sd
	MRVBF[5]	2 m	-	-	mean, sd	-	-
	MRVBF[6]	10 m	-	-	-	-	mean, sd
Soil map	soil type	1:20 000	-	-	majority	-	as such / majority
Bedrock map	bedrock type	1:200 000	-	-	-	-	as such
Topographic database	land use type	1:10 000	-	-	as such / majority	-	as such
SLICES	land use type	1:50 000	-	-	as such / majority	-	as such / majority
Forest Act habitats	habitat type	NA	-	-	-	-	as such
Forestry planning	habitat type	-	-	-	-	-	as such
MS-NFI	habitat type	20 m	-	-	as such / majority	-	majority

[1] Calculated with window sizes of 3×3, 7×7, 11×11

[2] Calculated with a window size of 3×3

[3] Calculated with radiuses of 10, 25, 50, 100

[4] Calculated with a radius of 100

[5] Calculated with threshold values of 16, 75

[6] Calculated with a threshold value of 28

(Guisan *et al.* 1999). A multiresolution index for valley bottom flatness (MRVBF) indicates if areas are relatively low or flat (Gallant and Dowling 2003). Features from layers that were not derived from the imagery or ALS data were not employed in the classification *per se*, but used to make post-classification adjustments. These data are called ancillary data. Finally, multisource National Forest Inventory (MS-NFI) data was used only in alternative HTCs.

In explaining species richness, several features were calculated from 1 km^2 quadrats. Features were divided into three groups: landscape features, geodiversity features and topographic features. Landscape features included HTC, CLC, and SH. In the SH calculation, the principal component 1 was derived from aerial imagery. The principal component image was then generalized per segment and quantized to 64 classes. The number of patches and variety of types of each landscape classification were calculated. Geodiversity features included soil and bedrock type diversity and hydrological feature diversity. Topographic features included altitude, slope, SWI and solar radiation. From all four of these topographic features, mean value, standard deviation and range were calculated.

3.6 Segmentation and classification

3.6.1 Segmentation methods

Two different segmentation methods were used. Watershed segmentation was used in I and region based segmentation in I, II, III, and IV.

The watershed segmentation used was the segmentation method of IDRISI Taiga software. In this method, watersheds are delineated by means of a weighted average of variance images derived from input layers using a moving window analysis. Both the size of the moving window and the weights of averaging can be adjusted by the user. After watershed delineation, watersheds are merged if they are the most similar to each other in the neighborhood and if their difference is smaller than a user-adjusted similarity tolerance. Difference is determined by the mean value and the standard deviation, whose weights are set by the user.

Region-based segmentation used was Fractal Net Evolution Approach (FNEA) (Baatz and Schäpe 2000, Benz *et al.* 2004). At the outset, each pixel is treated as a region and pixels are merged into larger regions. Three user parameters can be adjusted. Scale parameter controls average object size. The weighting of color and shape can be adjusted. If more weight is given to color homogeneity, less weight is given to a specific shape. Shape is composed of smoothness and compactness, whose relative weights can be adjusted.

3.6.2 Segmentation evaluation (I)

After initial evaluations in which a set of segmentation alternatives was selected, two different segmentation methods with different parameter options and layer combinations were evaluated by means of visual interpretation and 23 different supervised segmentation evaluation measures. Four layer combinations were used: only imagery, only ALS, and two combinations of imagery and ALS. The following user parameters were modified: similarity tolerance, mean and variance weights in IDRISI, scale parameter, and color/shape parameter in FNEA. Overall, 33 IDRISI and 30 FNEA parameter combinations were used with all of the different layer combinations. Evaluation measures included both area-based measures and location-based measures. Some of the measures were combined measures (i.e. they were combinations of other measures). The details of the evaluation measures used are given in Clinton *et al.* (2010), and they are also summarized in publication I. The evaluation was performed with eight different reference polygon sets, three of which were different datasets (our own fieldwork data, forestry planning data, and biotope inventory data) and five of which were subsets of the fieldwork data.

3.6.3 Classification of habitat types (I, II, III, IV)

The classification of segments was performed with a random forest classifier (Breiman 2001, Breiman and Cutler 2007; I, II, III, IV). A random forest is a forest (or an ensemble) of several bootstrapped classification trees. When a classification tree is built, approximately two-thirds of the data is used for training the classifier. At each node of the classification tree, the best split is chosen from a randomly selected subset of features. For each case, in our case segment, a majority vote of individual trees is obtained.

In I, habitat types were classified only to test different segmentation evaluation methods (i.e. if the best segmentation gives the highest classification accuracy). Hence, only a minimum set of features was used in the classification.

In II, the goal was to find an optimal classification by comparing different alternatives. Additionally, different sets of features were tested in the analysis to determine how omitting or including different features might affect classification accuracy. These features included WV-2 features, ALS features, GLCM and wavelet features, and additional topographic features (DTW, TPI, TRI, MRVBF). This analysis was complemented with random forest feature importance metrics using three different measures: permutation importance, gini importance (Breiman 2001, Breiman and Cutler 2007) and Boruta feature selection (Kursa and Rudnicki 2010). Furthermore, it was tested if a data split based on a one meter vertical distance to the nearest stream helps in mapping streamside habitats and mires, as well as in balancing the data. Classification accuracies were quantified using a confusion matrix, user's accuracy (error of commission) and producer's accuracy (error of omission), as well as allocation and quantity disagreement (Pontius and Millones 2011). Finally, the heterogeneity inside and between different habitat types was examined using Sammon's mapping (Sammon 1969) and a multi-dimensional scaling (Torgerson 1958) plot based on a random forest proximity measure, which quantified the proximities between segments.

In III and IV, HTC was performed in a different context, using a different set of features and habitat types. In these publications, the main focus was not on classification, but on using the classification for conservation value mapping (III) or to explain vascular plant species richness (IV).

In II, III, and IV, ancillary data was used after the random forest classification. In II, one of the aims was to test if ancillary data increases classification accuracy. Different habitat types were either derived directly from the ancillary data or a majority value per segment was calculated. In II and III, an alternative classification was performed in which all classes were derived from pre-existing datasets. Forest and peatland habitat types were obtained from MS-NFI data, while other habitat types were obtained from the same data as in the primary classification.

3.7 Mapping conservation value (III)

Conservation values were mapped using HTC as a basis, and different methods for conservation value assessment and mapping were compared. Each habitat type was valued by means of nine different valuation methods based on the habitat type preferences of species (Rossi and Kuitunen 1996). The first valuation method was the potential number of species in the habitat type. The second valuation method was range-size rarity (Williams *et al.* 1996) corrected number of species. In the third valuation method, potential species were weighted based on their Red List status (Rassi *et al.* 2010, Ryttäri *et al.* 2012). In the second and the third methods, primary habitat type preferences were given more value. The fourth, fifth and sixth valuation methods were modifications of the first three methods, in which the value of non-natural or degraded habitat types was less. In the last three methods (valuation methods 7–9), human habitats were not given value.

The complementarity and connectivity of different habitat types were mapped using two different methods: landscape metrics and spatial conservation prioritization. In calculations using landscape metrics calculations, two units of measurement were used: patch area and patch neighborhood similarity. The highest values were given to patches with a large area, similar neighborhood and high habitat type value. In some of the calculations, habitat type complementarity was taken into account, giving high values to each different habitat type. Overall, 30 different landscape metrics-based maps were created. In spatial conservation prioritization using Zonation software in core-area Zonation mode (Moilanen *et al.* 2012), cells were weighted by habitat type values. Successional stages of forests and different drainage statuses of mires were calculated using a condition layer. Connectivity was calculated using matrix connectivity. In half of the calculations, a habitat type connectivity matrix was also used as a habitat type similarity matrix for defining effective occurrences of habitat types. In all, 36 different spatial conservation prioritization maps were calculated.

To complement conservation value maps based on habitat types, three ecosystem services were mapped. The monetary value of timber was mapped using MS-NFI data and the current prices for lumber and pulpwood. Carbon storage was calculated using MS-NFI biomass calculations for the carbon in trees and average soil carbon estimates for different forest and peatland habitat types, based on Liski and Westman (1997) and Turunen *et al.* (2002). Forest and peatland habitat types were mapped using the MS-NFI. The landscape value for recreation was calculated using a viewshed analysis from the most important recreation routes and resting places (measured with an ALS-based DSM). The areas were valued based on habitat type scenic beauty, with the highest values given to natural areas and in relation to how many times a specific point can be seen.

3.8 Species richness modeling (IV)

Statistical modeling analyses were performed separately for total species richness and for native species richness. Before these statistical analyses, multicollinear features were removed from the models using variance inflation factors calculated following the approach used by Zuur *et al.* (2009).

Two different partitioning methods based on generalized linear models (GLMs, Nelder and Wedderburn 1972) were used. In variation partitioning (Heikkinen *et al.* 2004), the variation of the response variable (i.e. species richness) was divided into seven components: the discrete effects of landscape, topography, and geodiversity, as well as combinations of two or three discrete components. GLMs were processed with a lasso penalty (Tibshirani 1996), and quadratic terms of explanatory features were added into the models. In hierarchical partitioning (Walsh and Mac Nally 2013), the importance of each explanatory feature was measured. The mean Z-Scores of 100 randomizations were calculated and the statistical significance of each variable was tested.

In addition to GLMs, a semi-parametric extension of GLMs (i.e. generalized additive models, GAMs, Hastie and Tibshirani 1986) were used to explain and predict species richness. Variables were selected for the models by implementing an extra penalty term, as suggested by Marra and Wood (2011). Eight different GAMs were compared in order to test the importance of geodiversity features and different landscape features. GAMs were calculated using a six-fold cross-validation. Due to overdispersion of data, a quasipoisson distribution with a log-link function was used in GLMs and GAMs, following the approach given in Zuur *et al.* (2009).

3.9 Used software

Several different software products were used in the analyses. Different software products were used for different remote sensing, GIS, and statistical analyses. The programs used are listed in Table 4. Programs that were only used in initial analyses are not included in the table. Details of program usage are given in I–IV.

TABLE 4 Different software products used in the study.

Software	Publication	Purpose	Source
ArcGIS	I, II, III, IV	GIS analyses	Esri, Redlands, CA, USA
eCognition	III, IV	segmentation & GLCM feature calculation	Trimble, Sunnyvale, CA, USA
ERDAS Imagine	I, II, III, IV	image orthorectification & feature calculation	Intergraph, Huntsville, AL, USA
FRAGSTATS	III	landscape metrics	McGarigal and Ene (2012)
IDRISI Taiga	I, II	segmentation & GIS analysis	Clark Labs, Worcester, MA, USA
Intersector	I	segmentation evaluation	Clinton *et al.* (2010)
LAStools	I, II, III, IV	ALS data preprocessing	rapidlasso, Gilching, Germany
R	I, II, III, IV	feature calculation, statistical analysis	R Core Team (2012)
including e.g. the following packages			
randomForest	I, II, III, IV	random forest classification	Liaw and Wiener (2002)
EBImage	II	GLCM feature calculation	Pau *et al.* (2010)
waveslim	II	wavelet feature calculation	Whitcher (2012)
Boruta	II	random forest feature selection	Kursa and Rudnicki (2010)
vegan	III	ordination analysis	Oksanen *et al.* (2013)
lqa	IV	GLM shrinkage using lasso	Ulbricht (2012)
hier.part	IV	hierarchical partitioning	Walsh and Mac Nally (2013)
mgcv	IV	GAM modeling	Wood (2006, 2011)
SAGA-GIS	I, II, III, IV	topographic feature calculation	SAGA (2011)
TauDEM	II, III, IV	stream network mapping	Tarboton (2012)
TerraLib	I, II	segmentation	Câmara *et al.* (2008)
Zonation	III	spatial conservation prioritization	Moilanen *et al.* (2012)

4 RESULTS

4.1 Overview of the results

An illustration of the results is given in Fig. 4. First, different segmentations are evaluated and, second, the best one possible is selected for HTC. Third, the classification is tuned by adding different features and ancillary data. Fourth, the classification is applied to a second study area. Fifth, with the help of HTC, different conservation value maps are drawn and compared. Sixth, HTC is used as one of the explanatory features to explain and predict vascular plant species richness patterns.

4.2 Segmentation evaluation (I)

Supervised segmentation evaluation measures gave inconsistent results. In other words, different measures ranked different segmentations as the best. Additionally, results differed if the reference polygon set was changed. Some measures were consistent (i.e. the same segmentations were ranked as being among the best, irrespective of the reference polygon set that was used), but other measures varied widely in their results. Overall, FNEA segmentations were generally valued more highly than IDRISI segmentations, and segmentations with only ALS-based layers were most often ranked as the best. It was also found that undersegmentation measures preferred the lowest scale parameter values and oversegmentation measures the highest scale parameter values. Intermediate-scale parameter values were given the lowest ranks in only some of the combined measures.

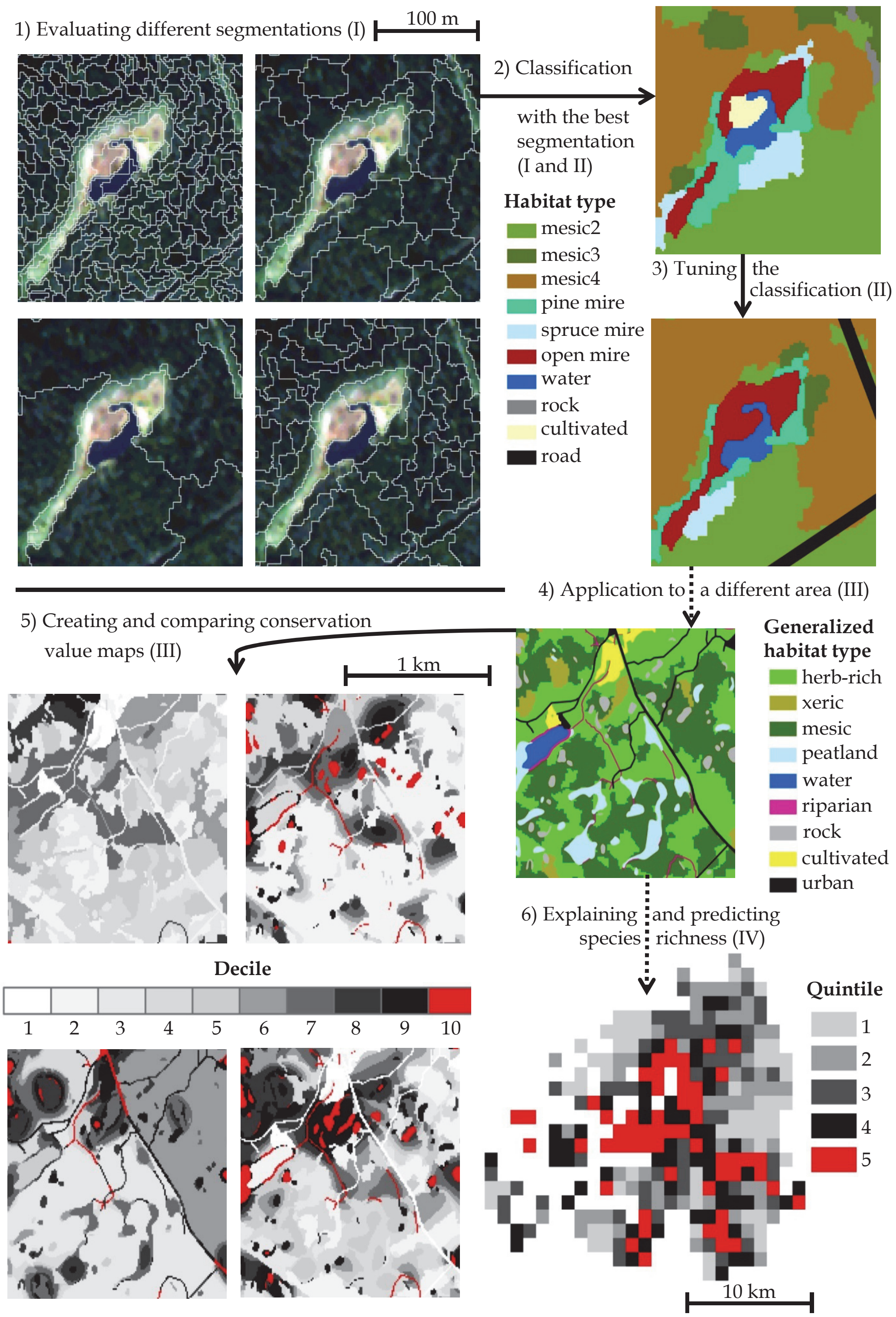
1) Evaluating different segmentations (I)
100 m
2) Classification
with the best segmentation (I and II)
Habitat type
mesic2
mesic3
mesic4
pine mire
spruce mire
open mire
water
rock
cultivated
road
3) Tuning the classification (II)
4) Application to a different area (III)
5) Creating and comparing conservation value maps (III)
1 km
Generalized habitat type
herb-rich
xeric
mesic
peatland
water
riparian
rock
cultivated
urban
Decile
1
2
3
4
5
6
7
8
9
10
6) Explaining and predicting species richness (IV)
Quintile
1
2
3
4
5
10 km

FIGURE 4 The figure is on the previous page. A simplified illustration of the results. 1) Four segmentations with a WV-2 image (© Digital Globe) in the background: top left: too fine (FNEA with all layers, scale parameter 5, color parameter 0.75), top right: the best (FNEA with WV-2 bands blue, green, red, and near infra-red 1, scale 10, color 0.5), bottom left: too coarse (FNEA with all layers, scale 50, color 0.5), bottom right: too complex (IDRISI with WV-2 bands blue, green, red, and near infra-red 1, similarity 40, mean 0.5, variance 0.5); 2) habitat type classification with the best segmentation and with the features used in I; 3) classification with the 100 highest scoring features, according to Boruta and adjusted with ancillary data; 4) classification in the second study area; 5) four conservation value maps divided into deciles with the highest conservation value in the 10th decile: top left: FRAGSTATS with complementarity and habitat valuation 9, top right: Zonation with similarity and habitat valuation 9, bottom left: Zonation without similarity, habitat valuation 1, bottom-right: Zonation without similarity, habitat valuation 9; 6) species richness in the studied quadrats predicted by GAM with all features and a six-fold cross-validation.

Visual interpretation was time-consuming and it was not thoroughly reliable for determining the best segmentation or the best parameter values. Nevertheless, it was found that both ALS and WV-2 layers were needed when delineating mire areas and water bodies, for example. FNEA segmentations with a low or intermediate weight for color were the most appealing segmentations, since segments were not too complex. Additionally, a rather low scale parameter value was needed to delineate the smallest habitat patches. Hence, the best segmentation was arguably chosen to be FNEA segmentation with the second lowest scale parameter (10), intermediate weight for color (0.5), and a combination of four spectral layers and two ALS layers.

4.3 Habitat type classification (I, II, III, IV)

In total, 12 classifications were performed for objects delineated with 12 selected segmentations (I, Table 5). The highest classification accuracy was achieved with the segmentation that was considered the best in visual interpretation. The differences in classification accuracies were, nevertheless, quite small. Finer segmentations led to higher classification accuracies overall, but the finest segmentations led to a salt-and-pepper effect and slightly worse classification accuracy. Additionally, the best classification accuracies were achieved by merging ALS and WV-2.

When the number of features was increased from 12 (used in segmentation evaluation and I) to 328 (used in II), classification accuracy increased from 72 % to 78 % (Tables 5, 6). The classification accuracy further increased to 79 % when the 100 highest scoring features according to Boruta were used and ancillary datasets were included. A data split did not increase classification accuracy, but ancillary data increased it a little. When only ALS or WV-2 features were used, classification accuracy was significantly lower. Additionally, omitting extra topographic features or both GLCM and wavelet features decreased classification accuracy. However, omitting either GLCM or wavelet features did not decrease classification accuracy. Finally, the accuracies of benchmark classifications based on MS-NFI and ancillary data were significantly lower than the accuracies based on random forest classification and WV-2 and ALS data.

TABLE 5 Classification accuracies of classifications based on 12 different segmentations. Segmentations are named as follows: text refers to segmentation method, a–d to layer combinations (a: only WV-2; b: only ALS; c: WV-2 blue, green, red, and near infra-red 1 bands, and ALS; d: all layers), s to similarity or scale parameter, c to color parameter, m to weight given to mean and v to weight given to variance.

Segmentation	Classification accuracy	Why segmentation was selected to the classification
FNEA_b_s5_c.75	0.69	The best in avoiding undersegmentation based on supervised evaluation measures, the worst based on the Combined measure (combination of all other measures) and full fieldwork reference polygon set
FNEA_c_s5_c.75	0.69	Comparison against FNEA_b_s5_c.75
FNEA_a_s10_c.5	0.71	WV-2 layers only, comparison against FNEA_c_s10_c.5
FNEA_c_s10_c.5	0.72	The best segmentation based on visual interpretation
FNEA_d_s15_c.25	0.71	Good in visual interpretation
FNEA_a_s20_c.75	0.69	Segmentation based on WV-2 data only, OK visually
FNEA_b_s25_c.5	0.66	The best segmentation based on the Combined measure and full reference polygon set, OK visually
FNEA_d_s35_c.5	0.66	The best segmentation based on measure D (a combined measure) and full reference polygon set
FNEA_c_s50_c.75	0.65	Good in avoiding oversegmentation based on supervised measures, good in visual boundary evaluation
IDRISI_d_s30_m9v1	0.70	OK in visual interpretation, small segments
IDRISI_c_s40_m5v5	0.69	OK in visual interpretation, quite small segments
IDRISI_b_s70_m9v1	0.60	The best based on OverUnder (a combined measure) and full reference polygon set, poor in visual interpretation

When the user's and producer's accuracies of single classes were examined (Tables 6, 7), it was found that a data split reduced omission errors in mapping mires and streamside habitats. Ancillary data reduced the omission errors of yards, roads, spruce mires, and rocky areas, for example. On the other hand, a data split increased the commission error of mapping spruce and pine mires, while ancillary data increased the commission error of meadows, rocky areas and herb-rich forests, for example. While some of the classes could not be classified at all and some had very low classification accuracies, classification accuracies of some of the classes were high across the board.

TABLE 6 Producer's accuracy and total classification accuracy (in %) of different habitat type classes (rows) with different classification methods (columns). Accuracies were calculated at a pixel level.

	all	only WV-2	only ALS	no extra topography	no GLCM	no wavelet	no GLCM & wavelet	boruta 100	boruta 100 & ancillary	split	split & Boruta	split & Boruta 100	split & Bor100 & ancillary	msNFI & ancillary	segmentation & msNFI
xeric0	78	76	61	78	72	78	71	77	77	66	69	67	67	0	0
xeric1	61	53	67	65	55	60	55	59	59	64	64	64	64	0	0
xeric2	77	72	65	76	76	76	76	79	78	73	73	78	77	16	14
xeric3	0	0	0	0	0	0	0	0	0	0	0	0	0	10	6
mesic0	58	48	0	58	57	55	53	58	56	54	54	56	55	0	0
mesic1	83	79	78	83	82	82	80	83	79	82	82	81	79	54	58
mesic2	90	88	90	90	91	91	91	90	88	89	90	89	86	51	59
mesic3	73	55	65	71	68	71	62	73	70	71	71	71	68	23	17
mesic4	89	84	88	89	91	90	87	89	89	88	88	89	88	52	59
herb-rich1	5	2	4	0	0	14	22	10	9	2	2	7	7	0	0
herb-rich2	12	7	5	7	14	17	13	13	30	13	15	17	21	2	1
herb-rich3	5	2	0	10	2	9	0	9	9	3	3	1	1	0	0
herb-rich4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
rocky area	46	48	31	41	48	48	49	47	57	45	45	41	55	50	51
pine mire	48	42	44	46	44	50	45	51	52	65	67	66	66	54	56
pine mire d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
spruce mire	7	0	15	2	6	12	12	11	19	18	17	23	29	14	13
spruce mire d	0	0	0	0	0	0	0	0	27	0	0	0	40	39	39
open mire	47	10	12	22	71	64	64	75	73	39	61	73	69	61	61
open mire d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
water	97	97	97	97	98	98	98	98	98	97	97	97	98	99	99
small creeks	4	0	1	0	4	6	6	6	6	27	28	26	25	0	0
spring	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
meadow	20	14	11	17	18	28	23	23	35	31	31	35	32	34	34
cultivated	94	92	88	94	94	93	90	95	96	93	93	94	96	97	96
road	51	48	41	50	52	48	48	51	94	50	51	49	94	91	91
yard	54	53	31	53	57	46	41	54	69	57	57	55	69	66	64
sand pit	97	94	91	96	97	97	95	98	98	97	97	97	99	75	75
total	78.0	73.1	73.8	77.1	78.0	78.1	76.0	78.6	79.1	77.9	78.0	78.2	78.7	51.8	54.6

TABLE 7 User's accuracy of different habitat type classes (rows) with different classification methods (columns). Accuracies were calculated at a pixel level.

CLASS	all	only WV-2	only ALS	no extra topography	no GLCM	no wavelet	no GLCM & wavelet	boruta 100	boruta 100 & ancillary	split	split & Boruta	split & Boruta 100	split & Bor100 & ancillary	msNFI & ancillary	segmentation & msNFI
xeric0	91	88	92	91	92	90	92	91	91	93	93	92	92	100	100
xeric1	90	96	93	88	94	89	96	92	92	92	92	89	89	0	0
xeric2	84	82	85	82	85	90	86	83	84	86	86	85	85	22	30
xeric3	100	100	100	100	100	100	100	100	100	100	100	100	100	1	0
mesic0	91	86	100	89	89	88	90	89	91	88	86	87	91	0	0
mesic1	81	75	67	80	81	80	79	81	84	81	81	80	83	61	68
mesic2	70	64	68	69	70	71	69	71	75	72	72	72	76	49	50
mesic3	64	59	60	62	64	65	60	65	67	66	66	66	69	17	14
mesic4	88	80	85	87	87	88	84	88	90	88	88	88	89	55	57
herb-rich1	100	100	74	100	100	98	84	71	29	100	67	100	36	0	0
herb-rich2	66	69	59	50	84	66	60	66	72	74	74	68	54	6	4
herb-rich3	62	93	100	79	100	58	100	74	76	85	85	66	5	100	100
herb-rich4	100	100	100	100	100	100	100	100	100	100	100	100	0	100	100
rocky area	91	83	87	87	87	86	83	89	76	89	89	91	78	68	46
pine mire	75	66	69	82	69	72	62	72	71	58	57	63	64	57	56
spruce mire	55	0	42	46	58	57	54	50	43	42	44	39	38	22	35
spruce mire d	100	100	100	100	100	100	100	100	19	100	100	100	25	42	42
open mire	55	71	38	58	60	62	57	61	57	62	66	66	61	34	35
water	99	98	95	99	99	99	98	99	99	99	99	99	99	97	98
small creeks	50	100	28	100	42	50	37	52	53	34	36	44	45	100	100
spring	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
meadow	67	73	57	72	74	75	70	75	46	88	80	81	45	45	45
cultivated	87	84	84	85	88	87	87	87	89	88	89	90	89	86	94
road	70	68	62	69	69	66	63	67	55	67	66	66	55	46	48
yard	55	50	50	54	58	49	50	60	59	56	56	56	58	61	60
sand pit	97	94	95	98	97	98	95	100	100	99	99	99	100	100	100

Overall, the allocation disagreement (13.3 %) was higher than the quantity disagreement (7.6 %) in the classification with the highest accuracy. In other words, most of the errors were caused by allocating habitat types to incorrect locations. Most of the total disagreement was caused by different mesic forest classes, partly because these were among the most common classes.

Different feature importance measures considered different features as important in the classification. Moreover, different features were regarded as being important in mapping different habitat types. Overall, mean features had higher importance values than standard deviation features, and 2 m features higher importance than 10 m features. In terms of the different WV-2 bands, band 1 (coastal blue) got the highest importance scores. Regarding the different ALS features, CHM, DTW and MRVBF got high importance values. The different GLCM and wavelet features were ranked as both important and not important.

When the heterogeneity within and between habitat type classes was examined by means of a Sammon's map (II, Fig. 5), it could be seen that different mesic forest classes had a huge heterogeneity and that other classes were often included in this heterogeneity. Additionally, only some of the classes (mesic2, mesic4, water) were easily distinguishable in the random forest multi-dimensional scaling plot (II, Fig. 6). In both of these plots, however, multi-dimensional data was reduced to two dimensions.

The accuracy of HTC in study area 2 (III, IV) was 47 % for the habitat type classes that were classified with the random forest classifier. These classes included mesic, xeric, and herb-rich forest classes. Other habitat types were derived straight from ancillary, or reference, data. An alternative classification based on MS-NFI, used in III, gave an accuracy of 23 % for the same habitat types. Two classification alternative maps were 50 % similar overall and 22 % similar in the mesic, xeric, and herb-rich forest areas.

4.4 Habitat type valuation and conservation value mapping (III)

The values given to different habitat types were approximately similar in relative terms in different habitat type valuation methods (i.e. between methods 1–3 or 4–6). Yet, if naturalness correction (valuation methods 4–6) was taken into account, values of some habitat types changed drastically. More specifically, when naturalness correction was not used, the most valuable habitat types were dry meadows, industrial and urban areas, and cultivated areas. When naturalness correction was taken into account, the most valuable habitat types were riparian habitats, non-calcareous rocky areas and mature herb-rich forests.

In the conservation value maps, all habitat types were seen as valuable. However, if complementarity was not taken into account in the landscape metrics calculations, large mature herb-rich forest patches had the highest values. If naturalness correction was not used, large human habitat patches also got high values. In landscape metric maps, a patch received one value, whereas in spatial conservation prioritization approaches, high-value areas were found inside a patch or divided between neighboring patches.

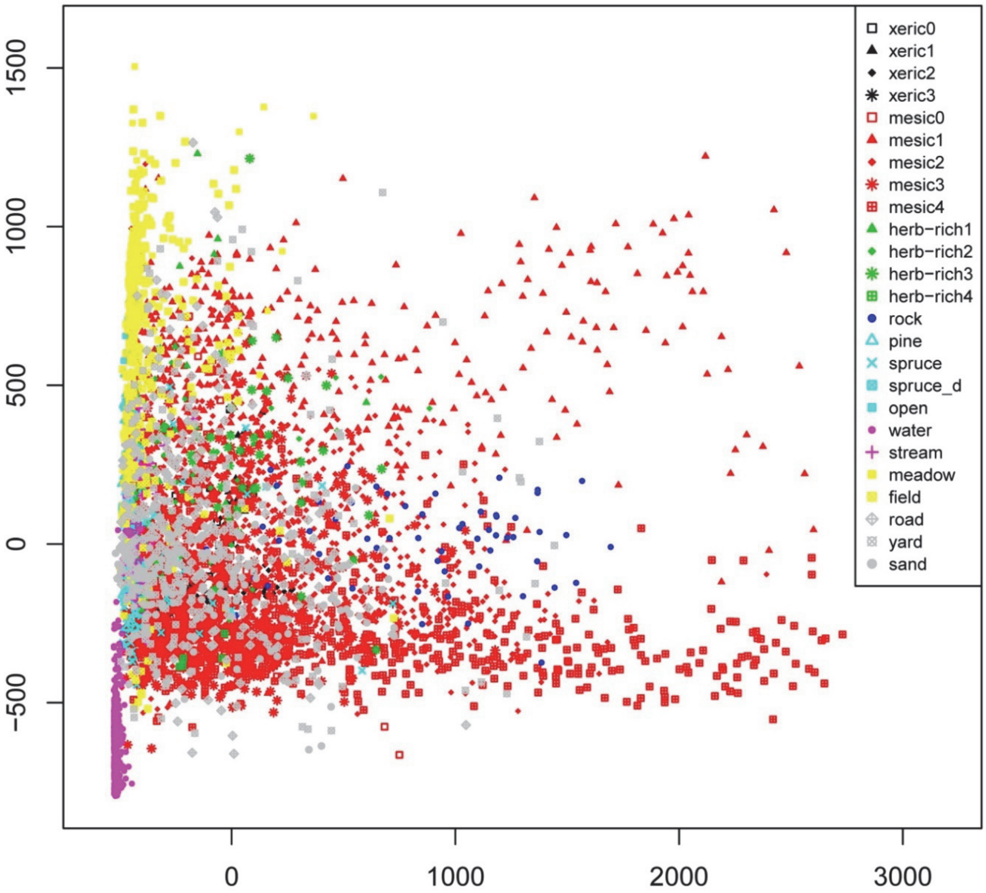

FIGURE 5 A Sammon's map of all segments in the training data. The whole study area and the 100 most important features (selected with Boruta) were used. The x and y axes represent the first two dimensions of the Sammon's projection.

Conservation value maps differed from each other, but the degree of difference varied (Table 8). In general, the differences were small if only habitat type valuation was changed (i.e. inside valuations 1–3 or 4–6). Differences were a little larger if naturalness correction was used or not used (i.e. between valuations 1–3 and 4–6), or if the HTC method was changed. And differences were even larger if different connectivity and complementarity calculations were used in the calculation method (landscape metric vs. spatial conservation prioritization). The differences were the largest when the calculation method for connectivity and complementarity was changed.

The maps of three ecosystem services were very different from the conservation value maps (Table 8). The conservation value maps that were the most similar with the ecosystem service maps either had naturalness correction or were calculated without considerations of complementarity. In terms of the ecosystem service maps themselves, the maps of carbon and timber were rather similar, whereas the maps of recreational value were different from the other two ecosystem service maps (Table 8). Open mires, other mires and older forests had the highest values in the carbon map, whereas the potential for timber production was greatest in older forests. Recreational value was the highest in natural habitats near recreational routes.

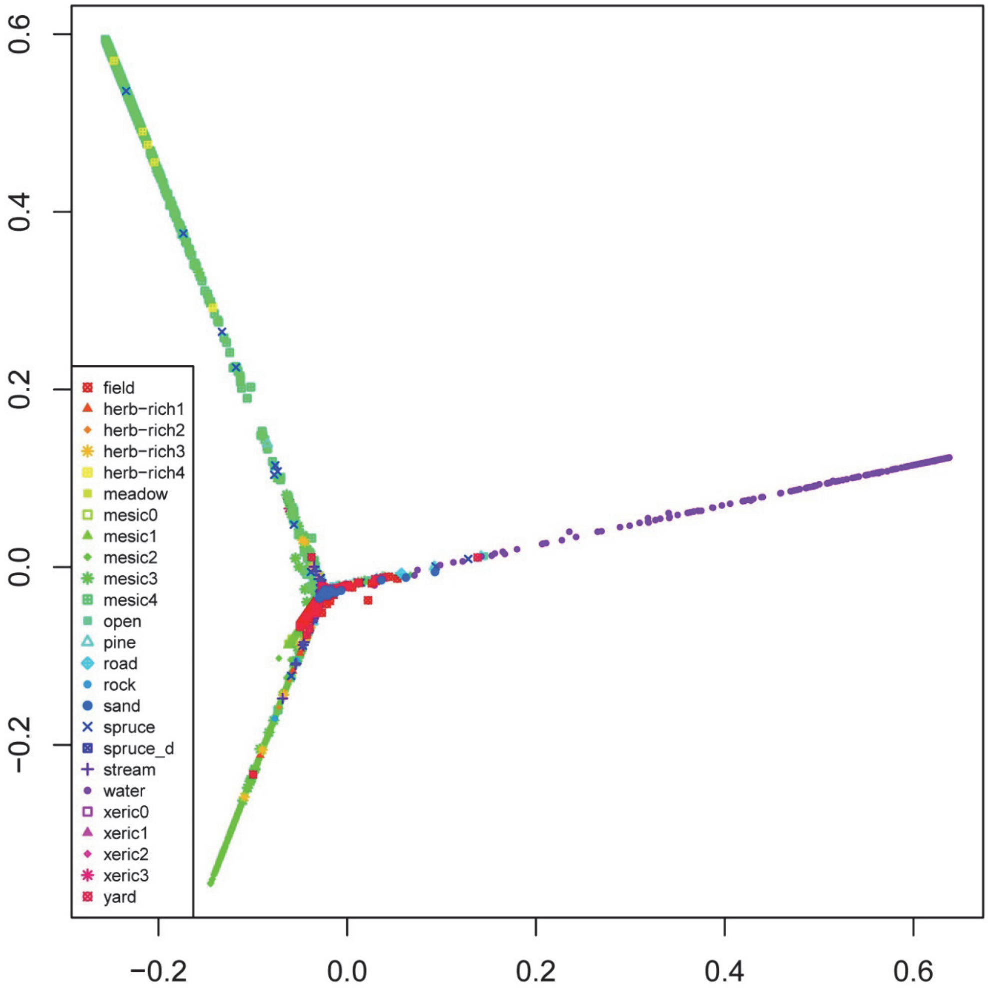

FIGURE 6 A multidimensional scaling plot of classified data based on a random forest proximity matrix. In the random forest run, the whole study area and the 100 most important features (based on Boruta results) were used. The x and y axes represent the first two scale coordinates of the proximity matrix.

TABLE 8 Pearson's correlation coefficients between conservation value and ecosystem service maps. The abbreviation F refers to landscape metric (FRAGSTATS) maps, Z to spatial conservation prioritization (Zonation) maps and complem to complementarity. Different valuation comparisons are: inside group: valuations 1–3 (or 4–6); groups compared: naturalness or not; human or not: human habitats were given or not given value. Column classification refers to the two classification methods.

Maps compared				Correlation	
Comparison 1	Comparison 2	Valuation	Classification	Lower	Upper
F w/o complem	F w/o complem	inside group	same in both	0.93	0.97
F w/o complem	F w/o complem	groups compared	same in both	0.64	0.91
F w/o complem	F w/o complem	same in both	different	0.57	0.64
F with complem	F with complem	inside group	same in both	1.00	1.00
F with complem	F with complem	groups compared	same in both	0.96	0.98
F with complem	F with complem	same in both	different	0.54	0.57
F with complem	F with complem	human or not	same in both	0.53	0.70
F w/o complem	F with complem	same in both	same in both	0.31	0.48
Z w/o similarity	Z w/o similarity	inside group	same in both	0.79	0.96
Z w/o similarity	Z w/o similarity	groups compared	same in both	0.45	0.75
Z w/o similarity	Z w/o similarity	same in both	different	0.58	0.77
Z w/o similarity	Z w/o similarity	human or not	same in both	0.49	0.78
Z with similarity	Z with similarity	inside group	same in both	0.85	0.92
Z with similarity	Z with similarity	groups compared	same in both	0.38	0.64
Z with similarity	Z with similarity	same in both	different	0.58	0.72
Z with similarity	Z with similarity	human or not	same in both	0.99	1.00
Z w/o similarity	Z with similarity	same in both	same in both	0.44	0.60
F w/o complem	Z w/o similarity	same in both	same in both	0.10	0.28
F w/o complem	Z with similarity	same in both	same in both	0.17	0.37
F with complem	Z w/o similarity	same in both	same in both	0.12	0.34
F with complem	Z with similarity	same in both	same in both	-0.15	0.03
F, no value to human habitat	Z, no value to human habitat	same in both	same in both	0.16	0.44
Recreational value	Other services	-	-	0.25	0.36
Recreational value	Recreational value	-	different	0.96	0.96
Timber	Carbon	-	-	0.84	0.84
Recreational value	All conservation value maps	-	-	-0.08	0.33
Timber	All conservation value maps	-	-	-0.24	0.47
Carbon	All conservation value maps	-	-	-0.29	0.46

Correlation color codes	< 0.1	0.1 to 0.29	0.3 to 0.49	0.5 to 0.69	0.7 to 0.89	> 0.89

4.5 Explaining species richness (IV)

In terms of explaining and predicting vascular plant species richness, HTC had equal or slightly worse explanatory and predictive power than the two other landscape classification schemes (i.e. SH and CLC) (Fig. 7, Table 9). In explaining and predicting native species richness, its relative power was slightly better than in explaining and predicting total species richness. The result differed a bit if results of hierarchical partitioning (Fig. 7) or different GAMs (Table 9) were assessed. Overall, all of the landscape type features and mean altitude had the strongest explanatory significance. Of these features, mean altitude had a negative relationship with species richness, whereas landscape heterogeneity had a positive relationship.

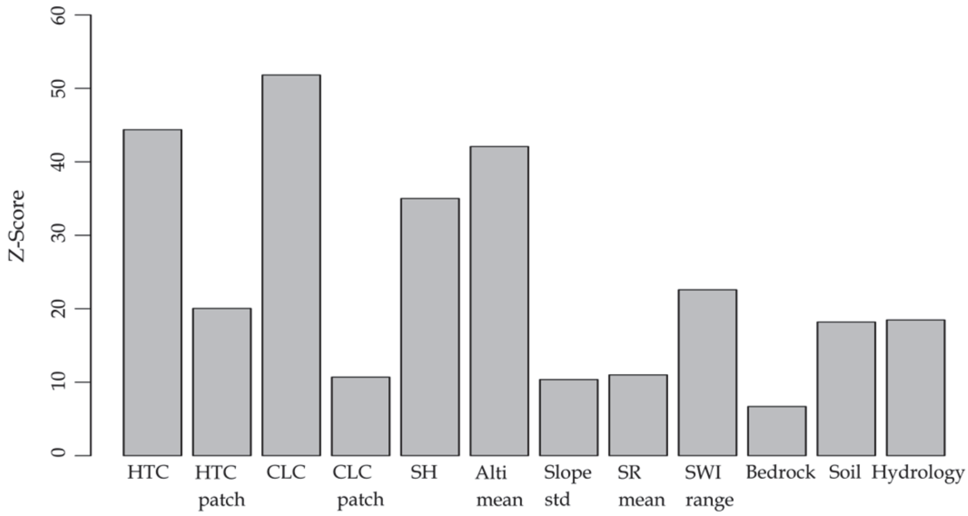

FIGURE 7 The relative importance of each used feature in explaining the total species richness based on hierarchical partitioning Z-Scores. HTC is an abbreviation of habitat type classification, CLC of Corine Land Cover, SH of spectral heterogeneity, Alti of altitude, SR of solar radiation, and SWI of SAGA wetness index.

Overall, most of the variation in species richness could be explained using landscape and topographic features (Fig. 8). Geodiversity *per se* could explain or predict little of the variation (Table 9, Fig. 8), but the combined effect of geodiversity and landscape as well as all three feature groups was considerable. The relative importance of topography and geodiversity were greater in explaining native species richness than in explaining total species richness. Additionally, the degree of unexplained variation was greater in explaining native species richness.

TABLE 9 Explained deviance, r^2 values and Spearman rank correlation coefficients between surveyed and predicted species richness for the different GAMs.

	all species				native species			
	calibration			test	calibration			test
	explained deviance	r^2 adjusted	r_s	r_s	explained deviance	r^2 adjusted	r_s	r_s
all variables	74.6 %	0.743	0.860	0.828	68.3 %	0.673	0.826	0.794
topo+GD+HTC	64.6 %	0.643	0.809	0.780	62.2 %	0.611	0.787	0.759
topo+GD+CLC	71.5 %	0.706	0.845	0.818	62.8 %	0.613	0.817	0.791
topo+GD+SH	68.8 %	0.677	0.830	0.793	63.4 %	0.625	0.789	0.763
patches omitted	74.3 %	0.742	0.858	0.827	67.9 %	0.670	0.826	0.797
topography only	46.4 %	0.448	0.690	0.663	49.0 %	0.476	0.476	0.660
topo+GD	59.4 %	0.583	0.778	0.745	57.8 %	0.561	0.747	0.718
topo+landscape	73.8 %	0.736	0.855	0.838	66.0 %	0.652	0.818	0.808

topo: topographic features

GD: geodiversity features

HTC: habitat type classification, number of habitat types

CLC: Corine Land Cover, number of land use / land cover types

SH: spectral heterogeneity, number of types

patches omitted: the number of patches features omitted from the model

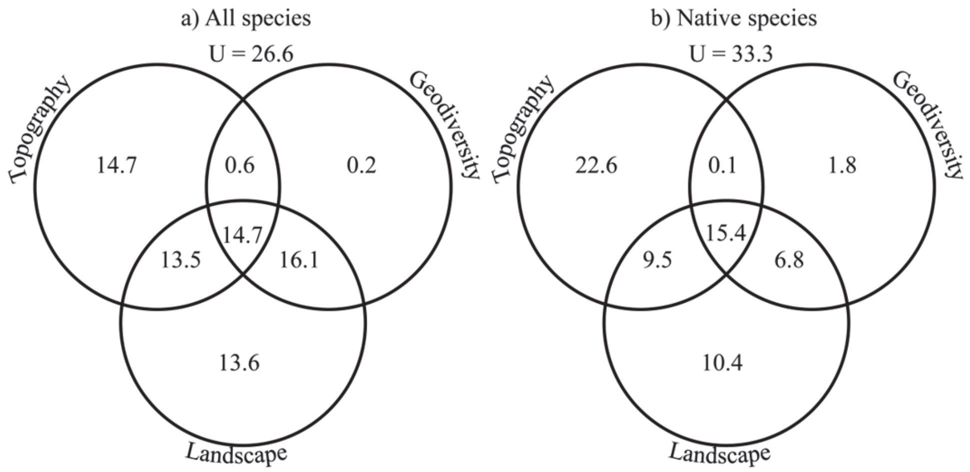

FIGURE 8 Variation of (a) total species richness and (b) native species richness divided into different fractions based on variation-partitioning analysis. Numbers inside circles are percentage fractions explained by different feature groups. For instance, landscape explained 13.6 % and landscape and topography together explained 13.5 % of the variation in total species richness. The unexplained variation is marked with the symbol U.

5 DISCUSSION

5.1 Uncertainties in the segmentation evaluation (I)

According to the results, supervised segmentation evaluation methods were inconsistent. Furthermore, when it is not possible to rank segmentations by means of supervised methods, it cannot be tested if better segmentations lead to higher classification accuracies, as has been previously argued by Kim *et al.* (2009), Clinton *et al.* (2010), Ke *et al.* (2010), and Gao *et al.* (2011). Therefore, it is vital to define in each case what segmentation is good and how quality could be measured.

In our case, we defined that good segmentation should have (1) meaningful and not too complex segments, (2) boundaries parallel to the reference polygon boundaries, even for the smallest reference polygons, and, when these two conditions are taken into account, (3) segmentation should be as coarse as possible. Since none of the available supervised measures quantified the conditions we wanted, we could find the best segmentation more easily by using visual interpretation. This was also confirmed by means of classification accuracy analysis: the highest accuracy was achieved with the segmentation that was thought to be the best in visual interpretation. Visual interpretation was, nevertheless, far from easy. We could not examine all 252 segmentations in detail, and accordingly only some general trends could be observed. More detailed observations were restricted to a few selected sites, such as small lakes or mires, inside the whole study area.

Previously, some have also argued that supervised methods are not applicable, since ground truth is subjective, inaccurate or does not exist (Wulder *et al.* 2008, Corcoran *et al.* 2010). In other words, a comparison of segmentations against reference polygons drawn by human interpreters has been criticized as not being useful since reference polygons might also be biased. In these critiques, either unsupervised methods or visual interpretation are often recommended. In these and all other approaches, it should be decided what is being sought by the analysis. It could be discussed endlessly which segmentation is the best. And overall, it is not straightforward to judge when segmentation is good (Hay and Castilla 2008), as different segmentations are good in different contexts. In other words, it can be stated that there is no right solution or truth, but that the truth is context-dependent. Therefore, it would be better to try to find an appropriate (as opposed to "right") segmentation (Lang *et al.* 2010). Furthermore, it is often enough to find an appropriate segmentation not the most appropriate. For instance, in our case, classification accuracies between some of the classifications (Table 5) were rather small. Additionally, segmentations can be altered after the initial segmentation (Blaschke *et al.* 2014), for example, by modifying the segmentation result in the parts that are not as appropriate.

In the future, segmentation evaluation could be further tested using unsupervised methods. Unsupervised methods using homogeneity or heterogeneity measures within and between segments have been widely used (e.g. Kim *et al.* 2009, Gao *et al.* 2011, Yue *et al.* 2012, Hou *et al.* 2013). These methods have not, however, been compared to supervised measures or to visual interpretation. Furthermore, it would be interesting to analyze how unsupervised measures perform when they are compared with reference polygons. In other words, would optimal segmentation based on unsupervised measures mimic human-drawn polygons in different contexts?

Another possible new research direction could include finding new methods and measures for segmentation evaluation. Such new methods could be based on landscape metrics (e.g. Neubert and Meinel 2003) or polygon boundary evaluation (e.g. Lucieer and Stein 2002). In the latter approaches, it could be assessed if the boundaries of segmentation polygons are parallel with reference polygon boundaries. However, these approaches are no panacea, since real boundaries do not always exist in nature or are complex (Wulder *et al.* 2008), and different boundaries are visible in different data and in the field.

Third, multi-scale segmentations have been suggested and performed in different contexts (e.g. Hay *et al.* 2003, Kim *et al.* 2011). Multi-scale segmentations could also be good segmentations, according to segmentation evaluation. However, in multi-scale segmentation approaches, it needs to be decided when or in what area each scale should be used.

5.2 Discussing how to map habitat types (II)

In HTC, it became evident that different mesic forests both were among the most common habitat types in the study area and had a large heterogeneity. Other forest habitats, as well as other habitat types, were often found in this heterogeneity and thus were easily confused with mesic forests. Hence, the evaluation of the heterogeneity within and between different habitat types showed its usefulness in the analysis.

One reason behind the heterogeneity problem might be that we followed the Finnish forest type classification system (Cajander 1949), in which habitat types are not based on tree species but on ground vegetation, nutrient status and soil characteristics. However, an initial analysis revealed that that classification accuracies were not better if forests were classified on the basis of the most common tree species (i.e. not based on ground-level information). Hence, spruce-dominated forests can also be heterogeneous. This reveals the difficulties of interpreting nature, which are evident both observing it from above using remote sensing or conducting fieldwork on the ground. For instance, when Cherrill and McClean (1995, 1999) compared habitat type maps drawn by expert fieldwork mappers with each other, they got low accuracy rates. This is also in line with our results in HTC in the second study area (III, IV), in which accuracy rates of forest habitat mapping were low.

To map different habitat types, different approaches are needed. For instance, mire and streamside habitats were mapped better when the studied area was split into two parts, based on one meter of vertical distance to the nearest stream. Additionally, topographic features were vital when mapping mires. These kinds of results have been found before. Random forest works slightly better when the data is more balanced (i.e. when rare classes have more cases or segments) (e.g. Breidenbach *et al.* 2010, Smith 2010). In the data of our study, uncommon mire habitats were more abundant near streams, relatively speaking. To map mires effectively, it has been found that topographic features – and potentially soil information as well – are needed (Ozesmi and Bauer 2002, Tomppo and Halme 2004, Wright and Gallant 2007, Maxa and Bolstad 2009, Corcoran *et al.* 2011). However, some rare habitat types, such as springs, could not be mapped at all. For these habitat types, another type of approach, such as more manual expert analysis (Thompson and Gergel 2008), is needed.

Different datasets and features are needed for effective mapping of forest and other habitat types. First, both ALS data and spectral images were needed to get the highest classification accuracy. This has also been found before (e.g. Geerling *et al.* 2007, Ke *et al.* 2010, Sasaki *et al.* 2012). Second, including both texture features and topographic features increased classification accuracy. Third, all types of features were regarded as important, depending on different feature importance measures. However, some of the features gave overlapping information and had large correlations. For instance, including both wavelet and GLCM texture features did not increase classification accuracy. Accordingly, it might be enough to use only one of these, not both.

In terms of habitat type classification, new research could target different comparisons. First, different classifiers could be compared. In forest inventories, for instance, improved k-NN classifications together with genetic algorithms have been used for feature selection and estimation (e.g. Tomppo and Halme 2004). A second possible comparison could be between pixel- and object-based classifications. In this comparison, hybrid or combined approaches could also be included (e.g. Bernardini *et al.* 2010, Aguirre-Gutierrez 2012). A third comparison could be between single- and multi-scale approaches. In the multi-scale approach, some areas or habitat types could be classified using a finer or coarser scale (e.g. Benz *et al.* 2004), or results could be combined using data fusion methods, such as majority vote.

Habitat types could be mapped using soft or fuzzy approaches. In other words, segments could be given probabilities of which class they most likely belong to (e.g. Benz *et al.* 2004). This could also be performed using the random forest classifier. In the random forest standard solution, a majority vote is taken. This majority vote could be replaced with proportions of how many times each class is given a vote. The heterogeneity results based on Sammon's mapping or other approaches could also be used in soft classification approaches. In these soft approaches, rarer and potentially high conservation value habitat types could be better mapped. In other words, slight indications that a segment belongs to a rare habitat type could be followed up on.

Another new research topic could be analysis of how different training datasets affect classification accuracy. Training datasets could be assessed, for example, in terms of homogeneity (i.e. if the feature information in training polygons and classes is homogenous). A subsequent test could determine if the use of more homogenous training datasets leads to higher classification accuracies. Finally, in this research, it should also be discussed if homogenous training datasets are realistic representations and if the ensuing classification results are actually found in the natural environment.

5.3 Uncertainties in conservation value mapping (III)

Many of the problems evident in segmentation evaluation are evident also in defining, assessing and mapping conservation values. If segmentation quality is a problematic concept, so is conservation value. As was discussed in the introduction, many different factors have an influence on conservation values. Furthermore, in different contexts, different issues should be targeted for conservation.

In mapping conservation values, several different uncertainties exist. We were able to show that different habitat type valuation, habitat type classification, and complementarity and connectivity calculations give different results. For habitat type valuation, it was found that if consideration of naturalness is included in the valuation, the results differ from when naturalness is not valued. Using naturalness correction is somewhat problematic, however, since arbitrary decisions must be made when differentiating the value between natural and non-natural habitat types. There are also some uncertainties in habitat type valuation. For instance, calcareous rocky areas had a lower conservation value than non-calcareous rocky areas, even though calcareous rocky areas are regarded as having a high conservation value in Finland. Of the threatened species that live on rock outcrops, three-fourths live on calcareous areas (Rassi *et al.* 2010). The reason why calcareous rocky areas were not highly valued in our analysis might be that species literature only identifies species that require calcareous habitats, even though several other species live on these habitats as well.

Different complementarity and connectivity mapping methods yielded rather different results. First, the differences between landscape metric and spatial conservation prioritization approaches were significant. The landscape metric approach was dependent on patches (i.e. delineated segments). Patches can suffer from an artefact effect: landscape metric results change when patches change (Mas *et al.* 2010) and the inner heterogeneity of patches is not taken into account (McGarigal *et al.* 2009). Hence, it is crucial that patches are mapped realistically. However, as was discussed above in evaluation of segmentation, differently delineated segments are only different representations of reality.

Because the spatial conservation prioritization approach that was used was not dependent on patches, it can be considered as more realistic. Additionally, maps produced by means of spatial conservation prioritization were visually more appealing. Nevertheless, it is necessary to choose if habitat type similarities are taken into account. When similarity was included, high-value habitat types were preferred. It is entirely context-specific if more high-value habitats or other different habitat types are identified as needing to be protected. Additionally, the uncertainties of habitat valuation increase when habitat type connectivity and complementarity are mapped.

The conservation value maps were different from the ecosystem service maps. All mapped ecosystem services were mostly produced in forest areas, however, whereas habitat type values were assigned to all habitat types in the area. Ecosystem services were mapped more realistically than by using simple look-up tables (see Costanza 1997, Troy and Wilson 2006, Burkhard *et al.* 2012). Comparisons between ecosystem services and other conservation values might be important. Nonetheless, when conservation decisions are made, conservation could target areas that produce many different ecosystem services and are also valuable on the basis of species composition and naturalness.

We recommend that at least three types of questions be asked when conservation values are mapped. First, it should be discussed why nature should be conserved in the first place. In different contexts, there are different reasons to conserve nature. It is necessary in some cases to protect the most natural areas and in other places the areas that provide most benefits to humans over the long term. Second, it is important to discuss what should be mapped. Naturalness or biodiversity cannot be mapped as such; instead, some surrogate must be used. Third, the method of mapping should be chosen. Even if what needs to be mapped is known, there are different alternatives of how to map (e.g. as regards datasets and specific methodologies).

In future studies, areas that are the most valuable based on different approaches could be mapped and thus targeted for conservation. Alternatively, potential exists in fuzzy approaches in which uncertainties are taken into account. For instance, based on a certain measure, an area might be assessed in terms of potential conservation value and likelihood of the area actually having that value. Additionally, more dynamic approaches, which take into account ecological processes and changes in landscapes, ought to be developed. In terms of comparisons and combinations, additional ecosystem services should be mapped. Approaches for mapping many different ecosystem services already exist, but most of these approaches are not spatially explicit in a fine scale. The usefulness of habitat type maps for mapping conservation values could be tested. For instance, fieldwork might be targeted at areas with high conservation values. Another potential research topic is comparisons between conservation value maps based on habitat type and SDM-based maps of showing areas that are potentially species rich.

5.4 Discussing what explains species richness (IV)

Most of the variation in species richness was explained through mean altitude and different landscape type features. Mean altitude does not, however, have a climatic effect on the study area, since altitudinal differences are less than 100 m. Instead, most of the nutrients and productive soil are washed and eroded from hilltops into lower altitudes. Additionally, human influence has been most pronounced at lower altitudes, due to more productive soils and closer proximity to water bodies. In other studies (e.g. Honnay *et al.* 2003, Wania *et al.* 2006), human influence has been found to have a positive relationship with plant species richness. Furthermore, areas with higher landscape variability often also have stronger human influence (e.g. Honnay *et al.* 2003, Wania *et al.* 2006).

Perhaps surprisingly, HTC was not a better explanatory feature than CLC or SH. In some of the comparisons, it was actually a bit worse. This is surprising since HTC has an ecological background (i.e. it is based on the habitat type preferences of species). On the other hand, different landscape variability measures gave complementary information, all of which was needed for the model to have the highest explanatory and predictive capability.

Geodiversity features did not have as strong explanatory power as topographic and landscape features. In a previous study, geodiversity measures were better explanatory features than climatic and topographic features (Hjort *et al.* 2012). In that study, however, landscape features were not included and the study areas were predominantly natural. The inclusion of landscape features is somewhat problematic, since humans have a significant influence on landscape but not on topography or geodiversity. Geodiversity and topography may, therefore, set the stage for and be the ultimate factors behind species richness (see Anderson and Ferree 2010, Beier and Brost 2010). Landscape, on the other hand, is a more proximate cause. In other words, landscape composition and human influence are not random but found in a framework set by topography and geodiversity. This can partly be seen in the difference between native and total species richness: the relative importance of topography and geodiversity was greater when it came to explaining native species richness. Moreover, in a previous study, geodiversity was positively linked with habitat richness (Jačková and Romportl 2008).

In future studies, instead of explaining species richness, explanations can target specific species, rare or threatened species richness, or rarity corrected species richness. New explanatory features that were not included here could also be added to the model. One of these is a geodiversity measure of geomorphological diversity, included in the analysis of Hjort *et al.* (2012). Additionally, other models than GLMs and GAMs could be used and compared with each other. For instance, although there are some comparisons between different approaches when taking spatial autocorrelation into account (e.g. Beale *et al.* 2010), these comparisons are not thorough. In SDMs, fuzzy approaches can also be used with uncertainties being taken into account, for example by maps that show the reliability of mapped distributions (Rocchini *et al.* 2011b).

5.5 Synthesizing results

Habitat type or other thematic maps, such as land cover maps, are widely used in different tasks, for instance in mapping ecosystem services (Egoh *et al.* 2012, Maes *et al.* 2012, Crossman *et al.* 2013) or biodiversity patterns (e.g. Kerr and Ostrovsky 2003, Turner *et al.* 2003, Gillespie *et al.* 2008).

According to Newton *et al.* (2009), the use of remotely sensed thematic land cover maps is widespread in landscape ecology (i.e. studies in which the relationship between spatial pattern and ecological processes is analyzed). However, maps are used uncritically; for instance, their accuracy is not discussed and other remotely sensed variables are not used. In this study, a critical approach has been taken in how successfully patches can be delineated (I) and habitat types can be classified (II), how different conservation value maps are produced when different habitat type maps are used as a starting point (III), and if different thematic maps work to explain biodiversity patterns (IV). Additionally, in the ecosystem service mapping (III), habitat type maps were only used in part, as direct biomass and timber volume estimates were also used. In landscape ecology, the role of discrete thematic patches has been partially questioned. As an alternative, the analysis of continuous surfaces of DTMs or remotely sensed datasets, for example, has been suggested (McGarigal *et al.* 2009). In this study, continuous surfaces were used in the form of DTM to explain species richness (IV) and also in the form of different layers to map habitat types (I, II, III). As was already suggested above, fuzziness or the internal variability of thematic classes should be taken into account in future studies. Otherwise, continuous surfaces could also be more widely used (e.g. to produce conservation value maps). Although patches were not visible in output spatial conservation prioritization maps (III), the input habitat type maps were composed of patches.

When thematic maps and patches are produced, OBIA methods are superior compared to pixel-based methods (e.g. Blaschke *et al.* 2014). The delineation of patches is initially performed in the segmentation phase. However, segmentations only delineate prototypes of landscape objects or patches. Furthermore, segments do not exist autonomously in the images; they are produced by human cognition. While delineated segments are prototypes of landscape objects, they are also only initial representations of image objects. For segments to be more meaningful image objects, they are often combined or modified. To act as landscape objects, segments are thematically combined and associated with classes such as habitat types (Castilla and Hay 2008). As was discussed in publication I, segments are very different according to different segmentations. Moreover, different segmentations produce different habitat type maps (I), which respectively entail different conservation value maps (III). Therefore, when conservation value maps are produced, efforts should be made at every stage (i.e. segmentation, classification, conservation value mapping) to be rigorous. Nevertheless, human cognition and decision-making play a part at every point. Therefore, while subjectivity cannot be removed from the mapping process, there should be awareness of how it affects the process.

In III and IV, species richness, conservation values and ecosystem services were considered separately. This differs from what was discussed in the introduction, however. As Fig. 1 points out, both ecosystem services and species richness (or diversity) are constituents of conservation values. In this manner, ecosystem service maps (III), together with species richness maps (IV), are also conservation value maps.

Yet the species richness maps drawn in publication IV are drastically different from the conservation value maps drawn in publication III. The reason behind this difference is threefold. First, the spatial scale differs. The grain size (i.e. the finest spatial resolution (see Turner *et al.* 2001)) in the species richness maps (IV) is 1000 m, whereas in the conservation value maps (III) the grain size is 10 m. Therefore, a single species richness survey quadrat contains 10,000 conservation value map cells. In different scales, different processes occur (Levin 1992). Thus, there exist problems of how to combine, or even to compare, these two maps. Second, in the conservation value maps, other considerations than species richness were taken into account. These include habitat type connectivity and complementarity, together with habitat type valuation based on naturalness and rarity. Third, it is problematic if species richness can be used as a proxy for conservation values. As discussed before, species richness is often highest in human-dominated cultural landscapes (e.g. Honnay *et al.* 2003), whereas conservation values, depending how they are defined, are often highest in more natural landscapes. Additionally, rarity of species is not taken into account in pure species richness, even though rarity is a major consideration in many of the conservation value assessments (see e.g. Margules and Usher 1981, Williams *et al.* 1996). In brief, the species richness maps and conservation value maps drawn in this study tell different stories.

Nonetheless, there are ways to combine conservation value and species richness. First, according to species richness modeling, landscape diversity explains species richness. Hence, it is advisable to take habitat type complementarity into account when mapping conservation values. In this way, the maximum number of species and habitats can be protected. Second, high-conservation value areas can also be areas with high species richness or with elements that can cause high species richness. For instance, areas which have high geodiversity have a high conservation value (see e.g. Anderson and Ferree 2010, Beier and Brost 2010, Ruddock *et al.* 2013). Third, species richness and other factors that contribute to conservation values have connections. For instance, there is a plethora of literature that discusses the relationship between species richness (or, more generally speaking, biodiversity) and ecosystem services. For a good synthesis, see Mace *et al.* (2012). Overall, it is important to sort out factors that explain species richness. Although it is problematic to use species richness *per se* as a conservation value, considerations of species richness help in mapping the different elements that are important for conservation.

When mapping conservation values, it is important to choose if the current conservation value or the future conservation value should be targeted. In a changing climate, it has been argued that areas with the elements to provide high biodiversity in the future should be targeted (Anderson and Ferree 2010, Beier and Brost 2010, Ruddock *et al.* 2013). The same could also be applied to other constituents of conservation value. For example, if there is a desire to protect nature so that the supply of ecosystem services is secured, it should be analyzed where the supply of services will be high in the future. Habitat type maps can also be used in these kinds of mapping tasks. Instead of mapping current habitat types based on present vegetation, maps of potential future habitat type distribution could be drawn.

Ideally the spatial distribution of conservation values could be mapped by mapping all the constituents of conservation values. This is a daunting task, however, since there are numerous constituents, and creating visual representations of them on maps is difficult. Moreover, several (often arbitrary) decisions need to be made when weighting different constituents. All in all, new tools for mapping different conservation values (e.g. species richness and ecosystem services) in various landscapes are needed. Many of these tools can be based on habitat type maps or on object-based image analysis methodologies.

5.6 Putting the findings into a planning context

In this research, several limitations and conditions of the results and methodologies used have been raised. While it is indeed problematic to use a conservation value map drawn in publication III for planning decisions *per se*, the use of a conservation value map, for example, would not be as problematic if it were combined with fieldwork. In other words, a conservation value map can be used for the purpose of determining fieldwork targets (see e.g. Rossi and Kuitunen 1996, Hilli and Kuitunen 2005). Instead of using random sampling, fieldwork could be performed in the most valuable locations shown on a conservation value map or in other interesting locations. Therefore, although the developed methods can be used to predict potential locations for conservation, they should always be accompanied by expert evaluation.

The limitations and conditions raised in this thesis should be taken into account when planning decisions are made. It should be understood that maps are not the truth, but they represent and simplify the truth. In map production, arbitrary decisions often have and will be made, and these have an effect on the map. The effect of these decisions should be understood and also documented when maps are produced. Furthermore, maps from intermediate phases are almost as important as the final maps. For instance, instead of using only a final conservation value map or habitat type map, SWI and CHM maps or satellite images can be used to make conservation decisions and so forth.

When maps are used in real-world decisions, the adopted workflow could be something similar to the workflow sketched out in Fig. 9. This kind of workflow could also be an ideal workflow for research purposes. The workflow is not intended to be used in planning *per se*; instead, the workflow should provide information that can aid in making planning decisions. The workflow is, in addition, rudimentary. In other words, it has not been tested, it could be thought out more thoroughly, and it has not been sketched out in a process involving other researchers or stakeholders. That is, problems related to land use, such as the degradation of the ecosystem functioning, are "no technical solution problems" (see Hardin 1968). In other words, there are no technical solutions of how to solve these problems, and using only natural science in the solutions is not enough. In research, therefore, approaches should be transdisciplinary (i.e. there should be an integration of academic researchers from the humanities or social sciences and the natural sciences, as well as non-academic participants). In transdisciplinary research, integrative theory and knowledge between science and society are created; hence, disciplines are merged (see Fry *et al.* 2007).

In Fig. 9, the workflow is framed inside a context. Three types of contexts are mentioned: spatial context refers to the location where research is performed, environmental context refers to the carrying capacity of the Earth (see e.g. Rockström *et al.* 2009), and socio-political context includes the planning context, together with local and wider society, culture, and economics. These contexts should be taken into account and analyzed when mapping is performed. The workflow starts with three questions: why, what and how (III, Section 5.3). After this, the workflow resembles what was sketched in Fig. 3 and is the most applicable to conservation value mapping based on habitat types. However, it can also be applied to other types of conservation value mapping tasks. After it has been decided how the mapping should be done, all relevant spatial and non-spatial data from the area should be collected. If needed, some fieldwork should be performed for the data to be used, for example, in accuracy assessments. After data gathering, the following phases are segmentation, classification and conservation value mapping based on the valuation of habitat types. As an alternative to habitat types, some other features (such as habitats or habitat quality) can also be mapped. In these three phases, uncertainties should be taken into account. Uncertainties are included in the workflow in each phase. Segmentations should be evaluated, checked if they resemble the needs of the study, and potential problematic aspects should be acknowledged. The accuracy of classification should be assessed overall, in relation to habitat types and in relation to different locations. Habitat type classification should also be subsequently adjusted if the accuracy in regard to some parts or habitat types is not reasonable. To perceive uncertainties in conservation value maps, some external reference data can be used. These data can be, for instance, species richness inventory data. In both habitat type classification and conservation value mapping, fuzzy approaches should be adopted (see Sections 5.2, 5.3 and 5.4). Finally, when conservation value maps are drawn, they need to be evaluated by experts and verified with fieldwork.

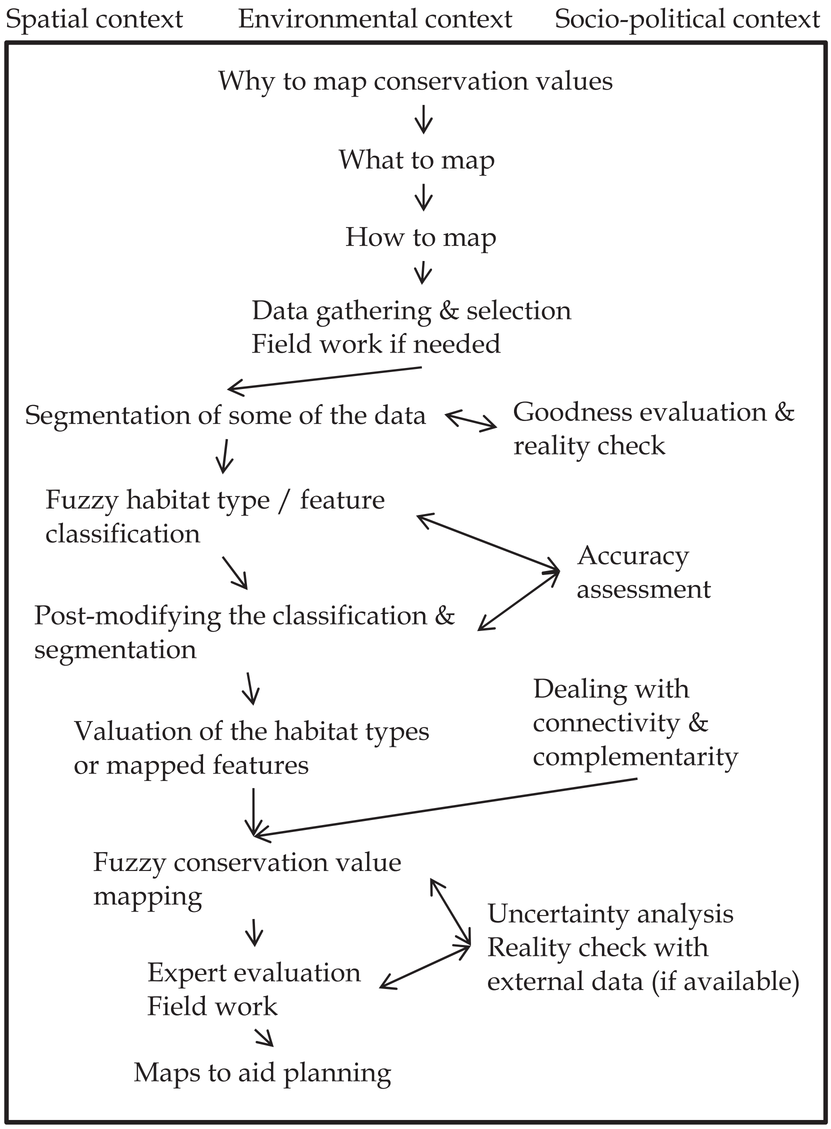

FIGURE 9 A simplified and ideal workflow showing how conservation values could be mapped in a research project whose results are intended for planning purposes.

6 CONCLUSION

In this research, new methods for mapping habitat types and conservation values were developed. The approach was critical and comparative. Uncertainties and potential limitations were discussed in each of the different tasks. In addition, comparisons were made between methodologies, datasets and approaches.

At the core of the thesis was an object-based classification approach for habitat types. It was found that this classification works relatively well. Nevertheless, there remain problems of how to delineate and classify areas in a natural environment. First, boundaries in nature are not always self-evident and their locations are open to interpretations. It was argued that choosing the best segmentation is often arbitrary and supervised evaluation measures are sub-optimal for judging segmentation quality in large natural areas. Second, thematic classification of natural environment is also open to interpretation. The heterogeneity within and between different habitat types should be taken into account when habitat type maps are constructed. Nevertheless, different datasets, layers, features and approaches are required when mapping all of the relevant habitat types and for the sake of the highest level of classification accuracy. Both ALS data and spectral images are needed, and several types of features must be calculated from them. When the different uncertainties are taken into account, habitat type maps can be used widely for different purposes.

When habitat type maps are converted into conservation value maps, several decisions need to be made, such as if species richness, rarity, landscape naturalness or ecosystem services should be targeted in the mapping. It was shown that maps where only potential species richness and rarity, species richness, rarity and naturalness, or ecosystem services were mapped were different. Additionally, other less significant decisions must be made (e.g. regarding mapping methodology and weighting of different issues). In the analysis, conservation value maps were different if habitat type mapping or connectivity and complementarity mapping methods were changed. Hence, it was pointed out that different decisions can have large effects on the resulting maps. Therefore, reasoning is needed for why, what and how to map conservation values. Nonetheless, conservation value maps are potentially effective tools for guiding conservation decisions.

It was discovered that most of the variation in vascular plant species richness could be explained by landscape and topography, while the relative role of geodiversity was small. Although habitat type maps performed slightly worse than two other landscape type classifications in terms of explaining and predicting species richness, different landscape type classifications provided complementary information. Species richness had a positive relationship with habitat type heterogeneity. Hence, when current levels of species richness are sought to be maintained, habitat type heterogeneity and complementarity should be taken into account in planning decisions. On the other hand, because habitat type heterogeneity is often highest in human-dominated areas, other aspects (e.g. naturalness and ecosystem services) should be considered at the same time. Furthermore, since climate change is expected to have an effect on species distribution patterns, it is vital to acknowledge ultimate factors that control species distribution, such as geodiversity and topography. This was also partially suggested by the fact that the relative roles of topography and geodiversity were greater in explaining native species richness

The methods developed here can be used for planning, but with care. The uncertainties must be taken into account and ground reference data should be used. Finally, it is important to remember that all planning and mapping occurs in a context that limits and directs how the mapping should be performed.

Acknowledgements

The research was funded by the Maj and Tor Nessling foundation and the Department of Biological and Environmental Science of the University of Jyväskylä. Additionally, Finnish Doctral Programme in Environmental Science and Technology provided some traveling funds.

I would like to thank my supervisors Anssi Lensu and Markku Kuitunen. First and foremost, thank you for putting your faith in me and recruiting me to do a PhD after I had barely started my MSc project. You both have also played a major role during the thesis. Although most of the final decisions of what to study and to write were mine, all the ideas were originally yours. Anssi, thank you for helping with statistics, programming, and geoinformatics, and for your extraordinary attention to detail. Sometimes I was almost deliberately sloppy with all the details, since I knew you would correct them. Markku, it has always been nice to discuss with you, and you gave me a lot of good inspiration, ideas and ecological insights during this process.

I thank my co-authors Erkki Tomppo, Jan Hjort, Antti Rusanen, Asta Vaso, and Tuomo Kuitunen. Erkki and Jan, it has been vital to have collaborations with other senior scientists than my supervisors. Antti and Asta, although I acted as a supervisor in your MSc theses, you both also supervised me and gave needed peer support. Tuomo, your work in mapping the flora of Luopioinen has been enormous, and papers III and IV would have been impossible to write without your work.

I am grateful to the members of the thesis follow-up group, Mikko Mönkkönen and Risto Kalliola. Although we only had two meetings during this process, you both provided me with good ideas and confidence in my work. Additionally, I thank reviewers of this thesis, Timo Kumpula and Duccio Rocchini for well-thought-out feedback, as well as Albion M. Butters and the University of Jyväskylä Language Services for proofreading the thesis.

I thank Riku Sinervo and Aino Voutilainen for doing a BSc thesis under my supervision and helping with some of the calculations. Some of the fieldwork data produced by Paula Kinanen was used in this thesis, and for that I thank you. I also want to acknowledge Antti Peltonen, Panu Kuokkanen, Merja Kytömäki, Janne Hartman, Raino Lampinen, and others for helping me to get data.

I taught at two nice field courses during my PhD studies. Thank you to Elisa Vallius, Lotta Jaakkola, and Markku for giving me freedom when teaching and for building a great atmosphere in the courses. Sirkku Juhola, thank you for encouraging me to do a PhD. Niina Käyhkö, thanks for adopting me at the University of Turku for a while. I had a brief stay at the Gund Institute of the University of Vermont during my PhD process. I acknowledge Austin Troy, Taylor Ricketts, and others for allowing me to come there and for sharing in enlightening discussions. There are also others with whom I have had insightful discussions during my work, and thank you all.

Finally, I am grateful to my family: Tuula, Reeta and Tilda. Without you, I would probably have quit this project some time ago.

YHTEENVETO (RÉSUMÉ IN FINNISH)

Elinympäristötyyppien ja suojeluarvojen kartoitus kaukokartoitusaineistojen ja paikkatietomenetelmien avulla

Tutkimuksessa kehitettiin uusia menetelmiä elinympäristötyyppien ja suojeluarvojen kartoitukseen. Toisin sanoen elinympäristötyyppien ja suojeluarvojen alueellisesta jakaumasta luotiin karttoja. Uuden kehittämisen lisäksi tutkimuksessa hyödynnettiin jo kehitettyjä menetelmiä ja kehitettiin niitä tarkoitukseen sopiviksi. Näkökulma oli kriittinen ja vertaileva, eli eri lähestymistapoja, aineistoja ja menetelmiä vertailtiin.

Tutkimusalueita oli kaksi: Keljon alue Jyväskylän lounaispuolella ja Luopioisten alue Tampereen kaakkoispuolella. Alueet ovat pääosin metsäisiä mutta sisältävät myös muita maankäyttötyyppejä kuten vesistöjä, maatalousalueita ja soita. Käytetyt aineistot olivat kaukokartoitus- ja paikkatietoaineistoja, joista tärkeimpiä olivat laserkeilausaineistot, ilmakuvat ja tarkkaresoluutioiset satelliittikuvat.

Tutkimus jakautui neljään osatutkimukseen. Kahdessa ensimmäisessä osatutkimuksessa tutkimusmenetelmät perustuivat kaukokartoitusaineistojen objektiperustaiseen kuva-analyysiin, jossa yksittäisten kuva-alkioiden sijaan tarkastellaan homogeenisia objekteja. Objektit luotiin analyysin ensimmäisessä vaiheessa segmentoinnissa eli automaattisesti tapahtuvassa yhtenäisten kuvaobjektien kuvioinnissa. Toisessa vaiheessa objektit luokiteltiin elinympäristötyyppeihin. Kolmannessa ja neljännessä osatutkimuksessa objektiperustaisen kuva-analyysin avulla kehitettyä elinympäristöluokittelua sovellettiin suojeluarvojen kartoittamisessa ja lajirunsauden selittämisessä.

Ensimmäisessä osatutkimuksessa tarkoituksena oli löytää segmentointimenetelmä, joka on käyttökelpoinen boreaalisessa metsämaisemassa. Lisäksi artikkelissa vertailtiin erilaisia tapoja, joilla segmentoinnin hyvyyttä voidaan arvioida ja verrata kenttätyönä tehtyyn maisemakuviointiin. Kaiken kaikkiaan vertailussa oli mukana 252 erilaista segmentointia. Tutkimuksessa havaittiin, että segmentoinnin hyvyyden arviointiin kehitetyt ohjatut arviointimittarit antoivat ristiriitaisia tuloksia eikä niiden avulla voitu päätellä, mikä segmentointi olisi paras. Vastaavasti segmentoinnin visuaalinen ihmistyönä tehtävä arviointi oli aikaa vievää eikä sen avulla pystytty tarkastelemaan erilaisia segmentointeja kattavasti. Kuitenkin visuaalisesti parhaalta näyttävä segmentointi antoi parhaan luokittelutarkkuuden elinympäristöluokittelussa, vaikkakin myös muilla mielekkäillä segmentoinneilla saatiin lähes yhtä hyviä tuloksia.

Toisessa osatutkimuksessa kehitettiin elinympäristötyyppien luokittelumenetelmää eteenpäin. Jokaisesta segmentistä eli kuvaobjektista laskettiin 328 erilaista piirrettä, jotka kuvasivat esimerkiksi kohteen heijastavuutta, heijastavuuden eroja, puuston rakennetta ja maaston muotoja. Tämän jälkeen kuvaobjektit luokiteltiin eri elinympäristötyyppeihin käyttämällä automaattista luokittelumenetelmää. Osatutkimuksessa vertailtiin erilaisia luokitteluja, joissa osa piirretyypeistä oli jätetty pois. Parhaassa luokittelussa 79 % alueesta luokittui niihin elinympäristötyyppeihin, jotka kyseisissä sijainneissa maastokartoituksen perusteella tulkittiin olevan. Osatutkimuksessa havaittiin, että luokittelussa tarvitaan paljon erilaisia piirteitä, jotta luokittelutarkkuus olisi korkein mahdollinen ja jotta kaikki elinympäristötyypit saataisiin mahdollisimman hyvin luokiteltua.

Kolmannessa osatutkimuksessa kehitettyä elinympäristöluokittelua käytettiin suojeluarvojen kartoittamisessa. Kartoitetuille elinympäristöille määritettiin arvo potentiaalisen lajirunsauden, lajiston harvinaisuuden ja elinympäristötyypin luonnonmukaisuuden avulla. Suojeluarvojen kartoittamisessa otettiin lisäksi huomioon elinympäristötyyppien sisäinen kytkeytyvyys ja toisiaan täydentävyys. Osatutkimuksessa vertailtiin, miten suojeluarvokartat muuttuvat, kun elinympäristöluokittelua, elinympäristötyyppien arvotustapaa tai kytkeytyvyyden ja täydentävyyden laskentatapaa vaihdetaan. Lisäksi suojeluarvokarttoja vertailtiin kolmeen ekosysteemipalvelukarttaan, jotka kuvastivat potentiaalista puun ja varastoidun hiilen määrää sekä maiseman virkistysarvoa. Osatutkimuksessa havaittiin, että erilaisilla kartoitustavoilla saadaan luotua hyvinkin erilaisia suojeluarvokarttoja. Vastaavasti kaikki suojeluarvokartat poikkesivat melko paljon ekosysteemipalvelukartoista. Siten erityistä huomiota on kiinnitettävä, miksi, mitä ja miten suojeluarvoja kartoitetaan.

Neljännessä osatutkimuksessa selitettiin ja ennustettiin neliökilometriruutujen putkilokasvilajirunsautta. Selittävinä tekijöinä käytettiin aikaisemmissa osatutkimuksissa luotua elinympäristöluokittelua, muita maiseman vaihtelevuutta kuvaavia muuttujia, maan muotojen vaihtelevuutta ja geologista monimuotoisuutta. Kokonaislajirunsaudesta pystyttiin selittämään noin 75 % ja alkuperäisten kasvien lajirunsaudesta noin 68 %. Maisemamuuttujat ja maanmuotojen vaihtelevuus selittivät runsauden vaihtelusta suurimman osan. Geologisen monimuotoisuuden rooli oli selvästi pienempi. Yksittäisistä muuttujista tärkeimmiksi osoittautuivat ruudun keskimääräinen korkeus merenpinnasta ja maisematyyppien lukumäärä ruudussa. Korkeuden lisäys vähensi ja maisematyyppien runsaus lisäsi lajirunsautta. Kehitetty elinympäristöluokittelu selitti runsauden vaihtelusta hieman vähemmän kuin vertailussa mukana olleet muut maisemamuuttujat. Osatutkimuksessa pohdittiin, että maan muodot ja geologinen monimuotoisuus ovat mahdollisesti lajirunsautta sääteleviä perimmäisiä syitä ja ne määrittelevät lajirunsauden lisäksi myös maisematekijöiden alueellista jakautumista.

Kaiken kaikkiaan tutkimuksessa havaittiin, että objektiperustainen elinympäristöluokittelu toimii hyvin ja sitä pystyy käyttämään laajasti erilaisissa tehtävissä. Sen käyttämisessä tulee kuitenkin huomioida epävarmuustekijät, jotka osin aiheutuvat luonnon monitulkintaisuudesta. Lisäksi suojeluarvokarttoja luotaessa tulee ottaa huomioon elinympäristötyyppien ja maanmuotojen vaihtelevuus, jos tarkoituksena on säilyttää nykyisenkaltainen lajirunsaus. Toisaalta lajirunsauden lisäksi usein halutaan huomioida myös muita tekijöitä, kuten lajiston harvinaisuus, maiseman luonnonmukaisuus ja potentiaalinen ekosysteemipalveluiden tuotanto.

Kehitettyjä menetelmiä karttojen luomiseksi voidaan käyttää maankäytön suunnittelun tukena, jos lopullisten karttojen ja välivaihekarttojen lisäksi asiantuntijat arvioivat epävarmuutta aiheuttavia tekijöitä ja suorittavat riittävät maastokartoitukset. Maankäytön suunnittelussa tulee luonnollisesti ottaa huomioon myös kulttuurinen, yhteiskunnallinen ja alueellinen asiayhteys, jossa karttoja käytetään.

REFERENCES

Aguirre-Gutierrez J., Seijmonsbergen A.C. & Duivenvoorden J.F. 2012. Optimizing land cover classification accuracy for change detection, a combined pixel-based and object based approach in mountainous area in Mexico. *Appl Geogr* 34: 29–37.

Ahti T., Hämet-Ahti L. & Jalas J. 1968. Vegetation zones and their sections in Northwestern Europe. *Ann Bot Fenn* 5: 169–211.

Altmoos M. & Henle K. 2007. Different in characteristics of reserve network selection using population data versus habitat surrogates. *Biodivers Conserv* 16: 113–135.

Anderson M.G. & Ferree C.E. 2010. Conserving the stage: climate change and the geophysical underpinnings of species diversity. *PLoS ONE* 5(7): e11554, doi:10.1371/journal.pone.0011554.

Angermeier P.L. 2000. The natural imperative for biological conservation. *Conserv Biol* 14: 373–381.

Angermeier P.L. & Karr J.R. 1994. Biological integrity versus biological diversity as policy directives. *Bioscience* 44: 690–697.

Antonorakis A.S., Richards K.S. & Brasington J. 2008. Object-based land cover classification using airborne LiDAR. *Remote Sens Environ* 112: 2988–2998.

Araújo M. & Guisan A. 2006. Five (or so) challenges for species distribution modelling. *J Biogeogr* 33: 1677–1688.

Arivazhagan S. & Ganesan L. 2003. Texture classification using wavelet transform. *Pattern Recogn Lett* 24: 1513–1521.

Arponen A., Cabeza M., Eklund J., Kujala H. & Lehtomäki J. 2010. Costs of integrating economics and conservation planning. *Conserv Biol* 24: 1198–1204.

Austin M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol Model* 200: 1–19.

Baatz M. & Schäpe A. 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl J., Blaschke T. & Griesebner G. (eds.), *Angewandte geographische Informations-verarbeitung XII*, Wichmann, Heidelberg, pp. 12–23.

Baatz M., Hoffmann C. & Willhauck G. 2008. Progressing from object-based to object-oriented image analysis. In: Blaschke T., Lang S. & Hay G.J. (eds.), *Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications*, Springer, Berlin, pp. 29–42.

Balaguer A., Ruiz L.A., Hermosilla T. & Recio J.A. 2010. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. *Comput Geosc* 36: 231–240.

Bar Massada A., Kent R., Blank L., Perevolotsky A., Hadar L. & Carmel Y. 2012. Automated segmentation of vegetation units in a Mediterranean landscape. *Int J Remote Sens* 33: 346–364.

Beale C.M., Lennon J.J., Yearsley J.M., Brewer M.J. & Elston D.A. 2010. Regression analysis of spatial data. *Ecology* 13: 246–264.

Beier P. & Brost B. 2010. Use of land facets to plan for climate change: conserving the arenas, not the actors. *Conserv Biol* 24: 701–710.

Benz U.C., Hofmann P., Willhauck G., Lingenfelder I. & Heynen, M. 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J Photogramm* 58: 239–258.

Bernardini A., Frontoni E., Malinverni E.S., Mancini A., Tassetti A.N. & Zingaretti P. 2010. Pixel, object and hybrid classification comparisons. *J Spat Sci* 55: 43–54.

Beven K.J. & Kirkby M.J. 1979. A physically based variable contributing area model of catchment hydrology. *Hydrolog Sci Bull* 24: 43–69.

Blaschke T. 2010. Object based image analysis for remote sensing. *ISPRS J Photogramm* 65: 2–16.

Blaschke T., Hay G.J., Kelly M., Lang S., Hofmann P., Addink E., Feitosa R.Q., van der Meer F., van der Werff H., van Coillie F. & Tiede D. 2014. Geographic object-based image analysis – towards a new paradigm. *ISPRS J Photogramm* 87: 180–191.

Bock M., Xofis P., Mitchley J., Rossner G. & Wissen M. 2005. Object-oriented methods for habitat mapping at multiple scales – case studies from Northern Germany and Wye Downs, UK. *J Nat Conserv* 13: 75–89.

Botequilha Leitão A. & Ahern J. 2002. Applying landscape ecological concepts and metrics in sustainable landscape planning. *Landscape Urban Plan* 59: 65–93.

Braun-Blanquet J. 1932. *Plant sociology. The study of plant communities*. McGraw-Hill, New York.

Breidenbach J., Næsset E., Lien V., Gobakken T. & Solberg S. 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sens Environ* 114: 911–924.

Breiman L. 2001. Random forests. *Mach Learn* 45: 5–32.

Breiman L. & Cutler A. 2007. *Random forests – classification description.* Available at http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed 23.12.2013).

Burkhard B., Kroll F., Nedkov S. & Müller F. 2012. Mapping ecosystem service supply, demand, and budgets. *Ecol Indic* 21: 17–29.

Cajander A.K. 1949. Forest types and their significance. *Acta For Fenn* 56: 1–71.

Callicott J.B. 1997. Conservation values and ethics. In: Meffe G.K., Carroll C.R. (eds.), *Principles of conservation biology*, 2nd ed., Sinauer Associates, Sunderland, MA, USA, pp. 29–55.

Callicott J.B., Crowder L.B. & Mumford K. 1999. Current normative concepts in conservation. *Conserv Biol* 13: 22–35.

Câmara G., Vinhas L., Ferreira K.R., De Queiroz G.R., De Souza R.C.M., Monteiro A.M.V., De Carvalho M.T., Casanova M.A. & De Freitas U.M. 2008. TerraLib: An open source GIS library for large-scale environmental and socio-economic applications. In Hall G.B. & Leahy M.G. (eds.), *Open source approaches in spatial data handling,* Springer, Berlin, pp. 247–270.

Carleer A.P., Debeir O. & Wolff E. 2005. Assessment of very high resolution satellite image segmentations. *Photogramm Eng Rem S* 71: 1285–1294.

Castilla G. & Hay G.J. 2008. Image objects and geographic objects. In: Blaschke T., Lang S. & Hay G.J. (eds.), *Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications,* Springer, Berlin, pp. 91–110.

Chan K.M.A., Shaw R.M., Cameron D.R., Underwood E.C. & Daily G.C. 2006. Conservation planning for ecosystem services. *PLoS Biology* 4 (11): e379.

Chan K.M.A., Satterfield T. & Goldstein J. 2012. Rethinking ecosystem services to better address and navigate cultural values. *Ecol Econ* 74: 8–18.

Cherrill A. & McClean C. 1995. An investigation of uncertainty in field habitat mapping and the implications for detecting land cover change. *Landscape Ecol* 10: 5–21.

Cherrill A. & McClean C. 1999. The reliability of 'Phase 1' habitat mapping in the UK: the extent and types of observer bias. *Landscape Urban Plan* 45: 131–143.

Clinton N., Holt A., Scarborough J., Yan L. & Gong P. 2010. Accuracy assessment measures for object-based image segmentation goodness. *Photogramm Eng Rem S* 76: 289–299.

Coburn C.A. & Roberts A.C.B. 2004. A multiscale texture analysis procedure for improved forest stand classification. *Int J Remote Sens* 25: 4287–4308.

Connor E.F. & McCoy E.D. 1979. The statistics and biology of the species area relationship. *Am Nat* 113: 791–833.

Corcoran P., Winstanley A. & Mooney P. 2010. Segmentation performance evaluation for object-based remotely sensed image analysis. *Int J Remote Sens* 31: 617–645.

Corcoran J., Knight J., Brisco B., Kaya S., Cull A. & Murnaghan K. 2011. The integration of optical, topographic, and radar data for wetland mapping in northern Minnesota. *Can J Remote Sens* 37: 564–582.

Corsi F., de Leeuw J. & Skidmore A. 2000. Modeling species distribution with GIS. In: Boitani, L. & Fuller, T.K. (eds.), *Research techniques in animal ecology: controversies and consequences,* Columbia University Press, New York, pp. 389–434.

Costanza R., d'Arge R., de Groot R., Farber S., Grasso M., Hannon B., Limburg K., Naeem S., O'Neill R.V., Paruelo J., Raskin R.G., Sutton P. & van den Belt M. 1997. The value of the world's ecosystem services and natural capital. *Nature* 387: 253–260.

Crossman N.D., Burkhard B., Nedkov S., Willemen L., Petz K., Palomo I., Drakou E.G., Martín-Lopez B., McPhearson T., Boyanova K., Alkemade R., Egoh B., Dunbar M.B. & Maes J. 2013. A blueprint for mapping and modelling ecosystem services. *Ecosystem Services* 4: 4–14.

Daubechies I. 1992. *Ten lectures on wavelets (CBMS-NSF regional conference series in applied mathematics 61)*. Society for Industrial and Applied Mathematics, Philadelphia.

De Groot R., Fisher B., Christie M., Aronson J., Braat L., Gowdy J., Haines-Young R., Maltby E., Neuville A., Polasky S., Portela R. & Ring I. 2010. Chapter 1: integrating the ecological and economic dimensions in biodiversity and ecosystem service valuation. In: Kumar P. (ed.), *The economics of ecosystems and biodiversity: ecological and economic foundations*, Earthscan, London, pp. 9–40.

Díaz Varela R.A., Rego P.R., Iglesias S.C. & Sobrino C.M. 2008. Automatic habitat classification methods based on satellite images: a practical assessment in the NW Iberia coastal mountains. *Environ Monit Assess* 144: 229–250.

Dingle Robertson L. & King D.J. 2011. Comparison of pixel- and object-based classification in land cover change mapping. *Int J Remote Sens* 32: 1505–1529.

Duelli P. & Obrist M.K.. 2003. Biodiversity indicators: the choice of values and measures. *Agr Ecosyst Environ* 98: 87–98.

Duro D.C., Franklin S.E. & Dubé M.G. 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens Environ* 118: 259–272.

Egoh B.N., Reyers B., Rouget M. & Richardson D.M. 2011. Identifying priority areas for ecosystem service management in South African grasslands. *J Environ Manage* 92: 1642–1650.

Egoh B., Drakou E.G., Dunbar M.B., Maes J. & Willemen L. 2012. *Indicators for mapping ecosystem services: a review*. JRC Scientific and Policy Reports, EUR 25456 EN. Publications Office of the European Union, Luxembourg.

Elith J. & Leathwick J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Evol S* 40: 677–697.

Espindola G.M., Camara G., Reis I.A., Bins L.S. & Monteiro A.M. 2006. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *Int J Remote Sens* 27: 3035–3040.

Eurola S., Huttunen A. & Kukko-Oja K. 1995. *Suokasvillisuusopas,* 2nd ed. Oulanka Reports 14. Oulanka Biological Station, University of Oulu, Oulu.

Evans K.L., Warren P.H. & Gaston K.J. 2005. Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biol Rev* 80: 1–25.

Ferrier S. & Guisan A. 2006. Spatial modelling of biodiversity at the community level. *J Appl Ecol* 43: 393–404.

Ferrier S., Watson G., Pearce J. & Drielsma M. 2002a. Extended statistical approaches to modeling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodivers Conserv* 11: 2275–2307.

Ferrier S., Drielsma M., Manion G. & Watson G. 2002b. Extended statistical approaches to modeling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodivers Conserv* 11: 2309–2338.

Fraser R.H. & Currie D.J. 1996. The species richness-energy hypothesis in a system where historical factors are thought to prevail: coral reefs. *Amer Nat* 148: 138–159.

Fry G., Tress B. & Tress G. 2007. Integrative landscape research: facts and challenges. In: Wu J. & Hobbs R.J. (eds.), *Key topics in landscape ecology,* Cambridge University Press, Cambridge, UK, pp. 246–268.

Gallant J.C. & Dowling T.I. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour Res* 39: 1347–1360

Gao Y., Mas J.F., Kerle N. & Navarrete Pacheco J.A. 2011. Optimal region growing segmentation and its effect on classification accuracy. *Int J Remote Sens* 32: 3747–3763.

Geerling G.W., Labrador-Garcia M., Clevers J.G.P.W., Ragas A.M.J. & Smits A.J.M. 2007. Classification of floodplain vegetation by data fusion of spectral (CASI) and LiDAR data. *Int J Remote Sens* 28: 4263–4284.

Geneletti D. 2003. Biodiversity impact assessment of roads: an approach based on ecosystem rarity. *Environ Impact Asses* 23: 343–365.

Geneletti D. 2007. An approach based on spatial multicriteria analysis to map the nature conservation value of agricultural land. *J Environ Manage* 83: 228–235.

Gillespie T.W., Foody G.M., Rocchini D., Giorgi A.P. & Saatchi S. 2008. Measuring and modelling biodiversity from space. *Prog Phys Geog* 32: 203–221.

Gonzales R.C. & Woods R.E. 2002. *Digital Image Processing*, 2nd ed. Prentice Hall, Upper Saddle River, NJ, USA.

Gray M. 2013. *Geodiversity: Valuing and conserving abiotic nature*, 2nd ed. John Wiley & Sons, Chichester.

Groves C.R., Jensen D.B., Valutis L.L., Redford K.H., Shaffer M.L., Scott J.M., Baumgartner J.V., Higgins J.W., Beck M.W. & Anderson M.G. 2002. Planning for biodiversity conservation: putting conservation science into practice. *Bioscience* 52: 499–512.

Guisan A. & Theurillat J. 2000. Equilibrium modeling of alpine plant distribution: How far can we go? *Phytocoenologia* 30: 353–384.

Guisan A. & Zimmermann N.E. 2000. Predictive habitat distribution models in ecology. *Ecol Model* 135: 147–186.

Guisan A., Weiss S.B. & Weiss A.D. 1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol* 143: 107–122.

Guisan A., Lehmann A., Ferrier S., Austin M., Overton J.M., Aspinall R. & Hastie T. 2006. Making better biogeographical predictions of species' distributions. *J Appl Ecol* 43: 386–392.

Hämet-Ahti L., Suominen J., Ulvinen T., Uotila P. & Vuokko S. (eds.). 1986. *Retkeilykasvio*. Suomen Luonnonsuojelun Tuki, Helsinki.

Han N., Wang K., Yu L. & Zhang X. 2012. Integration of texture and landscape features into object-based classification for delineating Torreya using IKONOS imagery. *Int J Remote Sens* 33: 2003–2033.

Haralick R.M. 1979. Statistical and structural approaches to texture. *P IEEE* 67: 786–804.

Haralick R.M., Shanmugam K. & Dinstein I. 1973. Textural features for image classification. *IEEE T Sys Man Cyb* 3: 610–621.

Hardin G. 1968. The tragedy of the commons. *Science* 162: 1243–1248.

Hastie T. & Tibshirani R. 1986. Generalized additive models. *Stat Sci* 1: 297–318.

Hay G.J. & Castilla G. 2008. Geographic object-based image analysis (GEOBIA): a new name for a new discipline. In: Blaschke T., Lang S. & Hay G.J. (eds.), *Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications,* Springer, Berlin, pp. 75–89.

Hay G.J., Blaschke T., Marceau D.J. & Bouchard A. 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS J Photogramm* 57: 327–345.

He K.S., Rocchini D., Neteler M. & Nagendra H. 2011. Benefits of hyperspectral remote sensing for tracking plant invasions. *Divers Distrib* 17: 381–392.

Heikkinen R.K. & Neuvonen S. 1997. Species richness of vascular plants in the subarctic landscape of northern Finland: modelling relationships to the environment. *Biodivers Conserv* 6: 1181–1201.

Heikkinen R.K., Luoto M., Virkkala R. & Rainio K. 2004. Effects of habitat cover, landscape structure and spatial variables on the abundance of birds in an agricultural–forest mosaic. *J Appl Ecol* 41: 824–835.

Hilli M. & Kuitunen M. 2005. Testing the use of a land cover map for habitat ranking in Boreal Forests. *Environ Manage* 35: 505–516.

Hjort J., Heikkinen R.K. & Luoto M. 2012. Inclusion of explicit measures of geodiversity improve biodiversity models in a boreal landscape. *Biodivers Conserv* 21: 3487–3506.

Honkanen M., Roberge J., Rajasärkkä A. & Mönkkönen M. 2010. Disentangling the effects of area, energy and habitat heterogeneity on boreal forest bird species richness in protected areas. *Global Ecol Biogeogr* 19: 61–71.

Honnay O., Piessens K., Van Landuyt W., Hermy M. & Gulinck H. 2003. Satellite based land use and landscape complexity indices as predictors for regional plant diversity. *Landscape Urban Plan* 63: 241–250.

Hou Z., Xu Q., Nuutinen T. & Tokola T. 2013. Extraction of remote sensing-based forest management units in tropical forests. *Remote Sens Environ* 130: 1–10.

Jačková K. & Romportl D. 2008. The relationship between geodiversity and habitat richness in Šumava National Park and Křivoklátsko Pla (Czech Republic): a quantitative analysis approach. *J Landscape Ecol* 1: 23–38.

Johansen K., Coops N.C., Gergel S.E. & Stange Y. 2007. Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sens Environ* 110: 29–44.

Justus J., Colyvan M., Regan H. & Maguire L. 2009. Buying into conservation: intrinsic versus instrumental value. *Trends Ecol Evol* 24: 187–191.

Kalamandeen M. & Gillson L. 2007. Demything "wilderness": implications for protected area designation and management. *Biodivers Conserv* 16: 165–182.

Karr J.R. 1990. Biological integrity and the goal of environmental legislation: lessons for conservation biology. *Conserv Biol* 4: 244–250.

Ke Y., Quackenbush L.J. & Im J. 2010. Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification. *Remote Sens Environ* 114: 1141–1154.

Kerr, J.T. & Ostrovsky M. 2003. From space to species: ecological applications for remote sensing. *Trends Ecol Evol* 18: 299–305.

Kim M., Madden M. & Warner T.A. 2008. Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. In: Blaschke T., Lang S. & Hay G.J. (eds.), *Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications,* Springer, Berlin, pp. 291–307.

Kim M., Madden M. & Warner T.A. 2009. Forest type mapping using object-specific texture measures from multispectral Ikonos imagery: segmentation quality and image classification issues. *Photogramm Eng Rem S* 75: 819–829.

Kim M., Warner T.A., Madden M. & Atkinson D.S. 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *Int J Remote Sens* 32: 2825–2850.

Kukkala A.S. & Moilanen A. 2013. Core concepts of spatial prioritisation in systematic conservation planning. *Biol Rev* 88: 443–464.

Kursa M.B. & Rudnicki W.R. 2010. Feature selection with the Boruta Package. *J Stat Softw* 36: 1–13.

Lang S., Albrecht F., Kienberger S. & Tiede D. 2010. Object validity for operational tasks in a policy context. *J Spat Sci* 55: 9–22.

Lefsky M.A., Cohen W.B., Parker G.G. & Harding D.J. 2002. Lidar remote sensing for ecosystem studies. *Bioscience* 52: 19–30.

Lehtomäki J., Tomppo E., Kuokkanen P., Hanski I. & Moilanen A. 2009. Applying spatial conservation prioritization software and high-resolution GIS data to a national-scale study in forest conservation. *For Ecol Manage* 258: 2439–2449.

Levin S.A. 1992. The problem of pattern and scale in ecology: The Robert H. MacArthur Award lecture. *Ecology* 73: 1943–1967.

Li Z., Hayward R., Walker R. & Liu Y. 2011. A biologically inspired object spectral-texture descriptor and its application to vegetation classification in power-line corridors. *IEEE Geosci Remote S* 8: 631–635.

Liaw A. & Wiener M. 2002. Classification and regression by randomForest. *R News* 2: 18–22.

Liski J. & Westman C.J. 1997. Carbon storage in forest soil of Finland: 2. Size and regional patterns. *Biogeochemistry* 36: 261–274.

Lobo J.M., Castro I. & Moreno J.C. 2001. Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. *Biol J Linn Soc* 73: 233–253.

Lucieer A. & Stein A. 2002. Existential uncertainty of spatial objects segmented from satellite sensor imagery. *IEEE T Geosci Remote* 40: 2518–2521.

Mace G.M., Norris K. & Fitter A.H. 2012. Biodiversity and ecosystem services: a multi-layered relationship. *Trends Ecol Evol* 27: 19–26.

Machado A. 2004. An index of naturalness. *J Nat Conserv* 12: 95–110.

Maes J., Egoh B., Willemen L., Liquete C., Vihervaara P., Schägner J.P., Grizzetti B., Drakou E.G., La Notte A., Zulian G., Bouraoui F., Paracchini M.L., Braat L. & Bidoglio G. 2012. Mapping ecosystem services for policy support and decision making in the European Union. *Ecosystem Services* 1: 31–39.

Margules C.R. & Pressey R.L. 2000. Systematic conservation planning. *Nature* 405: 243–253.

Margules C.R. & Usher M.B. 1981. Criteria used in assessing wildlife conservation potential: a review. *Biol Conserv* 21: 79–109.

Marpu P.R., Neubert M., Herold H. & Niemeyer I. 2010. Enhanced evaluation of image segmentation results. *J Spat Sci* 55: 55–68.

Marra G. & Wood S. 2011. Practical variable selection for generalized additive models. *Comput Stat Data An* 55: 2372–2387.

Mas J.-F., Gao Y. & Navarrete Pacheco J.A. 2010. Sensitivity of landscape pattern metrics to classification approaches. *For Ecol Manage* 259: 1215-1224.

Maxa M. & Bolstad P. 2009. Mapping northern wetlands with high resolution satellite images and LiDAR. *Wetlands* 29: 248–260.

McDermid G.J., Franklin S.E. & LeDrew E.F. 2005. Remote sensing for large-area habitat mapping. *Prog Phys Geog* 29: 449–474.

McGarigal K. & Ene E. 2012. *FRAGSTATS v4.1: spatial pattern analysis program for categorical and continuous maps*. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at http://www.umass.edu/landeco/research/fragstats/fragstats.html (accessed 23.12.2013).

McGarigal K., Tagil S. & Cushman S.A. 2009. Surface metrics: an alternative to patch metrics for the quantification of landscape structures. *Landscape Ecol* 24: 433–450.

McGarigal K., Compton B.W., Jackson S.D., Plunkett E., Rolih K., Portante T. & Ene E. 2011. *Conservation assessment and prioritization system (CAPS): statewide Massachusetts assessment: november 2011,* available at http://www.masscaps.org/pdf/CAPS2011MassachusettsAssessment.pdf.

Moilanen A., Franco A.M.A., Early R.I., Fox R., Wintle B. & Thomas C.D. 2005. Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *P Roy Soc B-Biol Sci* 272: 1885–1891.

Moilanen A., Wilson K.A. & Possingham H.P. 2009. *Spatial conservation prioritization: quantitative methods and computational tools*. Oxford University Press, Oxford.

Moilanen A., Meller L., Leppänen J., Montezino Pouzols F., Arponen A. & Kujala H. 2012. *Spatial conservation planning framework and software Zonation*. Version 3.1. User manual. University of Helsinki, Department of Biosciences, Helsinki. Available at http://cbig.it.helsinki.fi/files/zonation/ZONATION_v3.1_Manual_120416.pdf (accessed 23.12.2013).

Möller M., Lymburner L. & Volk M. 2007. The comparison index: a tool for assessing the accuracy of image segmentation. *Int J Appl Earth Obs* 9: 311–321.

Morgan J.L. & Gergel S.E. 2010. Quantifying historic landscape heterogeneity from aerial photographs using object-based analysis. *Landscape Ecol* 25: 985–998.

Morgan J.L., Gergel S.E. & Coops N.C. 2010. Aerial photography: a rapidly evolving tool for ecological management. *Bioscience* 60: 47–59.

Murphy P.N.C., Ogilvie J., Connor K. & Arp P. 2007. Mapping wetlands: a comparison of different approaches for New Brunswick, Canada. *Wetlands* 27: 846–854.

Murray H., Lucieer A. & Williams R. 2010. Texture-based classification of sub-Antarctic vegetation communities on Heard Island. *Int J Appl Earth Obs* 12: 138–149.

Mustonen J., Packalén P. & Kangas A. 2008. Automatic segmentation of forest stands using a canopy height model and aerial photography. *Scand J Forest Res* 23: 534–545.

Nagendra H. 2001. Using remote sensing to assess biodiversity. *Int J Remote Sens* 22: 2377–2400.

Nagendra H., Lucas R., Pradinho Honrado J., Jongman R.H.G., Tarantino C., Adamo M. & Mairota P. 2013. Remote sensing for conservation monitoring: assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecol Indic* 33: 45–59.

Nelder J.A. & Wedderburn R.W.M. 1972. Generalized linear models. *J Roy Stat Soc A* 135: 370–384.

Nelson E., Mendoza G., Regetz J., Polasky S., Tallis H., Cameron D.R., Chan K.M.A., Daily G.C., Goldstein J., Kareiva P.M., Lonsdorf E., Naidoo R., Ricketts T.H. & Shaw M.R. 2009. Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Front Ecol Environ* 7: 4–11.

Neubert M. & Meinel G. 2003. Evaluation of segmentation programs for high resolution remote sensing applications. In: Schroeder M., Jacobsen K. & Heipke C. (eds.), *Proceedings Joint ISPRS/EARSeL Workshop "High Resolution Mapping from Space 2003*, Hannover, October 6–8. Available at http://www.ipi.uni-hannover.de/fileadmin/institut/pdf/neubert.pdf (accessed 10.1.2014).

Newton A.C., Hill R.A., Etcheverría C., Golicher D., Rey Benayas J.M., Cayuela L. & Hinsley S.A. 2009. Remote sensing and the future of landscape ecology. *Prog Phys Geog* 33: 528–546.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H. & Wagner H. 2013. *vegan: community ecology package*. R package version 2.0-7. Available at http://CRAN.R-project.org/package=vegan (accessed 23.12.2013).

Orme C.D.L., Davies R.G., Burgess M., Eigenbrod F., Pickup N., Olson V.A., Webster A.J., Ding T., Rasmussen P.C., Ridgely R.S., Stattersfield A.J., Bennett P.M., Blackburn T.M., Gaston K.J. & Owens I.P.F. 2005. Global hotspots of species richness are not congruent with endemism or threat. *Nature* 436: 1016–1019.

Ouma Y.O., Tetuko J. & Takeishi R. 2008. Analysis of co-occurrence and discrete wavelet transform textures for differentiation of forest and non-forest vegetation in very-high-resolution optical-sensor imagery. *Int J Remote Sens* 29: 3417–3456.

Ozesmi S.L. & Bauer M.E. 2002. Satellite remote sensing of wetlands. *Wetl Ecol Manag* 10: 381–402.

Parks K.E. & Mulligan M. 2010. On the relationship between a resource based measure of geodiversity and broad scale biodiversity patterns. *Biodivers Conserv* 19: 2751–2766.

Parviainen M., Luoto M. & Heikkinen R.K. 2009. The role of local and landscape level measures of greenness in modelling boreal plant species richness. *Ecol Model* 220: 2690–2701.

Pau G., Fuchs F., Sklyar O., Boutros M. & Huber W. 2010. EBImage – an R package for image processing with applications to cellular phenotypes. *Bioinformatics* 26: 979–981.

Pausas J.G., Carreras J., Ferré A. & Font X. 2003. Coarse-scale plant species richness in relation to environmental heterogeneity. *J Veg Sci* 14: 661–668.

Pekkarinen A. 2002. Image segment-based spectral features in the estimation of timber volume. *Remote Sens Environ* 82: 349–359.

Pontius R.G. & Millones M. 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int J Remote Sen* 32: 4407–4429.

Pressey R.L. 2004. Conservation planning and biodiversity: assembling the best data for the job. *Conserv Biol* 6: 1677–1681.

R Core Team. 2012. *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0. Available at http://www.R-project.org/ (accessed 23.12.2013).

Raes N., Roos M.C., Slik J.W.F., van Loon E.E. & ter Steege H. 2009. Botanical richness and endemicity patterns of Borneo derived from species distribution models. *Ecography* 32: 180–192.

Rassi P., Kaipainen H., Mannerkoski I. & Ståhls G. (eds.) 1992. *Uhanalaisten eläinten ja kasvien seurantatoimikunnan mietintö*. Ympäristöministeriö, Helsinki.

Rassi P., Hyvärinen E., Juslén A. & Mannerkoski I. (eds.) 2010. *The 2010 Red List of Finnish species*. Ympäristöministeriö & Suomen ympäristökeskus, Helsinki.

Raudsepp-Hearne C., Peterson G.D. & Bennett E.M. 2010. Ecosystem service bundles for analyzing tradeoffs in diverse landscapes. *Proc Natl Acad Sci USA* 107: 5242–5247.

Riley S.J., DeGloria S.D. & Elliot R. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain J Sci* 5: 23–27.

Rocchini D., Balkenhol N., Carter G.A., Foody G.M., Gillespie T.W., He K.S., Kark S., Levin N., Lucas K., Luoto M., Nagendra H., Oldeland J., Ricotta

C., Southwort J. & Neteler M.. 2010. Remotely sensed spectral homogeneity as a proxy of species diversity: recent advances and open challenges. *Ecol Inform* 5: 318–329.

Rocchini D., McGlinn D., Ricotta C., Neteler M. & Wohlgemuth T. 2011a. Landscape complexity and spatial scale influence the relationship between remotely sensed spectral diversity and survey-based plant species richness. *J Veg Sci* 22: 688–698.

Rocchini D., Hortal J., Lengyel S., Lobo J.M., Jiménez-Valverde A., Ricotta C., Bacaro G. & Chiarucci A. 2011b. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Prog Phys Geog* 35: 211–226.

Rockström J., Steffen, W., Noone K., Persson Å., Chapin F.S., Lambin E.F., Lenton T.M., Scheffer M., Folke C., Schellnhuber H.J., Nykvist B., de Wit C.A., Hughes T., van der Leeuw S., Rodhe H., Sörlin S., Snyder P.K., Costanza R., Svedin U., Falkenmark M., Karlberg L., Corell R.W., Fabry V.J., Hansen J., Walker B., Liverman D., Richardson K., Crutzen P. & Foley J.A. 2009. A safe operating space for humanity. *Nature* 461: 472–475.

Rossi E. & Kuitunen M. 1996. Ranking of habitats for the assessment of ecological impact in land use planning. *Biol Conserv* 77: 227–234.

Ruddock K., August P.V., Damon C., LaBash C., Rubinoff P. & Robadue D. 2013. Conservation in the context of climate change: practical guidelines for land protection at local scales. *PLoS ONE* 8 (11): e80874.

Ruiz L.A., Fernandez-Sarría A. & Recio J.A. 2004. Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (Part B4), 1109–1115.

Ryttäri T., Kalliovirta M. & Lampinen R. (eds.) 2012. *Suomen uhanalaiset kasvit*. Tammi, Helsinki.

SAGA. 2011. *SAGA: System for Geoscientific Analyses*. Available at http://www.saga-gis.org/ (accessed 19.12.2013).

Sammon J.W. 1969. A nonlinear mapping for data structure analysis. *IEEE T Comput* C-18: 401–409.

Samu F., Csontos P. & Szinetár C. 2008. From multi-criteria approach to simple protocol: Assessing habitat patches for conservation value using species rarity. *Biol Conserv* 141: 1310–1320.

Sasaki T., Imanishi J., Ioki K., Morimoto Y. & Kitada K. 2012. Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data. *Landsc Ecol Eng* 8: 157–171.

Smith A. 2010. Image segmentation scale parameter optimization and land cover classification using the random forest algorithm. *J Spat Sci* 55: 69–79.

Smith P.G.R. & Theberge J.B. 1986. A review of criteria for evaluating natural areas. *Environ Manage* 10: 715–734.

Söderman T. 2012. *Biodiversity and ecosystem services in impact assessment – from components to services*. Diss. University of Helsinki, Department of Geosciences and Geography, Helsinki.

Spanhove T., Vanden Borre J., Delalieux S., Haest B. & Paelinckx D. 2012. Can remote sensing estimate fine-scale quality indicators of natural habitats? *Ecol Indic* 18: 403–412.

Spellerberg I.F. 1992. *Evaluation and assessment for conservation*. Chapman & Hall, London.

Su W., Zhang C., Yang J., Wu H., Deng L., Ou W., Yue A. & Chen M. 2012. Analysis of wavelet packet and statistical textures for object-oriented classification of forest-agriculture ecotones using SPOT 5 imagery. *Int J Remote Sens* 33: 3557–3579.

Tarboton D.G. 2012. *Terrain analysis using digital elevation models (TauDEM) 5.0*. Available at http://hydrology.usu.edu/taudem/taudem5/index.html (accessed 23.12.2013).

Thompson S.D. & Gergel S.E. 2008. Conservation implications of mapping rare ecosystems using high spatial resolution imagery: recommendations for heterogeneous and fragmented landscapes. *Landscape Ecol* 23: 1023–1037.

Thompson S.D., Gergel S.E. & Coops N. C. 2008. Classification of late seral coastal temperate rainforests with high spatial resolution QuickBird imagery. *Can J Remote Sens* 34: 460–470.

Thuiller W., Araújo M.B & Lavorel S. 2004. Do we need land-cover data to model species distributions in Europe? *J Biogeogr* 31: 353–361.

Tian J. & Chen D.-M. 2007. Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. *Int J Remote Sens* 28: 4625–4644.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J Roy Stat Soc B* 58: 267–288.

Tischendorf L. & Fahrig L. 2000. On the usage and measurement of landscape connectivity. *Oikos* 90: 7–19.

Titeux N., Maes D., Marmion M., Luoto M. & Heikkinen R.K. 2009. Inclusion of soil data improves the performance of bioclimatic envelope models for insect species distributions in temperate Europe. *J Biogeogr* 36: 1459–1473.

Tomppo E. & Halme M. 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. *Remote Sens Environ* 92: 1–20.

Torgerson W.S. 1958. *Theory & methods of scaling*. Wiley, New York.

Troy A. & Wilson M.A. 2006. Mapping ecosystem services: practical challenges and opportunities in linking GIS and value transfer. *Ecol Econ* 60: 435–449.

Tuomisto H. 2010. A consistent terminology for quantifying species diversity? Yes, it does exist. *Oecol* 164: 853–860.

Turner M.G., Gardner R.H. & O'Neill R.V. 2001. *Landscape ecology in theory and practice: pattern and process.* Springer, New York.

Turner W., Spector S., Gardiner N., Fladeland M., Sterling E. & Steiniger M. 2003. Remote sensing for biodiversity science and conservation. *Trends Ecol Evol* 18: 306–314.

Turunen J., Tomppo E., Tolonen K. & Reinikainen A. 2002. Estimating carbon accumulation rates of undrained mires in Finland–application to boreal and subarctic regions. *Holocene* 12: 69–80.

Ulbricht J. 2012. *lqa: Penalized Likelihood Inference for GLMs*. R package version 1.0-3. Available at http://CRAN.R-project.org/package=lqa (accessed 23.12.2013).

Uuemaa E., Mander Ü. & Marja R. 2013 Trends in the use of landscape spatial metrics as landscape indicators: a review. *Ecol Indic* 28: 100–106.

Vanden Borre J., Paelinckx D., Mücher C.A., Kooistra L., Haest B., De Blust G. & Schmidt A.M. 2011. Integrating remote sensing in Natura 2000 habitat monitoring: prospects on the way forward. *J Nat Conserv* 19: 116–125.

Vierling K.T., Vierling L.A., Gould W.A., Martinuzzi S. & Clawges R.M. 2008. Lidar: shedding new light on habitat characterization and modeling. *Front Ecol Environ* 6: 90–98.

Vihervaara P., Kumpula T., Tanskanen A. & Burkhard B. 2010. Ecosystem services – a tool for sustainable management of human–environment systems. Case study Finnish Forest Lapland. *Ecol Complex* 7: 410–420.

Villarroya A. & Puig J. 2012. Valuation of residual impacts of roads on landscape ecological units in Navarre, Spain. *Journal of Environ Plann Man* 55: 339–353.

Walsh C. & Mac Nally R. 2013. *hier.part: hierarchical partitioning*. R package version 1.0-4. Available at http://CRAN.R-project.org/package=hier.part (accessed 23.12.2013).

Wang Z., Jensen J.R. & Im J. 2010. An automatic region-based image segmentation algorithm for remote sensing applications. *Environ Modell Softw* 25: 1149–1165.

Wang Z., Boesch R. & Ginzler C. 2012. Forest delineation of aerial images with Gabor wavelets. *Int J Remote Sens* 33: 2196–2213.

Wania A., Kühn I. & Klotz S. 2006. Plant richness patterns in agricultural and urban landscapes in Central Germany – spatial gradients of species richness. *Landscape Urban Plan* 75: 97–110.

Weidner U. 2008. Contribution to the assessment of segmentation quality for remote sensing applications. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 Part B7: 479–484.

Whitcher B. 2012. *waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing*. R package version 1.7.1. Available at http://CRAN.R-project.org/package=waveslim (accessed 23.12.2013).

Whiteside T.G., Boggs G.S. & Maier, S.W. 2011. Comparing object-based and pixel-based classifications for mapping savannas. *Int J Appl Earth Obs* 13: 884–893.

Whittaker R.H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol Monograph* 30: 279–338.

Whittaker R.J., Willis K.J. & Field R. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. *J Biogeogr* 28: 453–470.

Willemen L., Verburg P.H., Hein L. & van Mensvoort M.E.F. 2008. Spatial characterization of landscape functions. *Landscape Urban Plan* 88: 34–43.

Williams P., Gibbons D., Margules C., Rebelo A. & Humphries C. 1996. A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds. *Conserv Biol* 10: 155–174.

Wilson K.A., McBride M.F., Bode M. & Possingham H.P. 2006. Prioritizing global conservation efforts. *Nature* 440: 337–340.

Wilson K.A., Cabeza M. & Klein C.J. 2009. Chapter 2: fundamental concepts of spatial conservation prioritization. In: Moilanen A., Wilson K.A. & Possingham H.P. (eds.), *Spatial conservation prioritization: quantitative methods and computational tools*, Oxford University Press, Oxford, pp. 16–27.

Wood S.N. 2006. *Generalized additive models: an introduction with R*. CRC Press, Boca Raton, FL. USA.

Wood S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J Roy Stat Soc B* 73: 3–36.

Wright C. & Gallant A. 2007. Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. *Remote Sens Environ* 107: 582–605.

Wulder M.A., White J.C., Hay G.J. & Castilla G. 2008. Towards automated segmentation of forest inventory polygons on high spatial resolution satellite imagery. *Forest Chron* 84: 221–230.

Yan G., Mas J.-F., Maathuis B.H.P., Xiangmin Z. & van Dijk P.M., 2006. Comparison of pixel-based and object-oriented image classification approaches – a case study in a coal fire area, Wuda, Inner Mongolia, China. *Int J Remote Sens* 27: 4039–4055.

Yu Q., Gong P., Clinton N., Biging G., Kelly M. & Schirokauer D. 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogramm Eng Rem S* 72: 799–811.

Yue A., Yang J., Zhang C., Su W., Yun W., Zhu D., Liu S. & Wang Z. 2012. The optimal segmentation scale identification using multispectral WorldView-2 images. *Sens Lett* 10: 285–291.

Zhan Q., Molenaar M., Tempfli K. & Shi W. 2005. Quality assessment for geo-spatial objects derived from remotely sensed data. *Int J Remote Sens* 26: 2593–2974.

Zhang H., Fritts J.E. & Goldman S.A. 2008. Image segmentation evaluation: a survey of unsupervised methods. *Comput Vis Image Und* 110: 260–280.

Zhang Y.J. 1997. Evaluation and comparison of different segmentation algorithms. *Pattern Recogn Lett* 18: 963–974.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. 2009. *Mixed effects models and extensions in ecology with R*. Springer, New York.

ORIGINAL PAPERS

I

WHAT MAKES SEGMENTATION GOOD? A CASE STUDY IN BOREAL FORES HABITAT MAPPING

by

Aleksi Räsänen, Antti Rusanen, Markku Kuitunen & Anssi Lensu 2013

International Journal of Remote Sensing 34: 8603–8627.

International Journal of Remote Sensing, 2013
Vol. 34, No. 23, 8603–8627, http://dx.doi.org/10.1080/01431161.2013.845318

What makes segmentation good? A case study in boreal forest habitat mapping

Aleksi Räsänen*, Antti Rusanen, Markku Kuitunen, and Anssi Lensu

Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

(*Received 30 April 2013; accepted 27 August 2013*)

Segmentation goodness evaluation is a set of approaches meant for deciding which segmentation is good. In this study, we tested different supervised segmentation evaluation measures and visual interpretation in the case of boreal forest habitat mapping in Southern Finland. The data used were WorldView-2 satellite imagery, a lidar digital elevation model (DEM), and a canopy height model (CHM) in 2 m resolution. The segmentation methods tested were the fractal net evolution approach (FNEA) and IDRISI watershed segmentation. Overall, 252 different segmentation methods, layers, and parameter combinations were tested. We also used eight different habitat delineations as reference polygons against which 252 different segmentations were tested. The ranking order of segmentations depended on the chosen supervised evaluation measure; hence, no single segmentation could be ranked as the best. In visual interpretation among the several different segmentations that we found rather good, we selected only one as the best. In the literature, it has been noted that better segmentation leads to higher classification accuracy. We tested this argument by classifying 12 of our segmentations with the random forest classifier. It was found out that there is no straightforward answer to the argument, since the definition of good segmentation is inconsistent. The highest classification accuracy (0.72) was obtained with segmentation that was regarded as one of the best in visual interpretation. However, almost similarly high classification accuracies were obtained with other segmentations. We conclude that one has to decide what one wants from segmentation and use segmentation evaluation measures with care.

1. Introduction

Since the early 2000s, with the rise of object-based image analysis (OBIA) methodology (Blaschke 2010), segmentation goodness evaluation has been an emerging topic within the remote-sensing literature (Clinton et al. 2010; Marpu et al. 2010). Evaluation has been concentrated on segmentation method development and comparison as well as parameter optimization.

Generally in segmentation, the objective is to partition imagery into regions that are meaningful and thus either mimic real-world objects (Zhang, Fritts, and Goldman 2008; Clinton et al. 2010) or minimize intra-segment and maximize inter-segment heterogeneity (Zhang, Fritts, and Goldman 2008; Hou et al. 2013). Remote-sensing segmentation methods are a special case of more general image segmentation methods, which can be divided into two complementary groups: similarity or region-based segmentation and discontinuity-based segmentation. In region-based segmentation, a similarity measure is

*Corresponding author. Email: t.aleksi.rasanen@jyu.fi

used to determine suitable regions. In discontinuity-based segmentation, discontinuities of the images, usually boundaries, are detected (Zhang 1997; Gonzales and Woods 2002). Some methods combine concepts from both groups. For instance, in watershed segmentation, dividing lines between basin areas are sought by flooding the image (Gonzales and Woods 2002).

Within remote sensing, several segmentation implementations have been region based. Arguably, the most widely used remote-sensing segmentation method has been the fractal net evolution approach (FNEA) developed by Baatz and Schäpe (2000), and implemented in eCognition software (Trimble, Sunnyvale, CA, USA). FNEA has been used as a benchmark segmentation against which other methods have been compared. Although some authors have claimed to have developed better methods (Derivaux et al. 2010; Li et al. 2010; Li, Huo, and Fang 2010; Wang, Jensen, and Im 2010), FNEA has been a good performer in method comparisons (Neubert and Meinel 2003; Meinel and Neubert 2004; Carleer, Debeir, and Wolff 2005; Neubert, Herold, and Meinel 2008; Marpu et al. 2010) and seems to still be the standard method (e.g. Bar Massada et al. 2012; Duro, Franklin, and Dubé 2012). In addition, many other standard remote-sensing analysis products such as ENVI (Exelis, McLean, VA, USA), ERDAS Imagine (Intergraph, Huntsville, AL, USA), and IDRISI Selva (Clark Labs, Worcester, MA, USA) have included segmentation methods in their newer versions. Yet, there are also numerous other methods, algorithms, and software applications for segmenting remotely sensed data. To analyse the goodness of these methods, segmentation evaluation has been performed (Zhang 1996; Clinton et al. 2010; Marpu et al. 2010).

Segmentation evaluation can be divided into two major categories: subjective (visual) evaluation and objective evaluation. Objective evaluation can be further divided into system-level, which evaluates the overall system in which segmentation is performed, and direct evaluation (Zhang, Fritts, and Goldman 2008). As an example, final classification output can be regarded as a system when assessing segmentation quality (Smith 2010; Wang, Jensen, and Im 2010; Gao et al. 2011). Direct evaluation can be either analytical or empirical, of which the former evaluates the method itself and the latter its results. Empirical methods consist of supervised and unsupervised methods, i.e. if ground truth is used as a reference or not (Zhang, Fritts, and Goldman 2008).

In remote sensing, both analytical methods (Hay et al. 2003) and unsupervised methods (Espindola et al. 2006; Corcoran, Winstanley, and Mooney 2010; Drăguţ, Tiede, and Levick 2010; Yue et al. 2012; Hou et al. 2013) have been used in segmentation evaluation. For instance, Corcoran, Winstanley, and Mooney (2010) evaluated segmentation goodness by measuring the contrast between segments that share a boundary. However, evaluation has largely been performed using supervised methods, more specifically either area-based or location-based measures. From these two, area-based measures evaluate whether either segmentation is too coarse (under-segmentation) or too fine (over-segmentation). Over- and under-segmentation measures can also be combined. Location-based measures, on the other hand, are based on distances between segment centroids and reference polygon centroids or distances between boundary pixels. For a good review of these measures and an evaluation of different measures, see Clinton et al. (2010). These goodness measures are applicable especially when mapping clearly bordered urban features (Tian and Chen 2007; Zhan et al. 2005; Weidner 2008; Clinton et al. 2010), agricultural areas (Lucieer and Stein 2002; Möller, Lymburner, and Volk 2007; Wang, Jensen, and Im 2010), or larger land-use/land-cover types (Weidner 2008). Yet, goodness measures have also been used in natural area segmentations (Carleer, Debeir, and Wolff 2005; Ke, Quackenbush, and Im 2010; Bar Massada et al. 2012). Some supervised segmentation evaluation methods use a larger set of reference polygons inside a larger area (e.g. Clinton et al. 2010), whereas others use only a couple of distinct reference polygons and semi-automated approaches (e.g. Marpu et al. 2010).

Segmentation evaluation can be used to compare different types of segmentation methods, i.e. region-based segmentation against discontinuity-based segmentation (Carleer, Debeir, and Wolff 2005). In addition, evaluation can be made between different segmentation methods or software, or inside a segmentation method as parameter optimization (Marpu et al. 2010). Although similar methods can be used in all these problems, task-specific methodology for the problems has also been developed, especially for parameter optimization. For instance, genetic algorithms have been used in optimizing segmentation to match reference delineation (Feitosa et al. 2006; Chabrier et al. 2008). Optimization has also been used without supervised goodness measures based on unsupervised evaluation. For instance, Drăguţ, Tiede, and Levick (2010) developed a scale parameter optimization tool that measures the rate of change of local variance inside a scene. Optimal scale parameters are those that have a local maximum of rate of change. Similarly, Espindola et al. (2006), Gao et al. (2011), and Yue et al. (2012) tried to combine low intra-segment variance and low inter-segment autocorrelation. On the other hand, Kim, Madden, and Warner (2008, 2009) hypothesized that optimal scales should only have low spatial autocorrelation between segments; Wang, Sousa, and Gong (2004) maximized Battacharya distance between candidate segments. Finally, Smith (2010) optimized the segmentation scale by minimizing classification error in a random forest classifier.

Forest inventory or forest habitat mapping is only one instance of where segmentation is often used. Yet, forest inventories are increasingly dependent on automatic segmentations (Pekkarinen 2002; Hay et al. 2005; Castilla, Hay, and Ruiz-Gallardo 2008; Mustonen, Packalén, and Kangas 2008; Wulder et al. 2008; Falkowski et al. 2009; Kim, Madden, and Warner 2009; Ke, Quackenbush, and Im 2010; Hou et al. 2013). Segmentations used in forest inventory are usually made with feature values calculated from aerial or satellite images; however, the usage of light detection and ranging (lidar) data has recently become popular (Mustonen, Packalén, and Kangas 2008; Ke, Quackenbush, and Im 2010; Breidenbach et al. 2011; Eysn et al. 2012; Hou et al. 2013). It has been noted that the synergy between imagery and lidar provides promising segmentation results and the selection of input data affects segmentation quality (Geerling et al. 2007, 2009; Mustonen, Packalén, and Kangas 2008; Ke, Quackenbush, and Im 2010; Hou et al. 2013). Despite this, studies incorporating different data sets remain scarce, both in forest inventory and in other applications.

In forest inventory or habitat mapping, segmentation goodness evaluations have been performed both qualitatively (Leckie et al. 2003; Wulder et al. 2008) and quantitatively using unsupervised (Kim, Madden, and Warner 2008, 2009; Hou et al. 2013) or supervised methods (Radoux and Defourny 2007; Ke, Quackenbush, and Im 2010). Some evaluations have been based on the thematic quality of segments against reference polygons (Pekkarinen 2002; Mustonen, Packalén, and Kangas 2008). Wulder et al. (2008) criticize quantitative evaluation because the used ground truth is a subjective delineation of forest patches; hence, real truth does not exist. Therefore, objects in forests are not as clearly separable as, e.g., urban features. Furthermore, in urban features, supervised evaluation has been criticized because of inaccurate ground truth (Corcoran, Winstanley, and Mooney 2010). In this work, we wanted to test whether supervised segmentation evaluation methods are applicable to forested areas on a larger set of reference polygons.

It has been stated and tested that segmentation goodness affects classification accuracy directly in OBIA classification (Kim, Madden, and Warner 2009; Clinton et al. 2010; Ke, Quackenbush, and Im 2010; Gao et al. 2011). Although there are numerous approaches and measures to evaluate segmentation goodness, the evaluation of goodness measures has not been thorough. In this article, we will test whether the widely used supervised segmentation goodness measures are applicable in boreal forest habitat-type mapping and whether best segmentation leads to best classification accuracy. We test supervised methods instead of unsupervised methods, because our objective is to determine a segmentation that matches habitat-type patches that are delineated using fieldwork. Furthermore, we try to find a segmentation method, parameter value, and image/data layer combination that best suits our purposes. To do this, we tested two different methods (FNEA region-based segmentation and IDRISI watershed segmentation), several parameter combinations, and different layers derived from WorldView-2 imagery and lidar data. We also tested whether different methods are habitat type, reference polygon, or area sensitive.

2. Methods

2.1. Study area and reference polygons

We studied a 7 km^2 rural-forested area southwest of the city of Jyväskylä located in Southern Finland. The area belongs to southern boreal vegetation zone (Ahti, Hämet-Ahti, and Jalas 1968). The geographic coordinates (WGS84) of the site are 62° 10′ 30″ N–62° 13′ 30″ N and 25° 29′ 0″ E–25° 38′ 0″ E. The study area mainly consists of both coniferous and deciduous forest habitats, mires, and agricultural area. The main tree species of the study area are the Scots pine (*Pinus sylvestris*), Norwegian spruce (*Picea abies*), and birches (*Betula pubescens* and *Betula verrucosa*). The study area was divided into three subareas with slightly varying land cover. The subareas were classified into 25 different habitat types (Table 1), which were mapped by fieldwork during June–August 2011. Habitat patches were used as reference polygons in segmentation goodness measures.

Most of the studied forest area is under heavy forestry and clear-cuts, which create a human-induced dynamic. Yet, two of the three delineated subareas included also one protected area, which covered 100 and 25 ha of these subareas, respectively. Protected areas were dominated by a semi-natural, over 100 year-old, forest. The larger protected area is part of a NATURA 2000 area. Inside the NATURA area and our study area, several different

Table 1. Different habitat types that were mapped during fieldwork when the reference polygons were drawn and that were used in the classification part of the research.

Habitat type	Number of age groups/management possibilities
Xeric (pine-dominated) forests	4: clear-cut, sapling stand, young, mature
Mesic (spruce-dominated) forests	5: clear-cut, sapling stand, young, mature, natural
Herb-rich (mixed/deciduous) forests	4: sapling stand, young, mature, natural
Bare rock	1
Pine mires	1
Spruce mires	2: not drained, drained
Open mires	1
Water (lakes and streams)	1
Small creeks	1
Springs	1
Grasslands	1
Fields	1
Roads	1
Yards	1
Sand pits	1

NATURA 2000 habitats are found. NATURA 2000 habitats were not mapped *per se*, but they were included in some of our mapped habitat types.

The three studied subareas were selected from different parts of a larger area southwest of Jyväskylä so that they included several different habitat types and different landscape configurations. Each subarea was delineated into habitat patches, resulting in a total of 628 habitat patches. Subarea 1 (Sallaajärvi) included small- to medium-sized patches of different-aged, mostly mesic forests, some spruce mires, small streams, meadows, lakes, and yards. The area also comprised some old fields, which have been afforested. Moreover, in the middle of the area, there is a 250 ha conservation area with a semi-natural forest. Subarea 2 (Kuusimäki) included large areas of protected old mesic forests with spruce mire patches. Moreover, this subarea had some open and pine mires, small lakes and fields, yards, and forests of different ages around the old forest. Subarea 3 (Lapinmäki) included areas of bare rock surrounded by mesic forest, and with yards, fields, and lakes on the fringes.

During fieldwork, patches were drawn into paper printouts of orthophotographs, which included 5 m contour lines derived from a topographic map. Additionally, Trimble GeoXT and Juno SB Global Positioning System (Trimble, Sunnyvale, CA, USA) devices with differential location corrections were used to check accurate location and to delineate patches, which were difficult to distinguish from aerial images. ArcGIS 9.3.1 (Esri, Redlands, CA, USA) editor was used when patches were manually drawn into a digital format. Patches were initially mapped as they were in the terrain. Later, some recent clear-cuts were modified to be of the same age and forest type as the neighbouring forest patches to match the state of the forest in the used satellite image and lidar data.

As alternative reference polygons, we used a forestry planning data set created for the City of Jyväskylä and a biotope classification data set created by the Finnish Forest and Park Service (FFPS) (Vesterbacka 2010). In the forestry planning data set, polygons are drawn first from aerial imagery, and after initial drawing, polygons are double-checked using fieldwork. This data was from subarea 1 only. FFPS biotope data is also generated using fieldwork and aerial imagery and it was from subarea 2 only.

2.2. Remotely sensed data

Our primary data consisted of eight-band multispectral 2 m resolution WorldView-2 (WV-2) satellite image taken on 14 July 2010 and lidar data with a minimum of 0.5 points per m^2 from May 2010. Additionally, we used 20 cm resolution aerial images (orthophotographs) taken in 2007 in assisting the drawing of reference polygons.

The WV-2 image, taken by Digital Globe Inc., consists of eight bands: coastal blue (band 1, 400–450 nm), blue (2, 450–510 nm), green (3, 510–580 nm), yellow (4, 585–625 nm), red (5, 630–690 nm), red-edge (6, 705–745 nm), NIR1 (7, 770–895 nm), and NIR2 (8, 860–1040 nm) in 2 m resolution, and a panchromatic band (450–800 nm) in 50 cm resolution. The image was delivered radiometrically and sensor corrected, projected to a plane with average terrain elevation. In our pre-processing phase, the image was first orthorectified using a 5 m resolution digital elevation model derived from lidar data. In georeferencing, 13 ground control points from block features (buildings, etc.), which were scattered all over the study area, were taken from orthophotographs and the nearest neighbour sampling method was used. In visual interpretation, the differences between orthophotographs, lidar data, and orthorectified WV-2 were at maximum a couple of metres. From WV-2, we used all multispectral bands in 2 m resolution.

Lidar data was created by the National Land Survey of Finland. Flying altitude is on average 2000 m. The used scan angle was ±20° and the laser pulse footprint on the ground was approximately 50 cm. The mean error in elevation information is at maximum 15 cm and in planar information at maximum 60 cm. Data was delivered automatically classified to ground hits, low vegetation hits, low error hits, and unclassified hits.

Lidar point clouds were first triangulated and after that rasterized using LAS tools (Rapidlasso, Gilching, Germany) (Isenburg 2011). We first derived two layers in 2 m resolution from lidar: digital terrain model (DTM) and digital surface model (DSM). In DTM, only ground hits were used; whereas in DSM, the point cloud was first thinned to 1 m resolution to include the highest hits. Then we subtracted DTM from DSM to create a canopy height model (CHM). CHM was further manipulated to include values only between 0 and 40 m to filter out unrealistic values. The CHM still had some wrong values below 40 m, but these could not be corrected easily.

From DTM, using SAGA GIS (System for Automated Geoscientific Analyses geographical information system), we calculated the SAGA wetness index (SWI) in 2 m resolution to model soil moisture, and thus, potential places for mires. SWI is a modification of the topographic wetness index (TWI). It has been noted that in wetland mapping, standard TWI performed worse than some other models, mainly because in these studies TWI underestimated the extent and contiguity of wetlands (Grabs et al. 2009; Murphy, Ogilvie, and Arp 2009). This might be because standard TWI concentrates large values to stream networks where water flow is concentrated. This underestimation problem is overcome in SWI, which assumes homogeneous hydrologic conditions in flat areas and predicts larger moisture values for cells with small vertical distance to streams (Böhner and Selige 2006, Equations (1) and (2)):

$$\alpha_{\mathrm{M}} = \alpha_{\max} t^{-\beta \exp\left(t^{\beta}\right)} \quad \text{for} \quad \alpha < \alpha_{\max} t^{-\beta \exp\left(t^{\beta}\right)}. \tag{1}$$

The specific catchment area (α) used in TWI is defined as the pixel's upslope contributing area per contour unit width, whereas α_{M} is the modified catchment area used in SWI. In calculating α, slope angle β (in radians) and neighbouring cell maximum $\alpha_{\max}$ are taken into account unless the results remain unchanged. Parameter t is a value for suction, so that lower values, e.g. under 10, lead to stronger suction and stronger spreading of large α values, and conversely higher values lead to weaker suction. After counting α_{M}, SWI is calculated with the standard equation given in the following:

$$\mathrm{SWI} = \ln\left(\frac{\alpha_M}{\tan\beta}\right). \tag{2}$$

Before calculations, DTM was filled to remove uncertainties, missing values, and false values from the data. Before the filling, values in DTM in known and evident places of bridges and culverts were manipulated to let imagined water to flow through road banks in those locations. To angle β, 0.0174532 rad was added so that division by 0 was avoided. In flow direction calculations, we used the multiple flow direction method proposed by Freeman (1991). In this method, the slope value is raised to the power of 1.1. Thus, steeper slopes are weighted only a little. It has been noted that in relatively flat areas, multiple flow direction methods, in which the slope value is raised by a low exponent (e.g. 0.5–2), give good results in TWI calculation (Güntner, Seibert, and Uhlenbrook 2004; Sørensen, Zinko, and Seibert 2006; Kopecký and Čížková 2010). Furthermore, parameter t in Equation (1) was decided to be default 10 after visual interpretation of SWI with different t values.

Before segmentation, SWI was quantized to 32 classes using equal intervals and CHM was quantized to 40 classes (nearest integer). WV-2 layers were first filtered using a 3×3 window and a median filter. After filtering, layers were quantized to 256 classes.

2.3. *Segmentation methods*

Data were segmented using different data sets, methods and parameters. To compare different types of segmentation methods, two segmentation methods were used: the watershed segmentation method and the region-based segmentation method. Brief introductions of the segmentation methods used and their parameters are given next.

Watershed segmentation was implemented in IDRISI Taiga software (Clark Labs, Worcester, MA, USA). In IDRISI segmentation, a variance image is derived from each layer by moving window analysis. A weighted average of variance images is the final surface image for watershed delineation. Both the size of the moving window as well as the weights of averaging can be adjusted by the user. The values of this surface image are treated as elevation values, as in a digital elevation model (DEM), and pixels are grouped into watersheds. After watershed delineation, watersheds are merged iteratively. Pairs of segments are merged if they are the most similar segments to each other in the neighbourhood and if their difference is smaller than a similarity tolerance adjusted by the user. The difference is evaluated by two aspects: the mean value and the standard deviation. The weights for the mean and for the standard deviation are set by the user.

Our region-based segmentation method was the widely used segmentation method of eCognition software, FNEA (Baatz and Schäpe 2001; Benz et al. 2004). FNEA segmentation was carried out using TerraLib 4.2.0 C++ GIS library (Câmara et al. 2008). In FNEA, regions are formed by merging pixels, i.e. in the beginning, each pixel is treated as a region. In segmentation, three user parameters can be adjusted: scale parameter and weights between colour and shape ($w_{\text{colour}} + w_{\text{shape}} = 1$) as well as between smoothness and compactness ($w_{\text{smooth}} + w_{\text{compt}} = w_{\text{shape}}$). The scale parameter controls the average object size. The more weight is given to colour (or spectral) homogeneity, the less weight is given to a specific shape, i.e. spatial homogeneity. Smoothness and compactness define the shape as follows. Smoothness is the ratio of the border length of the segment and border length of the bounding box of the segment. Compactness, on the other hand, is the ratio of the border length of the segment and the square root of the number of pixels in the segment. Hence, they are not antagonistic, but the weight is defined between them. Finally, the weights for the different layers are set by the user.

2.4. *Initial work for segmentation goodness evaluation*

In segmentation, several issues affect the final segmentation goodness: segmentation method used, parameterization including weights for the layers (e.g. Marpu et al. 2010), used layers (e.g. Ke, Quackenbush, and Im 2010), (re)classification of the layers, transformations made for the layers, and filtering of the layers (e.g. Carleer, Debeir, and Wolff 2005). Easily thousands of different combinations can be tested. Therefore, we first carried out initial trial-and-error testing and visual interpretation for different types of segmentations. We segmented single layers, reclassified and filtered the layers, and tried different parameter combinations and segmentation methods. In our initial analysis, the objective was to determine good segmentation methods that could be further evaluated using the evaluation measures. In addition, we wanted to scale our layers so that they could be used in the same segmentations; in other words, the segments that are produced in single-layer segmentations should be approximately of the same size. Needless to say, our initial evaluation was not thorough but it was good enough to find good segmentation methods. It was not possible to test all of the possible combinations, but we found a set of segmentations that were probably among the best that are available.

2.5. *Used parameter and layer combinations*

Four different layer combinations were tested: (a) WV-2 layers only, (b) lidar layers only, (c) WV-2 bands 2, 3, 5, 7 (blue, green, red, NIR1) and lidar layers, and (d) all layers. IDRISI segmentation was performed using a window size of 5. Similarity tolerance was varied between 20 and 70 with intervals of 5. Three different combinations of mean and variance weights were used: mean 0.5, variance 0.5; mean 0.9, variance 0.1; and mean 0.1, variance 0.9. Hence, overall, 33 IDRISI segmentations were performed for all layer combinations. FNEA segmentation was performed by varying the scale parameter between 5 and 50 with intervals of 5, and using colour parameter values of 0.25, 0.5, and 0.75. Therefore, 30 different FNEA segmentations were performed for all layer combinations. In all segmentations, all layers were given equal weight.

2.6. *Different reference polygons*

We tested different segmentation methods using eight different reference polygon sets. First, all reference polygons from the whole study area were used. Second, three sets including all reference polygons from three different subareas separately. Third, two sets including reference polygons of only one habitat type: one set included all mires and one set included water. Finally, we tested segmentation quality against two other reference polygon sets (FFPS biotope and forestry planning data) (Figure 1).

2.7. *Goodness evaluation measures*

Segmentation goodness was evaluated using several different supervised measures (Table 2), reviewed by Clinton et al. (2010) with a Java tool that they developed. For more clarification and equations, refer to Clinton et al. (2010) and the original publications listed in Table 2. All measures were calculated as a mean of all reference polygons inside a reference polygon set. The value for a specific reference polygon was calculated as a mean (or standard deviation) of the values of those segments that met at least one out of the following four criteria: (1) the centroid of the segment is inside the reference polygon; (2) the centroid of the reference polygon is inside the segment; (3) the shared area of the segment and the reference polygon is over 0.5 of the segment area; and (4) the shared area of the segment and the reference polygon is over 0.5 of the reference polygon area (Clinton et al. 2010). Some of the measures were weighted by the reference objects (Table 2). Furthermore, we calculated combined measures as proposed by Clinton et al. (2010) and which all included measures from single authors only (Table 3). Some of the combined measures were calculated as root mean square (RMS) individual criterion values, whereas some of them were simple sum calculations. In RMS calculations, all measures were adjusted so that ideal segmentation was set to 0. Finally, a combined measure, COMBINED, was calculated, which was an RMS of all basic area and location-based measures as suggested by Clinton et al. (2010). However, QLoc was not included since it was the same measure as RPsub. Before

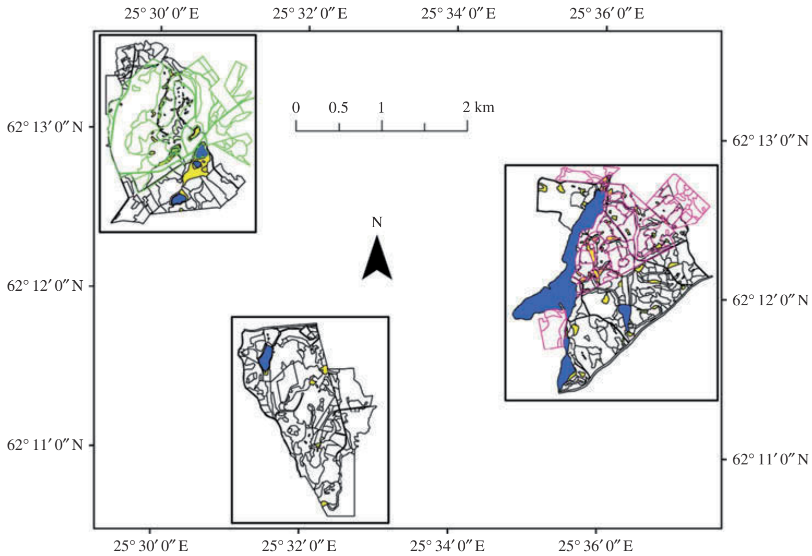

Figure 1. Different reference polygons used. Reference polygons drawn in our fieldwork are marked with black borders. Subarea 1 is located in the eastern, subarea 2 in the northwestern and subarea 3 in the southern part of the whole area. Waterbodies are coloured blue and mires yellow. Forestry planning polygons are marked with magenta outlines and FFPS polygons with green outlines. FNEA segmentations were performed inside the black rectangles, whereas IDRISI segmentations were also performed in the areas between the black rectangles.

RMS calculation in COMBINED, all measures were scaled to [0,1] by dividing each value with the maximum and setting the ideal segmentation to 0.

Furthermore, we measured segmentation goodness using visual interpretation. In visual interpretation, we focussed especially on whether the segmentation methods find the boundaries of some reference polygons and habitat types. Hence, we were more worried about under-segmentation than over-segmentation. Additionally, we checked whether the different kinds of habitat types are segmented and whether the segmentation produces objects that are meaningful entities and can be easily used in classification and planning (Hay et al. 2005). Therefore, segments should not be too complex (Mustonen, Packalén, and Kangas 2008). Owing to the large number of different segmentations, our visual interpretation was not thorough; instead, we tried to determine some general trends from different segmentation methods as well as from layer and parameter combinations.

2.8. Classifications

After segmentation evaluation, we selected 12 segmentations for classification. Segmentations were selected using subjective evaluation, so that meaningful evaluation of segmentation performance *versus* classification accuracy could be made and some of the segmentations could be compared to each other. Both good and not as good segmentations, based on evaluation measures and visual interpretation, were selected. In classification, we calculated the mean values of each layer per segment. In all classifications, all layers were

Table 2. Simple segmentation goodness evaluation measures that were used in segmentation evaluation. Column MEASURES refers to what the method should measure. Column SOURCE refers to the article where the measure was first used. Column WEIGHTED refers to whether the measure is weighted by a reference object.

Method	Measures	Source	Weighted	Note
Under-merging	Under-segmentation	Yang et al. (1995)		
Over-merging	Over-segmentation	Yang et al. (1995)		
AFI	Area match	Lucieer and Stein (2002)		
Count-over	Over-segmentation	Lucieer and Stein (2002)		Based on AFI
Count-under	Under-segmentation	Lucieer and Stein (2002)		Based on AFI
SimSize mean	Area match	Zhan et al. (2005)	X	
SimSize sd	Area match	Zhan et al. (2005)	X	
RAsuper	Under-segmentation	Möller, Lymburner, and Volk (2007)	X	
RAsub	Over-segmentation	Möller, Lymburner, and Volk (2007)	X	
QR	Area match	Weidner (2008)	X	
Over-segmentation	Over-segmentation	Clinton et al. (2010)	X	
Under-segmentation	Under-segmentation	Clinton et al. (2010)	X	
RPsuper	Distance to centroid	Möller et al. (2007)	X	
RPsub	Distance to centroid	Möller et al. (2007)	X	
QLoc mean	Distance to centroid	Zhan et al. (2005)	X	= RPsub
QLoc sd	Distance to centroid	Zhan et al. (2005)	X	

Table 3. Combined segmentation goodness evaluation measures that were used in segmentation evaluation. Column INCLUDES refers to the simple measures that are included in the respective combined measure. Column CALCULATION refers to how the combined measure was calculated.

Method	Includes	Calculation
M	RAsuper, RAsub, RPsuper, RPsub	RMS
ZH1	SimSize mean, SimSize std, QLoc mean, QLoc std	RMS
ZH2	SimSize mean, QLoc mean	RMS
D	Over-segmentation, Under-segmentation	RMS
Over-under	Count-over, Count-under	SUM
MergeSum	Over-merging, Under-merging	SUM
COMBINED	All other simple measures than QLoc mean	RMS, normalized

always used irrespective of which layer combination a–d was used in the segmentation phase.

Supervised classification was performed using the random forest classifier (Breiman 2001) with R package randomForest (Liaw and Wiener 2002) in R version 2.15.2 (R Development Core Team 2012). Random forest classification has been used in remote sensing and OBIA with good results (Lawrence, Wood, and Sheley 2006; Rodriguez-Galiano et al. 2012). Random forest is an ensemble classifier that combines several bootstrapped classification trees. In the final classification, there is a majority vote over all trees. Trees are randomized at each node by selecting only a subset of variables, of which the best split is chosen. When a tree is built, approximately two-thirds of the data are selected for training the classifier and the rest are called out of bag (OOB) test data. OOB data are used to estimate the error rate, which is averaged over all trees to obtain an error rate for the entire classification. Because of OOB, independent test data or cross-validation are not required when random forest is used (Breiman 2001; Breiman and Cutler 2007), which has been confirmed in remote-sensing studies (Lawrence, Wood, and Sheley 2006; Rodriguez-Galiano et al. 2012).

When random forest was performed, 500 trees were built and the number of features at each split was given the default value of the square root of all features. We used our own reference polygons over all three subareas as training data, not the FP or the FFPS data. The training set in the random forest run consisted of all those segments that had a minimum of 60 % coverage of one reference habitat type. Classification accuracies were calculated using all reference polygons with simple cross-tabulation matrices.

3. Results

3.1. Segmentation goodness based on evaluation measures

Based on all area and COMBINED measures, the best segmentation was obtained with FNEA with layer set b, scale parameter 25, and colour parameter 0.5 (Tables 4 and 5). However, choosing this segmentation as the best was contradictory, since no other segmentation evaluation measure ranked it as the best method. Besides, it ranked between 4 and 94 when COMBINED measure and other reference polygons other than all area were used. Hence, different segmentations were chosen as the best or being among the best when different goodness measures or reference polygons were used (Table 4). Some of the measures (Under-merging, Over-merging, Count-over, SimSize sd, RAsuper, RAsub, Over-segmentation, Under-segmentation), nonetheless, provided rather consistent results, i.e. the same segmentation proved to be the best or one of the best using different reference polygon sets. Other measures, on the other hand, had larger variation in their results. Consistency in results can be considered a downside, since reference sets were different, as illustrated in Figure 1. For instance, individual measures may prefer the segmentation result, which is as fine or as coarse as possible. On the other hand, consistency can also be considered an asset if some of the segmentations are truly better despite the reference set used, i.e. those segmentations contain almost all meaningful patch boundaries.

Some overall evaluations can be made from segmentations ranked as the best (Table 6). First, FNEA segmentation outperformed IDRISI segmentation, since FNEA was ranked best 140 times as against the 44 times of IDRISI. Second, layer set b outperformed other layer sets. Therefore, it could be considered that FNEA with layer set b provides the best results. To avoid under-segmentation, low scale parameters should be used. On the contrary, high scale parameters should be selected when over-segmentation is not desired. Segmentations with an intermediate scale or similarity parameter value were not ranked as best as often as segmentations with high or low parameter values. Yet, different combined measures as well as AFI, RP measures, SimSize mean, QLoc mean, QLoc sd, and QR usually preferred intermediate-scale parameter values. For instance, when all area reference polygons were used, the COMBINED measure favoured intermediate-scale parameter values; however, it ranked those segmentations with low scale parameter values as the worst (Table 5). In FNEA segmentations, a high value for the colour parameter provided more often best segmentations than low or intermediate values for colour. In IDRISI segmentations, on the other hand, a high mean and low variance combination resulted in the largest number of best segmentations.

When correlations between different goodness measure results were evaluated (Table 7), it was found that correlations range from large negative correlations to high

Table 4. Best segmentations according to different measures and reference polygons. Segmentation methods are marked as follows. Text refers to method, letter after the text to layer combination a–d (see text), s to scale (FNEA) or similarity parameter (IDRISI), c to colour parameter, m to weight given to mean, and v to weight given to variance.

Measure	All area	SA1	SA2	SA3	Water	Mires	FP	FFPS
Under-merging	FNEA_d_s5_c.75	FNEA_d_s5_c.75	FNEA_a_s5_c.75	FNEA_a_s5_c.75	FNEA_a_s5_c.5	FNEA_d_s5_c.75	FNEA_c_s5_c.75	FNEA_a_s5_c.5
Over-merging	FNEA_b_s50_c.25	FNEA_b_s50_c.25	FNEA_b_s50_c.5	FNEA_b_s50_c.5	IDRISI_b_s70_m1_v9	FNEA_b_s50_c.25	FNEA_b_s50_c.25	FNEA_c_s50_c.75
AFI	IDRISI_a_s25_m9_v1	IDRISI_a_s25_m1_v9	IDRISI_d_s20_m9_v1	IDRISI_c_s35_m1_v9	IDRISI_d_s70_m5_v5	FNEA_b_s5_c.25	IDRISI_a_s45_m5_v5	IDRISI_b_s35_m9_v1
Count-over	FNEA_b_s50_c.25	FNEA_b_s50_c.25	FNEA_b_s50_c.25	FNEA_c_s50_c.25*	FNEA_b_s50_c.5	FNEA_d_s45_c.25*	FNEA_b_s50_c.5*	FNEA_b_s50_c.75*
Count-under	IDRISI_a_s20_m1_v9*	IDRISI_a_s25_m1_v9*	IDRISI_a_s20_m9_v1*	IDRISI_b_s50_m9_v1	FNEA_b_s25_c.25*	IDRISI_a_s40_m1_v9*	IDRISI_a_s70_m1_v9*	IDRISI_a_s70_m1_v9*
SimSize mean	FNEA_b_s30_c.75	FNEA_b_s30_c.75	FNEA_d_s30_c.75	IDRISI_c_s65_m9_v1	FNEA_b_s45_c.75	IDRISI_c_s55_m9_v1	FNEA_c_s30_c.25	FNEA_d_s40_c.25
SimSize sd	FNEA_a_s5_c.75	FNEA_b_s5_c.75	FNEA_a_s5_c.75	FNEA_a_s5_c.75	FNEA_a_s5_c.5	FNEA_b_s50_c.25	FNEA_a_s5_c.75	FNEA_d_s5_c.75
RAsuper	FNEA_b_s5_c.75	FNEA_b_s5_c.75	FNEA_b_s5_c.75	FNEA_a_s5_c.75	FNEA_c_s5_c.5	FNEA_d_s5_c.75	FNEA_b_s5_c.75	FNEA_b_s5_c.75
RAsub	FNEA_b_s50_c.25	FNEA_b_s50_c.25	FNEA_b_s50_c.5	FNEA_b_s50_c.25	FNEA_b_s50_c.5	FNEA_b_s50_c.25	FNEA_b_s50_c.25	FNEA_b_s50_c.25
QR	FNEA_b_s35_c.75	FNEA_b_s30_c.75	FNEA_c_s30_c.75	FNEA_a_s45_c.75	IDRISI_b_s65_m9_v1	IDRISI_a_s45_m1_v9	FNEA_c_s45_c.75	FNEA_d_s40_c.75
Over-segmentation	FNEA_b_s50_c.5	FNEA_b_s50_c.25	FNEA_b_s50_c.5	FNEA_b_s50_c.5	IDRISI_b_s70_m1_v9	FNEA_b_s50_c.25	FNEA_b_s50_c.25	FNEA_c_s50_c.75
Under-segmentation	FNEA_b_s5_c.75	FNEA_b_s5_c.75	FNEA_b_s5_c.75	FNEA_b_s5_c.75	IDRISI_a_s20_m5_v5	FNEA_b_s5_c.75	FNEA_b_s5_c.75	FNEA_b_s5_c.75
RPsuper	IDRISI_d_s55_m9_v1	FNEA_b_s20_c.5	FNEA_d_s20_c.5	FNEA_b_s15_c.5	IDRISI_b_s65_m9_v1	FNEA_d_s10_c.25	FNEA_d_s30_c.75	IDRISI_d_s60_m9_v1
RPsub	FNEA_b_s35_c.75	IDRISI_b_s55_m5_v5	FNEA_a_s40_c.75	FNEA_c_s40_c.75	IDRISI_b_s65_m9_v1	FNEA_b_s15_c.25	FNEA_c_s30_c.75	FNEA_d_s40_c.75
QLoc mean	FNEA_b_s35_c.75	IDRISI_b_s55_m5_v5	FNEA_a_s40_c.75	FNEA_c_s40_c.75	IDRISI_b_s65_m9_v1	FNEA_b_s15_c.25	FNEA_c_s30_c.75	FNEA_d_s40_c.75
QLoc sd	FNEA_a_s45_c.25	IDRISI_b_s65_m9_v1	FNEA_c_s45_c.75	FNEA_c_s40_c.75	IDRISI_b_s65_m1_v9	FNEA_b_s20_c.25	FNEA_c_s30_c.75	FNEA_c_s50_c.25
M	FNEA_c_s25_c.75	IDRISI_b_s45_m9_v1	FNEA_a_s5_c.5	FNEA_b_s20_c.75	FNEA_c_s45_c.75	FNEA_d_s5_c.75	FNEA_b_s5_c.75	FNEA_a_s25_c.5
ZH1	FNEA_d_s50_c.25	FNEA_d_s50_c.25	FNEA_c_s5_c.25	FNEA_b_s45_c.5	IDRISI_b_s40_m5_v5	FNEA_b_s50_c.25	FNEA_a_s5_c.25	FNEA_a_s45_c.25
ZH2	FNEA_b_s35_c.75	IDRISI_b_s60_m9_v1	FNEA_d_s30_c.75	FNEA_c_s40_c.75	FNEA_b_s45_c.75	IDRISI_c_s55_m9_v1	FNEA_c_s30_c.25	FNEA_d_s45_c.25
D	FNEA_d_s35_c.5	FNEA_d_s35_c.5	FNEA_d_s35_c.5	FNEA_b_s35_c.75	IDRISI_b_s70_m1_v9	FNEA_b_s20_c.75	FNEA_c_s35_c.75	FNEA_d_s40_c.5
Over-under	IDRISI_b_s70_m9_v1	IDRISI_b_s65_m1_v9	IDRISI_a_s70_m5_v5	IDRISI_b_s70_m9_v1	FNEA_b_s50_c.5	IDRISI_a_s70_m9_v1	FNEA_c_s50_c.75	IDRISI_a_s70_m1_v9
MergeSum	FNEA_b_s15_c.5	FNEA_b_s15_c.5	FNEA_a_s15_c.75	FNEA_b_s20_c.75	IDRISI_b_s65_m9_v1	FNEA_b_s10_c.25	FNEA_c_s25_c.5	FNEA_d_s25_c.25
COMBINED	FNEA_b_s25_c.5	IDRISI_b_s50_m5_v5	FNEA_b_s45_c.75	FNEA_b_s30_c.75	IDRISI_b_s70_m1_v9	FNEA_a_s40_c.5	FNEA_b_s45_c.5	FNEA_a_s35_c.25

Note: * = tie with other segmentations that are not indicated here. Segmentation that is shown ranked best using the OverUnder evaluation method.

Table 5. Twenty best and 10 worst segmentations based on all area reference polygons and COMBINED measure. Segmentation methods are named as in Table 4.

	Segmentation	Combined
1	FNEA_b_s25_c.5	0.646
2	FNEA_b_s25_c.75	0.649
3	FNEA_b_s30_c.75	0.650
4	FNEA_c_s25_c.75	0.650
5	FNEA_d_s25_c.75	0.651
6	IDRISI_b_s55_m9_v1	0.652
7	FNEA_a_s30_c.75	0.652
8	IDRISI_b_s65_m9_v1	0.652
9	IDRISI_d_s65_m9_v1	0.652
10	FNEA_d_s30_c.75	0.652
11	IDRISI_b_s50_m5_v5	0.653
12	FNEA_b_s20_c.5	0.653
13	IDRISI_b_s55_m5_v5	0.653
14	IDRISI_b_s60_m9_v1	0.653
15	IDRISI_b_s50_m9_v1	0.653
16	IDRISI_c_s60_m5_v5	0.653
17	IDRISI_a_s60_m1_v9	0.653
18	IDRISI_a_s70_m9_v1	0.653
19	IDRISI_d_s65_m5_v5	0.654
20	IDRISI_a_s60_m5_v5	0.654
. . .		
243	IDRISI_c_s20_m9_v1	0.770
244	FNEA_a_s5_c.25	0.771
245	FNEA_c_s5_c.5	0.782
246	FNEA_d_s5_c.5	0.789
247	FNEA_a_s5_c.5	0.796
248	FNEA_b_s5_c.5	0.797
249	FNEA_c_s5_c.75	0.801
250	FNEA_d_s5_c.75	0.805
251	FNEA_a_s5_c.75	0.811
252	FNEA_b_s5_c.75	0.813

positive correlations. Hence, the measures provided different results and preferred different issues in segmentation. It can also be seen that measures that measure over-segmentation had positive correlations with the COMBINED measure, whereas under-segmentation measures had negative correlations (for over- and under-segmentation measures, see Table 2). Some measures (RPsuper, MergeSum, M, ZH1) had even both positive and negative correlations. Correlations depended on the reference polygons used; however, correlations between the COMBINED measure based on different reference polygons were rather high and positive (Table 8). Only water has correlations below 0.75.

3.2. Segmentation goodness based on visual interpretation

In visual interpretation, it was found that segmentations based on layer set b (lidar data only) were especially successful in delineating mires and small streams. In addition, the problem of shadow effect in WV-2 imagery was overcome when lidar data were used. On the other hand, the shorelines of waterbodies were insufficiently delineated with lidar

Table 6. Best segmentations based on different measures sorted by segmentation method, layer combinations, and different parameter options. Numbers refer to the number of segmentations that were regarded as best.

All segmentations			FNEA segmentations			IDRISI segmentations		
	Value	Count		Value	Count		Value	Count
Methods	FNEA	140	Scale	5	36	Similarity	20	4
	IDRISI	44		10	2		25	3
				15	6		30	0
Layers	a	35		20	6		35	2
	b	93		25	6		40	2
	c	28		30	13		45	3
	d	28		35	10		50	2
				40	12		55	5
				45	13		60	2
				50	36		65	9
							70	12
			Colour	0.25	40			
				0.5	30	Mean/var	0.1/0.9	15
				0.75	70		0.5/0.5	8
							0.9/0.1	21

Table 7. Correlations between the COMBINED goodness measure and individual goodness measures based on different reference polygons. SA refers to subarea, FP refers to forestry planning data, and FFPS refers to FFPS data.

Measure	ALL	SA1	SA2	SA3	Water	Mires	FP	FFPS
Under-merging	−0.24	−0.39	−0.63	−0.13	−0.63	−0.61	−0.69	−0.33
Over-merging	0.89	0.90	0.95	0.90	0.88	0.97	0.91	0.90
AFI	−0.22	−0.36	−0.63	−0.10	−0.40	−0.61	−0.54	−0.24
Count-over	0.69	0.78	0.91	0.73	0.87	0.91	0.88	0.71
Count-under	−0.25	−0.31	−0.73	−0.31	−0.53	−0.62	−0.53	−0.10
SimSize mean	0.89	0.91	0.91	0.89	0.85	0.29	0.92	0.82
SimSize sd	−0.91	−0.93	−0.93	−0.91	−0.94	−0.34	−0.93	−0.85
RAsuper	−0.79	−0.86	−0.94	−0.83	−0.95	−0.93	−0.96	−0.86
RAsub	0.43	0.54	0.80	0.53	0.86	0.75	0.85	0.52
QR	0.91	0.94	0.93	0.92	0.95	0.35	0.95	0.90
Over-segmentation	0.48	0.60	0.79	0.56	0.81	0.84	0.84	0.50
Under-segmentation	−0.52	−0.65	−0.82	−0.58	−0.79	−0.86	−0.87	−0.63
RPsuper	0.85	0.67	0.52	0.78	0.88	−0.55	0.74	0.96
RPsub	0.86	0.94	0.88	0.94	0.95	0.01	0.92	0.87
QLoc mean	0.86	0.94	0.88	0.94	0.95	0.01	0.92	0.87
QLoc sd	0.63	0.68	0.80	0.79	0.80	0.30	0.80	0.72
M	0.98	0.98	−0.88	0.97	−0.50	−0.72	−0.90	0.97
ZH1	0.66	0.60	−0.83	0.84	0.76	−0.29	−0.87	0.83
ZH2	0.89	0.96	0.92	0.94	0.94	0.29	0.92	0.89
D	0.77	0.82	0.88	0.81	0.83	0.36	0.90	0.65
OverUnder	0.76	0.84	0.90	0.78	0.91	0.86	0.89	0.75
MergeSum	0.41	0.50	−0.34	0.69	0.86	−0.52	0.85	0.90

data only. Moreover, boundaries between deciduous and coniferous forests were better delineated using WV-2 imagery. However, more gradual boundaries, for instance, between mesic and xeric forests, could not be easily segmented using any method or layer combination. In visual interpretation, we could not decide between layer sets c and d. Although

Table 8. Correlations between the COMBINED measure values based on different reference polygons.

	ALL	SA1	SA2	SA3	Water	Mires	FP	FFPS
ALL	1.00	0.89	0.85	0.94	0.64	0.84	0.78	0.94
SA1	0.89	1.00	0.83	0.84	0.59	0.83	0.79	0.88
SA2	0.85	0.83	1.00	0.87	0.82	0.97	0.97	0.89
SA3	0.94	0.84	0.87	1.00	0.73	0.87	0.83	0.88
Water	0.64	0.59	0.82	0.73	1.00	0.80	0.80	0.63
Mires	0.84	0.83	0.97	0.87	0.80	1.00	0.95	0.85
FP	0.78	0.79	0.97	0.83	0.80	0.95	1.00	0.84
FFPS	0.94	0.88	0.89	0.88	0.63	0.85	0.84	1.00

segmentation outputs were slightly different, the differences were minor. Similar observations were made when different IDRISI mean/variance weight alternatives were compared. Furthermore, to delineate some small objects, small values of scale or similarity parameters were required. In finding meaningful and simple entities, it was found that FNEA segmentation with low (0.25) or intermediate (0.5) weight for colour had better results compared to the other segmentation methods. Putting a little weight to colour had its downside, on the other hand. In other words, segment boundaries did not necessarily follow natural or data boundaries but segments were equally sized objects with often arbitrary boundaries. Nevertheless, FNEA segmentations with a large weight for colour and IDRISI segmentations were unnecessarily complex. Additionally, in IDRISI segmentations, boundaries often criss-crossed the reference polygon boundaries. Using visual interpretation, we chose FNEA segmentation with layer combination c, scale parameter 10, and colour parameter 0.5 as the best one (Figure 2(*c*)). This selection was, yet, more or less arbitrary, since many different segmentation options provided quite similar results. Furthermore, since there were numerous segmentation options, visual interpretation was not thoroughly reliable in determining the best parameter values. Hence, choosing the best segmentation method using visual interpretation was tricky.

3.3. Classification results

Classification accuracies between classifications derived from different segmentations varied to a certain extent (Table 9, some of the segmentations in Figure 2). Best accuracy (0.72) was achieved using the best segmentation in visual interpretation (Figure 2(*c*)), whereas the worst accuracy (0.60) was obtained using segmentation that was ranked high using some of the measures (Figure 2(*h*)). Several segmentations produced reasonably good results compared to the best classification method. Some of these segmentations were selected based on measures and some by using visual interpretation. On the other hand, best segmentation based on the COMBINED measure and all area (Figure 2(*e*)) was not among the best segmentations in classification accuracy analysis. It can be observed that fine or moderately fine segmentations led to better classification accuracies. *Vice versa*, coarse segmentation led to poorer accuracies. On the other hand, too fine segmentations can lead to a salt-and-pepper effect (Figure 2(*b*)) and thus possibly also worsen classification accuracy. Moreover, in visual interpretation, it became evident that classifications performed with FNEA segmentations and scale parameter value 5 suffered from this effect more than classifications performed with segmentations with scale value 10. The best classification

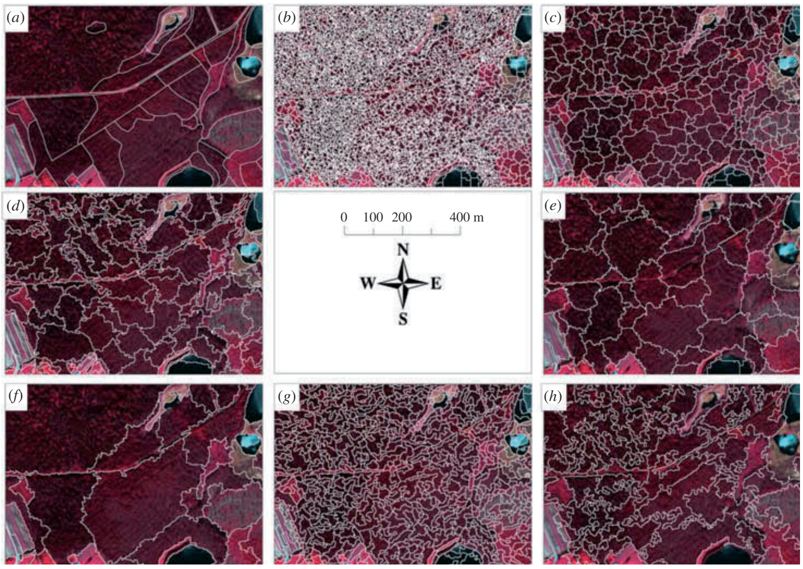

Figure 2. Reference polygons and a visually chosen set of different segmentations drawn on a WV-2 false colour image (red, band 7/NIR1; green, band 5/red; blue, band 3/green) in the background. (*a*) reference polygons, (*b*) FNEA_b_s5_c.75, (*c*) FNEA_c_s10_c.5, (*d*) FNEA_a_s20_c.75, (*e*) FNEA_b_s25_c.5, (*f*) FNEA_c_s50_c.75, (*g*) IDRISI_d_s30_m9v1, and (*h*) IDRISI_b_s70_m9v1. Images are from the southern part of subarea 2. Satellite imagery © 2010 Digital Globe Inc. Reproduced by permission.

Table 9. Classification accuracies derived from classifications based on different segmentations. Segmentations are marked as in Table 2. In addition, criteria why each segmentation was chosen to the classification analyses are given.

Segmentation	ACC	Why segmentation was selected to classification
FNEA_b_s5_c.75	0.69	BEST in avoiding under-segmentation, WORST based on COMBINATION and ALL AREA
FNEA_c_s5_c.75	0.69	Comparison against FNEA_b_s5_c.75
FNEA_a_s10_c.5	0.71	WV-2 layers only, comparison against FNEA_c_s10_c.5
FNEA_c_s10_c.5	0.72	BEST segmentation based on visual interpretation
FNEA_d_s15_c.25	0.71	GOOD in visual interpretation
FNEA_a_s20_c.75	0.69	Segmentation based on WV-2 data only, OK visually
FNEA_b_s25_c.5	0.66	BEST segmentation based on COMBINATION and ALL AREA, OK visually
FNEA_d_s35_c.5	0.66	BEST based on D and ALL AREA
FNEA_c_s50_c.75	0.65	GOOD in avoiding over-segmentation, GOOD in visual boundary evaluation
IDRISI_d_s30_m9v1	0.70	OK in visual interpretation, small segments
IDRISI_c_s40_m5v5	0.69	OK in visual interpretation, quite small segments
IDRISI_b_s70_m9v1	0.60	BEST based on OverUnder and ALL AREA, BAD in visual interpretation

accuracies were achieved using segmentations with both lidar and WV-2 layers. This might be because boundaries were best detected using both data types in segmentation. Yet, classification using only segmentations performed with lidar or WV-2 data were nevertheless not as accurate. In all, classification accuracy evaluation was not thorough; in other words, good classification accuracies can be obtained using segmentations that were not among the 12 segmentations tested here. Moreover, some measures may be good in selecting segmentations that maximize classification accuracy.

4. Discussion

4.1. Segmentation goodness compared to classification accuracy

One of our study objectives was to test whether better segmentation leads to better classification accuracy, as has been argued by others (Kim, Madden, and Warner 2009; Clinton et al. 2010; Ke, Quackenbush, and Im 2010; Gao et al. 2011). After our analysis, it is obvious that there is no straightforward answer to this question. Although it is self-evident that good segmentation is needed for good classification, there is no adequate definition of what makes a good segmentation. After classification analysis, one can easily state that the best segmentation was the segmentation with the best classification accuracy. There is no method, however, to test before classification which segmentation will provide the best classification accuracy. This is also illustrated by the studies of Kim, Madden, and Warner (2009) and Gao et al. (2011). Although they both claim that optimal segmentation produced the best classification output, their definitions of optimal segmentation were contradictory. Kim, Madden, and Warner (2009) minimized spatial autocorrelation between different segments, whereas Gao et al. (2011) sought for segmentations that combined low inter-segment autocorrelation and intra-segment variance. Furthermore, Gao et al. (2011) had the lowest inter-segment autocorrelation at the coarsest scale, which did not produce the best classification accuracy. On the other hand, the tasks in these studies were different, since Kim, Madden, and Warner (2009) used 4 m resolution IKONOS data in forest-type mapping, whereas Gao et al. (2011) used 25 m Landsat ETM+ data in a mixed mountainous shrub–forest–grassland landscape.

Based on the ambivalence of what segmentation is good, we propose that the goodness of segmentation should be defined in each case. In other words, one should know and clarify what one wants from segmentation, and critically evaluate whether the best segmentation can be selected based on the evaluation criteria. In our case, good segmentation was segmentation with (1) meaningful and not too complex segments, (2) boundaries parallel to reference polygon boundaries even for the smallest reference polygons, but (3) as coarse as possible. Considering the best segmentation based on some measure does not automatically lead to the best classification accuracy, as has been already noted by Verbeeck, Hermy, and van Orshoven (2012). Nevertheless, the classification accuracies between different classifications were rather small in our case study. This might point to the robustness of the OBIA methodology: good classification accuracy can be obtained even if the segmentation is not the best possible. On the other hand, the classification outputs that had better classification accuracies were visually more appealing. Boundaries were more often in the right places, different habitat types could be mapped, and patches were not too small.

Many authors have argued that over-segmentation is a smaller problem than under-segmentation in post-segmentation classification (e.g. Weidner 2008; Marpu et al. 2010). However, in the analysis by Verbeeck, Hermy, and van Orshoven (2012), it was found out that more under-segmented output gave better classification accuracy than more over-segmented output. Our results suggest that both arguments are partly correct. In other words, both over-segmentation and under-segmentation are problematic in classification, as found by Kim, Madden, and Warner (2009) and Gao et al. (2011). First, segmentation cannot be very coarse, since smaller objects are thus easily under-segmented. Second, if segmentation is too fine, a salt-and-pepper effect is obtained, which can lead to worse classification accuracy as has been found in OBIA *versus* pixel-based classification studies (e.g. Bock et al. 2005; Whiteside, Boggs, and Maier 2011). In addition, objects are more meaningful when they are not too small (Blaschke 2010). Yet, the better classification accuracy of OBIA is not automatic and some smaller, but rare, objects may be easily missed in OBIA classification (Dingle Robertson and King 2011). Overall, it has been noted that in single-scale segmentation, optimal segmentation is class-dependent, i.e. some classes can be poorly segmented even if the overall segmentation is optimal. Hence, multi-scale segmentation has been offered as a solution to this problem (Hay et al. 2003; Kim et al. 2011; dos Santos et al. 2012).

4.2. Object and patch delineation

We found that some habitat patches were not segmented properly using any of the methods, layers, or parameter combinations. For instance, stream-sided habitats or mires were often poorly delineated. Therefore, some extra analysis is required, such as stream network mapping (Räsänen et al., Forthcoming), other ancillary information or expert knowledge (Mustonen, Packalén, and Kangas 2008), or segmentation post-modification to delineate some of the patches correctly. It can even be asked whether the segmentation goodness over difficult patch delineation can even be calculated. For instance, Radoux and Defourny (2007) delineated only those patches that could be seen from the imagery. Therefore, it is not realistic to expect that segmentation delineates those objects that cannot be easily seen from the data that is segmented. Lidar data, however, helped in finding some of the tricky features, such as mires. On the other hand, segmentations with lidar data and four WV-2 layers were not significantly different compared to segmentations with lidar data and eight WV-2 layers. Besides, our study reasserted earlier studies that the problematic shadow effect of aerial or satellite imagery can be mitigated using lidar data (Geerling et al. 2007; Mustonen, Packalén, and Kangas 2008; Ke, Quackenbush, and Im 2010). Segmentations based on imagery only, nonetheless, produced classifications with almost as high classification accuracies as segmentations based on both imagery and lidar data. There can be at least two possible reasons for this small difference. First, segmentation based only on imagery may have other benefits compared to segmentation using both data types. Second, the proportion of shadow areas over all area can be rather small, especially with data resolution not higher than 2 m.

One major question in segmentation evaluation is whether it is better to delineate meaningful objects with meaningful thematic quality and maximum homogeneity (e.g. Mustonen, Packalén, and Kangas 2008) or to find segmentation that mimics field observations. Some authors (Wulder et al. 2008; Corcoran, Winstanley, and Mooney 2010) have questioned the rationality of supervised segmentation evaluation, especially in natural environments. It is true that nature is not easy to interpret. Different mappers classify habitat patches differently and also delineate patch boundaries differently. Yet, according to Cherrill and McClean (1995, 1999), the former type of error was more common in habitat mapping in the UK. Nevertheless, boundaries are not easy to draw and their locations depend on the study scale (Lang et al. 2010). In our analysis, there were differences between boundary locations when our field data was compared to either FP or FFPS data. There were some differences between optimal segmentations based on different reference polygons. These differences were mostly minor, and approximately the same kinds of segmentations were preferred irrespective of the reference data. Moreover, correlations between COMBINED measures based on different reference data were rather high (SA1 to FP 0.79 and SA2 to FFPS 0.89). Furthermore, segmentation evaluation based on thematic quality is not unproblematic either. Although the segmentation has good thematic quality, the segmentation boundaries do not necessarily match with habitat-type boundaries that exist in nature. This can lead to difficulties in habitat classification, if it is performed, and eventually to differences in planning decisions.

On a more general level, one can question whether automated segmentation is worse than manual delineation when it cannot find the boundaries that are manually delineated. Delineations are different, it is true, but it is not straightforward to judge either one of them better. For instance, automated delineation often produces more complex objects. Complexity of the objects, however, can be both good and bad. Although complexities hinder usage in the operational context, complexity can be reduced using GIS techniques. Furthermore, complex boundaries can be even truer, since natural boundaries are not always straight (Wulder et al. 2008). Therefore, automated and manual delineations are two different interpretations and both of them can be either good or bad depending on the segmentation method, mapper skills, or the operational context. In other words, the question is not necessarily whether one of them is correct or incorrect but whether it is appropriate or inappropriate (Lang et al. 2010).

4.3. How to evaluate and measure segmentation goodness?

According to our analysis, the FNEA was a better segmentation method than the watershed segmentation method in IDRISI Taiga. Still, IDRISI's segmentation method also provided good results. As already noted earlier, FNEA has produced good results in segmentation evaluations and is a standard method in OBIA studies. However, we cannot give any percentage or any other quantitative evaluation that indicates how much better FNEA is compared to IDRISI, contrary to values given, for instance, by Li, Huo, and Fang (2010). Li, Huo, and Fang (2010) classified different types of objects as correctly delineated, acceptably delineated, and wrongly delineated. From these classifications, they calculated the performances of different segmentations and also the percentage difference of performance. In our framework, such quantitative difference evaluation would be more or less artificial, since in our study different measures of segmentation goodness gave different results. This inconsistency has also been noted by Clinton et al. (2010). Partly, this inconsistency can be explained in terms of over- and under-segmentation; i.e. deliberately avoiding one of them often results in getting the other. However, evaluation measures that should quantify the same phenomenon can produce different results. One explanation of this is that we tested several different segmentations, of which some were rather similar to each other. Furthermore, these measures are somewhat dependent on the training data set used. One should, thus, be careful when selecting the reference data. On the other hand, some of the measures were robust, i.e. produced similar results irrespective of the reference data. Additionally, the general picture was more or less similar with different reference polygons.

According to classification accuracies derived in our study, the best segmentation was found using visual interpretation. Therefore, it could be argued that supervised segmentation goodness evaluation measures evaluated by Clinton et al. (2010) are not good. On the other hand, we knew what we wanted from visual interpretation and fixed our objectives based on these needs. Automated supervised segmentation goodness evaluation measures, on the contrary, were just selected based on what has been done before. Therefore, we knew better what we wanted from segmentation when we evaluated them visually: meaningful objects and boundaries. Yet, we would have included our own automated and supervised evaluation method in our analysis if we had found a successful way to carry out automated evaluation. One reason why supervised segmentation goodness evaluation measures partly failed in our analysis could be that we used continuous reference polygon data and objects that were difficult to delineate. On the other hand, segmentation goodness measures did not work that well on waterbodies either, although waterbodies are usually easy to delineate and are not bordered by each other. Based on measures, best segmentations for waterbodies were usually segmentations using lidar data or segmentations as fine as possible. In our visual interpretation, it was, nonetheless, found that waterbodies cannot be delineated using lidar data alone. A segmentation that could be good in waterbody delineation was ranked best only by evaluation measure M. Nevertheless, we cannot say that supervised evaluation measures are completely useless. On the contrary, one should know what evaluation measures favour and what one wants from segmentation before using evaluation measures. Moreover, visual interpretation is subjective, tedious and time-consuming (Zhang, Fritts, and Goldman 2008). It can be even practically impossible if several different segmentations over large areas are to be evaluated.

Automated segmentation goodness evaluation could be carried out using landscape or shape metrics and thus unsupervised evaluation (Neubert and Meinel 2003; Meinel and Neubert 2004; Neubert, Herold, and Meinel 2008; Li et al. 2010; Ji et al. 2012). This is problematic though, since for instance the FNEA method uses shape metrics as parameters which the user can modify. Hence, using shape metrics also in evaluation could lead to circular reasoning. Another possible solution in finding good segmentation evaluation measures could be focusing on boundaries. In other words, it could be examined whether boundaries drawn in the reference map are found in segmentation. For instance, Neubert and Herold (2008) measured what proportion of a segment's perimeter is inside a specific reference polygon's buffer zone. In a similar vein, Lucieer and Stein (2002) proposed a boundary-based measure, and Clinton et al. (2010) included a modification of this measure in their analysis. Whereas Lucieer and Stein (2002) calculated the shortest distances from reference polygons to any boundary pixel in segmentation, Clinton et al. (2010) averaged all distances to all segments inside the reference polygon. Of these measures, the original measure by Lucieer and Stein (2002) is more attractive, since boundaries inside a reference polygon can disappear in classification but boundaries near a reference polygon cannot be moved. However, Lucieer and Stein (2002) noted that the finest segmentations ranked the best using this evaluation. Taking this into account, they modified the original measure to take the length of boundary into account. These kinds of modifications, on the other hand, are difficult to design, because they easily favour either under-segmentation or over-segmentation. Furthermore, boundaries of natural objects are not exact. Hence, it is not always meaningful to find the 'real' boundaries but boundaries that are visible in data.

Finally, unsupervised segmentation evaluation methods that often measure inter-segment and intra-segment homogeneity or heterogeneity have been found useful in segmentation evaluation (Kim, Madden, and Warner 2009; Gao et al. 2011; Yue et al. 2012; Hou et al. 2013). Although in our case the objective was to find segmentation that mimics reference polygons, it could be interesting to test whether unsupervised methods work well in this kind of task.

5. Conclusion

We tested different supervised segmentation goodness evaluation measures and visual interpretation to determine a good segmentation for boreal forest habitat mapping. Although the different supervised segmentation goodness measures were fast to calculate from several segmentations, they provided inconsistent results. In other words, different segmentations stood out as being best when different measures were used. Visual interpretation, on the other hand, was tedious and segmentations could not be evaluated thoroughly in reasonable time. Although we selected only one segmentation as being the best based on visual interpretation, other segmentations were visually good. In classification analysis, the visually selected segmentation provided the best classification accuracy but differences between different segmentations were rather small. Better segmentation may lead to better classification, but there are several different definitions for good segmentation. Therefore, the relationship between segmentation and classification is not straightforward. We propose that the goodness of segmentation should be defined in each case separately and evaluation measures should be selected based on that definition. In our case, good segmentation was segmentation with (1) meaningful and not too complex segments, (2) boundaries parallel to reference polygon boundaries even for the smallest reference polygons, but (3) as coarse as possible. There were, however, no evaluation measures to determine these kinds of segmentations automatically. Overall, the best segmentations were FNEA segmentations with both imagery and lidar data. We conclude that different segmentation evaluation methods should be used with care, especially in natural environment mappings. When segmentation evaluation is rigorously used, however, it can assist in finding a more optimal segmentation. Quantitative segmentation evaluation might provide better results in urban environments; however, a more thorough testing is needed to support this claim.

Acknowledgements

This research was funded by the Maj and Tor Nessling foundation. We are grateful to the City of Jyväskylä for providing us with the reference polygon data set and the aerial imagery as well as to the Finnish Forest and Park Service for the reference polygon data set. MK received, from the Jenny and Antti Wihuri foundation, a sabbatical scholarship for the year 2013 that was partly funded also by the EU IMPERIA project (LIFE11 ENV/FI/905) and the University of Jyväskylä.

References

Ahti, T., L. Hämet-Ahti, and J. Jalas. 1968. “Vegetation Zones and their Sections in Northwestern Europe.” *Annales Botanici Fennici* 5: 169–211.

Baatz, M., and A. Schäpe. 2000. “Multiresolution Segmentation – An Optimization Approach for High Quality Multi-Scale Image Segmentation.” In *Angewandte Geographische Informationsverarbeitung XII*, edited by J. Strobl and G. Griesebner, 12–23. Heidelberg: Wichmann.

Bar Massada, A., R. Kent, L. Blank, A. Perevolotsky, L. Hadar, and Y. Carmel. 2012. “Automated Segmentation of Vegetation Units in a Mediterranean Landscape.” *International Journal of Remote Sensing* 33: 346–364.

Benz, U. C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. 2004. “Multi-Resolution, Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-Ready Information.” *ISPRS Journal of Photogrammetry and Remote Sensing* 58: 239–258.

Blaschke, T. 2010. “Object Based Image Analysis for Remote Sensing.” *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2–16.

Bock, M., P. Xofis, J. Mitchley, G. Rossner, and M. Wissen. 2005. “Object-Oriented Methods for Habitat Mapping at Multiple Scales – Case Studies from Northern Germany and Wye Downs, UK.” *Journal of Nature Conservation* 13: 75–89.

Böhner, J., and T. Selige. 2006. "Spatial Prediction of Soil Attributes Using Terrain Analysis and Climate Regionalisation." In *SAGA – Analysis and Modelling Applications*, Vol. 115, edited by J. Böhner, K. R. McCloy, and J. Strobl, 13–28. Göttingen: Goltze (Göttinger Geographische Abhandlungen 115).

Breidenbach, J., E. Næsset, V. Lien, T. Gobakken, and S. Solberg. 2010. "Prediction of Species Specific Forest Inventory Attributes Using a Nonparametric Semi-Individual Tree Crown Approach Based on Fused Airborne Laser Scanning and Multispectral Data." *Remote Sensing of Environment* 114: 911–924.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Breiman, L., and A. Cutler. 2007. "Random Forest: Classification Description." Accessed October 3, 2013. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Câmara, G., L. Vinhas, K. Reis Ferreira, G. Ribeiro de Queiroz, R. Cartaxo Modesto De Souza, A. Miguel Vieira Monteiro, M. Tílio de Carvalho, M. Antonio Casanova, and U. Moura De Freitas. 2008. "TerraLib: An Open Source GIS Library for Large-Scale Environmental and Socio-Economic Applications." In *Open Source Approaches in Spatial Data Handling*, edited by G. Brent Hall, and M. G. Leahy, 247–270. Berlin: Springer.

Carleer, A. P., O. Debeir, and E. Wolff. 2005. "Assessment of Very High Resolution Satellite Image Segmentations." *Photogrammetric Engineering & Remote Sensing* 71: 1285–1294.

Castilla, G., G. J. Hay, and J. R. Ruiz-Gallardo. 2008. "Size-Constrained Region Merging (SCRM): An Automated Delineation Tool for Assisted Photointerpretation." *Photogrammetric Engineering & Remote Sensing* 74: 409–419.

Chabrier, S., C. Rosenberger, B. Emile, and H. Laurent. 2008. "Optimization-Based Image Segmentation by Genetic Algorithms." *EURASIP Journal on Image and Video Processing* 2008: 842029. doi:10.1155/2008/842029.

Cherrill, A., and C. McClean. 1995. "An Investigation of Uncertainty in Field Habitat Mapping and the Implications for Detecting Land Cover Change." *Landscape Ecology* 10: 5–21.

Cherrill, A., and C. McClean. 1999. "The Reliability of 'Phase 1' Habitat Mapping in the UK: The Extent and Types of Observer Bias." *Landscape and Urban Planning* 45: 131–143.

Clinton, N., A. Holt, J. Scarborough, Li. Yan, and P. Gong. 2010. "Accuracy Assessment Measures for Object-Based Image Segmentation Goodness." *Photogrammetric Engineering & Remote Sensing* 76: 289–299.

Corcoran, P., A. Winstanley, and P. Mooney. 2010. "Segmentation Performance Evaluation for Object-Based Remotely Sensed Image Analysis." *International Journal of Remote Sensing* 31: 617–645.

Derivaux, S., G. Forestier, C. Wemmert, and S. Lefévre. 2010. "Supervised Image Segmentation Using Watershed Transform, Fuzzy Classification and Evolutionary Computation." *Pattern Recognition Letters* 31: 2364–2374.

Dingle Robertson, L., and D. J. King. 2011. "Comparison of Pixel- and Object-Based Classification in Land Cover Change Mapping." *International Journal of Remote Sensing* 32: 1505–1529.

dos Santos, J. A., P.-H. Gosselin, S. Phillipp-Foliguet, R. D. S. Torres, and A. X. Falcão. 2012. "Multiscale Classification of Remote Sensing Images." *IEEE Transactions on Geoscience and Remote Sensing* 50: 3764–3775.

Drăguţ, L., D. Tiede, and S. R. Levick. 2010. "ESP: A Tool to Estimate Scale Parameter for Multiresolution Image Segmentation of Remotely Sensed Data." *International Journal of Geographical Information Science* 24: 859–871.

Duro, D. C., S. E. Franklin, and M. G. Dubé. 2012. "A Comparison of Pixel-Based and Object-Based Image Analysis with Selected Machine Learning Algorithms for the Classification of Agricultural Landscapes Using SPOT-5 HRG Imagery." *Remote Sensing of Environment* 118: 259–272.

Espindola, G. M., G. Camara, I. A. Reis, L. S. Bins, and A. M. Monteiro. 2006. "Parameter Selection for Region-Growing Image Segmentation Algorithms Using Spatial Autocorrelation." *International Journal of Remote Sensing* 27: 3035–3040.

Eysn, L., M. Hollaus, K. Schadauer, and N. Pfeifer. 2012. "Forest Delineation Based on Airborne LIDAR Data." *Remote Sensing* 4: 762–783. doi:10.3390/rs4030762.

Falkowski, M. J., M. A. Wulder, J. C. White, and M. G. Gillis. 2009. "Supporting Large-Area, Sample-Based Forest Inventories with Very High Spatial Resolution Satellite Imagery." *Progress in Physical Geography* 33: 403–423.

Feitosa, R. Q., G. A. O. P. Costa, T. B. Cazes, and B. Feijo. 2006. "A Genetic Approach for the Automatic Adaptation of Segmentation Parameters." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (4/C42). http://www.isprs.org/proceedings/XXXVI/4-C42/Papers/11_Adaption%20and%20further%20development%20III/OBIA2006_Feitosa_et_al.pdf.

Freeman, T. G. 1991. "Calculating Catchment Area with Divergent Flow Based on a Regular Grid." *Computers and Geosciences* 17: 413–422.

Gao, Y., J. F. Mas, N. Kerle, and J. A. Navarrete Pacheco. 2011. "Optimal Region Growing Segmentation and its Effect on Classification Accuracy." *International Journal of Remote Sensing* 32: 3747–3763.

Geerling, G. W., M. Labrador-Garcia, J. P. G. W. Clevers, A. M. J. Ragas, and A. J. M. Smits. 2007. "Classification of Floodplain Vegetation by Data-Fusion of Spectral (CASI) and LiDAR Data." *International Journal of Remote Sensing* 28: 4263–4284.

Geerling, G. W., M. J. Vreeken-Buijs, P. Jesse, A. M. J. Ragas, and A. J. M. Smits. 2009. "Mapping River Floodplain Ecotopes by Segmentation of Spectral (CASI) and Structural (LiDAR) Remote Sensing Data." *River Research and Applications* 25: 795–813.

Gonzales, R. C., and R. E. Woods. 2002. *Digital Image Processing*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Grabs, T., J. Seibert, K. Bishop, and H. Laudon. 2009. "Modeling Spatial Patterns of Saturated Areas: A Comparison of the Topographic Wetness Index and a Dynamic Distributed Model." *Journal of Hydrology* 373: 15–23.

Güntner, A., J. Seibert, and S. Uhlenbrook. 2004. "Modeling Spatial Patterns of Saturated Areas: An Evaluation of Different Terrain Indices." *Water Resources Research* 40: W05114. doi:10.1029/2003WR002864.

Hay, G. J., T. Blaschke, D. J. Marceau, and A. Bouchard. 2003. "A Comparison of Three Image-Object Methods for the Multiscale Analysis of Landscape Structure." *ISPRS Journal of Photogrammetry and Remote Sensing* 57: 327–345.

Hay, G. J., G. Castilla, M. A. Wulder, and J. R. Ruiz. 2005. "An Automated Object-Based Approach for the Multiscale Image Segmentation of Forest Scenes." *International Journal of Applied Earth Observation and Geoinformation* 7: 339–359.

Hou, Z., Q. Xu, T. Nuutinen, and T. Tokola. 2013. "Extraction of Remote Sensing-Based Forest Management Units in Tropical Forests." *Remote Sensing of Environment* 130: 1–10.

Isenburg, M. 2011. *LAStools: Efficient Tools for LiDAR Processing*. Version 110815. http://lastools.org

Ji, X., Y. Bo, J. Cheng, Y. He, and X. Liu. 2012. "The Method of Assessing Shape Similarity of Object-Based Classification Result of Remote Sensing Imagery." In *Proceedings of the Second International Workshop on Earth Observation and Remote Sensing Applications (EORSA 2012)*, 157–160. IEEE. http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6248407/

Ke, Y., L. J. Quackenbush, and J. Im. 2010. "Synergistic Use of QuickBird Multispectral Imagery and LIDAR Data for Object-Based Forest Species Classification." *Remote Sensing of Environment* 114: 1141–1154.

Kim, M., M. Madden, and T. A. Warner. 2008. "Estimation of Optimal Image Object Size for the Segmentation of Forest Stands with Multispectral IKONOS Imagery." In *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, edited by T. Blaschke, S. Lang, and G. J. Hay, 291–307. Berlin: Springer.

Kim, M., M. Madden, and T. A. Warner. 2009. "Forest Type Mapping Using Object-Specific Texture Measures from Multispectral Ikonos Imagery: Segmentation Quality and Image Classification Issues." *Photogrammetric Engineering & Remote Sensing* 75: 819–829.

Kim, M., T. A. Warner, M. Madden, and D. S. Atkinson. 2011. "Multi-Scale GEOBIA with Very High Spatial Resolution Digital Aerial Imagery: Scale, Texture and Image Objects." *International Journal of Remote Sensing* 32: 2825–2850.

Kopecký, M., and Š. Čížková. 2010. "Using Topographic Wetness Index in Vegetation Ecology: Does the Algorithm Matter?" *Applied Vegetation Science* 13: 450–459.

Lang, S., F. Albrecht, S. Kienberger, and D. Tiede. 2010. "Object Validity for Operational Tasks in a Policy Context." *Journal of Spatial Science* 55: 9–22.

Lawrence, R. L., S. D. Wood, and R. L. Sheley. 2006. "Mapping Invasive Plants Using Hyperspectral Imagery and Breiman Cutler Classifications (RandomForest)." *Remote Sensing of Environment* 100: 356–362.

Leckie, D. G., F. A. Gougeon, N. Walsworth, and D. Paradine. 2003. "Stand Delineation and Composition Estimation Using Semi-Automated Individual Tree Crown Analysis." *Remote Sensing of Environment* 85: 355–369.

Li, H., H. Gu, Y. Han, and J. Yang. 2010. "Object-Oriented Classification of High-Resolution Remote Sensing Imagery Based on an Improved Colour Structure Code and a Support Vector Machine." *International Journal of Remote Sensing* 31: 1453–1470.

Li, N., H. Huo, and T. Fang. 2010. "A Novel Texture-Preceded Segmentation Algorithm for High-Resolution Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 48: 2818–2828.

Liaw, A., and M. Wiener. 2002. "Classification and Regression by randomForest." *R News* 2/3: 18–22.

Lucieer, A., and A. Stein. 2002. "Existential Uncertainty of Spatial Objects Segmented from Satellite Sensor Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 40: 2518–2521.

Marpu, P. R., M. Neubert, H. Herold, and I. Niemeyer. 2010. "Enhanced Evaluation of Image Segmentation Results." *Journal of Spatial Science* 55: 55–68.

Meinel, G., and M. Neubert. 2004. "A Comparison of Segmentation Programs for High Resolution Remote Sensing Data." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (B4): 1097–1105. http://www.isprs.org/proceedings/XXXV/congress/comm4/papers/506.pdf.

Möller, M., L. Lymburner, and M. Volk. 2007. "The Comparison Index: A Tool for Assessing the Accuracy of Image Segmentation." *International Journal of Applied Earth Observation and Geoinformation* 9: 311–321.

Murphy, P. N. C., J. Ogilvie, and P. Arp. 2009. "Topographic Modeling of Soil Moisture Conditions: A Comparison and Verification of Two Models." *European Journal of Soil Science* 60: 94–109.

Mustonen, J., P. Packalén, and A. Kangas. 2008. "Automatic Segmentation of Forest Stands Using a Canopy Height Model and Aerial Photography." *Scandinavian Journal of Forest Research* 23: 534–545.

Neubert, M., and H. Herold. 2008. "Assessment of Remote Sensing Image Segmentation Quality." *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (4/C1). http://www.isprs.org/proceedings/XXXVIII/4-C1/Sessions/Session2/6721_Neubert_Proc_pap.pdf.

Neubert, M., H. Herold, and G. Meinel. 2008. "Assessing Image Segmentation Quality – Concepts, Methods and Applications." In *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, edited by T. Blaschke, S. Lang, and G. J. Hay, 769–784. Berlin: Springer.

Neubert, M., and G. Meinel. 2003. "Evaluation of Segmentation Programs for High Resolution Remote Sensing Applications." In *Proceedings Joint ISPRS/EARSeL Workshop "High Resolution Mapping from Space 2003"*, edited by M. Schroeder, K. Jacobsen, and C. Heipke, Hannover, October 6–8. http://www.ipi.uni-hannover.de/fileadmin/institut/pdf/neubert.pdf

Pekkarinen, A. 2002. "Image Segment-Based Spectral Features in the Estimation of Timber Volume." *Remote Sensing of Environment* 82: 349–359.

R Development Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org

Radoux, J., and P. Defourny. 2007. "A Quantitative Assessment of Boundaries in Automated Forest Stand Delineation Using Very High Resolution Imagery." *Remote Sensing of Environment* 110: 468–475.

Räsänen, A., M. Kuitunen, E. Tomppo, and A. Lensu. Forthcoming. "Coupling High-Resolution Satellite Imagery with LiDAR-Based Canopy Height Model and Digital Elevation Model in Object-Based Boreal Forest Habitat Type Classification." Manuscript submitted for publication.

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. "An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93–104.

Smith, A. 2010. "Image Segmentation Scale Parameter Optimization and Land Cover Classification Using the Random Forest Algorithm." *Journal of Spatial Science* 55: 69–79.

Sørensen, R., U. Zinko, and J. Seibert. 2006. "On the Calculation of the Topographic Wetness Index: An Evaluation of Different Methods Based on Field Observations." *Hydrology and Earth System Sciences* 10: 101–112.

Tian, J., and D.-M. Chen. 2007. "Optimization in Multi-Scale Segmentation of High-Resolution Satellite Images for Artificial Feature Recognition." *International Journal of Remote Sensing* 28: 4625–4644.

Verbeeck, K., M. Hermy, and J. van Orshoven. 2012. "External Geo-Information in the Segmentation of VHR Imagery Improves the Detection of Imperviousness in Urban Neighborhoods." *International Journal of Applied Earth Observation and Geoinformation* 18: 428–435.

Vesterbacka, R. 2010. *Luontopalvelujen Luontotyyppi-Inventoinnin Maastotyöohje*. Vantaa: Metsähallitus.

Wang, L., W. Sousa, and P. Gong. 2004. "Integration of Object-Based and Pixel-Based Classification for Mapping Mangroves with IKONOS Imagery." *International Journal of Remote Sensing* 25: 5655–5668.

Wang, Z., J. R. Jensen, and J. Im. 2010. "An Automatic Region-Based Image Segmentation Algorithm for Remote Sensing Applications." *Environmental Modelling and Software* 25: 1149–1165.

Weidner, U. 2008. "Contribution to the Assessment of Segmentation Quality for Remote Sensing Applications." *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (B7): 479–484. http://www.isprs.org/proceedings/XXXVII/congress/7_pdf/4_WG-VII-4/01.pdf.

Whiteside, T. G., G. S. Boggs, and S. W. Maier. 2011. "Comparing Object-Based and Pixel-Based Classifications for Mapping Savannas." *International Journal of Applied Earth Observation and Geoinformation* 13: 884–893.

Wulder, M. A., J. C. White, G. J. Hay, and G. Castilla. 2008. "Towards Automated Segmentation of Forest Inventory Polygons on High Spatial Resolution Satellite Imagery." *The Forestry Chronicle* 84: 221–230.

Yang, L., F. Albregtsen, T. Lønnestad, and P. Grøttum. 1995. "A Supervised Approach to the Evaluation of Image Segmentation Methods." In *Computer Analysis of Images and Patterns: 6th International Conference*, *CAIP '95 Prague Czech Republic*, *September 6–8*, *1995 Proceedings*, edited by V. Hlaváč, and R. Šára, Vol. 970, 759–765. Lecture Notes on Computer Science. Heildelberg: Springer.

Yue, A., J. Yang, C. Zhang, W. Su, W. Yun, D. Zhu, S. Liu, and Z. Wang. 2012. "The Optimal Segmentation Scale Identification Using Multispectral WorldView-2 Images." *Sensor Letters* 10: 285–291.

Zhan, Q., M. Molenaar, K. Tempfli, and W. Shi. 2005. "Quality Assessment for Geo-Spatial Objects Derived from Remotely Sensed Data." *International Journal of Remote Sensing* 26: 2593–2974.

Zhang, H., J. E. Fritts, and S. A. Goldman. 2008. "Image Segmentation Evaluation: A Survey of Unsupervised Methods." *Computer Vision and Image Understanding* 110: 260–280.

Zhang, Y. J. 1996. "A Survey on Evaluation Methods for Image Segmentation." *Pattern Recognition* 29: 1335–1346.

Zhang, Y. J. 1997. "Evaluation and Comparison of Different Segmentation Algorithms." *Pattern Recognition Letters* 18: 963–974.

II

COUPLING HIGH-RESOLUTION SATELLITE IMAGERY WITH ALS-BASED CANOPY HEIGHT MODEL AND DIGITAL ELEVATION MODEL IN OBJECT-BASED BOREAL FOREST HABITAT TYPE CLASSIFICATION

by

Aleksi Räsänen, Markku Kuitunen, Erkki Tomppo & Anssi Lensu

Accepted to ISPRS Journal of Photogrammetry and Remote Sensing

III

COMPARISON OF DIFFERENT REMOTE SENSING AND GIS METHODS TO EVALUATE CONSERVATION VALUES AND ECOSYSTEM SERVICES IN SOUTHERN FINLAND

by

Aleksi Räsänen, Anssi Lensu, Erkki Tomppo & Markku Kuitunen

Submitted manuscript

IV

THE ROLE OF LANDSCAPE, TOPOGRAPHY, AND GEODIVERSITY IN EXPLAINING AND PREDICTING VASCULAR PLANT SPECIES RICHNESS IN A FRAGMENTED LANDSCAPE, SOUTHERN FINLAND

by

Aleksi Räsänen, Markku Kuitunen, Jan Hjort, Asta Vaso, Tuomo Kuitunen & Anssi Lensu

Submitted manuscript