# Monitoring biodiversity in cultural landscapes: development of remote sensing- and GIS-based methods

Dalmayne, Jonas

2016

*Citation for published version (APA):*
Dalmayne, J. (2016). *Monitoring biodiversity in cultural landscapes: development of remote sensing- and GIS-based methods*. [Doctoral Thesis (compilation), Dept of Physical Geography and Ecosystem Science]. Lund University, Faculty of Science, Department of Physical Geography and Ecosystem Science.

*Total number of authors:*
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

# Monitoring biodiversity in cultural landscapes: development of remote sensing- and GIS-based methods

# Monitoring biodiversity in cultural landscapes: development of remote sensing- and GIS-based methods

Jonas Dalmayne

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Science, Lund University, Sweden.
To be defended in Världen (room 111), Geocentrum I, Friday 17 June at 10:00

*Faculty opponent*

Dr. Nathalie Pettorelli,

Zoological Society of London

| | |
|---|---|
| Organization<br>LUND UNIVERSITY<br>Department of Physical Geography and Ecosystem Science, Sölvegatan 12, 22362 Sweden | Document name<br>Doctoral dissertation |
| | Date of issue<br>May 2016 |
| Author: Jonas Dalmayne | Sponsoring organization |

Title and subtitle

Monitoring biodiversity in cultural landscapes: development of remote sensing- and GIS-based methods.

Abstract

In this thesis, I explore the relationships between structural and compositional landscape properties, and species diversity, using remotely sensed data on a variety of spatial scales. The thesis shows that increased landscape heterogeneity, measured using environmental and spectral variables that were used both separately and combined, is generally positively related to plant species richness. I further found that plant species richness could be predicted with <20% deviance in species numbers, in approximately 80% of the study area within the province of Scania, using a combination of environmental and spectral descriptors of landscape heterogeneity. Further, I used Landsat satellite data, aided by ancillary data on topography and a spectral proxy of seasonal variation in vegetation phenology, to classify historical (ca 1975) and contemporary (ca 2001) land-cover data within the province of Scania, with +85% accuracy. The produced land use/land cover (LULC) data showed correlations with levels of plant species richness, with the proportion of cropland generally being negatively correlated to levels of plant species richness, and the proportion of LULC classes such as grazed grassland, wetland and deciduous forest being positively correlated to levels of plant species richness. Further, the positive change between the historical and contemporary landscapes in the proportion of deciduous forest, and in the number of unique LULC patches, were positively correlated with contemporary levels of plant species richness. I modeled the importance of non-crop habitat types for plant species richness within the province of Scania, and showed that for the promotion of plant species richness, the most wide-spread non-crop LULC types were most important within the most simplified landscapes, while the amount of non-crop small biotopes were most important in more complex landscapes.

In a series of studies on grazed grasslands on the Baltic island of Öland, I showed that dissimilarity in Worldview-2 satellite spectral reflectance was related to plant species dissimilarity within a set of grassland plots, and then used spectral dissimilarity to predict levels of plant species richness in other grassland plots. I used HySpex hyperspectral aerial reflectance data to predict plant species diversity (species richness and Simpson's diversity), using the full range of wavebands and also using a reduced set of wavebands. Finally, I classified grassland plots into age classes using reflectance data from the HySpex hyperspectral sensor, and achieved better classification results when using a reduced set of wavebands compared to using the full range of wavebands.

In summary, the findings of this thesis demonstrate that remote sensing and GIS-based methods can be useful tools in the monitoring of cultural landscapes, because of their combined ability to model landscape properties and relate those measures to species diversity, at a range of spatial scales and within a range of habitats.

Key words

Remote sensing, plant species diversity, landscape complexity, heterogeneity, Landsat, Worldview-2, HySpex, LULC, grazed grasslands, support vector machines, partial least squares, spatial scales, statistical modeling,

| | | |
|---|---|---|
| Classification system and/or index terms (if any) | | |
| Supplementary bibliographical information | | Language: English |
| ISSN and key title | | ISBN 978-91-85793-57-0 |
| Recipient's notes | Number of pages<br>225 | Price |
| | Security classification | |

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Jonas Dalmayne Date 2016-05-17

# Monitoring biodiversity in cultural landscapes: development of remote sensing- and GIS-based methods

Jonas Dalmayne

LUND
UNIVERSITY

Cover by Jonas Dalmayne

Faculty of Science
Department of Physical Geography and Ecosystem Science
Lund University

ISBN 978-91-85793-57-0

Printed in Sweden by Media-Tryck, Lund University
Lund 2016

# Thank you!

To family, friends and coworkers, for supporting me through this process, each in your own way.

To Honor C. Prentice, for your input, patience, and genuine interest in my work and well-being.

To Karin Hall. For everything. This didn’t turn out the way either of us had intended, or perhaps hoped for. But now it’s done, and I owe that to you. Thank you.

# Content

# List of papers

I. **Dalmayne, J.,** Irminger-Street, T., Olsson, O., Prentice, H. C., Hall, K. Modelling plant species richness in cultural landscapes using public-domain data on environmental and spectral heterogeneity. *(in preparation)*

II. **Dalmayne, J.,** Prentice, H. C., Hall, K. Comparisons of historical and contemporary Landsat data reveal relationships between present-day plant species richness and patterns of land-cover change in cultural landscapes. *(in preparation)*

III. Irminger-Street, T., **Dalmayne, J.,** Olsson, O., Prentice, H. C., Smith, H. G., Tyler, T. T., Hall, K. Small fragments of non-crop habitat make a significant contribution to landscape-scale vascular plant richness in agricultural landscapes. *(in preparation)*

IV. **Dalmayne, J.,** Möckel, T., Prentice, H. C., Schmid, B. C., Hall, K. (2013) Assessment of fine-scale plant species beta diversity using Worldview-2 satellite spectral dissimilarity. Ecological Informatics 18:1-9.

V. Möckel, T., **Dalmayne, J.,** Schmid, B. C., Prentice, H. C., Hall, K. (2016) Airborne hyperspectral data predict fine-scale plant species diversity in grazed dry grasslands. Remote Sensing. 8, 133.

VI. Möckel, T., **Dalmayne, J.,** Prentice, H. C., Eklundh, L., Purschke, O., Schmidtlein, S., Hall, K. (2014) Classification of grassland successional stages using airborne hyperspectral imagery. Remote Sensing, 6, 7732-7761.

# List of contributions

I. Dalmayne led the design of the study, performed data preparation and statistical analysis, led the interpretation of the results and led the writing.

II. Dalmayne led the design of the study, performed data preparation and statistical analysis, led the interpretation of the results and led the writing.

III. Dalmayne performed data preparation and GIS-based analysis and participated in general discussions.

IV. Dalmayne led the design of the study, participated in the field data sampling, performed data preparation and statistical analysis, led the interpretation of the results and led the writing.

V. Dalmayne contributed to the design of the study, participated in the field data sampling, contributed to the interpretation of the results and contributed to the writing.

VI. Dalmayne contributed to the design of the study, participated in the field data sampling, contributed to the interpretation of the results and contributed to the writing.

# Populärvetenskaplig sammanfattning

## Övervakning av biodiversitet i kulturlandskap – metodutveckling baserad på fjärranalys och GIS

Kulturlandskapet är starkt knutet till människan och hennes aktiviteter. Flera tusen år av samspel mellan människa och natur har skapat landskap som definierar levnadsvillkoren för en mångfald av arter. Men när samhället förändras, förändras även kulturlandskapet. De forna årtusendenas långsamma och gradvisa förändringar har accelererats av människans teknologiska framsteg. I södra Sverige har jordbruksreformer, konstgödsel, modernt skogsbruk och urbanisering omvandlat kulturlandskapet på bara 200 år. Men mycket av biodiversiteten i landskapet är format av och beroende av habitat, landskapsstruktur och skötsel som de såg ut för 200 år sedan. För att kunna bevara så mycket som möjligt av vår biodiversitet måste vi övervaka kulturlandskapet och nya kostnadseffektiva, storskaliga metoder måste tas fram för att underlätta bevarandearbetet. Att övervaka och samla in data över kulturlandskapet med hjälp av satellit- och flygbaserade sensorer (fjärranalys) och behandla dessa data i datormiljö, kan vara nyckeln till ett effektivt övervakningssystem.

I denna avhandling undersöker jag framförallt kopplingarna mellan artrikedomen av växter och kulturlandskapets egenskaper, med hjälp av datorbaserad behandling av flyg- och satellitdata, samt statistisk modellering. Jag har med hjälp av satellitdata och officiell rumslig och statistisk data utfört tre studier i delar av Skåne, med målsättningen att kunna koppla heterogeniteten i landskapet och landskapets sammansättning vad gäller markanvändning, till nivåer av artrikedom av växter. Med hjälp av högupplöst satellitdata och hyperspektral flygdata har jag även genomfört tre studier i ett gräsmarksområde på Öland, där jag har undersökt kopplingarna mellan betade gräsmarkers spektrala reflektans och deras växtdiversitet, samt möjligheterna att åldersklassificera betade gräsmarker med hjälp av deras spektrala signaturer.

I den första studien baserad i Skåne visade jag att landskapets heterogenitet, uppmätt med mått från officiell Svensk Marktäckedata (rumsliga mått) och heterogeniteten i den spektrala reflektansen från Landsat-satellitdata (spektrala mått), är kopplad till artrikedomen av växter. Jag visade också att kopplingen mellan artrikedom av växter och uppmätt landskapsheterogenitet är särskilt stark om man kombinerar rumsliga och spektrala mått. Med hjälp av kombinationer av rumsliga och spektrala heterogenitetsmått lyckades jag sedan bygga modeller som förutsade artrikedomen av växter med <20% felmarginal i ca 80% av studieområdet.

I den andra studien baserad i Skåne skapade jag först historisk (1975) och nutida (2001) markanvändningsdata, baserad på Landsat-satellitdata, officiell höjddata, samt årstidsskillnader i ett spektralt baserat mått på växtlighetens fenologi. Jag visade sedan att både historisk och nutida markanvändning och landskapsheterogenitet är kopplade till nutida artrikedom av växter. Jag visade att proportionen åkermark i landskapet var negativt kopplad till artrikedom av växter, medan proportionerna av våtmark, lövskog och betesmark generellt var positivt kopplade till växternas artrikedom. Jag visade också att en ökning i proportionen av lövskog mellan det historiska och nutida landskapet kan påverka artrikedomen av växter positivt och att detta också gällde när heterogeniteten i landskapet ökade. I den tredje studien i Skåne visade jag att artrikedomen av växter i landskap som är enkla i sin sammansättning och heterogenitet, bäst förklarades av proportionerna av de vanligaste markanvändningsklasserna som inte var åker. Jag visade även att i de mest komplexa landskapen så var det mängden av småbiotoper (t.ex. märgelgravar, häckar, vägrenar) inom landskapet som förklarade artrikedomen av växter bäst.

I den första av studierna på Öland visade jag att skillnader mellan betade gräsmarker i deras spektrala signaturer, uppmätta med data från Worldview-2-satelliten, var kopplade till skillnader i deras artsammansättning av växter. Jag förutsade även skillnader i gräsmarkernas artsammansättning av växter genom att mäta skillnaderna i deras spektrala signaturer. I den andra studien på Öland använde jag data från den hyperspektrala flygbaserade sensorn HySpex för att förutsäga artdiversitet (artrikedom och Simpson’s index) av växter inom betade gräsmarker. Jag visade att de bästa resultaten kom när man använde samtliga 245 spektrala band för att mäta gräsmarkernas reflektans. I den tredje studien på Öland använde jag återigen data från HySpex-sensorn för att klassificera betade gräsmarker i tre olika åldersklasser (5-15 år, 16-50 år och >50 år). Det bästa resultatet fick jag när jag började med att välja ut de spektrala band (177 band) som var viktigast för att kunna särskilja åldersklasserna.

Sammanfattningsvis så visar mina resultat att metoder baserade på fjärranalys och GIS (geografiska informationssystem) kan vara mycket användbara verktyg för att effektivt kunna övervaka biodiversitet inom kulturlandskapet. Detta både genom deras förmåga att modellera landskapets struktur och sammansättning och genom deras förmåga att relatera dessa modeller till faktisk biodiversitet i ett brett spektrum av rumsliga skalor, habitat och landskap.

# Introduction

## Cultural landscapes

The cultural landscape has been defined as the "cultural properties that represent the combined works of nature and of man." by the World Heritage Committee (UNESCO 2012). The cultural landscape is shaped by a strong relationship between land-use and natural resources, including abiotic factors such as e.g. climate, water availability, soil and bedrock. These resources and conditions have been the main drivers of the land-use composition in the Nordic countries during most of the last 6000 years (Ihse 1995). However, anthropogenic activities have played an increasingly important role as civilisation evolved, as thousands of years of interaction between man and environment have created landscapes that define the living conditions for a multitude of species. As such, cultural landscapes reveal much about our evolving relationship with the natural world, because as society changes, so does the cultural landscape. Cultural landscapes are intrinsically linked to human activities, which have modified, and continue to modify, the general structure and function of landscape features, and their impacts on habitat conditions and the species inhabiting these landscapes (Martínez et al. 2010). Much of the present biodiversity within the European cultural landscapes is tied to habitats developed centuries ago, and to the agricultural practices of those times, whose management regimes are no longer economically viable (Vos & Meekes 1999). The intensification and extensification of agricultural practices have re-shaped the cultural landscapes, and have led to a transformation towards a simpler landscape structure (Benton et al. 2003). To monitor the cultural landscape has become increasingly important in order to quantify its composition and structural complexity, their change in space and time, and to learn their roles in the distribution of species, and the habitats that sustain those species. For this, we need to develop cost-effective yet efficient methods capable of large-scale assessments. Methods based on remote sensing and GIS may offer key toolsets for future monitoring systems.

# The development of cultural landscapes in the south of Sweden

The earliest signs of anthropogenic activity affecting the land-cover in the south of Sweden was seen during the Neolithic age (ca. 4000 BC), when hunter-gatherer communities started to adopt slash-and-burn and clearcutting agricultural techniques (Eriksson et al. 2002). This led to patchiness in the pristine vegetation, which was probably mainly dense broad leaved deciduous forest (Eriksson et al. 2002). Eventually, as more permanent residences and agricultural practices started to develop, the concept of the village took on a new importance, as the processes surrounding it started to shape the countryside landscape (Emanuelsson et al. 2002). A system of “infields”, which mainly consisted of arable fields and meadows, developed close to the villages, while so-called “outfields”, positioned further out from the populated areas, were used as pastures and for collecting fodder and wood for fuel (Widgren 1983). In the early stages of the system, outfields were generally used as commons, with no real ownership attached and free general use. As the exploatation of the commons increased and the amount of forest decreased, boundaries and borders between counties and villages were established (Emanuelsson et al. 2002). This general system of land use within the cultural landscape persisted into the late 18th century. Starting in the early 19th century, the cultural landscape changed dramatically (Emanuelsson et al. 2002). Villages were broken up, and farmsteads were spread out over the landscape. Arable fields increased in size, and the development of artificial fertilisers increased production capacity (Emanuelsson et al. 2002). An increased production of ley reduced the need for pastures and haymaking in semi-natural grasslands and meadows, thereby diminishing the areas of these habitats (Eriksson et al 2002). Many of the traditionally managed grasslands that survived this initial period of change were abandoned after the 1940’s and developed into forest (Eriksson et al. 2002). Generally, forests were re-established on the former outfields, either through active plantation activities or by the cessation of grazing activities (Emanuelsson et al. 2002), and active forestry as a means of income was increased (Emanuelsson et al. 2002). Due to economic factors, fast growing coniferous species, mainly spruce, came to dominate the new plantations. As with the agricultural areas, the forested areas changed towards a simpler, more homogenic landscape structure.

## Grasslands

Grasslands in the south of Sweden can generally be divided into two categories; improved grasslands and semi-natural grasslands. Both have their origins in anthropogenic activities, and are also maintained through further anthropogenic involvement (Ihse & Lindahl 2000). Semi-natural grasslands were an integral part of the traditional Swedish cultural landscape as long as the outfields were still used for grazing, and their remnants are still of particular importance to the biodiversity within the cultural landscape. They are generally characterised by low nutrient availability, as grazing activites remove nutrients, and as they are not improved by artificial fertilising, although they may have their origin on former fertilised arable fields (Bullock 2011). They are also not improved by agricultural practices such as plowing or seed-sowing (Bullock 2011). These conditions favour grassland specialist species and other species with poor ability to compete in the more common high-nutrient environments. Consequently, semi-natural grasslands generally have a high level of plant species richness, as well as a high proportion of grassland specialist species (Reitalu et al. 2009), making them habitats of high conservation value (Reitalu et al 2010). They also contribute to regulatory ecosystem services, socio-cultural ecosystem services (recreation, education and esthetic values) (Bullock 2011), and can also be considered as a cultural heritage, owing to their importance within the concept of the historical agricultural landscape. Levels of plant species richness within semi-natural grasslands are dependent on both former (Lindborg & Eriksson 2004) and present land use, as well as other physical properties of the landscape such as topography, soil type, and landscape context. Stochastic processes can also have an impact on plant species richness (Öster et al. 2007) in semi-natural grasslands.

## Changes to the cultural landscape, and their effects on biodiversity

Landscape heterogeneity (complexity) is a direct result of the number of land use/land cover (LULC) classes and the distribution and configuration of these within an area (Turner et al. 2001). In order to study changes in the complexity of landscapes, we need information on change in LULC, which has been considered the most important variable of global change impacting ecological systems (Vitousek 1994, Foody 2002), and rivals climate change as the largest variable affecting the global environment (Skole 1994, Foody 2002). Change in LULC has also been described as the largest global threat to biological diversity (Chapin et al. 2000). It has been suggested that the main cause of the decline in biodiversity in agricultural areas is loss of spatial and temporal heterogeneity, i.e. farmland becoming more simplified (Benton et al. 2003, Persson et al. 2010).

Not only the overall landscape complexity, but also its composition, is of importance to biodiversity in cultural landscapes. Proportions of non-cropland LULC types have been shown to be positively associated with levels of species diversity in agricultural regions. The intensification of agricultural practices has also led to an added importance of smaller non-crop elements within the agricultural landscape, such as field margins, stone walls, and field islets. Non-crop elements are important to overall species diversity, as they provide refuge and resources vital to a variety of species (Duelli & Obrist 2003), and also aid in the dispersal of species within the landscape. Preserving and maintaining semi-natural elements within the agricultural landscape is also important for ecosystem services, such as those provided by pollinators or natural enemies of crop pests (Ricketts et al. 2004). In Sweden, field margins and other small biotopes of non-crop habitats have traditionally been held open by grazing activities and fodder-harvesting (Cousins & Eriksson 2001), and many management-intensive species are therefore tied to these habitats. As management has decreased with the cessation of traditional agricultural practices, ecologically important small biotopes have become increasingly overgrown by woody vegetation (Jordbruksverket 2006).

Semi-natural grasslands in Sweden have been greatly diminished through the large-scale cessation of grazing activities, resulting in a 90% loss in the area of traditionally managed semi-natural grasslands since the beginning of the 20th century (Bernes 1994). The rapid loss of semi-natural grassland habitats has been identified as a major reason for the decline seen in biodiversity throughout Europe (Tscharnke et al. 2005). Those traditionally managed grasslands that persist are mainly preserved through grazing on the prescription of various conservation agencies. Grazing is considered an essential part of preserving semi-natural grasslands (Reitalu et al. 2010), along with maintaining a certain amount of connectivity between grassland patches. However, remaining patches of semi-natural grasslands within the agricultural landscape have become increasingly isolated (Ihse 1995) and exposed to fertilisers and agrochemicals from the surrounding croplands, which may compromise their function as high-species richness habitats (Robinson & Sutherland 2002).

It has become increasingly clear that in order to fully understand the impact that changes to the complexity and composition of LULC has on biodiversity in the cultural landscape, we must focus on changes over space and time, and thus changes to landscape structure as a whole (Walz & Syrbe 2013). Analyses of species richness patterns have revealed that not only present land use, but also management continuity and land-use history has a strong influence on present day plant species richness (Eriksson & Eriksson 1997, Bruun et al. 2001, Eriksson et al. 2002). However, our knowledge of change in LULC and its dynamics is limited (Foody 2002). Accurate LULC data has not necessarily been readily available or easy to acquire (DeFries & Townsend 1994), and results from traditional field-based surveys quickly become outdated and irrelevant in the face of rapid environmental change (Zewdie & Csaplovics 2015). Reliable long term information on spatial and temporal changes in LULC brought on by anthropogenic activity is vital, as is the reliable evaluation of management actions (Pettorelli et al 2014).

# Remote sensing

Remote sensing, as defined for this thesis, relates to the detection of electromagnetic energy from the earth's surface from aircraft or satellite based sensors (Turner et al. 2003). Remote sensing offers a cost-effective way of monitoring the cultural landscape over a broad set of spatial and temporal scales.

## Remote sensing of vegetation

The reflectance properties of vegetation can vary between vegetation types, but their reflectance characteristics are built on variation within the same parameters; optical properties of plant litter, stem and leaves; canopy biophysical properties (e.g. area and orientation of leaves); density of the vegetation; illumination conditions; and viewing geometry (Asner 1998). The effect of the latter two, which are not properties of the vegetation itself but of the conditions in which we study them, can be minimised through correction procedures (Schaepman-Strub et al. 2006), leaving the former three as the main influences on the spectral properties that define spectral signatures of vegetation types.

Figure 1 shows a typical reflectance curve for green vegetation through the visible (0.4 µm – 0.7 µm), near infrared (NIR, 0.7 µm – 1.2 µm) and short wave infrared (SWIR, 1.2 µm – 2.6 µm) wavelengths. The visible wavelengths (blue, green, red) are characterised by low reflectance, mainly due to light absorption from leaf pigments, such as chlorophyll. The green colour of healthy vegetation is due to the chlorophylls absorption of more light in the blue and red wavelengths, which can be clearly seen in the reflectance peak in the green band. The reflectance in the NIR wavelengths are generally controlled by the structure of the vegetation, which affects air-cell interactions of photons within the mesophyll of the leaves (Woolley 1971), while reflectance in the SWIR wavelengths is mainly dependent on water absorption and non-pigment plant compounds such as nitrogen and lignin (Asner 1998). While these interactions are well understood and documented (e.g. Asner 1998, Kumar et al. 2001), actual canopy reflectance is harder to interpret due to interactions between multi-layered vegetation and the soil background (Homolova et al. 2013).

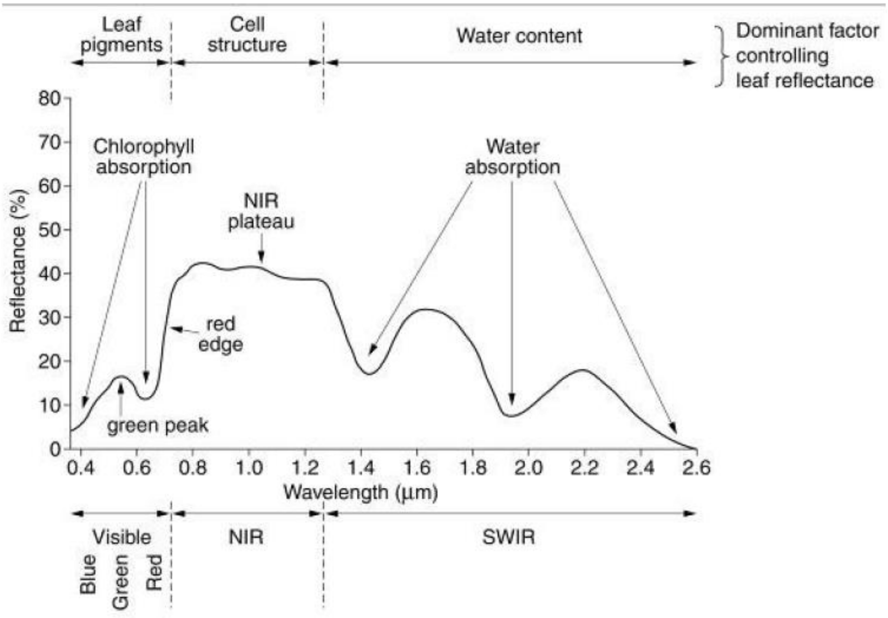


Figure 1. Typical spectral reflectance curve for vegetation. Modified from Medcalf et al. (2010).

The way vegetation properties are studied using remote sensing can broadly be categorised into two methods; physical methods, where simulations of plant-light interactions (Jaquemod et al. 2009) are used, and empirical methods, where statistical relationships between empirical data on vegetation properties and reflectance information is used (Ustin et al. 2009). Physical models are computationally intensive and generally require extensive parameterisation and calibration procedures before they can be applied (Liang 2005). Empirical methods are generally computationally faster, but are dependent on locally measured parameters from a given sensor, and are therefore hard to transfer to different spatial or temporal locations or to use with another sensor. It is also possible to combine physical and empirical methods (Liang 2005) in order to find the optimal solution for a given problem and situation.

## Vegetation indices

The aim of a vegetation index (VI) is to reduce the spectral response from a given number of spectral bands into a single number that can be used to find relationships between vegetation properties and spectral reflectance (Perry & Lautenschlager 1984). The most well-known VI is the Normalised Difference Vegetation Index (NDVI), which combines information from spectral bands in wavelengths corresponding to red and near infrared (NIR), calculated as NDVI =

(NIR – red) / (NIR + red). The NDVI was first used in the 1970's (Rouse et al. 1973) and is today established as the go-to VI for the study of vegetation properties (Pettoreli et al. 2005). It has e.g. been used as a proxy for vegetation productivity and biomass (e.g. Box et al. 1989, Pettoreli et al. 2005), and for the prediction of plant species richness (e.g. Gillespie 2005, Parviainen et al. 2010).

## Remote sensing for estimating species diversity of plants

The estimation and prediction of species diversity using remote sensing techniques can be accomplished using three types of methods; direct methods, where spectral reflectance is directly related to individual organisms or communities of organisms; indirect methods, where spectral reflectance is related to environmental properties that can act as proxies for species diversity (Turner et al. 2003); and an approach which involves the development of direct relationships between spectral radiance values, and species distribution patterns recorded from observations in the field (Nagendra 2001). Direct methods are generally limited by the mismatch between the spatial resolution of sensors and the size of species being studied, but advances in sensor technology has made it possible to study certain species assemblages and larger organisms, such as trees (Turner et al. 2003). Duro et al. (2007) identify four general methods for identifying biodiversity using indirect remote sensing techniques; through the study of (i) the earth's physical structure, such as climate and topography, (ii) vegetation productivity or function, (iii) habitat suitability with respect to structure and spatial arrangement, and, (iv) heterogeneity as a consequence of disturbance. The NDVI has been used frequently in method (ii) due to its ties to vegetation productivity and biomass (e.g. Rocchini 2009, Parviainen et al. 2010).

The structure and spatial arrangement of landscape features have been extensively used in species diversity studies using remote sensing (e. g. Deutschewitz et al. 2003, Honnay et al. 2003, Gillespie 2005), where remote sensing is generally used for the creation of spatial metrics based on spatially explicit landscape features, which are then related to species data. Related to this is using the remotely sensed measurements of the heterogeneity of landscape features as a proxy for species diversity, which has been done in a variety of studies at different spatial scales (c.f. Ricklefs 1977, Shmida & Wilson 1985, Huston 1994). Continuous spectral heterogeneity, where heterogeneity in spectral reflectance values is directly used, has also been related to species diversity (e.g. Rocchini et al. 2004, Parviainen et al. 2013). The origin for the use of continuous remotely sensed data for the production of heterogeneity measurements lies in the Spectral Variance Hypothesis (SVH, Palmer et al. 2000, Palmer et al. 2002), which suggests that spatial variation of reflectance should correlate with variation in habitat conditions, and thus with plant species richness. It is theorised that continuous remotely sensed data can add detail on habitat characteristics at the landscape level beyond what can be supplied by climate- and topography-variables and LULC classes (Parviainen et al. 2013).

# Aims

The general aim of this thesis has been to develop new methods and techniques based on remote sensing and GIS, in order to investigate relationships between plant species diversity and landscape parameters at varying spatial and temporal scales, within cultural landscapes in Southern Sweden.

I have specifically addressed the following questions:

**Paper I:** Can landscape-scale plant species richness be related to individual and combined landscape-scale measures of environmental and spectral heterogeneity? Can we predict landscape-scale plant species richness using combined measures of landscape-scale environmental and spectral heterogeneity?

**Paper II:** Can Landsat satellite data, supported by ancillary data on topography and seasonal differences in the NDVI, be used to measure LULC change between historical (ca 1975) and contemporary (ca 2001) landscapes? Can we relate changes in LULC between the historical and contemporary landscapes to contemporary levels of landscape-scale plant species richness?

**Paper III:** What is the relative importance of different non-crop habitat types for large-scale total plant species richness? Are the relationships influenced by landscape complexity?

**Paper IV:** Is Worldview-2 satellite spectral dissimilarity related to fine-scale plant species dissimilarity in semi-natural grasslands? Are the relationships affected by the spatial extents of grassland plots and spectral windows?

**Paper V:** Can HySpex hyperspectral aerial data be used to predict fine-scale plant species diversity, using i) spectral heterogeneity ii) the direct relationship between spectral reflectance and plant species diversity?

**Paper VI:** Can HySpex hyperspectral aerial data be used to classify dry grazed grasslands into three age classes? Can pre-selecting suitable wavebands improve the classification results?

# Material and Methods

Studies resulting in papers I, II and III were carried out in the province of Scania, in the south of Sweden. Studies resulting in papers IV, V and VI were carried out on the Baltic island of Öland. For clarity and to avoid confusion, the data and methods used for each study area will be presented separately.

## Scania

Scania, Sweden's southernmost province, covers an area of approximately 11 000 km2, which represents 2.5% of Sweden's total land surface The province is Sweden's most productive agricultural region, with almost half of its area dedicated to agricultural land use (Statistics Sweden 2010). For the studies in this thesis, I have adopted two different official partitions of the province, both based on the premise that Scania consists of three distinct districts. In papers I and II, definitions and borders defined by the Nordic Council of Ministers were used, which define the districts on the basis of differences in geological, edaphic, topographic and climatic conditions Nordiska Ministerrådet 1977) (Figure 2).

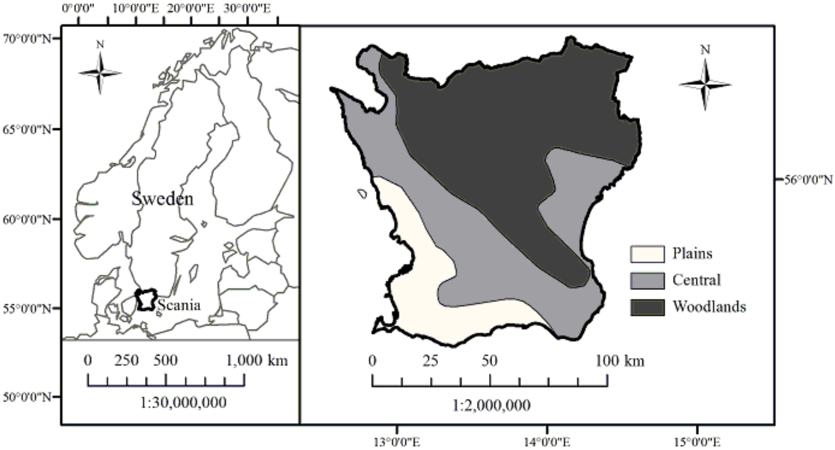


Figure 2. The location of the province of Scania (inset) and map showing the Plains, Central, and Woodland districts in Scania.

In paper III, definitions and borders defined by the Swedish Ministry of Agriculture were used, which divides the province into three agricultural regions, also known as production areas. The borders and physical makeups of the two different delineations are very similar, but there are some differences, as the geographical districts are mainly based on the physical geography of the province, and the agricultural regions are defined by intensity in agricultural practices. Both systems of delineation generally describe the districts/regions in the same way: The Plains district, found in the south-western part of the province, is dominated by arable cultivation on fertile clayey moraine soils, with patches of planted conifers and deciduous woodlands. The Central district has a similar physical makeup, but with lower agricultural intensity and a higher proportion of forest. The Woodlands district in the north-eastern part of the province is characterised by igneous rock that is typically overlaid by moraine soils with low clay content. Agricultural practices are less intense in the Woodlands district, and the main land-cover is commercially managed mixed forest, dominated by coniferous species (Germundsson and Schlyter 1999). The Plains, Central and Woodlands districts represent a gradient of decreasing land-use intensity and increasing landscape complexity.

## Plant species richness

The aim of paper I, II and III based on the Scania study area was to explain levels of plant species richness, and in paper I, also to predict levels of plant species richness.  Plant species occurrence data was derived from the project "Skånes Flora", a province-wide floristic survey coordinated by the Lund Botanical Society during the years 1987 – 2005 (Tyler et al. 2007). For the survey, Scania was divided into 2.5 × 2.5 km grid squares. I used the total number of plant species in paper I. In paper II, the (i)  total number of  plant species, (ii) the total number of red listed species (according to the International Union for conservation of Nature Red List), and (iii) the total number of grassland habitat specialist species (Ekstam and Forshed 1992) were used. In paper III, the (i) total number of plant species, and (ii) the total number of grassland habitat specialist species were used.

## Landsat satellite data

Paper I made use of a Landsat 7 ETM+ scene from the 1 July 2001, while paper II made use of Landsat 2 MSS satellite scenes from 27 February 1973 (dormant season), 3 July 1975 (growth season) and 27 August 1973 (senescence season), and Landsat 7 ETM+ satellite scenes from 23 March 2003 (dormant season), 1 July 2001 (growth season), and 12 September 2002 (senescence season). In paper

I, the Landsat 7 ETM+ scene was used to produce measures of spectral heterogeneity; the range and standard deviation of the red and near-infrared bands, and the normalised difference vegetation index (NDVI), calculated as ((NIR – red) / (NIR + red)), within each of the 2.5 × 2.5 km grid squares used in the study. In paper II, the Landsat 2 MSS data was used in Support Vector Machine-classifications to describe a historical landscape for the period around year 1975. A SVM (Vapnik 1995) is a machine learning technique, based on statistical learning theory, which delivers a supervised classification. The Landsat 7 ETM+ scenes were used in Support Vector Machine-classifications to describe a contemporary landscape for the period around year 2001, which temporally matched with the contemporary plant richness data. Reflectance values of individual bands from the growth season images were used directly in the classifications, while dormant season and senescence season images were used as ancillary data to calculate differences in NDVI between dormant season – growth season, and senescence season – growth season for the historical and contemporary landscapes, in order to aid in separating spectrally similar classes.

## Explanatory variables

*Proportions of LULC classes*

The proportion of different classes of LULC was used as descriptors of plant species richness in paper II and III. In paper II, I used LULC data that were created using supervised SVM-classifications, while I used data from official sources in paper III. In paper III, I used the Swedish Land Cover Database (SMD) (Metria), which incorporates 51 LULC classes at a spatial resolution of 25 m × 25 m (table S1 (supplemental material) in paper II). The original 51 classes were reclassified to 8 summary classes, before proportions of each class for each grid square was calculated. I also used "Blockdatabasen", an annually updated Integrated Administration and Control System (IACS) produced by the Swedish Board of Agriculture, in order to be able to define sub-classes of farmland and their administrative borders. "Blockdatabasen" was also used in order to approximate the amount of small biotopes found within the study area. I used the 2001-version of the database. In paper II, SVM-classifications were used to create representations of historical (around 1975) and contemporary (around 2001) LULC in Scania, using the same 8 classes as in the reclassified SMD-data. From these classifications, proportions of each LULC class for each grid square were calculated.

*Topography*

Topography is known to be linked to species distributions in mountainous regions (Hofer et al. 2008), but it may also influence plant species diversity in flatter regions (Moeslund et al. 2013), such as Scania, where elevation ranges from 0 to 212 meters above sea-level. I used measures of topographic heterogeneity in both paper I (range and standard deviation of the elevation and the slope) and paper III (standard deviation of the elevation) as descriptors of plant species richness within each grid square. I used elevation as ancillary data in the SVM classifications in paper II.

### *Soil and Bedrock*

Soil properties, which in large part are dependent on the properties of the bedrock on which the soil is formed, have a strong influence on the spatial distribution of plants (Tyler 2007). I used the number of unique soil types within grid squares in paper I, and the proportion of bedrock-categories within grid squares in paper III, as descriptors of plant species richness.

### *Environmental heterogeneity*

All three studies based on the Scania study area incorporated measures of environmental heterogeneity as descriptors of plant species richness, and in paper I also as predictors of plant species richness. Several data sources were used to produce measures of environmental heterogeneity, incorporating data on the spatial distribution and composition of LULC, as well as data on topography and soil diversity. Paper I and III focused on data from official sources, while paper II used environmental heterogeneity variables from LULC data created by the SVM classifications.

### *Spectral heterogeneity*

In paper I, six measures of spectral heterogeneity were calculated from the unclassified, continuous Landsat data, using the standard deviation and the range of the red (0.63 – 0.69 μm), and near infrared (NIR) (0.77 – 0.90 μm) wavebands, and the NDVI (NIR-red)/(NIR+red). Measures of spectral heterogeneity were related to levels of plant species richness.

### *Hybrid heterogeneity*

In paper I, environmental and spectral measures of landscape heterogeneity were used together in models (hybrid heterogeneity) and related to levels of plant species richness. Hybrid heterogeneity measures were also used to predict levels of plant species richness in paper I.

### Data analysis

In ecological studies, the distributions of the relationships between response and explanatory variables are often complex or unknown (e.g. Yee & Mitchell 1991, Guisan et al. 2002, Parviainen et al. 2008). There is also a need to adapt methods that can deal with collinearity among a potentially high number of explanatory variables. The potential of modeling non-linearity in the relationships between response and explanatory variables was also something to consider when choosing statistical methods.

In paper I, I used Generalised Additive Mixed Models (GAMM) to relate explanatory variables of environmental (EH) and spectral (SH) heterogeneity, and combinations of the two (hybrid heterogeneity (HH)), to plant species richness in training datasets, before testing the models on separate validation datasets by predicting levels of species richness, also using GAMM.

In paper II, I used Pearson's correlation coefficients to relate proportions of LULC classes and measurements of environmental heterogeneity to plant species richness in the historical and contemporary landscapes, respectively, as well as relating the changes in proportions of LULC classes and measurements of environmental heterogeneity between the historical and contemporary landscapes to plant species richness. I also tested for differences in change in LULC classes between the 30 least and most species rich squares between the historical and contemporary landscapes, using 2-tailed Student's t-tests.

In paper III, I used Generalised Least Squares (GLS) linear models and an information-theoretic approach with multi-model averaging to relate explanatory environmental variables to plant species richness.

Akaike Information Criteria (AIC)-values were used as the variable selection tool in models in both paper I and III. I also tested and included spatial correlation structures to correct for spatial autocorrelation within the data in both paper I and III.

## Öland

The study area (centred on 56°40′49″N, 16°33′58″E), located close to the village of Jordtorp, covers approximately 22.5 km2, and consists of a mosaic of grasslands of different ages, cropland, forest and small villages (Figure 3). The topography is generally flat, with a few low ridges built up of glacio-fluvial sediments. The area is generally dry, with a mean annual precipitation of 468 mm

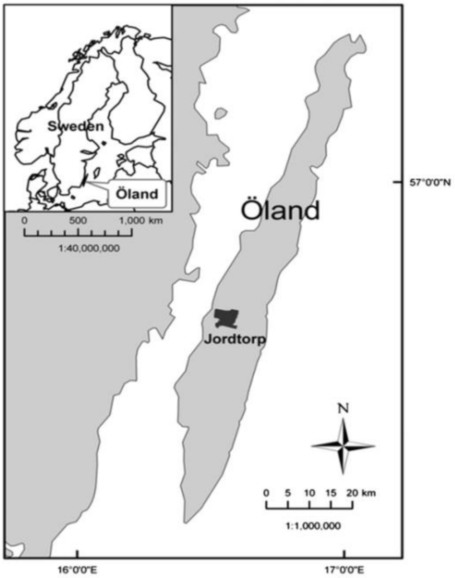


Figure 3. The Jordtorp study area on the Baltic island of Öland, Sweden

and a mean annual temperature of 7ºC (Forslund 2001). The grasslands in the area range from recently developed on former arable fields and forest sites (Johansson et al. 2008), to old semi-natural grasslands with a grazing continuity of over 280 years. The majority of the grasslands in the study area are grazed, mainly by cattle, with varying intensity. Grasslands currently cover approximately 10% of the study area, compared to roughly 80% at the beginning of the 19th century (Johansson et al. 2008).

## Dependent variables

Measuring and predicting grassland age and grassland plant diversity were the topics in the Öland studies. Species richness of plants and the inverse Simpson's diversity index (iSDI) within 52 dry grazed grassland sites of three different age classes (5–15y, 15–50y, >50y) were used in paper V. Within the same sites, I predicted age classes of grassland plots in paper VI. In paper IV, I calculated, and then predicted, pairwise dissimilarity in plant species composition (a measure of β-diversity) between 15 sites within the >50y grassland age class.

## Remote sensing data

I used a Worldview-2 multispectral satellite scene (2m spatial resolution, 8 bands) in paper IV, where I calculated pairwise spectral dissimilarity for individual spectral bands, as well as for three vegetation indices (VI's), between plots in 15 grassland sites. In paper V, I used information from individual bands and spectral heterogeneity and, in paper VI, I used information from individual bands, from the HySpex airborne hyperspectral sensor.

## Statistical analysis

In paper IV, I used Worldview 2 multispectral satellite data (2m spatial resolution, 8 bands) to relate pairwise measures of spectral dissimilarity to measures of plant species dissimilarity (β-diversity) between 15 old (>50y) dry grassland plots in training datasets, using univariate regression and vegetation indices (VI's), and partial least squares regression (PLSR) and the full set of Worldview-2 spectral bands. I used two vegetation plot sizes, (2m × 2m and 4m × 4m) and different sized spectral windows, centred over the vegetation plots, to capture the remote sensing data (smallest: 1 × 1 pixel, largest: 11 × 11 pixels). The models developed on the training datasets were then used to predict β-diversity between pairs of plots in validation datasets (15 plots), using RMA regression and PLSR, respectively.

In paper V, individual spectral bands from the HySpex sensor were related to levels of plant species diversity (species richness and the inverse Simpson's index (iSDI)) in training plots, using partial least square regression (PLSR). Two sets of models were built, model 1 used all spectral bands (245 bands) and model 2 used a subset of spectral bands, 25 for species richness and 35 for the iSDI. The bands were selected based on their importance for predicting species diversity. Spectral heterogeneity within training plots, calculated with the "mean distance to the spectral centroid"-method, were also related to levels of plant species diversity using ordinary least squares regression (OLR). Models developed on the training data were then used to predict levels of plant species diversity (species richness and the inverse Simpson's index (iSDI)) through the use of PLSR for PLSR-training models, and reduced major axis regression (RMA) for OLR-training models, in validation plots.

In paper VI, I used individual spectral bands from the HySpex sensor and partial least squares discriminant analysis (PLS-DA) in training plots to build models to predict age classes (5–15y, 15–50y, >50y) of dry grazed grasslands. Two sets of models were built, model 1 used all spectral bands (245 bands) and model 2 used a subset of 177 spectral bands, selected based on their importance for predicting grassland age classes. Models developed on the training data were then used to predict age classes (5–15y, 15–50y, >50y) of dry grazed grasslands in validation plots, using PLS-DA.

# Results and discussion

Studies resulting in paper I, II and III were carried out in the province of Scania, in the south of Sweden. Studies resulting in paper IV, V and VI were carried out on the Baltic island of Öland. For clarity and to avoid confusion, the results for each study area will be presented and discussed separately.

## Scania

The common general aim in paper I, II and III was to explain, and in paper I also to predict, levels of plant species richness in Scania and its three geographical districts.

### The influence of landscape heterogeneity on plant species richness

In paper I, I used generalised additive mixed models (GAMM) to show that plant species richness was related to, and could be predicted with, measures of environmental (EH) and spectral heterogeneity (SH), and that EH in general was a better descriptor of plant species richness than SH. I also found that the combination of EH and SH (hybrid heterogeneity, HH) generally could improve on models compared to when EH and SH were used separately (table 3 in paper I). The results also suggest that SH has a larger complementary role in models of heterogeneity when the EH has a low impact on plant species richness. These results are in agreement with e.g. Parviainen et al. (2013), who show that spectral data can complement environmental data and enhance the performance of models describing plant species richness. I found that measures of LULC heterogeneity generally were the best environmental descriptors of plant species richness. The importance of LULC heterogeneity in the variance of plant species richness in areas with low altitudinal ranges, such as Scania and its districts, is corroborated by earlies studies (e.g. Heikinnen et al. 2004, Ortega et al. 2004, Waldhardt et al 2004). I found that the red and near infrared individual spectral bands were generally better descriptors of plant species richness than the NDVI, except in the woodlands district. The NDVI is often the go-to vegetation index used in ecological studies (Pettorelli et al. 2014), but our results suggest that the use of the NDVI’s individual components (red and near-infrared) may be more suitable under certain environmental conditions, and that they should not be discarded in favour of the NDVI as spectral descriptors of plant species richness before their descriptive capabilities have been assessed. These findings are in accordance with other studies using spectral proxies to describe landscape heterogeneity (e.g. Wen et al 2012, Ding et al. 2014). I found that positive linear relationships generally described the relationships between heterogeneity variables and plant species richness, but that weakly unimodal, bimodal, and negative linear relationships also occurred (table 3 in paper I). In a review of 192 studies, Stein et al (2014) found similar general trends in the shape of the relationships between environmental heterogeneity and plant species richness.

The importance of LULC heterogeneity for promoting plant species richness was further shown in paper II, where measures of EH was significantly correlated with levels of plant species richness for the three species categories in both the historical and contemporary landscapes (Table S9 (supporting information) in paper II). The relatively high correlations for the historical landscape suggest that contemporary levels of plant species richness are dependent not only on contemporary environmental heterogeneity, but also on earlier levels of heterogeneity in the landscape. The lower correlations for red-listed plant species richness in both the historical and contemporary landscapes indicates that red-listed species may already have declined before the mid-1970s, as previous studies suggest that red-listed species may be affected by changes in the landscape configuration earlier than other categories of species (e.g. Cousins et al. 2015).

## Predicting plant species richness using hybrid heterogeneity and GAMMs

In paper I, using GAMMs to predict plant species richness in separate validation data sets resulted in prediction deviances of <20% in approximately 80% of validation data set squares for Scania and its three geographical districts (figure 3 in paper I). While I feel that this represents a successful attempt at predicting plant species richness, there is currently very little stated from stakeholders and practicing conservationists on their demands and needs for prediction accuracy (Skidmore et al. 2015). As far as I know, no common guidelines or practices exist on what “good” or “acceptable” prediction accuracy of species richness actually is, making it hard to properly judge and evaluate the impact of the prediction results.

## Land use/land cover and its influence on plant species richness

In paper II, I used SVM-classifiers utilising Landsat data and ancillary data on topography and seasonal differences in the NDVI, to create LULC information with high degrees of accuracy (+85% overall accuracy, see Foody 2002) for both the historical and contemporary landscapes (see tables 3 and 4 in paper II). Specifically, I found that the addition of the ancillary data improved the classification of spectrally similar classes, such as grazed grasslands and cropland, particularly within the historical landscape. This is especially encouraging considering the technical limitations, specifically in regards to spatial, radiometric and spectral resolution, of the older Landsat 2 MSS data used in the classification of the historical landscape. This suggests that historical Landsat data in conjunction with ancillary data and modern classification methods, such as SVM, can produce accurate LULC information for historical conditions. Further, I showed that proportions of LULC in both the historical and contemporary landscapes were related to contemporary plant species richness for grid squares in southern Scania and within its three geographical districts (table S8 (supporting information) in paper II). I found that correlations between proportions of LULC in the historical landscape and contemporary plant species richness were comparable to those of LULC in the contemporary landscape and contemporary plant species richness (table S8 (supporting information) in paper II), suggesting that contemporary levels of plant species richness are related also to earlier LULC composition in the landscape. Correlations between red-listed plant species richness and LULC proportions in the historical and contemporary landscapes were lower than for the other species-groups, further indicating that red-listed species may already have declined before the mid-1970s.

In paper II and paper III, I found that non-crop habitats were highly important for plant species richness within Scania’s three districts, but that which habitat was important depended on the district in question, and to some extent which method of analysis and which data was used. Both papers conclude that cropland has a negative effect on levels of plant species richness within the study areas. Proportions of semi-natural- or grazed grasslands and deciduous forest were important in both studies, being generally positively correlated to plant species richness categories in paper II (table S8 (supporting information) in paper II), and selected as variables of importance in paper III (tables 3 and 4 in paper III). In paper III, I also found that small biotopes were important in the Central and Woodlands districts, but not in the Plains district, which has a higher proportion of cropland. This suggests that there needs to be a certain level of non-crop habitats in the surrounding landscape for small biotopes to have a positive effect on plant species richness (Aavik & Liira 2009). In paper II, heterogeneity within the landscape had higher correlations than did proportions of LULC, with levels of plant species richness, while in paper III, LULC variables were generally more important than environmental heterogeneity in explaining levels of plant species richness. These somewhat contradictory results suggest that the choice of data and methodology can play an important role in the results of studies of how environmental variables affect plant species richness.

## Change in land use/land cover and landscape heterogeneity between historical and contemporary landscapes

In general, the measured change between historical and contemporary landscapes in proportions of LULC classes in paper II (table 5 in paper II), coincided well with data from official sources. For those LULC categories that were shown to be important to levels of plant species diversity within the individual landscapes in both paper II and paper III, I showed that between the historical and contemporary landscapes, cropland decreased, and grazed grassland and deciduous forest increased. LULC classes generally shown to be important to plant species richness thus seem to be increasing in area, while LULC classes generally shown to be detrimental to plant species richness are decreasing in area, within studied areas of the province of Scania. The supposed increase in new grazed grasslands, often the results of abandonment of arable fields, may in the future lead to an increase of species rich grassland vegetation if grazing management continues (Lengyel et al 2012), and a supposed general shift from low species richness planted coniferous forest to more species rich deciduous forest may also lead to a general increase in plant species richness and diversity.

## The influence of land use/land cover change on plant species richness

In paper II, I show that, generally, change between the historical and contemporary landscapes in the proportion of deciduous forest is positively correlated to levels of plant species richness in southern Scania (table S10 (supporting information) in paper II). I found significant differences in mean change between historical and contemporary landscapes in the proportion of cropland, coniferous forest and deciduous forest, when comparing the 30 least (coldspots) and 30 most (hotspots) species rich squares within Scania (figure 5 in paper II). In hotspots, the proportion of deciduous forest had increased more than in coldspots, while the proportion of cropland and coniferous forest had decreased more in hotspots than in coldspots of plant species richness.

In paper II, variables of landscape heterogeneity showed fairly high correlations with plant species richness for the historical and contemporary landscapes individually, but for the change between the historical and contemporary landscapes, only the number of patches (for Scania and the Plains district, for total and grassland sp. species richness) and Shannon's diversity index (for the Plains district for grasslands sp. species richness) showed significant correlation with plant species richness. There can be several reasons for this; change in heterogeneity may generally be too small to have a significant effect on levels of plant species richness; change in heterogeneity may not be modelled well with the used methodology; or as shown in paper I, the effect of heterogeneity on plant species richness may not have been exclusively linear, or positive.

## Öland

Papers IV, V and VI, based on studies in the Öland study area, have prediction techniques in grasslands using remote sensing and statistical modelling as the common theme.

In paper IV, I found significant positive relationships between spectral dissimilarity and plant species dissimilarity in the pairwise comparisons using RMA regression and VI's, for both vegetation plot sizes in the training datasets, with the strength of the relationships decreasing as the spectral windows used increased in size (figure 3 in paper IV). Using PLSR and Worldview-2 spectral bands in the training dataset, I found the same pattern, with the strength of the relationships decreasing, and the cross-validated RMSE increasing, with increasing size in spectral windows (figure 5 in paper IV). When applying the models on the validation datasets, correlations between predicted and field-observed β-diversity were similar for both vegetation plot sizes, and for when VI- and individual band measurements of spectral dissimilarity were used, respectively. Correlations between predicted and field-observed β-diversity generally decreased with increasing pixel window size (table 2 in paper IV). This suggests that spectral windows of approximately equal size to vegetation plots best captures the environmental conditions within the vegetation plots. The similarity in correlations between predicted and field-observed β-diversity for the two vegetation plot sizes suggests that the species present within the 2 m × 2 m plots represent a subset of species present in the 4 m × 4 m plots, as there is an expected correlation between measures of species diversity within semi-natural grasslands (Öster et al. 2007). While using PLSR with the full set of Worldview-2 spectral bands gave slightly higher correlations between predicted and field-observed β-diversity than did using VI's, the difference was small. While using the full of set of spectral bands in PLSR models can be expected to improve the relationships between dependent and explanatory variables (c.f. Fava et al. 2010), the NDVI-based VI's apparently captured the majority of the available spectral response, most likely due to the NDVI's correlation to vegetation productivity and biomass. I suggest that the observed relationships between β-diversity and spectral dissimilarity in paper IV is mainly due to differences in grazing conditions between plots, affecting levels of biomass, field layer height, plant litter and exposed soil, which will have an effect on both the species composition and the spectral response within plots. I also note that there are other variables other than heterogeneity that affect fine-scale plant species richness, such as grassland age (Johansson et al. 2008) and the availability of soil nutrients (Reitalu 2014).

In paper V, I predicted levels of plant species diversity in grassland plots of varying ages using PLSR and HySpex hyperspectral data, resulting in correlation coefficients (r2) of 0.43 (species richness) and 0.45 (iSDI) for model 1 (using all HySpex spectral bands), and 0.19 (species richness) and 0.40 (iSIDI) for model 2 (using 25 and 35 HySpex spectral bands for species richness and ISDI, respectively), with relative prediction errors (RMSEp) of approximately 20% for both species richness and iSDI for both models (table 3 in paper V). Using a subset of 25 HySpex spectral bands to predict species richness reduced correlation coefficients drastically. This is contradictory to the findings of e.g. Fava et al. (2010), who improved prediction accuracy by reducing the number of spectral bands in PLSR-based prediction of plant species richness in hay meadows in the European Alps. Using a subset of 35 HySpex spectral bands to predict iSDI did not reduce the prediction accuracy nearly as much, suggesting the extra ten bands used in the model 2 prediction of iSDI was enough to make a large impact on prediction accuracy. Predicting species richness and iSDI using spectral heterogeneity resulted in statistically non-significant results, with high (> 30%) prediction errors, and I must therefore conclude that the method was unsuccessful. It's likely that the 1m spatial resolution of the HySpex sensor did not manage to capture the fine-scale variation within environmental conditions in the grassland plots. Also, a spatial mismatch between the 4 m × 4 m field plots and the 8 m × 8 m spectral windows may have further influenced the prediction accuracy negatively.

In paper VI, I classified grassland plots into age classes (5 - 15y, 16 – 50y, > 50y), using partial least squares discriminant analysis (PLS-DA) and hyperspectral HySpex data. Models developed on the training dataset and evaluated with cross-validation, and models applied on the validation dataset, had very similar overall classification accuracies and kappa-values, which suggests that the models are robust. Model 1, using the full range of Hyspex spectral bands (269 bands), had an overall classification accuracy of 77% and a kappa-value of 0.65 for both the training and validation datasets. Model 2, using a subset of 177 HySpex spectral bands, had an overall classification accuracy of 81% and a kappa-value of 0.71 for the training dataset, and an overall classification accuracy of 85% and a kappa-value of 0.77 for the validation dataset. This shows that a pre-selection of wavebands using appropriate methods can improve the accuracy of grassland age classifications using hyperspectral data. I found that grassland plots belonging to the youngest age bracket were associated with the highest levels of nutrients in plots, and the oldest grasslands with the lowest levels of nutrients in plots. Young grasslands are typically established on recently abandoned arable fields, where residual nutrient levels can be expected to be high compared to those plots where grasslands have persisted for a longer time and grazing activities have reduced nutrient availability. Remotely sensed data suggested a higher amount of chlorophyll in the young plots than in the old plots, which correlates to the more vigorous vegetation of nutrient rich habitats.

# Conclusions

This thesis shows that plant species diversity patterns can be described with remotely sensed proxies of landscape properties, or using spectral reflectance directly, at a variety of spatial and temporal scales. The thesis has explored relationships between remotely sensed data and both α- and β-diversity of plants, thus incorporating both components of landscape γ-diversity. The thesis also highlights the difficulty and necessity of finding the relevant spatial scales where relationships between landscape properties and species diversity are at their strongest, and then finding suitable proxy data to model those relationships. The modeled relationships have not always been strong, or straightforward. Even though I have used relatively new multivariate statistical methods, and incorporated a variety of spectral sensors, including modern hyperspectral sensors, much of the variation within plant species diversity patterns have remained unexplained. Some of the loss in the strength of the relationships can be explained by the aims of the individual papers, as they generally focus on individual or limited aspects of the total variation within landscape properties that can be expected to affect plant species diversity. However, the complexity in how plant species diversity responds to landscape properties, and perhaps even more in the varying responses of the individual species themselves, makes the responses inherently difficult to model.

There are limitations with using simple measures of plant species diversity and landscape properties. Estimating habitat diversity through spectral heterogeneity, or using simple remotely sensed measures of landscape diversity, usually produces summary measures which do not say much, if anything, about the uniqueness or ecological importance of the different habitats (Rocchini et al 2016). Likewise, species richness and other simple measures of species diversity do not contain information on actual species composition, which can be of great importance in conservation efforts (Luoto 2005).

Despite these limitations, remote sensing allows for the inexpensive collection of environmental information over vast areas, and does so in a consistent and repeatable manner while disregarding borders, conflicts, and other difficulties that may affect field-based inventories (Skidmore et al., 2015) within cultural landscapes. It also allows for information gathered in elapsed time to be revisited and reexamined, and to put together data on temporal changes and thereby rapidly reveal where to direct conservation efforts (Skidmore et al. 2015).

New satellite systems are continually developed, and technological advances may bring about new possibilities for the remote sensing of cultural landscapes. The Worldview-3 sensor, the successor to the Worldview-2 sensor used in paper IV in this thesis, offers extremely detailed (0.31m spatial resolution, panchromatic band) satellite data for the monitoring of, for example, individual trees or small scale species assemblages, and data is now available for interested parties, although at a fee. The EnMAP satellite sensor, due for launch within the near future, will supply data with the same spatial resolution as the Landsat 7 satellite used in papers I and II in this thesis, but with hyperspectral resolution, using 244 spectral bands (in the 0.42 μm – 2.45 μm wavelengths) compared to the eight found on the Landsat 7 satellite sensor.

Remote sensing can be a crucial tool in ecological studies and conservation efforts, but in order for it to reach its full potential, there is a need for closer cross-disciplinary collaboration between ecologists and remote sensing specialists (Skidmore et al. 2015). Many landscape and vegetation properties can be measured through remote sensing, but there are few standards or guidelines in place on how to translate those measurements into metrics that are actually meaningful for biodiversity monitoring (Skidmore et al. 2015). However, there are promising initiatives to improve the acquisition, coordination and delivery of biodiversity observations to users, such as national level decision makers and the scientific community. These include the Group on Earth Observations Biodiversity Observation Network (GEO BON). The main aim of GEO BON is to aid in coordinating large-scale biodiversity monitoring, linked to environmental data, to ensure that the Convention on Biological Diversity (CBD) reaches its stated goals for the biodiversity targets implemented during the 2010 Aichi-meeting (GEO BON).

Perhaps, in some small way, this thesis can also contribute to that goal.

# References

Aavik, T., Liira, J. (2009) Agrotolerant and high nature-value species - Plant biodiversity indicator groups in agroecosystems. Ecological Indicators, 9, 892-901.

ArtDatabanken, Swedish Species Information Centre. http://www.artfakta.se/GetSpecies.aspx?SearchType=Advanced

Asner, G. P. (1998) Biophysical and biochemical sources of variability in canopy reflectance. Remote Sensing of Environment 64(3): 234-253.

Benton, T. G., Vickery, J. A., Wilson, J. D. (2003) Farmland biodiversity: is habitat heterogeneity the key? Trends in Ecology & Evolution 18(4): 182-188.

Bernes, C. (Ed.), (1994) Biological Diversity in Sweden.A Country Study. Monitor 14. Swedish Environmental Protection Agency, Solna, Sweden.

Box, E. O., Holben, B. N., Kalb, V. (1989) Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net CO2 flux. Vegetatio 80(2): 71-89.

Bruun, H. H., Fritzboger, B., Rindel, P. O., Hansen, U. L. (2001) Plant species richness in grasslands: the relative importance of contemporary environment and land-use history since the Iron Age. Ecography 24(5): 569-578.

Bullock, J. M., Jefferson, R. G., Blackstock, T. H., Pakeman, R. J., Emmett, B. A., Pywell, R. J., Grime, J. P., Silvertown, J. (2011) Semi-natural grasslands. Cambridge, UK, UNEP-WCMC. (In: Technical Report: The UK National Ecosystem Assessment, 162-195)

Chapin, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., Hooper, D. U., Lavorel, S., Sala, O., Hobbie, S. E., Mack, M. C., Diaz, S. (2000) Consequences of changing biodiversity. Nature 405(6783): 234-242.

Colwell, R. (1956) Determining the prevalence of certain cereal crop diseases by means of aerial photography. Hilgardia 26(5):223-286.

Cousins, S. A. O., Eriksson, O. (2001) Plant species occurrences in a rural hemiboreal landscape: effects of remnant habitats, site history, topography and soil. Ecography 24(4): 461-469.

Cousins, S. A. O., Auffret, A. G., Lindgren, J., Trank, L. (2015) Regional-scale land-cover change during the 20th century and its consequences for biodiversity. Ambio 44: 17-27.

Defries, R. S., Townshend, J. R. G. (1994) NDVI-derived land-cover classifications at a global-scale. International Journal of Remote Sensing 15(17):

Deutschewitz, K., Lausch, A., Kuhn, I., Klotz, S. (2003) Native and alien plant species richness in relation to spatial heterogeneity on a regional scale in Germany. Global Ecology and Biogeography 12(4): 299-311.

Ding, Y. L., Zhao, K., Zheng, X. M., Jiang, T. (2014) Temporal dynamics of spatial heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8 OLI imagery. International Journal of Applied Earth Observation and Geoinformation 30: 139-145.

Duelli, P., Obrist, M.K. (2003) Regional biodiversity in an agricultural landscape: the contribution of seminatural habitat islands. Basic and Applied Ecology 4(2): 129-138.

Duro, D., Coops, N. C., Wulder, M. A., Han, T. (2007) Development of a large area biodiversity monitoring system driven by remote sensing. Progress in Physical Geography 31(3): 235-260.

Ekstam, U., Forshed, N. (1992) Om hävden upphör: kärlväxter som indikatorarter i ängs- och hagmarker. (If Grassland Management Ceases: Vascular plants as Indicator Species in Meadows and Pastures). Naturvårdsverket, Stockholm, SE (in Swedish with English summary).

Emanuelsson, U., Bergendorff, C., Billqvist, M., Carlsson, B., Lewan, N. (2002) Det skånska kulturlandskapet, BTJ Tryck AB, Lund.

Eriksson, A., Eriksson, O. (1997) Seedling recruitment in semi-natural pastures: the effects of disturbance, seed size, phenology and seed bank. Nordic Journal of Botany 17(5): 469-482.

Eriksson, O., Cousins, S. A. O., Bruun, H. H. (2002). Land-use history and fragmentation of traditionally managed grasslands in Scandinavia. Journal of Vegetation Science 13(5): 743-748.

Evans, G. C., Coombe, D. E. (1959) Hemisperical and Woodland Canopy Photography and the Light Climate. Journal of Ecology 47(1) 103-113.

Fahrig, L. (1992) Relative Importance of Spatial and Temporal Scales in a Patchy Environment. Theoretical Population Biology 41(3): 300-314.

Fava, F., Parolo, G.,,Colombo, R., Gusmeroli, F., Della Marianna, G., Monteiro, A. T., Bocchi, S. (2010) Fine-scale assessment of hay meadow productivity and plant diversity in the European Alps using field spectrometric data. Agriculture Ecosystems & Environment 137(1-2): 151-157.

Foody, G. M. (2002) Status of land cover classification accuracy assessment. Remote Sensing of Environment 80(1): 185-201.

Forslund, M. (2001) Natur och Kultur på Öland; Länsstyrelsen i Kalmar: Kalmar, Sweden.

*GEO BON, Group on Earth Observations Biodiversity Observation Network* http://geobon.org/ 2016-05-10

Germundsson, T., Schlyter, P. (Eds.) (1999) Atlas över Skåne. Sveriges Nationatlas förlag. Stockholm, SE.

Gillespie, T. W. (2005) Predicting woody-plant species richness in tropical dry forests: A case study from south Florida, USA. Ecological Applications 15(1): 27-37.

Guisan, A., Edwards, T. C., Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157(2-3): 89-100.

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. Journal of Applied Ecology 41(5): 824-835.

Hofer, G., Wagner, H. H., Herzog, F., Edwards, P. J. (2008) Effects of topographic variability on the scaling of plant species richness in gradient dominated landscapes. Ecography 31(1): 131-139.

Homolova, L., Maenovsky, Z., Clevers, J. G. P. W., Garcia-Santos, G., Schaeprnan, M. E. (2013) Review of optical-based remote sensing for plant trait mapping. Ecological Complexity 15: 1-16.

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 species diversity. Landscape and Urban Planning 63(4): 241-250.

Huston, M.A. (1994) Biological diversity. The coexistence of species on changing landscapes. Cambridge University Press, Cambridge,UK.

Ihse, M. (1995) Swedish Agricultural Landscapes - Patterns and changes during the last 50 years, studied by aerial photos. Landscape and Urban Planning 31(1-3): 21-37.

Ihse, M., Lindahl, C. (2000) A holistic model for landscape ecology in practice: the Swedish survey and management of ancient meadows and pastures. Landscape and Urban Planning 50(1-3): 59-84.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., Francois, C., Ustin, S. L. (2009) PROSPECT plus SAIL models: A review of use for vegetation characterization. Remote Sensing of Environment 113: S56-S66.

Jones, H. G., Vaughan, R. A. (2010) Remote sensing of vegetation: principles, techniques and applications., Oxford University Press, Oxford, NY, 2010, ISBN978-0-19-920779-4.

Johansson, L. J., Hall, K., Prentice, H. C., Ihse, M., Reitalu, T., Sykes, M. T., Kindstrom, M. (2008) Semi-natural grassland continuity, long-term land-use change and plant species richness in an agricultural landscape on Oland, Sweden. Landscape and Urban Planning 84(3-4): 200-211.

Jordbruksverket (2006) Kulturhistoriska bidrag och särdrag – uppföljning och utvärdering av miljöersättningen till natur- och kulturmiljöer. Report 2006:10. Jönköping. SE.

Kleijn, D., Sutherland, W. J. (2003) How effective are European agri-environment schemes in conserving and promoting biodiversity? Journal of Applied Ecology 40(6): 947-969.

Kleijn, D., Kohler, F., Baldi, A., Batary, P., Concepcion, E. D., Clough, Y., Diaz, M., Gabriel, D., Holzschuh, A., Knop, E., Kovacs, A., Marshall, E. J. P., Tscharntke, T., Verhulst, J. (2009) On the relationship between farmland biodiversity and land-use intensity in Europe. Proceedings of the Royal Society B-Biological Sciences 276(1658): 903-909.

Kumar, L., Schmidt, K., Dury, S., Skidmore, A. (2001) Imaging spectrometry and vegetation science. 111-155 in Meer, F. V D., Jong, S. D (editors) Imaging spectrometry. Springer Netherlands, Dordrecht, The Netherlands.

Landsbygdsdepartementet (2003) Ett levande kulturlandskap – en halvtidsutvärdering av miljö- och landsbygdsprogrammet. Swedish Government Official Reports (SOU) 2003:105.

Lengyel, S., Varga, K., Kosztyi, B., Lontay, L., Déri, E., Török, P., Tóthmérész, B. (2012) Grassland restoration to conserve landscape-level biodiversity: a synthesis of early results from a large-scale project. Applied Vegetation Science 15: 264–276.

Liang, S. (2005) Quantitative remote sensing of land surfaces. Wiley, Hoboken, USA.

Lindborg, R., Eriksson, O. (2004) Historical landscape connectivity affects present plant species diversity. Ecology 85(7): 1840-1845.

Luoto, M., Pöyry, J., Heikkinen, R. K., Saarinen, K. (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. Glob. Ecol. Biogeogr. 14, 575–584.

Martinez, S., Ramil, P., Chuvieco, E. (2010) Monitoring loss of biodiversity in cultural landscapes. New methodology based on satellite data. Landscape and Urban Planning 94(2): 127-140.

Medcalf, K. A., Jarman, M. W., Keyworth, S. J. (2010) Assessing the Extent and Severity of Erosion on the Upland Organic Soils of Scotland Using Earth Observation and Object Orientated Classification Methods. Geobia 2010: Geographic Object-Based Image Analysis 38-4-C7

Metria, Swedish geographic information service http://gis-services.metria.se/nvfeed/atom/annex2.xml) 2016-05-10

Moeslund, J. E., Arge, L., Bocher, P. K., Dalgaard, T., Svenning, J. C. (2013) Topography as a driver of local terrestrial vascular plant diversity patterns. Nordic Journal of Botany 31(2): 129-144.

Nagendra, H. (2001) Using remote sensing to assess biodiversity, International Journal of Remote Sensing, 22:12, 2377-2400.

Nordiska Ministerrådet (1977) Naturgeografisk regionindelning av Norden. Gotab, Stockholm.

Ortega, M., Elena-Rosello, R., Del Barrio, J. M. G. (2004) Estimation of plant diversity at landscape level: A methodological approach applied to three Spanish rural areas. Environmental Monitoring and Assessment 95(1-3): 97-116.

Palmer, M.W., Wohlgemuth, T., Earls, P., Arévalo, J.R., Thompson, S.D., (2000) Opportunities for long-term ecological research at the Tallgrass Prairie Preserve, Oklahoma. In: Lajtha, K., Vanderbilt, K. (Eds.), Cooperation in long term ecological research in Central and Eastern Europe: Proceedings of ILTER regional workshop, Budapest, Hungary, 22–25 June, 1999, 123–128.

Palmer, M.W., Earls, P., Hoagland, B.W., White, P.S., Wohlgemuth, T., (2002) Quantitative tools for perfecting species lists. Environmetrics 13, 121–137.

Parviainen, M., Luoto, M., Ryttari, T., Heikkinen, R. K. (2008) Modelling the occurrence of threatened plant species in taiga landscapes: methodological and ecological perspectives. Journal of Biogeography 35(10): 1888-1905.

Parviainen, M., Luoto, M., Heikkinen, R. K. (2010) NDVI-based productivity and heterogeneity as indicators of plant-species richness in boreal landscapes. Boreal Environment Research 15(3): 301-318.

Parviainen, M., Zimmermann, N. E., Heikkinen, R. K., Luoto, M (2013) Using unclassified continuous remote sensing data to improve distribution models of red-listed plant species. Biodiversity and Conservation 22(8): 1731-1754.

Perry, C. R., Lautenschlager, L. F. (1984) Functional equivalence of spectral vegetation indexes. Remote Sensing of Environment 14(1-3): 169-182.

Persson, A. S., Olsson, O., Rundlof, M., Smith, H. G. (2010) Land use intensity and landscape complexity-Analysis of landscape characteristics in an agricultural region in Southern Sweden. Agriculture Ecosystems & Environment 136(1-2): 169-176.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., Stenseth, N. C. (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology & Evolution 20(9): 503-510.

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., Turner, W. (2014) Satellite remote sensing for applied ecologists: opportunities and challenges. Journal of Applied Ecology 51(4): 839-848.

Regeringskansliet (2012) Rural development programme for Sweden – the period 2007-2013.

Reitalu, T., Sykes, M. T., Johansson, L. J., Lonn, M., Hall, K., Vandewalle, M., Prentice, H. C. (2009) Small-scale plant species richness and evenness in semi-natural grasslands respond differently to habitat fragmentation. Biological Conservation 142(4): 899-908.

Reitalu, T., Johansson, L. J., Sykes, M. T., Hall, K., Prentice, H. C. (2010) History matters: village distances, grazing and grassland species diversity. Journal of Applied Ecology 47(6): 1216-1224.

Reitalu, T., Helm, A., Partel, M., Bengtsson, K., Gerhold, P., Rosen, E., Takkis, K., Znamenskiy, S., Prentice, H. C. (2014) Determinants of fine-scale plant diversity in dry calcareous grasslands within the Baltic Sea region. Agriculture Ecosystems & Environment 182: 59-68.

Ricketts, T. H., Daily, G. C., Ehrlich, P. R., Michener, C. D. (2004) Economic value of tropical forest to coffee production. Proceedings of the National Academy of Sciences of the United States of America 101(34): 12579-12582.

Ricklefs, R.E. (1977) Environmental heterogeneity and plant species diversity: a hypothesis. American Naturalist 111:376–381.

Robinson, R. A., Sutherland, W. J. (2002) Post-war changes in arable farming and biodiversity in Great Britain. Journal of Applied Ecology 39(1): 157-176.

Rocchini, D., Chiarucci, A., Loiselle, S. A. (2004) Testing the spectral variation hypothesis by using satellite multispectral images. Acta Oecologica-International Journal of Ecology 26(2): 117-120.

Rocchini, D., Ricotta, C., Chiarucci, A., de Dominicis, V., Cirillo, I., Maccherini, S. (2009) Relating spectral and species diversity through rarefaction curves. International Journal of Remote Sensing 30(10): 2705-2711.

Rocchini, D., Boyd, D. S., Feret, J-B., Foody, G. M., He, K. S., Lausch, A., Nagendra, H., Wegmann, M., Pettorelli, N. (2016). Satellite remote sensing to monitor species diversity: potential and pitfalls. Remote Sensing in Ecology and Conservation. (2)1, 25–36.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W. (1973) Monitoring vegetation systems in the great plains with ERTS. In: Proceedings of the Third ERTS Symposium NASA SP-351, 1, U.S. Government Printing Office Washington DC, 309–317.

Schaepman-Strub, G., Schaepman, M.E., Painter, T.H., Dangel, S., Martonchik, J.V. (2006) Reflectance quantities in optical remote sensing—definitions and case studies. Remote Sensing of Environment 103: 27–42

Shmida, A., Wilson, M.V. (1985) Biological determinans of species diversity. Journal of Biogeography 12: 1–20.

Skidmore, A. K.,Pettorelli, N., Coops, N. C., Geller, G. N., Hansen, M., Lucas, R., Mucher, C. A., O'Connor, B., Paganini, M., Pereira, H. M., Schaepman, M. E., Turner, W., Wang, T. J., Wegmann, M. (2015) Agree on biodiversity metrics to track from space. Nature 523(7561): 403-405.

Skole, D. L. (1994) Data on global land-cover change: acquisition, assessment and analysis. In: W. B. Meyer, B. L. Turner II (Eds.), Changes in land use and land cover: a global perspective (437–471). Cambridge: Cambridge University Press.

Statistics Sweden (2010) Markanvändningen i Sverige, SCB-tryck, Örebro.

Stein, A., Gerstner, K., Kreft, H. (2014) Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. Ecology Letters 17(7): 866-880.

Tscharntke, T., Klein, A. M., Kruess, A., Steffan-Dewenter, I., Thies, C. (2005) Landscape perspectives on agricultural intensification and biodiversity - ecosystem service management. Ecology Letters 8(8): 857-874.

Turner, W., Sterling, E. J., Janetos, A. C. (2001) Special section: Contributions of remote sensing to biodiversity conservation: a NASA approach. Conservation Biology 15(4): 832-834.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M. (2003) Remote sensing for biodiversity science and conservation. Trends in Ecology & Evolution 18(6): 306-314.

Tyler, G. (2007) Markegenskapernas betydelse för arternas fördelning. In Floran i Skåne. Arterna och deras utbredning. (Tyler, T., Olsson, K.-A., Johansson, H., Sonesson, M., Eds.) Lunds Botaniska Förening, Lund.

Tyler, T., Olsson, K.-A., Johansson, H., Sonesson, M. (Eds) (2007) Floran i Skåne. Arterna och deras utbredning. Lunds Botaniska Förening, Lund.

Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., Zarco-Tejada, P. (2009) Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sensing of Environment 113: 67-77.

Vapnik, V.N. (1995) The nature of statistical learning theory (New York: Springer-Verlag).

Vitousek, P. M. (1994) Beyond Global Warming - Ecology and Global Change. Ecology 75(7): 1861-1876.

Vos, W., Meekes, H. (1999) Trends in European cultural landscape development: perspectives for a sustainable future." Landscape and Urban Planning 46(1-3): 3-14.

Waldhardt, R., Simmering, D., Otte, A. (2004) Estimation and prediction of plant species richness in a mosaic landscape. Landscape Ecology 19(2): 211-226.

Walz, U., Syrbe, R. U. (2013) Linking landscape structure and biodiversity. Ecological Indicators 31: 1-5.

Wen, Z. F., Zhang, C., Zhang, S. Q., Ding, C. H., Liu, C. Y., Pan, X., Li, H. P., Sun, Y. (2012) Effects of normalized difference vegetation index and related wavebands' characteristics on detecting spatial heterogeneity using variogram-based analysis. Chinese Geographical Science 22(2): 188-195.

Widgren, M. (1983) Settlement and farming systems in the early Iron Age – A study of fossil agrarian landscapes in Östergötland, Sweden

Woolley, J. T. (1971) Reflectance and transmittance of light by leaves. Plant Physiology 47(5): 656-662.

Yee, T. W., Mitchell, N. D. (1991) Generalized additive-models in plant ecology. Journal of Vegetation Science 2(5): 587-602.

Zewdie, W. E., Csaplovics, E. (2015) Remote Sensing based multi-temporal land cover classification and change detection in northwestern Ethiopia. European Journal of Remote Sensing 48: 121-139.

Öster, M., Cousins, S. A. O., Eriksson, O. (2007) Size and heterogeneity rather than landscape context determine plant species richness in semi-natural grasslands. Journal of Vegetation Science 18(6): 859-868.

# Paper I

# Modelling plant species richness in cultivated landscapes using public-domain data on environmental and spectral heterogeneity

Jonas Dalmayne, Therese Irminger-Street, Ola Olsson, Honor C. Prentice, Karin Hall

## Abstract

Theoretical and empirical studies indicate that plant species richness is likely to be promoted by environmental heterogeneity (at different spatial scales), with a diversity of habitats allowing the coexistence of species that have differing ecological requirements. The present study investigates relationships between vascular plant species richness (data from 965 2.5 × 2.5 km grid squares) and landscape-scale habitat heterogeneity in the Swedish province of Scania. Habitat heterogeneity was characterized with the help of variation in Land Use Land Cover (LULC) data in the public domain (environmental heterogeneity) and/or variation in Landsat spectral data (spectral heterogeneity). Generalized Additive Mixed Models (GAMMs) were used to analyse relationships between species richness and habitat heterogeneity: in the data set for the whole of Scania and, separately, for the subsets of squares representing the three geographic districts within Scania (the "Plains", "Central" and "Woodland" districts). Each of the four data sets was divided into similarly-sized training and validation sets of squares before analysis. Three types of model (in which the explanatory variables were based, respectively, on descriptors of environmental heterogeneity, spectral heterogeneity and the combined environmental and spectral variables) were used to explain levels of within-square species richness (response variable) in each of the four training data sets. A combination of environmental and spectral descriptors of heterogeneity provided the best models for species richness in the Scanian, Plains and Woodland training data sets, whereas the environmental descriptors gave the best model for the Central district. In all four data sets, the models based on spectral heterogeneity alone were inferior to the models based only on environmental heterogeneity. Models based on a combination of environmental and spectral heterogeneity predicted within-square species richness with +/- 20% accuracy in ca. 80% of the squares in the independent validation data set for Scania. Squares where species richness was underestimated by >20% had a significantly higher mean observed species richness, and squares where species richness was overestimated by >20% had a significantly lower mean observed species richness, than the overall mean observed richness within the squares in the validation data set. The study shows that landscape-scale habitat heterogeneity is related to plant species richness in the province of Scania and its three districts. The relationship is generally best characterized by measures of habitat heterogeneity that are based on a combination of environmental and spectral variables.

# 1. Introduction

Individual plant species have particular habitat preferences and ecological amplitudes and, if different species are favoured by different environmental conditions, then a high level of spatial environmental heterogeneity (EH) is expected to be accompanied by a higher level of species diversity than that found under more uniform environmental conditions at the same spatial scale (Whittaker 1972; Silvertown 2004). The traditional European cultural landscapes, which developed over centuries of human land-use, contained extensive areas of species-rich semi-natural habitats, such as pastures, meadows and forest margins, and were characterized by heterogeneous environmental conditions and high levels of biodiversity (Bignal and McCracken 1996; Poschlod and WallisDeVries 2002). The progressive intensification of agriculture and forestry (e.g. Niemelä et al. 2005) during the last century has led to the loss and fragmentation of semi-natural habitats, and to a loss of spatial heterogeneity in the environmental conditions of many European landscapes (Bignal and McCracken 1996; Benton et al. 2003; Tscharntke et al. 2005; Krauss et al. 2010). For example, in Sweden landscapes with extensive areas of species rich semi-natural habitats (such as grasslands with a long continuity of grazing management) have been progressively transformed into landscapes that are dominated by either intensive arable cultivation or large-scale commercial forestry

(Bernes 1994; Wramner et al. 2003; Johansson et al. 2008). The remaining fragments of species-rich semi-natural habitats support a substantial proportion of the present-day species diversity in many European rural landscapes (Billeter et al. 2007). As the relict patches of semi-natural habitat become progressively more fragmented, disjunct and isolated (Ihse 2005), habitat quality is increasingly affected by mineral fertilizers and agrochemicals from the surrounding landscapes (Robinson and Sutherland 2002). The loss and fragmentation of high quality semi-natural habitats, followed by a decrease in EH, has been identified as a major reason for the dramatic decline of species diversity in many modern European cultural landscapes (Storkey et al. 2012). Concern about the loss of species diversity has increased the demand for a better understanding of the mechanisms that drive the loss or maintenance of diversity, and there is a need for more information on the way in which the relationship between EH and levels of species diversity varies between different types of landscape. It has been suggested that, in the future, landscapes should be managed so as to enhance the spatial heterogeneity at several spatial scales (Benton et al. 2003; Belfrage et al. 2015), and methods that allow the evaluation of species diversity over large areas need to be developed.

In the ecological literature, the term environmental heterogeneity has been used to refer to spatial variation in many types of abiotic and biotic environmental conditions (Stein et al. 2014; Stein and Kreft 2015). Stein et al. (2014) conclude that, separate effects of spatial heterogeneity in climate, soil, topography, land-cover, and vegetation may influence the richness of terrestrial plants and animals, and that variation in vegetation and topography have a particularly strong impact on the levels of species richness at a variety of spatial scales. Studies of different groups of organisms, carried out within different types of habitat and at a range of spatial scales, generally report a positive relationship between EH and species diversity (e.g. Stein et al. 2014; Stein and Kreft 2015). Stein and Kreft (2015) argue that EH is expected to promote levels of species diversity because it favours species coexistence, species persistence, and species diversification, and that the processes driving the positive heterogeneity-diversity relationships may act at different spatial and temporal scales. Some studies have, however, also shown that the association between EH and species richness may be negative (particularly at fine spatial scales) or non-linear, or that species diversity is not significantly associated with environmental heterogeneity (e.g. Yang et al

2015; Stein et al. 2014), highlighting the importance of choosing modelling techniques that allow for the description of both linear and non-linear heterogeneity-diversity relationships in studies of the associations between EH and species diversity.

It is becoming increasingly clear that the use of Earth-observation satellite data in ecological studies can contribute to a deeper understanding of the mechanisms influencing the patterns of biodiversity (e.g. Pettorelli et al. 2014a). Unclassified continuous satellite data have shown to be particularly useful in ecological research, providing additional, detailed information on environmental characteristics that complements data on, for example, climate, topography and land-cover (Parviainen et al. 2013; Sheeren et al. 2014). The use of unclassified continuous reflectance values avoids the loss of information that is involved in the process of classification into simple land-cover classes (Parviainen et al. 2013). The spectral variation of remotely sensed data is expected to be associated with the heterogeneity in the environment (Rocchini et al. 2010), with higher levels of both species diversity and spectral heterogeneity (SH) being found in more heterogeneous habitats (the spectral variation hypothesis (SVH); Palmer et al. 2002). Measures of SH, in combination with measures of EH, have been used successfully in the assessment of patterns of species diversity within a range of different habitats and geographic regions (e.g. Zimmermann et al. 2007; Camathias et al. 2013; Parviainen et al. 2013). For example, Zimmermann et al. (2007) showed that a combination of Landsat data and data on topography and climate could be used to explain and predict the distribution of tree species in Utah, USA.

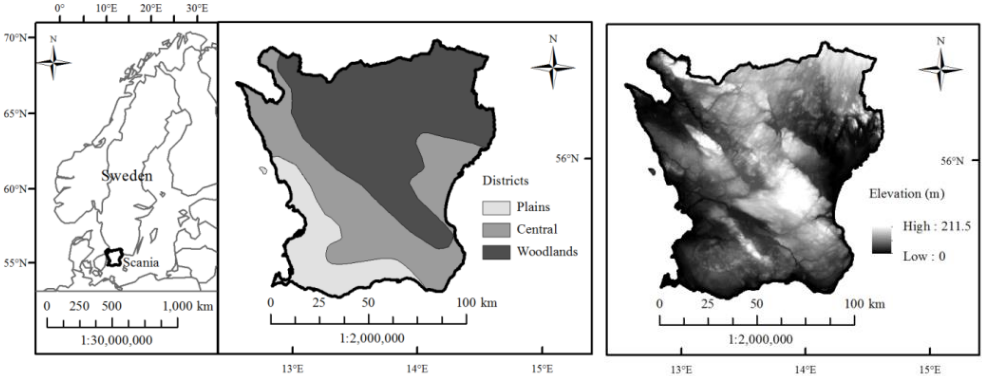
At present, the potential use of satellite imagery in biodiversity research has not been fully explored, and several problems have hindered the use of satellite data as a tool in conservation management (Turner et al. 2015). For example, difficulties with the retrieval of open-access satellite data have had a negative impact on the applicability of satellite data in research and operational applications (Turner et al. 2015). However, several recently launched internet-based platforms, such as the United States Geological Survey (USGS) Landsat archive, now provide free satellite data (Woodcock et al. 2008; Turner et al. 2015). In Sweden, the Saccess satellite archive (https://saccess.lantmateriet.se/portal/saccess_se.htm) provides free retrospective satellite data.

In the present study, we developed predictive models for vascular plant species richness within the Swedish region of Scania and its three geographic districts. The three Scanian districts represent a gradient in farming intensity and landscape heterogeneity – ranging from the intensively farmed and homogenous landscapes of the Plains district and the more heterogeneous agricultural landscapes of the Central district, to the Woodland district which is dominated by intensively managed forests. Data on vascular plant species richness (the response variable) within 2.5 km × 2.5 km grid squares were available from a province-wide floristic inventory in Scania (Tyler et al. 2007). Each of the grid squares was characterized by three sets of explanatory variables: EH variables (including topographic, soil, and land-cover variables), SH variables derived from Landsat satellite data, and a combination of both environmental and spectral heterogeneity variables. Both the EH and SH variables were developed from databases, in the public domain, that can be accessed free-of-charge. We ask the following questions. (1) Can landscape-scale (2.5 km × 2.5 km) vascular plant species richness be predicted from the sets of EH and SH variables, respectively? (2) Is the predictive ability of a combination of EH and SH variables superior to that of the separate sets of variables? (3) Does the accuracy of the environmental and spectral heterogeneity-based predictions of landscape-scale species richness differ between the province of Scania as a whole and the three geographic districts (the Plains, Central and Woodland districts) within Scania? We also examined which EH and SH variables were important for the prediction of landscape-scale species richness in the whole province of Scania and in each of the three Scanian geographic districts.

# 2. Methods

## 2.1. Study area

The study area was the province of Scania in southern Sweden which covers approximately 11 000 km$^2$ and is divided into three geographic districts (Fig. 1a) that are characterized by different geological, edaphic, topographic and climatic conditions (Germundsson and Schlyter 1999). The geology of the plains of Southwestern Scania (the "Plains district") and Central Scania (the "Central district") (Fig. 1a) is dominated by sedimentary bedrock that is mainly overlain by fertile, clayey moraine soils (Germundsson and Schlyter 1999). The northeastern district – the Woodlands of Northeastern Scania (the "Woodland district") (Fig. 1a) – is characterized by igneous rock that is mainly overlain by moraine soils that are less clayey and less fertile than those in the Plains and Central districts (Germundsson and Schlyter 1999). The overall topography in the Plains and Central districts is flat or gently undulating but, particularly in the Central district, several horsts contribute to local differences in elevation (Fig. 1b). The topography in the Woodland district is hilly, with some hills reaching more than 200 m (Fig. 1b). The mean January temperature in Scania ranges from -2 °C (in the Woodland district) to -0.5 °C (in the Plains district), and the mean July temperature ranges from 15.5 °C (Woodland district) to 16.5 °C (Plains district). The mean annual precipitation ranges from 550 mm (Plains district) to 900 mm (Woodland district) (Germundsson and Schlyter 1999). Both the Plains and Central districts are characterized by intensive arable (mainly cereal) cultivation. The agricultural intensity is higher in the Plains district than in the Central district, which contains a higher proportion of planted stands of conifers and patches of deciduous forests than the Plains district (Germundsson and Schlyter 1999). The land-cover in the Woodland district consists mainly of intensively managed mixed forests dominated by coniferous species (Germundsson and Schlyter 1999).



a) b)

**Figure 1**. (a) The location of the province of Scania (inset) and map showing the locations of the Plains, Central, and Woodland districts within Scania (b) map showing elevation (metres above sea-level).

## 2.2. Plant species richness

We used species data from a province-wide floristic inventory in Scania (coordinated by the Lund Botanical Society), in which presence/absence data for all vascular plant species were collected, between 1987 and 2005, from 1560 2.5 km × 2.5 km squares (Tyler et al. 2007). In the present study, species richness was represented by the total number of vascular plant species sampled in each of the 1406 2.5 km × 2.5 km squares (together covering ca. 77 % of Scania) that were regarded as being fully inventoried (Tyler et al. 2007).

## 2.3. Explanatory variables

Each of the 1406 2.5 km × 2.5 km squares was characterized by three sets of heterogeneity variables; (i) environmental heterogeneity (EH) variables, (ii) spectral heterogeneity (SH) variables (Table 1), and (iii) a combination of EH and SH variables, "hybrid heterogeneity" (HH) (cf. Parviainen et al. 2013). The heterogeneity variables were extracted from official databases in the public domain, using ArcGIS 10.1 (ESRI) and FRAGSTATS ver. 4 (McGarigal et al. 2012).

### *2.3.1. Environmental heterogeneity variables*

Ten variables were used to characterize EH: four topographic variables, one soil and five land-cover variables (Table 1). We derived the topographic variables (standard deviation (SD) of the elevation ($EH_{SD\_elevation}$), range of the elevation ($EH_{R\_ elevation}$), SD of the slope ($EH_{SD\_slope}$) and range of the slope ($EH_{R\_slope}$)) from a digital elevation model (DEM), with a spatial resolution of 50 m and an accuracy of ± 2 m, supplied by the National Land Survey of Sweden (Table 1). We used the soil patch-richness ($EH_{PR\_soil}$; Table 1), representing the number of soil types, as a proxy for soil heterogeneity. $EH_{PR\_soil}$ was estimated on the basis of data extracted from the Swedish Soil Cover Map (1:50 000), supplied by the Swedish Geological Survey (http://www.sgu.se). The digital soil map was transformed from vector to raster format with a spatial resolution of 5 m.

The land-cover EH was characterized in terms of the heterogeneity of different categories of land-cover (types of land-use and vegetation). We calculated five landscape indices based on different descriptors of patch

**Table 1**. The explanatory variables (environmental heterogeneity and spectral heterogeneity) used in the prediction of vascular plant species richness at the landscape scale (in 2.5 km × 2.5 km grid squares) in the province of Scania and in its three districts.

| Explanatory variables | Abbreviations | Units | Description | Source, data provider | Spatial resolution |
|---|---|---|---|---|---|
| **Environmental heterogeneity** | | | | | |
| ***Topography*** | | | | | |
| SD of elevation | $EH_{SD\_elevation}$ | m | | DEM, NLS | 50 m |
| $EH_{SD\_elevation}$ | $EH_{R\_elevation}$ | m | | DEM, NLS | 50 m |
| SD of slope | $EH_{SD\_slope}$ | degree | | DEM, NLS | 50 m |
| Range of slope | $EH_{R\_slope}$ | degree | | DEM, NLS | 50 m |
| ***Soil*** | | | | | |
| Soil patch richness | $EH_{Soil\_PR}$ | number | number of soil types | SSM, SGU | 1:50000 |
| ***Land-cover*** | | | | | |
| Number of patches | $EH_{NP}$ | number | number of individual patches | SMD, NLS | 25 m |
| Total edge | $EH_{TE}$ | m | total length of patch-edges | SMD, NLS | 25 m |
| Patch richness | $EH_{PR}$ | number | number of land-cover classes | SMD, NLS | 25 m |
| Shannon's diversity index | $EH_{SHDI}$ | | $H' = -\sum_{i=1}^{R} p_i \ln p_i$ | SMD, NLS | 25 m |
| Shannon's evenness index | $EH_{SHEI}$ | | $J' = \frac{H'}{H'_{max}}$ | SMD, NLS | 25 m |
| **Spectral heterogeneity** | | | | | |
| SD of red | $SH_{SD\_red}$ | | Landsat ETM+ band 3, 0.63 – 0.69 µm | USGS | 30 m |
| Range of red | $SH_{R\_red}$ | | Landsat ETM+ band 3, 0.63 – 0.69 µm | USGS | 30 m |
| SD of NIR | $SH_{SD\_NIR}$ | | Landsat ETM+ band 4, 0.77 – 0.90 µm | USGS | 30 m |
| Range of NIR | $SH_{R\_NIR}$ | | Landsat ETM+ band 4, 0.77 – 0.90 µm | USGS | 30 m |
| SD of NDVI | $SH_{SD\_NDVI}$ | | (NIR – red) / (NIR + red) | USGS | 30 m |
| Range of NDVI | $SH_{R\_NDVI}$ | | (NIR – red) / (NIR + red) | USGS | 30 m |

SD = standard deviation, R = Range, DEM = Digital Elevation Model, NIR = Near-infrared wavelength band, NDVI = Normalized Difference Vegetation Index, NLS = National Land Survey of Sweden, SSM = Swedish Soil Cover Map, SGU = Swedish Geological Survey, SMD = Swedish Land Cover Database, USGS = United States Geological Survey.

complexity; total edge ($EH_{TE}$), number of patches ($EH_{NP}$), patch richness ($EH_{PR}$), Shannon's diversity index ($EH_{SHDI}$) and Shannon's evenness index ($EH_{SHEI}$) (Table 1). The landscape indices were estimated on the basis of data for the 51 land-cover classes, included in the Swedish Land Cover Database – "Svenska Marktäckedata" (SMD) (http://gis-services.metria.se/nvfeed/atom/annex2.xml) at a spatial resolution of 25 m × 25 m (Table S1). The SMD database (provided by the Swedish Environmental Protection Agency) is a further development of the CORINE Land Cover project (http://www.eea.europa.eu/publications/COR0-landcover). In addition to the satellite-based CORINE data, the SMD land-cover classification incorporates data from terrain, road and vegetation maps

(Swedish National Land Survey), and forest-inventory data (Swedish University of Agricultural Sciences), as well as information from the Swedish Meteorological and Hydrological Institute, Statistics Sweden, the Swedish Environmental Protection Agency, and the various Swedish county administrative boards.

### *2.3.2. Spectral heterogeneity variables*

A Landsat ETM+ scene from 7 July 2001, covering the province of Scania, was acquired from the Landsat archives of the United States Geological Survey. Landsat ETM+ provides eight-band multispectral imagery in the spectral range 0.45 − 12.5 µm, with a spatial resolution of 30 m (with the exception of the thermal (60 m) and panchromatic (15 m) bands, which were not included in the present study). The Landsat scene was chosen on the basis of its quality (a minimum of haze and cloud cover) and to provide a temporal match with the growing season. The Landsat data were supplied as a level 1T product, with systematic radiometric and geometric corrections applied with the help of ground control points (GCPs), and topographic corrections carried out using a DEM. We transformed digital numbers (DNs) to top-of-atmosphere reflectance values (see Chander et al. 2009), and we also carried out a dark object subtraction to correct the Landsat scene for atmospheric disturbance.

Different earth-surface features reflect and absorb solar radiation differently at different wavelengths, and the ability of remote sensing data to detect spatial heterogeneity across landscapes may be influenced by the wavelengths or wavelength combinations that are used (e.g. Ding et al. 2014). For example, the red reflectance has been shown to be sensitive to the soil properties of land surfaces, and the spatial variation of the red reflectance has been shown to be useful in studies of landscape heterogeneity in areas with sparse vegetation cover (Garrigues et al. 2006; Ding et al. 2014). The near infrared (NIR) reflectance is sensitive to the cellular structure of leaves in the vegetation, and the spatial variation in the NIR reflectance has been used to describe the overall variation in vegetation properties across landscapes (Garrigues et al. 2006, Ding et al. 2014). The normalized difference vegetation index (NDVI) (Rouse et al. 1973, Tucker 1979) is a frequently used spectral vegetation index, which characterizes the contrast between the absorption in the red waveband by chlorophyll pigments and the reflectance in the NIR waveband caused by leaf cellular structure. The spatial variability of the NDVI is often used to describe landscape heterogeneity (e.g. Oliver et al. 2005) and, for example, Parviainen et al. (2010) used the standard deviation (SD) and range of the NDVI to characterize the heterogeneity in greenness in boreal forests in Finland. The separate use of the red and NIR wavebands may provide information that complements that of the NDVI (Garrigues et al. 2008, Ding et al. 2014). In the present study, a total of six measures of spectral heterogeneity were calculated from the unclassified, continuous Landsat data, using the SD and range of the red (0.63 – 0.69 µm) ($SH_{SD\_red}$, $SH_{R\_red}$), and near infrared (NIR) (0.77 – 0.90 µm) ($SH_{SD\_NIR}$, $SH_{R\_NIR}$) wavebands, and the NDVI (NIR-red)/(NIR+red) ($SH_{SD\_NDVI}$, $SH_{R\_NDVI}$) (Table 1).

## 2.4 The data set representing the whole province of Scania and the subsets representing the three geographic districts within Scania

Each of the 1406 fully inventoried squares in the floristic inventory of the province of Scania (Tyler et al. 2007) was assigned to one of the three geographic districts; the Plains district (224 squares), the Central district (501 squares), and the Woodland district (681 squares) (Fig. 1a). Each of the four data sets (the full Scanian data set, and each of the three district subsets) was divided into two data sets (a training set and a validation set) by randomly assigning the squares to one or the other of the two data sets – giving a total of four training data sets and four validation data sets. Squares, for which the combined coverage of the SMD classes “surface water” and “built up area” (Table S1) exceeded 50%, and squares associated with spectral data considered to be affected by atmospheric disturbance were removed from the training and validation data sets. The removal of squares from the data sets reduced the total number of squares representing (i) the full Scanian data set to 965 squares (481 and 484 squares for training and validation data sets, respectively), (ii) the Plains district subset to 155 squares (84 and 71 squares for training and validation data sets, respectively), (iii) the Central district subset to 379 squares (184 and 195 squares for training and validation data sets, respectively), and (iv) the Woodland district subset to 431 squares (213 and 218 squares for training and validation data sets, respectively) that were used in the statistical analysis.

## 2.5 Statistical analysis

### *2.5.1 Training data sets*

For each of the four training data sets, the values of within-square species richness (response variables) were related to each of the three sets of explanatory heterogeneity variables (EH, SH, and HH sets of variables) with the help of the generalized additive mixed model (GAMM) regression-modelling approach (Chen 2000). The GAMM approach is an extension of general additive models (GAMs) (Hastie and Tibshirani 1986) and has the advantage that it is able to account for spatial autocorrelation within data sets (Zuur et al. 2009). Both GAMs and GAMMs allow the simultaneous fitting of both parametric and non-parametric response shapes within the same model (Chen 2000). It is argued that both approaches are appropriate for use in ecological studies, where the distributions of the relationships between response and explanatory variables are often complex or unknown (e.g. Yee and Mitchell 1991; Guisan et al. 2002; Drexler and Ainsworth 2013; Parviainen et al. 2013).

In the present study, model development was carried out on the training data, using Akaike information criterion (AIC) (Akaike 1974) values to compare and select variables for candidate models in a forward stepwise model selection procedure. AIC values take into consideration both goodness-of-fit and model complexity and, unlike likelihood ratio tests, allow the comparison of non-nested models. In the forward stepwise model selection procedure, each explanatory heterogeneity variable is first individually tested, and the variable generating the model with the lowest AIC value is entered as the first variable in the candidate model. At each subsequent step, the remaining variables are individually added to the variable(s) already present in the model, and the variable that produces a candidate model with the lowest AIC value is selected, while also ensuring that all the included smoothing terms are statistically significant ($p \leq 0.05$). The procedure is repeated until the AIC is not lowered by > 4 by the addition of a further variable: the model is then considered final. The fact that the addition of a variable to a GAMM automatically adds 2 to the AIC needs to be taken into consideration when using AIC values to compare models (Arnold 2010). To minimize the effect of multicollinearity in the statistical analyses, we disqualified an explanatory variable from entering a model if it showed a ≥ 0.7 correlation (cf. Zimmermann et al. 2007) with a variable that had entered the model at an earlier step. Smoothing selection (identification of the best-fitted distribution) for the model terms was carried out using a maximum likelihood (ML) method. The maximum number of degrees of freedom (df) was restricted to four, to avoid over-fitting and to retain interpretability of the response shapes. A Gaussian probability distribution with an identity function was assumed for the response (species-count) data, because of the large size of the data sets and in order to be able to use AIC values for model comparison (cf. the mgcv package, R Development Core Team 2016).

If spatial autocorrelation is not accounted for in models analysing the relationships between environmental conditions and species abundance, environmental factors may incorrectly be identified as being important or unimportant for the observed pattern of species diversity (cf. Lennon et al. 2000, Keitt et al. 2002, Wagner and Fortin 2005). To check, and account, for spatial patterns in the model residuals for the training data sets, we constructed five spatial residual correlation structures (exponential, Gaussian, linear, rational, spherical) that were included in the GAMMs (see Zuur et al. 2009). When the five models including spatial residual correlation structures were compared with each other and with the uncorrected models, at each step in the forward selection procedure, the models including the exponential correction structure consistently gave the lowest AIC values and, thus, showed the best ability to model the spatial autocorrelation within the training data sets (Zuur et al. 2009). Consequently, all model selection was performed with an exponential spatial residual correlation structure built into the candidate models.

For each of the four training data sets, the ability of the final model (for each of the three sets of explanatory variables) to explain within-square species richness was investigated by calculating the adjusted r-squared (r2) values, the root mean square error (RMSE), and the normalized RMSE (nRMSE) of the relationships between the species richness assessed from the model and the field-observed species richness. Finally, for each of the four training data sets, the final GAMM models derived from each of the three sets of explanatory variables (EH, SH and HH variables) were compared by calculating the differences in AIC values (ΔAIC) between the models, and the best final model was identified. If ΔAIC between two models is two, or less than two, both models are plausible candidates for the best model (Burnham and Anderson 2002, Monteiro et al. 2013). A ΔAIC value between four and seven indicates that, compared with the model with the lower AIC value, the model with the higher AIC value has "considerably less support" (in terms of its ability to explain the variation in the response data). A ΔAIC greater than ten indicates that, compared with the model with the lower AIC value, the model with the higher value has "essentially no support" as the best model (Burnham and Anderson 2002). GAMM model building was done using the mgcv package (Wood 2011) in the R statistical environment (R Development Core Team 2016).

### *2.5.2 Validation data sets*

The validation data sets were used to evaluate the best final regression models for the training data sets. Each of the four final models developed from the training data sets (for the province of Scania, and the Plains, Central and Woodland districts, respectively) was fitted to its respective validation data set and used to predict species richness in the squares belonging to the validation data set. The predictive capability of the final model, for each of the four validation data sets, was assessed with the help of Pearson correlation coefficient ($r$) values, the root mean square error of the prediction (RMSEP), and the normalized RMSEP (nRMSEP) of the relationship between the species richness predicted from the final model and the field-observed species richness. In addition, we calculated the percentage difference between the species richness predicted from the final model and the field-observed species richness for the squares within each of the four validation data sets.

## 3. Results

Levels of within-square species richness in the full Scanian data set and in each of the three geographic districts are summarized in Table 2, and the levels of within-square species richness for the full Scanian data set are also presented in Fig. 2. Summary statistics for the EH and SH explanatory variables for the full Scanian data set and each of the three geographic districts are shown in Tables S2, S3. The explanatory variables characterizing topographic heterogeneity (EHSD_elevation, EHR_elevation EHSD_slope, EHR_slope) were intercorrelated ($r \geq 0.70$, $p \leq 0.05$) within each of the eight data sets (the training and validation sets for Scania and each of the geographic districts, respectively), as were the land-cover variables (EHNP, EHTE, EHSHDI, and EHSHEI), and two of the SH variables (SHSD_red and SHSD_NDVI). The correlation coefficients between all other pairs of explanatory variables were below 0.70 ($p \leq 0.05$).

**Table 2.** Summary statistics for vascular plant species richness within 2.5 km × 2.5 km grid squares for the training and validation data sets representing the whole province of Scania, and the subsets of squares representing the Plains, Central, and Woodland districts. SD = standard deviation.

| | | Species richness | | | | |
|---|---|---|---|---|---|---|
| **Training data set** | **Number of squares** | **Mean** | **SD** | **Range** | **Min** | **Max** |
| Scania | 481 | 424 | 84 | 640 | 191 | 831 |
| Plains | 84 | 378 | 79 | 371 | 191 | 562 |
| Central | 184 | 436 | 86 | 624 | 207 | 831 |
| Woodland | 213 | 432 | 78 | 442 | 245 | 687 |
| **Validation data set** | | | | | | |
| Scania | 484 | 426 | 82 | 560 | 212 | 772 |
| Plains | 71 | 384 | 81 | 296 | 224 | 520 |
| Central | 195 | 440 | 88 | 560 | 212 | 772 |
| Woodland | 218 | 428 | 72 | 341 | 248 | 589 |

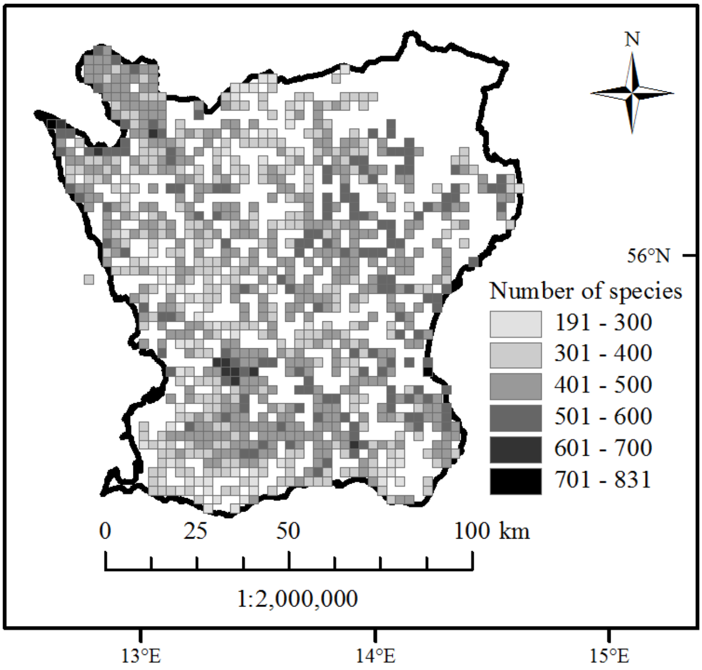


**Figure 2.** Species richness (pooled into six classes) within the 965 2.5 km × 2.5 km grid squares representing the whole province of Scania. The white squares were not included in the present study.

## 3.1 Relationships between species richness and the heterogeneity variables in the training data sets

The final EH, SH and HH models differed in their ability to explain levels of within-square species richness in the full Scanian training data set (Table 3). The adjusted $r^2$ values for the relationships between the species richness predicted from the final models and the field-observed species richness ranged between 0.385 (HH model) and 0.072 (SH model) (Table 3). The HH model had the lowest AIC value, followed by the EH and SH models (Table 3). The ΔAIC was greater than ten between the HH and EH models (ΔAIC = 16), between the HH and SH models (ΔAIC = 107), and between the EH and SH models (ΔAIC = 91) (Table 3). The results suggest that the best (final) EH and SH models have no support compared with the best HH model, and that the best SH model has no support when compared with the best EH model explaining the variation in within-square species richness for the Scanian training data set (cf. Burnham and Anderson 2002; Monteiro et al. 2013).

For the training data set representing the Plains district, the adjusted $r^2$ values for the relationships between the predicted and observed within-square species richness varied between 0.446 (HH model) and 0.212 (SH model) (Table 3). The final HH model had the lowest AIC value, followed by the final EH and SH models (Table 3). The ΔAIC scores for the comparisons between the HH and the EH models (ΔAIC = 6), between the HH and SH models (ΔAIC = 11), and between the EH and SH models (ΔAIC = 5) indicate that (i) the best (final) EH model had considerably less support than the best HH model, (ii) the best SH model had no support compared with the best HH model, and (iii) the best SH model had considerably less support than the final EH model explaining the variation in species richness for the Plains district training set of squares (cf. Burnham and Anderson 2002; Monteiro et al. 2013).

For the Central district training data set of squares, the adjusted $r^2$ values for the relationships between the assessed and field-observed species richness within squares varied between 0.454 (EH model) and 0.158 (SH model) (Table 3). There were no combinations of EH and SH variables that delivered a model that was as good, or better, than the best (final) EH model, (no HH model is presented in Table 3). The ΔAIC between the EH and SH models (ΔAIC = 54) suggests that the best SH model has no support when compared with the best EH model explaining the within-square species richness in the Central district training set of squares.

For the Woodland district training set, the adjusted $r^2$ value for the relationships between the predicted and observed within-square species richness was somewhat higher for the final EH model (0.206) than for the final HH model (0.171), whereas the $r^2$ value for the final SH model (0.113) model was lower than those of both the EH and HH models. However, the HH model had the lowest AIC value, followed by the values for the EH and SH models (Table 3). The differences in AIC scores between the HH and EH models ($\Delta$AIC = 21) and between the HH and SH models ($\Delta$AIC = 20) suggest that the best EH and SH models have no support compared with the best HH model for within-square richness. The difference in AIC score between the EH and the SH models ($\Delta$AIC = 1) implies that both models are possible candidates for the best model. However, the higher $r^2$ value of the relationship between the assessed and observed within-square species richness for the EH model compared with the SH model suggests that the performance of the best EH model was better than that of the best SH model.

**Table 3.** Adjusted r-squared values ($r^2$), AIC values, root mean square error (RMSE), and normalized RMSE (nRMSE) for the best (final) model developed from each of the training data sets (representing the whole province of Scania, and the three geographic districts within Scania; the Plains, Central, and Woodland districts, respectively) for each of the three sets of explanatory variables (environmental heterogeneity (EH) variables, spectral heterogeneity (SH) variables, and a combination of EH and SH variables (“hybrid heterogeneity”). Variable (var) 1 – 5 lists the order in which the variables enter the models.

| **Model type** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Scania** | **var 1** | **var 2** | **var 3** | **var 4** | **var 5** | $r^2$ | **AIC** |
| Environmental | EH (PR +) | EH (TE ∩∩) | EH (STD_slope +) | EH (soil PR +) | | 0.338 | 5313 |
| Spectral | SH (R_NIR +) | SH (R_NDVI +) | | | | 0.072 | 5404 |
| Hybrid | EH (PR +) | EH (TE ∩∩) | SH (STD_red ∩) | EH (STD_slope +) | SH (R_NIR +) | 0.385 | 5297 |
| **Plains** | | | | | | | |
| Environmental | EH (NP ∩) | | | | | 0.421 | 932 |
| Spectral | SH (R_NIR +) | SH (R_red ∩) | | | | 0.212 | 937 |
| Hybrid | EH (NP ∩) | SH (R_NIR +) | | | | 0.446 | 926 |
| **Central** | | | | | | | |
| Environmental | EH (SHDI +) | EH (R_elevation +) | EH (PR +) | | | 0.454 | 2015 |
| Spectral | SH (STD_NIR ∩) | | | | | 0.158 | 2069 |
| Hybrid | | | | | | | |
| **Woodland** | | | | | | | |
| Environmental | EH (soil PR +) | EH (PR +) | EH (NP -) | | | 0.206 | 2348 |
| Spectral | SH (R_NDVI +) | SH (R_NIR +) | | | | 0.113 | 2347 |
| Hybrid | SH (R_NDVI +) | EH (soil PR +) | EH (PR +) | SH (R_NIR +) | | 0.171 | 2327 |

Smoothing shapes: positive linear (+), negative linear (-), unimodal (∩), bimodal (∩∩)

## 3.2 The heterogeneity variables included in the final models for the training data sets

The best EH model for the Scanian training data set included land-cover, topographic, and soil heterogeneity explanatory variables, whereas the best HH model included land-cover, spectral and topographic heterogeneity variables (Table 3). For the Plains training data set, only one land-cover variable (EHNP) was included in the final EH model, while both the EHNP and a spectral heterogeneity variable (SHR_NIR) were included in the final HH model (Table 3). The final EH model for the Central district training data set included land-cover and topographic heterogeneity variables. For the Woodland training data set, the final EH model included soil and land-cover heterogeneity variables, whereas the final HH model included both spectral, soil and land-cover heterogeneity variables (Table 3).

Individual spectral wavebands were more commonly included in the final models than the NDVI, which was only included in the final SH and HH models for the Woodland district (Table 3).

The majority of the explanatory variables showed a positive, linear relationship with species richness (Table 3), indicating that an increase in the within-square environmental and spectral heterogeneity is associated with an increase in the level of within-square species richness. In the Scanian, Plains, and Central training data sets, a few explanatory variables (e.g. SHSTD_red, and EHNP, Table 3) showed unimodal relationships with species richness – indicating an initial positive trend in species richness at low levels of heterogeneity, but with a decreasing trend in species richness as heterogeneity increases. In the final EH and HH models for the Scanian training data set, the EHTE showed a bimodal relationship with species richness (Table 3). There was a single negative linear relationship (for EHNP in the final EH model for the Woodland training set) (Table 3).

## 3.3 Relationships between species richness predicted from the training data sets, and the observed species richness in the validation data sets

The final models that provided the best explanation of the within-square species richness in each of the four training data sets (Scania, and the Plains, Central and Woodland districts) were used to predict the within-square richness in the corresponding validation data sets. The HH models were fitted to, and used to predict species richness in, respectively, the Scanian, Plains and Woodland validation sets. The EH model was used to predict within-square richness in the Central district. The correlation coefficients between the predicted and field-observed species richness were significant ($p \leq 0.001$) for each of the four validation data sets (Table 4). The r values were lowest for the Woodland district ($r = 0.38$, $p \leq 0.001$), and highest for the Central district ($r = 0.61$, $p \leq 0.001$) (Table 4).

**Table 4.** Pearson's correlation coefficients (*r*), root mean square error (RMSE), and normalized RMSE (nRMSE (%)) between the species richness predicted from the best (final) models for each of the training data sets (484, 71, 195, 218 squares, representing the whole province of Scania and the Plains, Central, and Woodland districts, respectively), and the field-observed species richness for the squares in the independent validation data set. All the correlations are statistically significant ($p \leq 0.001$).

| Best model | *r* | RMSE | nRMSE (%) |
|---|---|---|---|
| **Scania** | | | |
| Hybrid model | 0.55 | 68.5 | 12.2 |
| **Plains** | | | |
| Hybrid model | 0.54 | 69.7 | 23.6 |
| **Central** | | | |
| Environmental model | 0.61 | 69.6 | 12.4 |
| **Woodland** | | | |
| Hybrid model | 0.38 | 66.8 | 19.6 |

The difference between the predicted and observed species richness was calculated for each of the four validation data sets. The individual grid squares within each validation data set were then assigned to one of five classes on the basis of the differences (0 - 5.0%, 5.1 - 10.0%, 10.1 - 15.0%, 15.1 - 20.0%, and > 20%) between predicted and observed richness within each square. The difference between predicted and observed species richness was less than 5.0% in 22.7% of the squares within the Scanian validation data set (Fig. 3). Within the validation data sets for the individual districts, 12.7% of the squares representing the Plains district, 30.3% of the squares representing the Central district, and 24.8% of the Woodland district squares had differences between predicted and observed richness that were less than 5.0% (Fig. 3). The differences between predicted and observed richness were lower than 20.0 % in approximately 80% of the squares within each validation data set (79.3% Scania, 75.7% Plains, 77.4 % Central, and 80.3% Woodland) (Fig. 3).

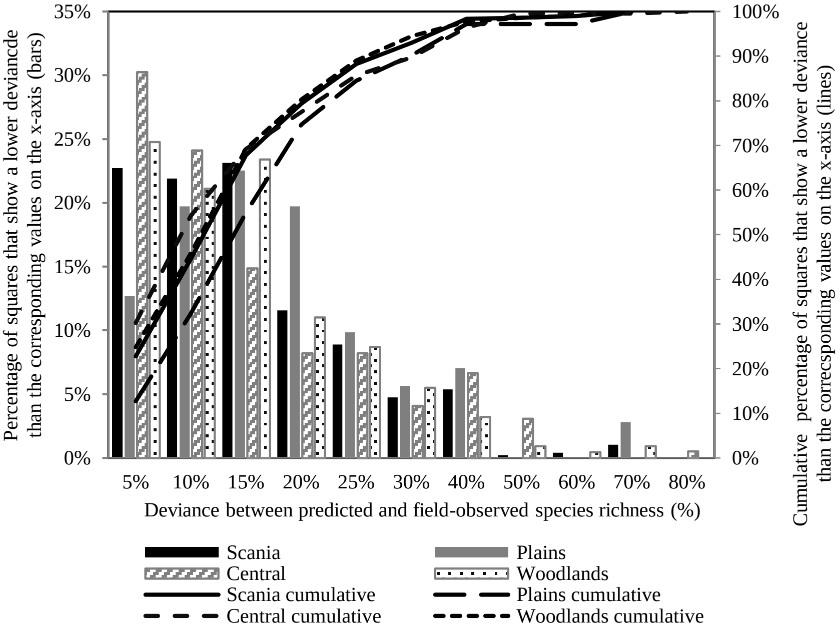


**Figure 3.** The deviance between the species richness predicted from the best (final) heterogeneity model for each of the training data sets (representing the whole province of Scania, and the Plains, Central, and Woodland districts, respectively) and the field-observed species richness in the validation data set squares, sorted into groups based on the deviance in percent for each square.

Within each of the four validation data sets, the predicted species richness tended to be underestimated for squares with high observed species richness and overestimated for squares with low observed species richness. For example, squares where species richness was underestimated by >20% had a significantly ($p \leq 0.05$) higher mean observed species richness, and squares where species richness was overestimated by >20% had a significantly ($p \leq 0.05$) lower mean observed species richness, than the overall mean observed richness within the squares in the validation data set. For the province of Scania, the observed species richness in the squares where species richness was overestimated by more than 20.0% ranged between 212 and 423 species, whereas the observed species richness for the squares where richness was underestimated by more than 20% ranged between 415 and 772 species (Fig. 4). Maps showing the locations of the validation data set squares associated with over- and underestimates of the predicted species richness, respectively, are presented in Fig. S1.

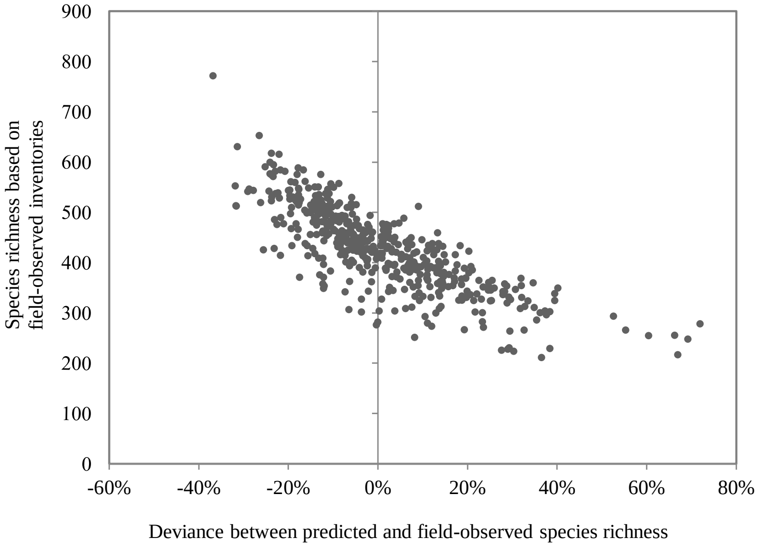


**Figure 4.** The deviance (%) between the species richness predicted from the hybrid heterogeneity model, and the field-observed species richness in the validation data set squares, plotted against the field observed species richness in the validation data set squares, for the whole province of Scania.

# 4. Discussion

The present study examined the ability of models based, respectively, on descriptors of environmental heterogeneity, spectral heterogeneity and a combination of environmental and spectral variables, to predict levels of vascular plant species richness within 2.5 km × 2.5 km grid squares in the Swedish province of Scania. The models were constructed using training data sets based on subsets of the grid squares, and the predictive ability of the models was tested with the help of species-richness data from an independent set of squares. The species richness models that included only variables characterizing environmental heterogeneity were superior to those based only on spectral heterogeneity in the data set for the whole of Scania, and in separate data sets for the Plains, Central and Woodland districts within Scania. A combination of EH and SH variables provided the best models for species richness within the Scanian, Plains and Woodland data sets, whereas the best model for the Central data set included only EH variables.

## 4.1 The best species richness models for the training data sets of squares

The study showed that a combination of descriptors of spectral heterogeneity (calculated from unclassified Landsat data) and environmental heterogeneity (based on data on land-cover, topography and soil properties) provided the best explanations of within-square species richness for the province of Scania as a whole, and for the Plains and Woodland districts (Table 3). The results of the study support those of previous studies which show that unclassified Landsat data can provide a valuable complement to environmental data in explanatory models for species richness. Parviainen et al. (2013) showed, for example, that the distributions of red-listed plant species in NE Finland were better explained by species richness models including Landsat-based explanatory variables than by models based on only data on climatic and topographic variables.

The extent to which remote sensing-based variables contribute to the explanation of species richness may vary between biogeographic regions. For example, Camathias et al. (2013) showed that the contribution of high-resolution remotely sensed data to models for species richness in Swiss forests increased with decreasing regional variability in the coarse-scale topographic and climatic conditions. Our study also shows that the contribution of remotely-sensed variables to the explanation of species richness varies between regions. Whereas the species richness models based on a combination of EH and SH variables were superior to the models based on only EH or SH variables for the full Scanian data set, as well as for the Plains and Woodland data sets, the best species richness model for the Central data set included only EH variables. At the same time, the EH variables made a somewhat greater contribution to the explanation of within-square species richness in the Central district than in the other data sets.

Variation in topography and climatic conditions has been shown to provide a good explanation of levels of species richness in mountainous regions (e.g. Moser et al. 2005), whereas land-cover provides a better explanation of species richness in regions with a limited altitudinal range (Waldhardt and Simmering 2004). Our study was carried out in a geographic area characterized by a relatively low altitudinal range and, in the final EH models for the squares within the province of Scania, and within the Plains, and Central districts, respectively, the explanatory variables with the lowest

AIC values were the land-cover EH variables. However, in the models for the Woodland district, the explanatory variable characterizing soil heterogeneity (EHSoil_PR) had the lowest AIC value. One possible explanation for the poorer contribution of the land-cover EH variables to the final species richness models for the Woodland data set compared with the other data sets may be related to the definition of the forest land-cover classes in the SMD classification (Table S1). The suite of forest land-cover classes in the SMD classification, from which the land-cover EH in our study was calculated, not only reflect the types of tree taxa that form the forest canopy but also include classes that represent different age-categories within managed coniferous forest. The SMD data were originally developed as a general tool for multi-disciplinary planning and may not be suitable for describing the environmental heterogeneity within forest regions.

For the Woodland district, the best explanatory variable − among all the variables in both the SH and HH models − was the range of the NDVI. The relationship between the NDVI and a number of vegetation properties (e.g. leaf area index, fraction of intercepted photosynthetic active radiation, absorbed photosynthetically active radiation, green biomass and primary productivity, and leaf chlorophyll content) is well-documented (e.g. Boegh et al 2013). The NDVI is often used to describe vegetation productivity or the "greenness" of vegetation (see e.g. Pettorelli et al. 2006; Parviainen et al. 2010). However, in vegetation that has a high canopy-density, the saturation of the relationship between the NDVI and vegetation productivity may affect the accuracies of NDVI-based biomass estimations (Pettorelli et al. 2005; Mutanga et al. 2012). Our results indicate that the within-square variation of the NDVI captures information on aspects of the within-square habitat heterogeneity that are associated with species richness but are not captured by the forest land-cover classes in the SMD classification. Our results are consistent with those of previous studies which show that heterogeneity measures based on the NDVI can be used in the assessment of biodiversity within forests (e.g. Parviainen et al. 2010; 2013). Parvianen et al. (2010) showed that a combination of NDVI-based measures of productivity and heterogeneity could be used as an indicator of plant-species richness in a boreal forest landscape in Finland. The present study revealed a positive association between the within-square species richness and variation in the NDVI, whereas the study by Parviainen et al. (2010) revealed both unimodal as well as positive relationships between species richness and NDVI-based measures of productivity and heterogeneity.

The positive relationships between the levels of species richness and the levels of both EH and SH generally shown in the present study are in accordance with the expectation that a diversity of habitats or niches will promote levels of species co-existence at a range of spatial scales (Whittaker 1972; Silvertown 2004). Increasingly higher levels of spatial heterogeneity may, however, also be accompanied by a decrease in the habitat area suitable for particular species, explaining the humped-shaped relationships between heterogeneity and species richness that has also been observed in some studies (cf. Stein et al. 2014). Negative heterogeneity-diversity relationships have occasionally been reported in the literature − particularly at fine spatial scales (e.g. Gazol et al. 2013). For the Woodland district, our results show a decrease in the level of species richness as the number of individual patches (irrespective of land-cover type) increases. The negative relationship between EHNP and species richness may, again, reflect the fact that the SMD land-cover classification of coniferous forest includes stand-age categories. Areas of forest where a mosaic of differently-aged coniferous stands contributes to higher values of EHNP may not necessarily represent areas that are characterized by a high number of patches representing fundamentally different environmental conditions.

### 4.2 Predicting species richness for the validation data sets with the help of the species richness models developed for the training data sets

The species richness models developed for the training data sets could be used successfully to predict the observed within-square levels of species richness within the independent validation data sets for Scania, and the Plains, Central and Woodland districts, respectively (Table 4). The differences between the predicted and observed richness were low (< 20%) for the majority (ca. 80%) of the squares for all four validation data sets (Fig. 3). In addition, the predicted species richness differed by less than five percent from the observed richness in approximately 20% of the squares in the Scanian validation data set. The squares that were associated with the largest differences between the predicted and observed species richness tended to be characterized by either high or low levels of observed species richness (Fig. 4). However, although the species richness models tended to underestimate the levels of species richness in species-rich squares, and overestimate the levels of species richness in species-poor squares, the sets of most and least species-rich squares were still correctly predicted. The tendency for predictive models for species richness to under- and overestimate the number of species in species-rich and species-poor areas, respectively, has been observed in other studies (e.g. Maes et al. 2005). Maes et al. (2005) pointed out that mismatches between the scales of the data source (from which the explanatory variables used in the predictive models are generated) and the scales on which the processes driving the species richness act, may limit the accuracy of predictive species richness models. In the present study, it is possible that the spatial resolution of the explanatory data we used (EH variables: 25 m × 25 m and 50 m × 50 m, SH variables: 30 m × 30 m, Table 1) may be too coarse to fully capture the contribution of relict patches of semi-natural habitats ("small biotopes") to the within-square heterogeneity. These small biotopes support a substantial proportion of the landscape-scale species richness in Scania (Irminger Street et al. 2015). Several additional factors may affect the performance of the models used to predict species richness in the independent validation data sets of squares in the present study. The explanatory variables used in the study only characterized within-square heterogeneity, and the levels of within-square plant species richness may also be influenced by a range of environmental variables that were not investigated in our study. The proportions of different land-cover types within squares may have an impact on the levels of within-square species richness. For example, the overall proportions of both cropland (Storkey et al. 2012) and forests (Cousins et al. 2015) in the landscape have been shown to have negative effects on the levels of plant species richness. The levels of within square-species richness may also be influenced by the proportions of land-cover types and the environmental heterogeneity in the surrounding squares (cf. Roschewitz et al. 2005, Tscharntke et al. 2005), and the degree of connectivity between habitats is also expected to have an impact on levels of species richness (Taylor et al. 1993; Donald and Evans 2006). In addition, many recent studies reveal that levels of present-day plant species richness (particularly at fine-scales) are better explained by the land-cover and spatial properties of the historical landscape than by the properties of the contemporary landscape (e.g. Reitalu et al. 2010; Hájek et al. 2016).

# 5. Conclusions

The results of the present study show that predictive models for species richness that were based only on descriptors of environmental heterogeneity were superior to models based on only variables characterizing spectral heterogeneity, and that predictive models based on a combination of environmental and spectral variables generally performed better than the models that were based on only EH or SH variables. Our results also indicate that the additional contribution of remote sensing-based explanatory variables to the explanation of species richness varies between different geographic regions.

The importance of maintaining and promoting environmental heterogeneity has been emphasized in recent studies of biodiversity and conservation-management in European landscapes (Benton et al. 2003; Tscharntke et al. 2005). To prevent further declines in species diversity it will be essential to develop information sources that support future decision-making in conservation and management planning (cf. Pettorelli et al. 2014b). Our results show that relationships between species richness and land-cover heterogeneity can be revealed with the help of simple descriptors of both environmental heterogeneity and satellite data-based spectral heterogeneity that are based on information that is in the public domain – and can be accessed free-of-charge. Such descriptors can potentially be used to inventory and monitor species richness over wide areas in cultural landscapes, and make a practical contribution to conservation planning.

# Acknowledgments

We would like to thank the Lund Botanical Society for granting us access to the database for the project Skånes Flora, and Oskar Löfgren and Barbara C. Schmid for valuable discussions. The study was financed by grants from The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) to Karin Hall and Honor C. Prentice.

# References

Akaike, H., (1974) A new look at the statistical model identification (1974). IEEE Transactions on Automatic Control 19: 716-723.

Arnold, T., (2010). Uninformative parameters and model selection using Akaike's information criterion. Journal of Wildlife Management 74:1175-1178.

Belfrage, K., Bjorklund, J., Salomonsson, L. (2015) Effects of farm size and on-farm landscape heterogeneity on biodiversity - case study of twelve farms in a Swedish landscape. Agroecology and Sustainable Food Systems 39: 170-188.

Benton, T. G., Vickery, J. A., Wilson, J. D. (2003) Farmland biodiversity: is habitat heterogeneity the key? Trends in Ecology & Evolution 18: 182-188.

Bernes, C. (Ed.) (1994) Biological Diversity in Sweden.  A Country Study. Monitor 14. Swedish Environmental Protection Agency, Solna, Sweden.

Bignal, E. M., McCracken, D. I. (1996) Low-intensity farming systems in the conservation of the countryside. Journal of Applied Ecology 33: 1591-1591.

Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek, R., Burel, F., Cerny, M., De Blust, G., De Cock, R., Diekotter, T., Dietz, H., Dirksen, J., Dormann, C., Durka, W., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., Le Coeur, D., Maelfait, J. P., Opdam, P., Roubalova, M., Schermann, A., Schermann, N., Schmidt, T., Schweiger, O., Smulders, M. J. M., Speelmans, M., Simova, P., Verboom, J., van Wingerden, W. K. R. E., Zobel, M., Edwards, P. J. (2008) Indicators for biodiversity in agricultural landscapes: a pan-European study. Journal of Applied Ecology 45: 141-150.

Boegh, E., Houborg, R., Bienkowski, J., Braban, C. F., Dalgaard, T., van Dijk, N., Dragosits, U., Holmes, E., Magliulo, V., Schelde, K., Di Tommasi, P., Vitale, L., Theobald, M. R., Cellier, P., Sutton, M. A. (2013) Remote sensing of LAI, chlorophyll and leaf nitrogen pools of crop-and grasslands in five European landscapes. Biogeosciences 10: 6279-6307.

Burnham, K. P., Anderson, D. R. (2002) Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach. Springer-Verlag New York, Inc.

Camathias, L., Bergamini, A., Kuchler, M., Stofer, S., Baltensweiler, A. (2013) High-resolution remote sensing data improves models of species richness. Applied Vegetation Science 16: 539-551.

Chander, G. Markham, B.L., Helder, D.L. (2009) Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sensing of Environment 113: 893–903.

Chen, C. (2000) Generalized additive mixed models. Communications in Statistics - Theory and Methods 29: 1257-1271.

CORINE Land Cover project. http://www.eea.europa.eu/publications/COR0-landcover. Downloaded 2016-05-10.

Cousins, S. A. O., Auffret, A. G., Lindgren, J., Trank, L. (2015) Regional-scale land-cover change during the 20th century and its consequences for biodiversity. Ambio 44: 17-27.

Ding, Y. L., Zhao, K., Zheng, X. M., Jiang, T. (2014) Temporal dynamics of spatial heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8 OLI imagery. International Journal of Applied Earth Observation and Geoinformation 30: 139-145.

Donald, P. F., Evans, A. D. (2006) Habitat connectivity and matrix restoration: the wider implications of agri-environment schemes. Journal of Applied Ecology 43: 209-218.

Drexler, M., Ainsworth, C. H. (2013) Generalized additive models used to predict species abundance in the Gulf of Mexico: An ecosystem modeling tool. PLoS ONE 8: e64458.

Garrigues, S., Allard, D., Baret, F., Weiss, M. (2006) Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. Remote Sensing of Environment 105: 286-298.

Garrigues, S., Allard, D., Baret, F., Morisette, J. (2008) Multivariate quantification of landscape spatial heterogeneity using variogram models. Remote Sensing of Environment 112: 216-230.

Gazol, A., Tamme, R., Price, J. N., Hiiesalu, I., Laanisto, L., Partel, M. (2013) A negative heterogeneity-diversity relationship found in experimental grassland communities. Oecologia 173: 545-555.

Germundsson, T., Schlyter, P. (Eds.) (1999) Atlas över Skåne. Sveriges Nationatlas förlag. Stockholm, SE.

Guisan, A., Edwards, T. C., Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157: 89-100.

Hajek, M., Dudova, L., Hajkova, P., Rolecek, J., Moutelikova, J., Jamrichova, E., Horsak, M. (2016) Contrasting Holocene environmental histories may explain patterns of species richness and rarity in a Central European landscape. Quaternary Science Reviews 133: 48-61.

Hastie, T., Tibshirani, R., (1986) Generalized additive models. Statistical Science 1: 297-318.

Ihse, M. (1995) Swedish agricultural landscapes - patterns and changes during the last 50 years, studied by aerial photos. Landscape and Urban Planning 31: 21-37.

Irminger-Street, T., Prentice, H. C., Hall, K., Smith, H. G., Olsson, O. (2015) Removal of woody vegetation from uncultivated field margins is insufficient to promote non-woody vascular plant diversity. Agriculture Ecosystems & Environment 201: 1-10.

Johansson, L. J., Hall, K., Prentice, H. C., Ihse, M., Reitalu, T., Sykes, M. T., Kindstrom, M. (2008) Semi-natural grassland continuity, long-term land-use change and plant species richness in an agricultural landscape on Oland, Sweden. Landscape and Urban Planning 84: 200-211.

Keitt, T., Bjørnstad, O., Dixon, P., Citron-Pousty, S. (2002) Accounting for spatial pattern when modeling organism-environment interactions. Ecography 25: 616-625.

Krauss, J., Bommarco, R., Guardiola, M., Heikkinen, R. K., Helm, A., Kuussaari, M., Lindborg, R., Ockinger, E., Partel, M., Pino, J., Poyry, J., Raatikainen, K. M., Sang, A., Stefanescu, C., Teder, T., Zobel, M., Steffan-Dewenter, I. (2010) Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. Ecology Letters 13: 597-605.

Lennon, J. (2000) Red-shifts and red herrings in geographical ecology. Ecography 23: 101-113.

Maes, D., Bauwens, D., De Bruyn, L., Anselin, A., Vermeersch, G., Van Landuyt, W., De Knijf, G., Gilbert, M. (2005) Species richness coincidence: conservation strategies based on predictive modelling. Biodiversity and Conservation 14: 1345-1364.

McGarigal, K., Cushman, S. A, Ene, E. (2012) FRAGSTATS v4: Spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. http://www.umass.edu/landeco/research/fragstats/fragstats.html. Downloaded 2016-05-10.

Monteiro, A. T., Fava, F., Goncalves, J., Huete, A., Gusmeroli, F., Parolo, G., Spano, D., Bocchi, S. (2013) Landscape context determinants to plant diversity in the permanent meadows of Southern European Alps. Biodiversity and Conservation 22: 937-958.

Moser, D., Dullinger, S., Englisch, T., Niklfeld, H., Plutzar, C., Sauberer, N., Zechmeister, H. G., Grabherr, G. (2005) Environmental determinants of vascular plant species richness in the Austrian Alps. Journal of Biogeography 32: 1117-1127.

Mutanga, O., Adam, E., Cho, M. A. (2012) High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. International Journal of Applied Earth Observation and Geoinformation 18: 399-406.

Niemela, J., Young, J., Alard, D., Askasibar, M., Henle, K., Johnson, R., Kurttila, M., Larsson, T. B., Matouch, S., Nowicki, P., Paiva, R., Portoghesi, L., Smulders, R., Stevenson, A., Tartes, U., Watt, A. (2005) Identifying, managing and monitoring conflicts between forest biodiversity conservation and other human interests in Europe. Forest Policy and Economics 7: 877-890.

Oliver, M. A., Shine, J. A., Slocum, K. R. (2005) Using the variogram to explore imagery of two different spatial resolutions. International Journal of Remote Sensing 26: 3225-3240.

Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S., Wohlgemuth, T. (2002) Quantitative tools for perfecting species lists. Environmetrics 13: 121-137.

Parviainen, M., Luoto, M., Heikkinen, R. K. (2010) NDVI-based productivity and heterogeneity as indicators of plant-species richness in boreal landscapes. Boreal Environment Research 15: 301-318.

Parviainen, M., Zimmermann, N. E., Heikkinen, R. K., Luoto, M. (2013) Using unclassified continuous remote sensing data to improve distribution models of red-listed plant species. Biodiversity and Conservation 22: 1731-1754.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., Stenseth, N. C. (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology & Evolution 20: 503-510.

Pettorelli, N., Gaillard, J. M., Mysterud, A., Duncan, P., Stenseth, N. C., Delorme, D., Van Laere, G., Toigo, C., Klein, F. (2006) Using a proxy of plant productivity (NDVI) to find key periods for animal performance: the case of roe deer. Oikos 112: 565-572.

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., Turner, W. (2014a) Satellite remote sensing for applied ecologists: opportunities and challenges. Journal of Applied Ecology 51: 839-848.

Pettorelli, N., Safi, K., Turner, W. (2014b) Satellite remote sensing, biodiversity research and conservation of the future. Philosophical Transactions of the Royal Society B-Biological Sciences 369: 20130190.

Poschlod, P., WallisDeVries, M. F. (2002) The historical and socioeconomic perspective of calcareous grasslands - lessons from the distant and recent past. Biological Conservation 104: 361-376.

R Development Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016.

Reitalu, T., Johansson, L. J., Sykes, M. T., Hall, K., Prentice, H. C. (2010) History matters: village distances, grazing and grassland species diversity. Journal of Applied Ecology 47: 1216-1224.

Robinson, R. A., Sutherland, W. J. (2002) Post-war changes in arable farming and biodiversity in Great Britain. Journal of Applied Ecology 39: 157-176.

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., Southworth, J., Neteler, M. (2010) Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. Ecological Informatics 5: 318-329.

Roschewitz, I., Gabriel, D., Tscharntke, T., Thies, C. (2005) The effects of landscape complexity on arable weed species diversity in organic and conventional farming. Journal of Applied Ecology 42: 873-882.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W. (1973) Monitoring vegetation systems in the great plains with ERTS. In: Proceedings of the Third ERTS Symposium NASA SP-351, 1, U.S. Government Printing Office Washington DC. 309-317.

Sheeren, D., Bonthoux, S., Balent, G. (2014) Modeling bird communities using unclassified remote sensing imagery: effects of the spatial resolution and data period. Ecological Indicators 43: 69-82.

Silvertown, J. (2004) Plant coexistence and the niche. Trends in Ecology & Evolution 19: 605-611.

Stein, A., Gerstner, K., Kreft, H. (2014) Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. Ecology Letters 17: 866-880.

Stein, A., Kreft, H. (2015) Terminology and quantification of environmental heterogeneity in species-richness research. Biological Reviews 90: 815-836.

Storkey, J., Meyer, S., Still, K. S., Leuschner, C. (2012) The impact of agricultural intensification and land-use change on the European arable flora. Proceedings of the Royal Society B-Biological Sciences 279: 1421-1429.

Svenska Marktäckedata (SMD). http://gis-services.metria.se/nvfeed/atom/annex2.xml. Downloaded 2016-05-10.

Taylor, P. D., Fahrig L., Henein, K., Merriam, G. (1993) Connectivity is a vital element of landscape structure. Oikos 68: 571–573.

Tscharntke, T., Klein, A. M., Kruess, A., Steffan-Dewenter, I., Thies, C. (2005) Landscape perspectives on agricultural intensification and biodiversity - ecosystem service management. Ecology Letters 8: 857-874.

Tucker, C. J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8: 127–150.

Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A. K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koh, L. P., Leimgruber, P., Taubenboeck, H., Wegmann, M., Wikelski, M., Woodcock, C. (2015) Free and open-access satellite data are key to biodiversity conservation. Biological Conservation 182: 173-176.

Tyler, T., Olsson, K.-A., Johansson, H., Sonesson, M. (2007) Floran i Skåne. Arterna och deras utbredning. Lunds Botaniska Förening, Lund.

Waldhardt, R., Simmering, D., Otte, A. (2004) Estimation and prediction of plant species richness in a mosaic landscape. Landscape Ecology 19: 211-226.

Wagner, H., Fortin, M-J. Spatial analysis of landscapes: concepts and statistics. Ecology 86: 1975-1987.

Whittaker, R. H. (1972) Evolution and measurement of species diversity. Taxon 21: 213-251.

Wood, S. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society B 73: 3-36.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P. S., Vermote, E. F., Vogelmann, J., Wulder, M. A., Wynne, R. (2008) Free access to Landsat imagery. Science 320: 1011.

Wramner, P. Levin, L., Milivojevic, M. (2003) Ett levande kulturlandskap. Statens offentliga utredningar 2003:105. Stockholm: Jordbruksdepartmentet.

Yang, Z. Y., Liu, X. Q., Zhou, M. H., Ai, D., Wang, G., Wang, Y. S., Chu, C. J., Lundholm, J. T. (2015) The effect of environmental heterogeneity on species richness depends on community position along the environmental gradient. Scientific Reports 5: 15723.

Yee, T. W., Mitchell, N. D. (1991) Generalized additive-models in plant ecology. Journal of Vegetation Science 2: 587-602.

Zimmermann, N. E., Edwards, T. C., Moisen, G. G., Frescino, T. S., Blackard, J. A. (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. Journal of Applied Ecology 44: 1057-1067.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., Smith, G. (2009) Mixed effects models and extensions in ecology with R. Springer, New York.

# Supplementary information

**Table S1**. ID-codes and class names for the 51 land-cover classes in the SMD data used in the present study.

| SMD ID | SMD class name | SMD ID | SMD class name |
|---:|---|---:|---|
| 141 | Green urban area | 3131 | Mixed woodland |
| 1422 | Airport | 3132 | Mixed woodland mire |
| 1424 | Golf course | 3133 | Mixed woodland rock |
| 1425 | Nonurban park | 3241 | Scrub |
| 1426 | Camping holiday homes | 211 | Arable land |
| 111 | City center | 222 | Permanent crop |
| 121 | Industrial area | 231 | Gr. grassl. |
| 122 | Road railroad | 321 | Natural grassland |
| 123 | Harbour | 322 | Moorland |
| 124 | Airport | 511 | Watercourse |
| 132 | Dumps | 521 | Coastal lagoon |
| 133 | Construction site | 522 | Estuary |
| 1122 | Village | 5121 | Open lake pond |
| 1123 | Rural settlement | 5122 | Covered lake pond |
| 1421 | Sport leisure facility | 5231 | Open sea |
| 11211 | City urban | 5232 | Covered sea |
| 11212 | City suburb | 331 | Beaches dunes sand planes |
| 3122 | Coniferous woodland mire | 332 | Bare rock |
| 3123 | Coniferous woodland rock | 1311 | Sand gravel extraction |
| 3242 | Clearing | 1312 | Mineral extraction |
| 3243 | Young forest | 411 | Marshland |
| 312121 | Coniferous woodland 5 - 15y | 421 | Saltmarsh |
| 312122 | Coniferous woodland >15y | 4121 | Wet mire |
| 3111 | Deciduous woodland | 4122 | Mire |
| 3112 | Deciduous woodland mire | 4123 | Peat extraction site |
| 3113 | Deciduous woodland rock | | |

**Table S2.** Summary statistics for the environmental heterogeneity variables within the 2.5 km × 2.5 km grid squares representing the whole province of Scania and the Plains, Central, and Woodland districts, respectively. SD = standard deviation.

| **Scania (n = 965)** | | | | |
|---|---|---|---|---|
| *Topography* | **Mean** | **Min** | **Max** | **SD** |
| $EH_{SD\_elevation}$ | 9.8 | 0.9 | 69.2 | 7.7 |
| $EH_{R\_elevation}$ | 46.1 | 4.9 | 187.5 | 28.6 |
| $EH_{SD\_slope}$ | 1.3 | 0.2 | 6.7 | 0.8 |
| $EH_{R\_slope}$ | 8.8 | 1.9 | 39.8 | 4.9 |
| *Soil* | | | | |
| $EH_{Soil\_PR}$ | 10.8 | 2 | 21 | 3.6 |
| *Land-cover* | | | | |
| $EH_{NP}$ | 63 | 1 | 173 | 44.8 |
| $EH_{TE}$ | 47854 | 0 | 110050 | 28689 |
| $EH_{PR}$ | 10.5 | 1 | 23 | 3.6 |
| $EH_{SHDI}$ | 1.33 | 0 | 2.57 | 0.66 |
| $EH_{SHEI}$ | 0.55 | 0 | 0.94 | 0.23 |

| **Central (n = 379)** | | | | |
|---|---|---|---|---|
| *Topography* | **Mean** | **Min** | **Max** | **SD** |
| $EH_{SD\_elevation}$ | 9.2 | 0.9 | 40.4 | 7.2 |
| $EH_{R\_elevation}$ | 43.7 | 4.9 | 185.2 | 28.4 |
| $EH_{SD\_slope}$ | 1.2 | 0.2 | 6.7 | 0.8 |
| $EH_{R\_slope}$ | 8.4 | 1.9 | 39.8 | 5.2 |
| *Soil* | | | | |
| $EH_{Soil\_PR}$ | 11.8 | 3 | 21 | 3.3 |
| *Land-cover* | | | | |
| $EH_{NP}$ | 40.9 | 2 | 133 | 27.2 |
| $EH_{TE}$ | 34729 | 800 | 95740 | 20449 |
| $EH_{PR}$ | 9.7 | 2 | 23 | 3.4 |
| $EH_{SHDI}$ | 1.1 | 0.22 | 2.44 | 0.58 |
| $EH_{SHEI}$ | 0.47 | 0.02 | 0.92 | 0.21 |

| **Plains (n = 155)** | | | | |
|---|---|---|---|---|
| *Topography* | **Mean** | **Min** | **Max** | **SD** |
| $EH_{SD\_elevation}$ | 6.8 | 2.2 | 21.7 | 3.6 |
| $EH_{R\_elevation}$ | 33.7 | 11.8 | 87.2 | 15.7 |
| $EH_{SD\_slope}$ | 1.2 | 0.4 | 3.8 | 0.6 |
| $EH_{R\_slope}$ | 7.8 | 3.2 | 22.8 | 3.8 |
| *Soil* | | | | |
| $EH_{Soil\_PR}$ | 13.7 | 6 | 18 | 2.9 |
| *Land-cover* | | | | |
| $EH_{NP}$ | 21.3 | 1 | 92 | 17.3 |
| $EH_{TE}$ | 19403 | 0 | 78275 | 15003 |
| $EH_{PR}$ | 8 | 1 | 19 | 3.7 |
| $EH_{SHDI}$ | 0.71 | 0 | 2.25 | 0.52 |
| $EH_{SHEI}$ | 0.32 | 0 | 0.8 | 0.2 |

| **Woodland (n = 431)** | | | | |
|---|---|---|---|---|
| *Topography* | **Mean** | **Min** | **Max** | **SD** |
| $EH_{SD\_elevation}$ | 11.4 | 2.9 | 69.2 | 8.7 |
| $EH_{R\_elevation}$ | 52.6 | 14.5 | 187.5 | 30.5 |
| $EH_{SD\_slope}$ | 1.5 | 0.3 | 5.7 | 0.8 |
| $EH_{R\_slope}$ | 9.6 | 2.4 | 36.1 | 4.8 |
| *Soil* | | | | |
| $EH_{Soil\_PR}$ | 8.9 | 2 | 19 | 2.9 |
| *Land-cover* | | | | |
| $EH_{NP}$ | 97.3 | 5 | 173 | 38.8 |
| $EH_{TE}$ | 69627 | 4675 | 110050 | 21868 |
| $EH_{PR}$ | 12.1 | 3 | 22 | 3 |
| $EH_{SHDI}$ | 1.75 | 0.14 | 2.57 | 0.46 |
| $EH_{SHEI}$ | 8.9 | 2 | 19 | 2.9 |

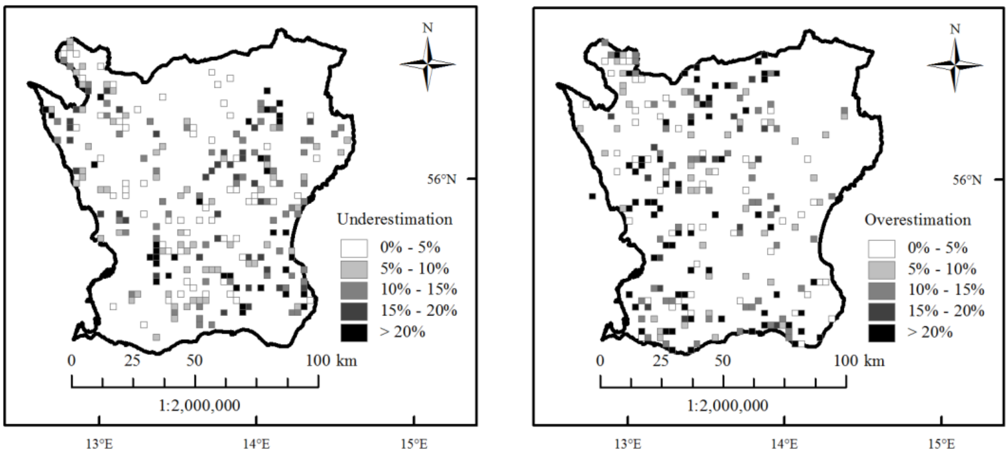
**Table S3.** Summary statistics for the spectral heterogeneity variables within the 2.5 km × 2.5 km grid squares representing the whole province of Scania and the Plains, Central, and Woodland districts, respectively. SD = standard deviation.

| **Scania (n = 965)** | | | | |
|---|---|---|---|---|
| *Spectral var.* | **Mean** | **Min** | **Max** | **SD** |
| $SH_{SD\_red}$ | 0.016 | 0.004 | 0.042 | 0.005 |
| $SH_{R\_red}$ | 0.13 | 0.044 | 0.355 | 0.046 |
| $SH_{SD\_NIR}$ | 0.055 | 0.029 | 0.088 | 0.008 |
| $SH_{R\_NIR}$ | 0.369 | 0.21 | 0.543 | 0.05 |
| $SH_{SD\_NDVI}$ | 0.109 | 0.028 | 0.227 | 0.032 |
| $SH_{R\_NDVI}$ | 0.721 | 0.333 | 1 | 0.126 |

| **Central (n = 379)** | | | | |
|---|---|---|---|---|
| *Spectral var.* | **Mean** | **Min** | **Max** | **SD** |
| $SH_{SD\_red}$ | 0.018 | 0.008 | 0.042 | 0.004 |
| $SH_{R\_red}$ | 0.142 | 0.06 | 0.355 | 0.047 |
| $SH_{SD\_NIR}$ | 0.055 | 0.035 | 0.088 | 0.009 |
| $SH_{R\_NIR}$ | 0.38 | 0.244 | 0.543 | 0.053 |
| $SH_{SD\_NDVI}$ | 0.124 | 0.061 | 0.223 | 0.029 |
| $SH_{R\_NDVI}$ | 0.764 | 0.483 | 1 | 0.108 |

| **Plains (n = 155)** | | | | |
|---|---|---|---|---|
| *Spectral var.* | **Mean** | **Min** | **Max** | **SD** |
| $SH_{SD\_red}$ | 0.019 | 0.012 | 0.038 | 0.004 |
| $SH_{R\_red}$ | 0.15 | 0.085 | 0.329 | 0.043 |
| $SH_{SD\_NIR}$ | 0.052 | 0.036 | 0.075 | 0.008 |
| $SH_{R\_NIR}$ | 0.369 | 0.251 | 0.528 | 0.052 |
| $SH_{SD\_NDVI}$ | 0.127 | 0.077 | 0.227 | 0.027 |
| $SH_{R\_NDVI}$ | 0.797 | 0.527 | 1 | 0.096 |

| **Woodland (n = 431)** | | | | |
|---|---|---|---|---|
| *Spectral var.* | **Mean** | **Min** | **Max** | **SD** |
| $SH_{SD\_red}$ | 0.13 | 0.004 | 0.041 | 0.004 |
| $SH_{R\_red}$ | 0.112 | 0.044 | 0.355 | 0.038 |
| $SH_{SD\_NIR}$ | 0.056 | 0.029 | 0.08 | 0.007 |
| $SH_{R\_NIR}$ | 0.36 | 0.21 | 0.495 | 0.045 |
| $SH_{SD\_NDVI}$ | 0.09 | 0.028 | 0.203 | 0.026 |
| $SH_{R\_NDVI}$ | 0.657 | 0.333 | 1 | 0.119 |



a)

b)

**Figure S1.** The spatial location of squares in which species richness, predicted from the hybrid heterogeneity model was (a) underestimated and (b) overestimated in the validation data set (484 squares) for the province of Scania.

# Paper II

# Comparisons of historical and contemporary Landsat data reveal relationships between present-day plant species richness and patterns of land-cover change in cultural landscapes

Jonas Dalmayne, Honor C. Prentice, Karin Hall

## Abstract

Changes in land-use/land-cover (LULC) are regarded as a global threat to biodiversity. Remote sensing techniques, based on satellite scenes, can be used to characterize patterns of LULC during different time-periods and can potentially be used to investigate the impact of changes in LULC on contemporary species diversity. The present study makes use of data on vascular plant species richness (collected between the years 1987 – 2005) within each of 473 2.5 × 2.5 km grid squares in the southern part of the Swedish province of Scania, together with Landsat scenes from the mid-1970s (historical Landsat scenes) and the period around the turn of the millennium (contemporary Landsat scenes) covering the same geographic area. We used growing-season Landsat data and support vector machines to provide a supervised classification of LULC classes in the historical and contemporary landscapes. We also investigated whether the inclusion of ancillary data (elevation data from a digital elevation model, and data on seasonal differences in the normalized difference vegetation index (NDVI) improved the classification accuracy. We also examined whether the LULC data could be used to assess changes in LULC composition and heterogeneity between the historical and contemporary landscapes. Finally we analysed the extent to which levels of contemporary within-square species richness could be explained by changes in the proportions of eight LULC classes and changes in levels of LULC heterogeneity, between the historical and contemporary landscapes. Satisfactory (≥85% overall accuracy) classifications for the LULC classes could be produced for both the historical and contemporary landscapes using a combination of growing-season Landsat data and ancillary data. Within-square proportions of cropland and coniferous forest decreased, while the proportions of grazed grassland and deciduous forest increased between 1975 and 2001. Levels of present-day species richness were negatively related to the proportion of cropland, and positively related to the proportions of deciduous forest and wetlands in both the historical and contemporary landscapes. An increase in the within-square proportion of deciduous forest between 1975 and 2001 was positively correlated with species richness, and the 30 most species-rich squares showed a higher increase in the proportion of deciduous forest than in the 30 squares with the lowest species richness. Levels of LULC heterogeneity were positively correlated with species richness, and the 30 most species-rich squares showed a greater increase in heterogeneity than the 30 squares with the lowest species richness. We conclude that satisfactory LULC and heterogeneity data can be produced for the southern part of the Swedish province of Scania for the years 1975 and 2001 using Landsat data acquired during the growing season in combination with ancillary data. Even on the relatively short time-scale of 26 years, the contemporary within-square species richness of vascular plants shows significant associations with changes in the proportion of deciduous forest and levels of LULC heterogeneity.

# 1. Introduction

Land-use change is one of the most important factors underlying the global loss of biodiversity (Lambin et al. 2001) and land-use change is expected to continue to drive the loss of biodiversity during the coming century (Chapin et al. 2000). In Europe, the traditional cultural landscapes evolved in response to hundreds of years of human land-use and were characterized by extensive areas of species-rich semi-natural habitats, such as pastures, meadows and forest margins, and high levels of species diversity (Poschlod and WallisDeVries 2002). Since the industrial revolution, land-use intensification has led to the loss, fragmentation and degradation of these species-rich semi-natural habitats in many areas of Europe (Poschlod and WallisDeVries, 2002), and to an overall reduction in the structural complexity of the landscape (landscape heterogeneity) (e.g. Benton et al. 2003). Many plant species are dependent on the remaining fragments of semi-natural habitat in the modern cultural landscape (Benton et al. 2003; Tscharntke et al. 2005) and, together with deciduous forests and wetland habitats (Brinson and Malvárez 2002; Gilliam 2007), these relict fragments support a substantial proportion of the present-day species diversity. As the relict patches of semi-natural habitat become increasingly more fragmented, disjunct and isolated (Ihse 1995), habitat quality is increasingly affected by mineral fertilizers and agrochemicals from the surrounding landscapes (Robinson and Sutherland 2002). Changes in land-use and land-cover are now recognized as the primary cause of the dramatic biodiversity decline in many modern European cultural landscapes (e.g. Storkey et al. 2012).

Although it is accepted that large scale land-cover change is a major driver of biodiversity loss (Haines-Young 2009), studies of the impact of habitat loss and fragmentation on plant species richness typically have a focus on particular types of habitat, such as semi-natural grasslands (e.g. Helm et al. 2006; Reitalu et al. 2012) and forests (e.g. Vellend et al. 2006). The majority of such studies are carried out on a relatively fine scale and examine the local loss of a specific habitat type. For example, in a fragmented agricultural landscape in Sweden, Reitalu et al. (2012) showed that the response of grassland species richness to the loss of semi-natural grassland habitats differed between spatial scales (polygon or plot scale) and between habitat specialists and generalists. Fewer studies have examined relationships between large-scale patterns of land-cover change and species diversity (but see Hooftman and Bullock 2012; Cousins et al. 2015). However, in order to evaluate the overall impact of land-cover change on biodiversity, we not only need information on the way in which the number of species is affected by the pattern of loss, over time, of particular type of habitat, but also a more integrated overview of the way in which present-day species richness is related to overall changes in the proportions of different types of land-cover – at larger scales within landscapes or regions.

Because different habitat types or land-use regimes provide suitable conditions for different suites of species, a decline in the proportion of a particular habitat type within the landscape is likely to be accompanied by the local or regional loss of the habitat specialist species that are dependent on that habitat (e.g. Reitalu et al. 2012). In addition, again because different suites of species are adapted to different habitat conditions, there is a general expectation that overall species richness at the regional level will be related to land-cover heterogeneity (Tscharntke et al. 2005). Land-cover heterogeneity can be characterized in terms of the numbers of different land-cover classes (e.g. semi-natural grasslands, deciduous forest, wetlands), and their spatial distribution and configuration in the landscape (cf. Li and Reynolds 1995). A high level of land-cover heterogeneity is expected to be accompanied by a higher level of species richness than that found under more uniform land-cover conditions at the same spatial scale (cf. Stein et al. 2014), and a decline in heterogeneity, over time, may be expected to have a negative impact on the overall species richness within a landscape (Tscharntke et al. 2005).

Studies of regional land-cover change require detailed information on land-use/land-cover (LULC) in both the contemporary landscape and in the historical landscape. And, if changes in LULC and landscape heterogeneity are to be related to levels of present-day species diversity, there is also a need for reliable, systematic inventories of biodiversity that cover the studied region. Local comparisons of LULC between historical and contemporary landscapes are generally based on data extracted from historical maps and/or aerial photographs (e.g. Gerard et al. 2010). However, large-scale comparisons based on these types of data may be challenging. A recent study (Gerard et al. 2010) of land-cover changes across Europe between 1950 and 2000 highlighted some of the difficulties associated with the use of aerial photographs in the analysis of LULC change over wide areas. The study emphasised the need for standardized methodology and nomenclature, as well as the importance of the choice of spatial scale.

Satellite-based data provide a remote sensing tool that is more appropriate than aerial photos for the assessment of LULC over wide geographic areas, and the advantages of using satellite-based remote sensing techniques for detecting, characterizing, and monitoring LULC reliably at different spatial and temporal scales are widely recognized (e.g. Honnay et al. 2003; Comber et al. 2016; Tarantino et al. 2016). A recent review (Willis 2015) identifies Landsat satellite data as being particularly useful for acquiring information on LULC and temporal changes in LULC. Landsat data are available (at no cost) from the 1970s to the present day (Willis 2015). However, there may be technical challenges associated with the use of historical and contemporary satellite scenes to classify LULC and assess temporal changes in land-cover. Historical satellite scenes have a poorer resolution than contemporary scenes (cf. Sloan 2012), and difficulties with discriminating between particular types of land-cover may reduce the classification accuracy of LULC classes – in both historical and contemporary satellite scenes.

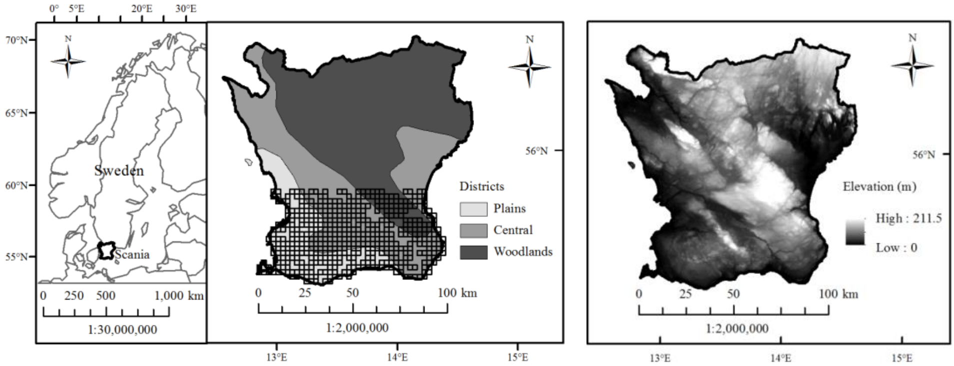
The generation of information on LULC from satellite data is commonly carried out with the help of supervised or unsupervised classification of the remote sensing data. Unsupervised classification approaches are directly based on the remote sensing data. Supervised approaches usually employ reference data (training data) derived from, for example, the interpretation of aerial photos or from field-based inventories, to build a model of the classification algorithm. The classification algorithm is then used to identify each LULC class within the geographic area covered by the entire remote sensing image. However, if LULC types are spectrally similar, the remote sensing-based information on land-cover types may be inaccurate. For example, it may be difficult to classify LULC classes such as cropland, pasture, and natural savanna vegetation, with the help of satellite data acquired within a single season – because of a lack of spectral dissimilarities between these types of land-cover during that season (Müller et al. 2015). Recent studies show that it may be possible to improve the classification accuracies for spectrally similar LULC classes by including spectral information on seasonal changes in the vegetation (e.g. Müller et al. 2015; Schuster et al. 2015; Zoungrana et al. 2015). For example, Müller et al. (2015) showed that the inclusion of multi-temporal Landsat-derived measures of the normalized difference vegetation index (NDVI) (Rouse et al. 1973, Tucker 1979) in the classification procedure captured phenological variation in the vegetation characteristics, and improved the classification accuracies of spectrally similar land-cover types in a savanna landscape in Brazil.

In the present study of the Swedish province of Scania, we investigate whether supervised classification of Landsat data can be used to produce reliable information on eight LULC classes, within both the historical (mid-1970s) and contemporary (turn-of-millennium) landscapes. The three geographic districts within Scania represent a gradient in farming intensity and landscape heterogeneity – ranging from intensively farmed and homogeneous landscapes, through more heterogeneous agricultural landscapes, to landscapes that are dominated by forests. High-quality data on vascular plant species richness within 2.5 km × 2.5 km grid squares were obtained from a modern floristic inventory of Scania (Tyler et al. 2007). We asked five main questions. 1) Can Landsat data be used to provide accurate classifications of LULC classes in both the historical and contemporary landscapes within the province of Scania and each of its three geographic districts?  2) Is the classification accuracy improved if the Landsat data are complemented with ancillary environmental data (topographic data) and data on seasonal changes in the NDVI? 3) Is present-day within-square vascular plant species richness (the total number of vascular plant species, red listed species, and grassland habitat specialists, respectively) related to Landsat-based information on the within-square proportions of LULC classes and measures of landscape heterogeneity in the historical and contemporary landscapes, respectively? 4) Are the three categories of present-day within-square vascular plant species richness related to changes in Landsat-based information on the proportions of LULC classes and land-cover heterogeneity between the historical and contemporary landscapes? 5) Does the level of change (between the historical and contemporary landscapes) in the proportions of LULC classes and measures of landscape heterogeneity differ between” hot” squares (squares with high present-day plant species richness) and “cold” squares (squares with low present-day species richness) in Scania?

# 2. Methods

## 2.1 Study area

The study area was the southern part of the province of Scania in southern Sweden (Fig. 1a), and represents approximately 3 000 km2 (out of the 11 000 km2 covered by the whole province). Scania can be divided into three geographic districts, on the basis of differences in geological, edaphic, topographic and climatic conditions (Germundsson and Schlyter 1999) (Fig. 1a). The geology of the plains of



a) b)

**Figure 1**. (a) the location of the province of Scania (inset) and map showing the study area grid squares in relation to the Plains, Central, and Woodland districts within Scania (b) map showing elevation (metres above sea-level).

Southwestern Scania (the "Plains district") and Central Scania (the "Central district") (Fig. 1a) is dominated by sedimentary bedrock, generally overlaid by fertile, clayey moraine soils (Germundsson and Schlyter 1999). The "Woodland district" (Woodlands of Northeastern Scania) (Fig. 1a) is characterized by igneous rock that is typically overlaid by moraine soils with a low clay content (Germundsson and Schlyter 1999). The overall topography in the Plains and Central districts is generally flat or gently undulating, but the Central district has several horsts that contribute to local topographic heterogeneity (Fig. 1b). The topography in the Woodland district is generally hilly, with local elevations reaching > 200 m above sea-level (Fig. 1b). The mean temperatures within the province range from -2 °C (in the Woodland district) to -0.5 °C (Plains district) in January and from 15.5 °C (Woodland district) to 16.5 °C (Plains district) in July. The mean annual precipitation varies between 550 mm (Plains district) and 900 mm (Woodland district) (Germundsson and Schlyter 1999). Whereas both the Plains and Central districts are characterized by arable cultivation (mainly cereal production), the agricultural intensity is lower in the Central district, which also contains a higher proportion of forest. The forest vegetation in the Plains and Central districts consists of planted stands of conifers and patches of deciduous forests. The main land-cover in the Woodland district is commercially managed mixed forest dominated by coniferous species (Germundsson and Schlyter 1999). The Plains, Central and Woodland districts represent a gradient of decreasing land-use intensity and increasing landscape complexity.

## 2.2 Plant species data

We used species data from a province-wide floristic inventory in Scania in which presence/absence data for all vascular plant species were collected, between 1987 and 2005, from 1560 2.5 km × 2.5 km squares (together covering ca. 77 % of

Scania) (Tyler et al. 2007). A subset of 1406 2.5 km × 2.5 km squares were regarded as being fully inventoried (Tyler et al. 2007). In the present study, species richness in the 1406 fully inventoried squares was represented by; (i) the total number of vascular plant species, (ii) the total number of red listed species (Artdatabanken), and (iii) the total number of species that are classed as grassland habitat specialists ("grassland specialists") by Ekstam and Forshed (1992).

## 2.3 Data used for the satellite data-based classification of land-use and land-cover classes

### *2.3.1 Satellite data used for the classification of land-use and land-cover classes*

We searched the United States Geological Survey (USGS) Landsat archive and the Swedish Saccess satellite archive for Landsat images covering the province of Scania during the 1970s (historical satellite images) and the period around the turn of the millennium (contemporary satellite images). We retrieved a set of historical satellite images comprising three Landsat 2 MSS scenes (acquired 27 August 1973, 27 February 1975, and 3 July 1975) consisting of four spectral bands with 60 m spatial resolution, and a set of contemporary satellite images comprising three Landsat 7 ETM+ scenes (acquired 1 July 2001, 12 September 2002, and 23 March 2003) consisting of eight spectral bands with 30 m spatial resolution (Table 1). The Landsat data were supplied by the USGS Landsat archive as a level 1T product, with systematic radiometric and geometric corrections applied with the help of ground control points, and topographic corrections carried out using a digital elevation model (DEM). We transformed digital numbers to top-of-atmosphere reflectance values (see Chander et al. 2009), and we also carried out a dark object subtraction to correct each Landsat scene for atmospheric disturbance. To match the spatial resolutions of the Landsat MSS and ETM+ images, the Landsat ETM+ images were resampled, using a cubic spline transformation, to a spatial resolution of 60 m. Out of the 1406 fully inventoried 2.5 km × 2.5 km squares in the floristic inventory in Scania (Tyler et al. 2007), we selected the 473 squares (Fig. 1a) that were free from cloud or haze in all six Landsat images for use in the present study (Table 1).

### *2.3.2 Ancillary data used to complement the Landsat data acquired during the growing season*

We used growing-season Landsat data to classify LULC classes in the historical (based on Landsat data acquired 3 July 1975) and contemporary (based on Landsat data acquired 1 July 2001) landscapes, respectively (Table 1). However, the fact that particular LULC classes such as, for example, cropland and grassland, may have similar spectral responses during the growing season (e.g. Vogelmann et al.

**Table 1**. Landsat satellite data used in the present study.

| **Historical satellite data**<br>Landsat 2 MSS | **Wavelength (micrometres)** | **Resolution (metres)1)** |
|---|---|---|
| | 0.5-0.6, green | 60 |
| | 0.6-0.7, red | 60 |
| | 0.7-0.8, near infrared 1 | 60 |
| | 0.8-1.1, near infrared 2 | 60 |
| **Acquisition date** | **Use in the present study** | |
| 27-Aug-73 | Calculation of the senescence-season NDVI | |
| 3-Jul-75 | Base for SVM-based classification<br>Calculation of the growth-season NDVI | |
| 27-Feb-75 | Calculation of the dormant-season NDVI | |
| **Contemporary satellite data**<br>Landsat 7 ETM+ | **Wavelength (micrometres)** | **Resolution (metres)** |
| | 0.45-0.52, blue | 30 |
| | 0.52-0.60, green | 30 |
| | 0.63-0.69, red | 30 |
| | 0.77-0.90, near infrared | 30 |
| | 1.55-1.75, short-wave infrared 1 | 30 |
| | 2.09-2.35, short-wave infrared 2 | 30 |
| | 10.40-12.50, thermal infrared2) | 60 |
| | 0.52-0.90, panchromatic2) | 15 |
| **Acquisition date** | **Use in the present study** | |
| 1-Jul-01 | Base for the SVM-based classification<br>Calculation of the growth-season NDVI | |
| 12-Sep-02 | Calculation of the senescence-season NDVI | |
| 23-Mar-03 | Calculation of the dormant-season NDVI | |

1) Original MSS pixel size was 79 × 57 metres, production systems now resample the data to 60 metres
2) Not used in the present study
NDVI = normalized difference vegetation index, SVM = support vector machine

1998) may affect the classification accuracies of the remote sensing-based classes. In the present study, we therefore examined whether the classification of LULC classes based on growing-season Landsat data could be improved by including additional data, on (i) seasonal changes in the NDVI (Rouse et al. 1973; Tucker 1979) and (ii) topography, in the classification procedure. The NDVI − a spectral vegetation index that is often used to describe vegetation productivity or the "greenness" of vegetation and changes in plant phenology (cf. Goetz et al. 2005; Pettorelli 2005; Parviainen et al. 2010) − is calculated as: NDVI = (NIR-VIS)/(NIR+VIS) where VIS and NIR represent reflectance measurements in the visible and near infrared spectral regions, respectively (e.g. Rouse et al. 1973; Tucker 1979).

We calculated the NDVI for the historical growing-season (3 July 1975), senescence-season (27 August 1973) and dormant-season (27 February 1975) images, using the red (0.6 - 0.7 μm) and the near infrared 2 (0.8 - 1.1 μm) wavebands of the Landsat MSS scenes (Table 1). The NDVI for the contemporary growth-season (1 July 2001), senescence-season (12 September 2002) and dormant-season (23 March 2003) images was calculated from the red (0.63 - 0.69 μm) and the near infrared (0.77 - 0.90 μm) wavebands of the Landsat ETM+ scenes (Table 1).

The differences in the NDVI between the growing-season, senescence-season and dormant-season images were calculated for the historical and contemporary sets of Landsat scenes, respectively. The senescence-season NDVI should be able to discriminate between post-harvest agricultural fields (with a low cover of vegetation) and grasslands (with continued active growth), and the fact that differences between the growing-season NDVI and the senescence-season NDVI are likely to be greater for agricultural fields than for grasslands, is expected to contribute to an improved spectral separation and classification of cropland and grasslands. The dormant-season NDVI characterizes the vegetation at a time when productivity is low and leaves are absent from deciduous trees. The fact that differences in the NDVI between growth-season images and dormant-season images are likely to be greater in areas with deciduous trees than those with evergreen conifers has the potential to contribute to the separation and classification of deciduous and coniferous species during LULC classification procedure.

Landscape relief affects patterns of land-use and vegetation-cover (Dorner et al. 2002, Wrbka et al. 2004), and the incorporation of topographic data may improve the accuracy of land-cover classifications based on satellite data. In the present study, data on the mean altitude (metres) for each pixel in each of the 473 2.5 km × 2.5 km squares were extracted from a DEM (supplied by the National Land Survey of Sweden) with a spatial resolution of 50 m and an accuracy of ± 2 meters.

### *2.3.3 Training and validation data for the Landsat based historical and contemporary classifications of land-use and land-cover classes*

We used LULC information from the Swedish Land Cover Data (Svenska Marktäckedata, SMD) (http://gis-services.metria.se/nvfeed/atom/annex2.xml) to select training pixels for classifications (without and with ancillary data) of historical land-cover based on the Landsat scene from July 1975. The SMD (provided by the National Land Survey of Sweden) is an extension of the CORINE Land Cover project (http://www.eea.europa.eu/publications/COR0-landcover), and records 51 LULC classes in the province of Scania at a spatial resolution of 25 m × 25 m for the reference year 2000 (Table S1). We pooled the 51 classes into eight classes: cropland, grazed grassland, deciduous forest, coniferous forest, water, wetland, barren areas and built up areas (Table S1).

A stratified random sampling procedure (based on the map of SMD classes), was used to select 100 sample (point) locations within each of the eight LULC classes:

the sampling procedure included the criterion that there should be a minimum distance of 300 m between locations. We then identified the positions of the selected points on black and white aerial photographs (0.5 m spatial resolution) from 1973 and 1975, and checked for dissimilarities between the SMD classes and the visually interpreted LULC categories in the historical aerial photographs. Where dissimilarities between the SMD and aerial photographs were detected, the position of the original training point was moved to the nearest polygon that represented the appropriate LULC class on the aerial images. A similar procedure was used to construct a training data set for classifications (without and with ancillary data) of the contemporary land-cover (based on the Landsat scene from 1 July 2001). Although the SMD data (reference year 2000) and the Landsat scene (2001) both represent the early 21st century, we checked for discrepancies between the SMD classes and the visually interpreted land-cover classes on colour-infrared (1.0 m spatial resolution, 2004) aerial photographs. Where discrepancies were detected, the position of the original training point was moved to the nearest polygon that represented the appropriate LULC class on the aerial image.

Validation data for the evaluation of the LULC classifications of the historical and contemporary landscapes, respectively, were acquired following the same procedures that were used to construct the training data sets. As with the training data sets, the validation sets for each LULC class were based on 100 randomly selected sample points for each LULC class in the historical and contemporary landscapes, respectively.

When the positioning of the training and validation sample locations was finalized, pixel values for each location were extracted for each of the four wavebands included in each of the three (historical) Landsat MSS images, and for each of the six wavebands included in each of the three (contemporary) Landsat ETM+ images (Table 1). Differences in the NDVI between the growing-season, senescence-season and dormant-season images, for the historical and contemporary sets of Landsat scenes, respectively, and information on the mean elevation (m) were also extracted for the pixel of each sample location. The data used for the satellite data-based classification of LULC classes were extracted using the ArcGIS 10.1 (ESRI), IDRISI Selva (Clarks Labs) and FRAGSTATS ver. 4 (McGarigal et al. 2012) software.

## 2.4 Classification procedure and assessment of classification accuracy

### *2.4.1 Models used to classify the land-cover in the study area*

A support vector machine (SVM) classification approach was used, with the help of the training data (without and with ancillary variables), to produce Landsat-based LULC classifications of the whole study area in the province of Scania (473 squares) (Fig. 1a) for the historical (mid-1970s) and contemporary (turn-of-millennium) landscapes, respectively. A SVM (Vapnik 1995) is a machine learning technique, based on statistical learning theory, which delivers a supervised classification. The SVM approach has been shown to be superior to Maximum Likelihood, Neural Network and Decision Tree classification approaches for the remote sensing data-based classification of land cover classes (e.g. Huang et al. 2002, Foody and Mathur 2004), while requiring comparatively small amounts of input data (Pal and Mather 2004). The LIBSVM library (Chang and Lin 2011) in the Python programming environment was used to produce the SVM classifications in the present study. We used a radial basis function-kernel, with C and $\gamma$ values produced by the cross-validation grid-search script "easy.py" included in the LIBSVM library.

The LULC classifications in the historical landscape were based on; (i) Landsat MSS data (3 July 1975; including all four wavebands) (Classification Model 1a), and (ii) Landsat MSS data (3 July 1975, including all four wavebands) and ancillary data (elevation data, and seasonal differences in the NDVI) (Classification Model 1b) (Table 1). The sets of input data for the classifications in the contemporary landscape were (i) Landsat ETM+ data (1 July 2001; including six wavebands) (Classification Model 2a), and (ii) Landsat ETM+ data (1 July 2001; including six wavebands) and ancillary data (elevation data, and seasonal differences in the NDVI) (Classification Model 2b) (Table 1).

### *2.4.2 Assessment of classification accuracies*

The respective abilities of the four models to classify the eight LULC classes in the whole study area were quantified with the help of confusion matrices (based on 100 validation pixels per class) for the classification results for each of the individual models (Classification Models 1a, 1b, 2a, 2b). Classification accuracies from the confusion matrices were used to calculate the producer's accuracy and the user's accuracy for each LUCL class for each classification model. The producer's accuracy refers to the probability that a pixel associated with a specific LULC class in the validation data set will be assigned to the corresponding LULC class in the SVM-based classification procedure. The user's accuracy represents the probability that a pixel classified (with the help of the SVM-based classification procedure) as belonging to a specific LUCL class is associated with the same class in the validation data. The overall accuracy and the Kappa statistic value (which assesses the inter-classifier agreement (Cohen 1960; Campbell 2002), were calculated from each confusion matrix.

## 2.5 Analyses

### *2.5.1 Proportions of land-use and land-cover classes, and measures of landscape heterogeneity and landscape change*

Each of the 473 squares within the whole Scanian study area (classified with the help of Classification Models 1a, 1b, 2a, 2b, respectively) was assigned to one of the three geographic districts in Scania; the Plains district (193 squares), the Central district (226 squares), or the Woodland district (54 squares) (Fig. 1a). The proportion (%) of each of the eight LULC classes was calculated for the squares within the whole study area and for squares within each district, for each of the classification models (Classification Models 1a, 1b, 2a, 2b). A set of five landscape heterogeneity indices were also estimated, for each of the squares for each classification model. The indices included descriptors of patch configuration (number of patches "NP"), patch complexity (total-patch-edge "TE"), and patch diversity (patch richness "PR", Shannon’s diversity "SHDI" and Shannon's evenness "SHEI") (e.g. Billeter et al. 2007; Plexida et al. 2014). The landscape heterogeneity indices were calculated using FRAGSTATS ver. 4 (McGarigal et al. 2012) software. Changes in the LULC class proportions and in the measures of landscape heterogeneity between the historical and contemporary landscapes were calculated for each square within the whole study area as well as for the squares within each district, using each of the four classification models.

### *2.5.2 Relationships between contemporary species richness, land-cover, landscape heterogeneity and the degree of landscape change*

Correlations between each of the species richness categories (total number of plant species, number of red listed species, and number of grassland specialists) and the proportions of LULC classes and measures of landscape heterogeneity in the historical and contemporary landscapes, respectively, were analysed separately for the squares within the whole study area and for the squares within each district. Species richness values for squares were also related to within-square changes in the proportions of the LULC classes and measures of landscape heterogeneity between the historical and contemporary landscapes. Because multiple tests may result in an increased risk of Type I error, the significance values of the correlations were assessed after Bonferroni correction.

Finally, we examined whether the levels of change (between the historical and contemporary landscapes) in the proportions of LULC classes and in the measures of landscape heterogeneity differed between” hot” squares (squares with high plant species richness) and “cold” squares (squares with low species richness) with the help of two-tailed Student’s t-tests. For each of the categories of species richness (total number of plant species, number of red-listed species, and number of grassland specialists), the 30 squares (out of the total of 473 squares) with the highest and lowest number of species were designated, respectively, as the hot and cold squares.

# 3. Results

Table 2 presents summary statistics for the within-square species richness for each of the species richness categories (total number of plant species, number of red listed species, and number of grassland specialists), for the whole study area in the province of Scania and for each of the three geographic districts within the study area (Fig. 1a). There were significant ($p \leq 0.05$ for all pairs) pairwise correlations between all three species richness categories, in the whole study area and within each of the three districts (Table S2).

**Table 2**. Summary statistics for species richness within each of three categories of vascular plants (total number of plant species, number of red-listed species, and number of grassland specialists) within 2.5 km × 2.5 km grid squares representing the whole province of Scania, and the subsets of squares representing the Plains, Central, and Woodland districts. SD = standard deviation.

| **Total number of plant species** | **Number of squares** | **Mean** | **SD** | **Range** | **Min** | **Max** |
|---|---|---|---|---|---|---|
| Scania | 473 | 389 | 89 | 688 | 58 | 746 |
| Plains | 193 | 357 | 80 | 458 | 75 | 533 |
| Central | 226 | 409 | 92 | 688 | 58 | 746 |
| Woodland | 54 | 422 | 70 | 283 | 277 | 560 |
| **Number of red-listed species** | | | | | | |
| Scania | 473 | 12 | 6 | 38 | 2 | 40 |
| Plains | 193 | 12 | 6 | 38 | 2 | 40 |
| Central | 226 | 13 | 6 | 34 | 3 | 37 |
| Woodland | 54 | 12 | 5 | 23 | 3 | 26 |
| **Number of grassland specialist species** | | | | | | |
| Scania | 473 | 72 | 19 | 133 | 12 | 145 |
| Plains | 193 | 63 | 15 | 92 | 15 | 107 |
| Central | 226 | 80 | 20 | 133 | 12 | 145 |
| Woodland | 54 | 84 | 15 | 57 | 54 | 111 |

## 3.1 Classification accuracies for the models based on Landsat data, with and without ancillary data

For the classification of the LULC classes in the whole study area (473 squares), the classification model based on Landsat data in combination with ancillary data performed better than the model based only on Landsat data – for both the historical (mid-1970s) and the contemporary (turn-of-millennium) landscapes. In the historical landscape, the producer’s and user’s classification accuracies for the individual LULC classes in the study area varied between 53% and 99% for the classification model based on only growing-season Landsat data (Classification

Model 1a) (Table S3), and between 69% and 98% for the model based on growing-season Landsat data together with ancillary data (Classification Model 1b) (Table 3). The overall classification accuracy and the Kappa statistic values were 77.8% and 0.746, respectively, for Model 1a and 86.4% and 0.844, respectively, for Model 1b (Tables 3, S3). In particular, the inclusion of ancillary data (Model 1b approach) gave better classification accuracies for cropland, grazed grassland, and wetlands than the Model 1a approach (Table 3, S3). For these three LULC classes, the inclusion of ancillary data led to, respective, increases of 10.0%, 17.0%, and 16.0% in the user's accuracy, and of 25.9%, 16.0%, and 10.4% in the producer's accuracy (Table 3, S3).

**Table 3**. Error matrix for the SVM classification for the historical landscape, based on 4 bands from a Landsat 2 MSS scene from 3 July 1975, with supporting data on seasonal changes in the NDVI, and data on topography. The diagonal indicating the correct class-assignments is shaded.

| | **Historical landscape** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Land-cover classes identified from aerial photos** | **Land-cover classes identified from the SVM-classified 4-band Landsat 2 MSS scene, with support from data on seasonal changes in the NDVI, and data on topography** | | | | | | | | | |
| | Cropland | Gr. grassl. | Deciduous | Conifer | Water | Wetlands | Barren | Built-up | Total | User's acc. |
| Cropland | 79 | 10 | 0 | 0 | 0 | 4 | 2 | 5 | 100 | 79.0% |
| Grazed grassland | 7 | 72 | 2 | 3 | 0 | 6 | 3 | 7 | 100 | 72.0% |
| Deciduous | 0 | 0 | 92 | 2 | 0 | 6 | 0 | 0 | 100 | 92.0% |
| Conifers | 0 | 0 | 0 | 98 | 0 | 2 | 0 | 0 | 100 | 98.0% |
| Water | 0 | 0 | 0 | 0 | 98 | 2 | 0 | 0 | 100 | 98.0% |
| Wetlands | 3 | 8 | 11 | 8 | 1 | 69 | 0 | 0 | 100 | 69.0% |
| Barren | 0 | 0 | 0 | 1 | 1 | 0 | 88 | 10 | 100 | 88.0% |
| Built-up | 3 | 0 | 0 | 0 | 0 | 0 | 2 | 95 | 100 | 95.0% |
| Total | 92 | 90 | 105 | 112 | 100 | 89 | 95 | 117 | 800 | |
| Producer's acc. | 85.9% | 80.0% | 87.6% | 87.5% | 98.0% | 77.5% | 92.6% | 81.2% | | |
| | | | | | | | | | Overall accuracy = 86.4% | |
| | | | | | | | | | Kappa = 0.844 | |

For the classification of LULC classes in the whole study area in the contemporary landscape, the producer's and user's classification accuracies for the individual classes varied between 75% and 100% for the classification model based on only growing-season Landsat data (Model 2a) (Table S4), and between 84% and 99% for the model based on Landsat data together with ancillary data (Model 2b) (Table 4). The overall classification accuracy and the Kappa statistic values were 89.3% and 0.877, respectively for Model 2a, and 91.6% and 0.904, respectively for Model 2b (Tables 4, S4). The inclusion of ancillary data (Model 2b) increased the user's accuracy for wetlands by 14 % compared with the Model 2a approach, while leaving the user's accuracy for the other classes relatively unchanged. The producer’s accuracy increased by 10%, 6.0 %, 5.5% and 5.1%, respectively, for cropland, grazed grassland, coniferous forest, and wetlands, whereas the producer’s accuracy decreased (relative to the 2a approach) by 7.5% for water in the Model 2b approach (Table 4, S4).

**Table 4**. Error matrix for the SVM classification for the contemporary landscape, based on 6 bands from a Landsat 7 ETM+ scene from 1 July 2001, with supporting data on seasonal changes in the NDVI, and data on topography. The diagonal indicating the correct class-assignments is shaded.

| | **Contemporary landscape** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Land-cover classes identified from aerial photos** | **Land-cover classes identified from the SVM-classified 6-band Landsat 7 ETM+ scene, with support from data on seasonal changes in the NDVI, and data on topography** | | | | | | | | | |
| | Cropland | Gr. grassl. | Deciduous | Conifer | Water | Wetlands | Barren | Built-up | Total | User's acc. |
| Cropland | 89 | 4 | 0 | 0 | 2 | 3 | 1 | 1 | 100 | 89.0% |
| Grazed grassland | 2 | 86 | 1 | 0 | 1 | 7 | 1 | 2 | 100 | 86.0% |
| Deciduous | 1 | 1 | 97 | 0 | 0 | 1 | 0 | 0 | 100 | 97.0% |
| Conifers | 0 | 1 | 6 | 88 | 0 | 5 | 0 | 0 | 100 | 88.0% |
| Water | 0 | 0 | 0 | 0 | 99 | 1 | 0 | 0 | 100 | 99.0% |
| Wetlands | 0 | 6 | 2 | 0 | 3 | 89 | 0 | 0 | 100 | 89.0% |
| Barren | 1 | 0 | 0 | 0 | 0 | 0 | 96 | 3 | 100 | 96.0% |
| Built-up | 0 | 2 | 0 | 1 | 2 | 0 | 6 | 89 | 100 | 89.0% |
| Total | 93 | 100 | 106 | 89 | 107 | 106 | 104 | 95 | 800 | |
| Producer's acc. | 95.7% | 86.0% | 91.5% | 98.9% | 92.5% | 84.0% | 92.3% | 93.7% | | |
| | | | | | | | | | Overall accuracy = 91.6%<br>Kappa = 0.904 | |

Because, the models in which the growing-season Landsat data were complemented by ancillary data gave better overall classification accuracies and kappa statistics than the models without ancillary data, for both the historical landscape (Model 1b) and the contemporary landscape (Model 2b), only the results from Models 1b and 2b were used as the basis for further analysis.

## 3.2 Changes in land-cover between the landscape in the mid-1970s and at the turn of the millennium

Each of the 473 squares within the whole study area (classified using the Model 1b and 2b approaches, respectively), was assigned to one of the three geographic districts in Scania; the Plains district, the Central district, or the Woodland district. The proportion of crop land within the whole study area, and within each of the three districts, decreased between the mid-1970s and the turn of the millennium (Table 5). In contrast, the proportions of grazed grassland and deciduous forest increased: the increase in deciduous forest was particularly pronounced in the Woodland district (Table 5). In Scania as a whole, the increase in the area of grazed grassland mainly reflected the conversion of cropland into grazed grassland (Table S5). At the same time, some areas of grazed grassland were transformed into cropland or deciduous forest (Table S5). In the Woodland district, the persistence of cropland was 61.6 % (61.6% of the pixels that were classified as cropland in the historical landscape were also classified as cropland in the contemporary landscape), while the persistence of cropland was 85.8% for the Plains district (Table 5). The persistence of grazed grassland, deciduous forest and water, in the whole study area, and in the individual districts, ranged between 52.6% and 96.4%. The lowest persistence was shown for barren in the Woodland district (7.2%) (Table 5).

**Table 5**. The proportions (%) of the land-cover classes in the historical and contemporary landscapes, respectively, and the levels of change and persistence of the land-cover classes between the historical and contemporary landscapes. Results are based on the SVM classifications using the Landsat 2 MSS scene from 3 July 1975, and the Landsat 7 ETM+ scene from 1 July 2001, with supporting data on seasonal changes in the NDVI, and data on topography.

| **Scania** | **Cropland** | **Gr. grassl.** | **Deciduous** | **Conifers** | **Water** | **Wetlands** | **Barren** | **Built-up** |
|---|---|---|---|---|---|---|---|---|
| Historical landscape (%) | 55.3 | 15.9 | 6.7 | 5.6 | 1.3 | 5.2 | 0.9 | 9.1 |
| Contemporary landscape (%) | 51.3 | 17.9 | 10.0 | 4.6 | 2.2 | 5.1 | 1.1 | 7.8 |
| Change (%) | -4.0 | 2.0 | 3.3 | -1.0 | 0.9 | -0.1 | 0.2 | -1.3 |
| Persistence (%) | 80.5 | 55.0 | 80.7 | 47.3 | 90.7 | 19.0 | 16.1 | 53.3 |
| **Plains** | | | | | | | | |
| Historical landscape (%) | 67.9 | 10.6 | 2.0 | 1.0 | 0.9 | 2.5 | 1.5 | 13.7 |
| Contemporary landscape (%) | 64.6 | 12.0 | 3.1 | 0.8 | 2.2 | 2.1 | 1.1 | 14.1 |
| Change (%) | -3.3 | 1.4 | 1.1 | -0.2 | 1.3 | -0.4 | -0.4 | 0.6 |
| Persistence (%) | 85.8 | 52.6 | 81.3 | 28.0 | 75.5 | 13.8 | 11.6 | 65.9 |
| **Central** | | | | | | | | |
| Historical landscape (%) | 50.2 | 18.5 | 7.4 | 7.2 | 2.0 | 6.6 | 0.6 | 7.5 |
| Contemporary landscape (%) | 46.4 | 21.1 | 10.7 | 6.8 | 2.6 | 6.6 | 1.1 | 4.6 |
| Change (%) | -3.8 | 2.6 | 3.3 | -0.4 | 0.6 | 0.0 | 0.5 | -2.9 |
| Persistence (%) | 78.1 | 56.7 | 77.6 | 54.6 | 96.4 | 21.1 | 24.8 | 35.8 |
| **Woodland** | | | | | | | | |
| Historical landscape (%) | 36.0 | 22.7 | 18.5 | 13.6 | 0.1 | 8.2 | 0.1 | 0.9 |
| Contemporary landscape (%) | 28.5 | 24.3 | 29.0 | 7.8 | 0.2 | 9.1 | 0.6 | 0.6 |
| Change (%) | -7.5 | 1.6 | 10.5 | -5.8 | 0.1 | 0.9 | 0.5 | -0.3 |
| Persistence (%) | 61.6 | 53.2 | 85.3 | 36.7 | 93.1 | 17.4 | 7.2 | 25.0 |

There were significant pairwise correlations between all five landscape heterogeneity indices (NP, TE, PR, SHDI, and SHEI) ($p \leq 0.05$ for all pairs), in the whole study area and in each of the three districts for both the historical and contemporary landscapes (Table S6). In general, the individual measures of landscape heterogeneity were higher for the contemporary landscape than for the historical landscape (Table S7).

## 3.3 Proportions of land-use classes, landscape heterogeneity and species richness

Squares characterized by low levels of contemporary (1987-2005) species richness (for all three categories of species) tended to be associated with higher proportions of cropland than squares with high species richness – in both the historical and contemporary landscapes (Fig. 2, Table S8). Positive relationships ($p \leq 0.05$ between all three categories of species richness and the proportion of LULC classes such as grazed grassland, deciduous and coniferous forests, wetlands, and built-up areas in the squares were found within the whole study area and within each district -- in both the historical and contemporary landscapes (Fig. 2, Table S8).

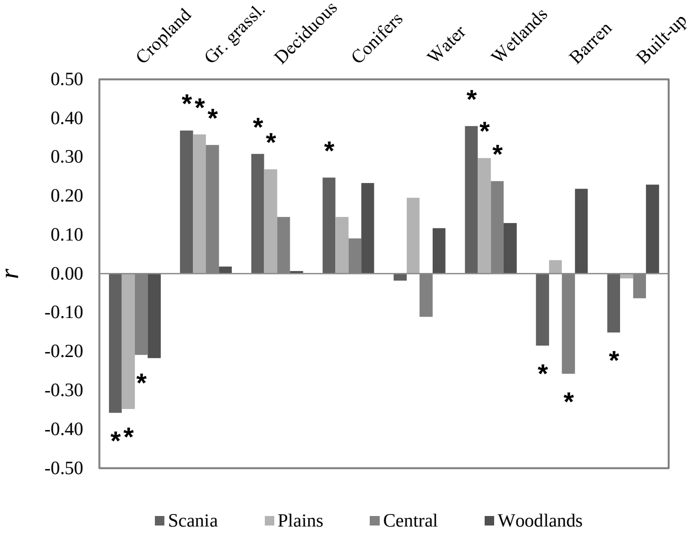


**Figure 2**. Pearson's correlation coefficients between proportions of land-cover classes in the historical landscape, and the contemporary species richness of grassland specialists within 2.5 km × 2.5 km grid squares representing the whole province of Scania (473 squares) and the subsets of squares representing the Plains (193 squares), Central (226 squares), and Woodland (54 squares) districts. Results are based on the SVM classification using 4 bands from Landsat 2 MSS scene from 3 July 1975, with supporting data on seasonal changes in the NDVI, and data on topography. Correlations marked * are statistically significant ($p \leq 0.05$).

Within-square species richness (for all three categories of species) was generally positively associated with the five measures of landscape heterogeneity in both the historical and contemporary landscapes (Fig. 3, Table S9).

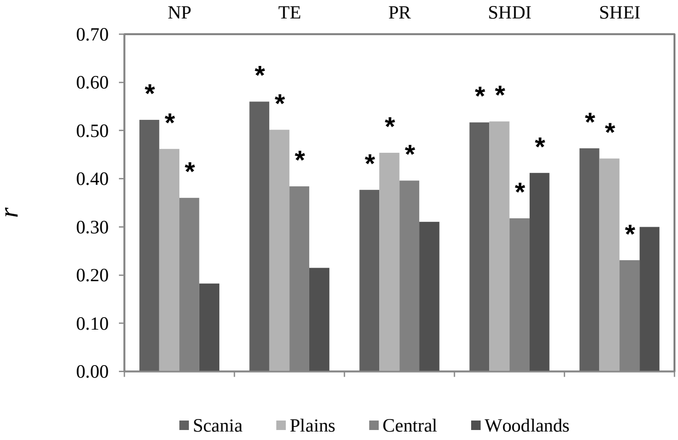


**Figure 3**. Pearson's correlation coefficients between measures of the heterogeneity of land-cover classes in the historical landscape, and the contemporary species richness of grassland specialists within 2.5 km × 2.5 km grid squares representing the whole province of Scania (473 squares) and the subsets of squares representing the Plains (193 squares), Central (226 squares), and Woodland (54 squares) districts. Results are based on the SVM classification using 4 bands from a Landsat 2 MSS scene from 3 July 1975, with supporting data on seasonal changes in the NDVI, and data on topography. NP = number of patches, TE = total edge (m), PR = patch richness, SHDI = Shannon's diversity index, SHEI = Shannon's evenness index. Correlations marked * are statistically significant ($p \leq 0.05$).

## 3.4 Landscape change and species richness

The within-square total species richness (for the whole study area and each of the three districts) was positively ($p \leq 0.05$) related to the level of increase in the proportion of deciduous forest between the mid-1970s and the turn of the millennium (Table S10). The levels of change in the proportion of deciduous forest were also generally positively correlated ($p \leq 0.05$) with the within-square number of red-listed species and the number of grassland specialists (Fig. 4, Table S10). There were significant ($p \leq 0.05$) negative correlations between levels of change in the proportion of coniferous forest, and both the number of red-listed species (for the squares within the Central and Woodland districts, respectively) and the number of grassland specialists (for the squares within the whole study area and the Central district, respectively) (Fig. 4, Table S10). Within-square species richness was not related to the degree of change (between the historical and contemporary landscapes) in the proportions of grazed grassland, water, wetlands or barren.

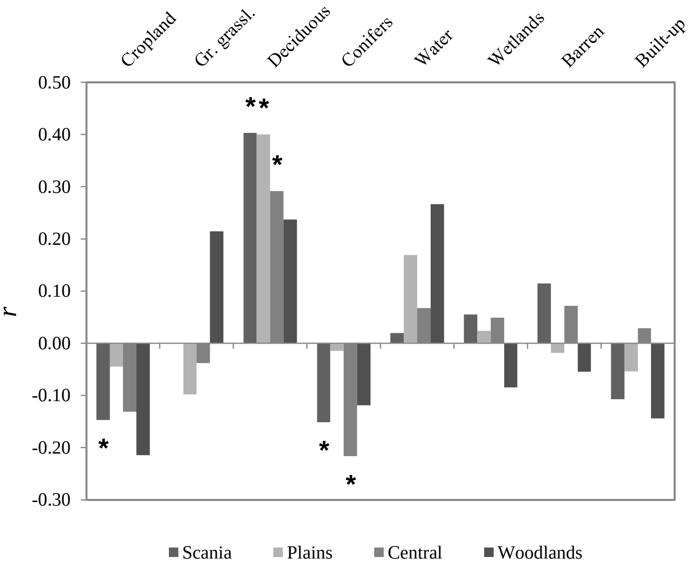


**Figure 4**. Pearson's correlation coefficients between change in the proportions of land-cover classes between the historical and contemporary landscapes, and the level of species richness of grassland specialists within 2.5 km × 2.5 km grid squares representing the whole province of (473 squares) and the subsets of squares representing the Plains (193 squares), Central (226 squares), and Woodland (54 squares) districts in the contemporary landscape. Results are based on the SVM classifications using the Landsat 2 MSS scene from 3 July 1975, and the Landsat 7 ETM+ scene from 1 July 2001, with supporting data on seasonal changes in the NDVI, and data on topography. Correlations marked * are statistically significant ($p \leq 0.05$).

Only changes in the landscape heterogeneity characterized with the help of the number of patches (NP) were correlated with the within-square species richness.

The within-square total number of plant species and number of grassland specialists were positively related to the changes in the NP ($0.13 \leq r \leq 0.20$, $p \leq 0.05$), within the whole study area and within the Plains district.

## 3.5 Landscape change, and species richness in hot and cold squares

The increase in the proportion of deciduous forest between the historical and contemporary landscapes was significantly ($p \leq 0.05$) higher in the hot squares than in the cold squares for both grassland specialists (Fig. 5) and the total number of species. The hot squares for grassland specialists were also significantly ($p \leq 0.05$) associated with a greater reduction in the proportion of cropland than the cold squares (Fig. 5). In addition, the hot squares for grassland specialists were associated with a reduction in the proportion of coniferous forest, whereas the cold squares were associated with an increase in the proportion of coniferous forest (Fig. 5). The levels of change in the proportions of LULC classes did not differ between the hot and cold squares for the red-listed plant species.

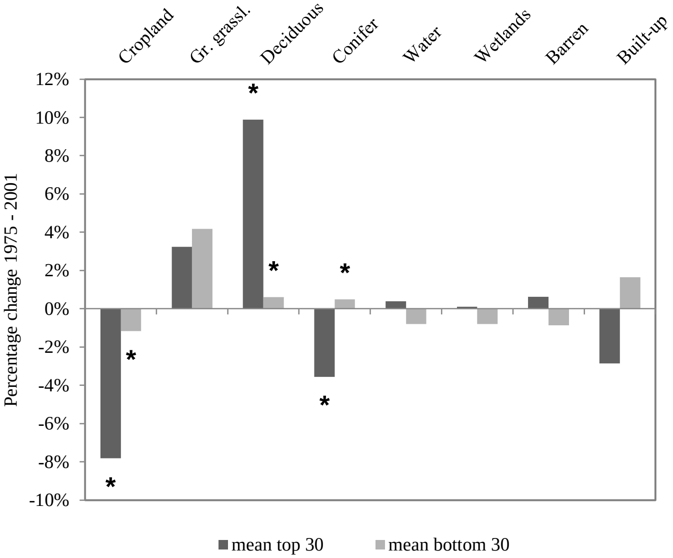


**Figure 5**. The mean change in proportions of land-cover classes between the historical and contemporary landscapes, in the 30 least and the 30 most species-rich 2.5 km × 2.5 km grid squares within Scania. Results are based on the SVM classifications using the Landsat 2 MSS scene from 3 July 1975, and the Landsat 7 ETM+ scene from 1 July 2001, with supporting data on seasonal changes in the NDVI, and data on topography. Land-cover classes marked * show significantly different mean levels of change between the 30 most ("mean top 30") and 30 least ("mean bottom 30") species-rich grid squares within that class (Student's 2-tailed t-test, $p \leq 0.05$).

The level of change in landscape heterogeneity (for NP) was higher for the hot squares than for the cold squares for each of the three categories of species richness. For the total number of plant species, the increase in the mean number of patches was 11.23 for the hot squares and 0.40 for the cold squares ($p \leq 0.05$). For the red-listed plant species, the increase in the mean NP was 12.33 for the hot squares and 2.75 for the cold squares ($p \leq 0.05$). For the number of grassland specialists, the increase in the mean NP was 8.77 for the hot squares and 1.20 for the cold squares ($p \leq 0.05$). For the total number of plant species, there was a significant difference ($p \leq 0.05$) in the level of change in the mean total-patch-edge (TE) between the hot and the cold squares. The mean TE increased by 2.73 km for the hot squares, while the mean TE decreased by 0.81 km for the cold squares between the mid-1970s and the turn of the millennium.

# 4. Discussion

## 4.1 Classification accuracies

In the present study of the southern Swedish province of Scania, the classification models that included both Landsat data acquired during the growing-season and ancillary data on topography and seasonal changes in the vegetation gave overall land-cover classification accuracies that were greater than 85% (cf. Foody 2002) (Tables 3,4). The study shows that a classification approach based on growing-season Landsat data, together with ancillary environmental data and satellite-based data on seasonal changes in reflectance, improves the separation between spectrally similar land-cover classes such as cropland and grazed grassland – particularly when using historical Landsat data. Our results are consistent with those of other satellite-based studies which show that a multi-temporal approach may make a valuable contribution to the classification of spectrally similar land-cover classes (e.g. Müller et al. 2015; Schuster et al. 2015; Zoungrana et al. 2015). Müller et al. (2015), for example, used Landsat-derived spectral/temporal variability metrics to distinguish between cropland, pasture and natural vegetation within a savanna landscape in the Brazilian Cerrado and, as in our study, showed that the inclusion of data on seasonal changes in the vegetation improved the classification accuracies of spectrally similar land-use classes. Schuster et al. (2015) carried out a land-cover classification of a German landscape that was dominated by woodlands but also contained a mosaic of heathlands, semi-natural grasslands and wetlands. The authors were able to produce a fine-scale classification of the vegetation types, with an overall accuracy of >90%, with the help of a multi-temporal set of satellite scenes (acquired over three years) that captured phenological variation in the characteristics of vegetation both within and between years.

Remote sensing data acquired with the help of sensors with high radiometric resolution (the capacity of the sensor to distinguish differences in light intensity or reflectance) are better at detecting small differences in reflected or emitted energy than sensors with low radiometric resolution. The accuracy of a remote sensing-based classification of land-cover classes is expected to increase as the spectral resolution of the remote sensing data increases (e.g. Xu and Gong 2007): the use of many wavebands delivers more detailed spectral information on land-cover classes than the use of few wavebands. In addition, remote sensing data collected with the help of sensors with a high spatial resolution (small pixel size) can be used to detect smaller features than sensors with a low spatial resolution (coarse pixel size). In the present study, the inclusion of ancillary data in the classification models improved the classification accuracies for the historical landscape more than for the contemporary landscape. The somewhat poorer classification accuracies for the historical compared with the contemporary landscape (for the classifications both with and without ancillary data) are likely to be explained by the fact that the technological capabilities of the Landsat missions have increased and the instruments onboard the satellites have been improved since the launch of the first Landsat satellite in 1972 (cf. Sloane 2012). For example, the radiometric and spectral resolution of the Landsat MSS data (6 bits, four wavebands) used to classify the historical landscape in the present study is lower than that of the Landsat ETM+ data (8 bits, six wavelength bands) used to classify the contemporary landscape. In addition, the Landsat MSS data used to classify the historical landscape in our study are provided with an original pixel size of 79 × 57 m, whereas the Landsat ETM+ data used to classify the contemporary landscape are provided with a pixel size of 30 × 30 m.

## 4.2 Land-cover change

As well as the intensification of agriculture in many regions, there has also been a progressive increase in the area of abandoned cultivated land in several areas of Europe over the past decades (Cramer et al. 2008; Pullin et al. 2009). The decrease in the proportion of cropland and the increase in the proportion of grazed grassland observed in our study (Table 5) are supported by official statistics from Statistics Sweden. Official records report a reduction in the area of cropland by 5.6% between 1980 and 2000, and an increase in the area of grazed grassland of 5.4% (between 1980 and 2000) for the whole province of Scania (Statistics Sweden, http://www.statistikdatabasen.scb.se). An increase in the area of grassland at the expense of cropland (Table S5) is also observed in other parts of Europe where, in many regions, arable fields are gradually being transformed into grazed grasslands (MacDonald et al. 2000; Török et al. 2012). The official statistics revealed a decrease of 6.0% in the total forest cover in Scania (between 1980 and 2000) (http://www.statistikdatabasen.scb.se), while our study showed that the proportion of coniferous forests decreased between 1975 and 2001 in Scania. Our study also indicates that the proportion of deciduous forest in Scania has increased between 1975 and 2001 – in agreement with the results from an inventory of beech (*Fagus sylvatica*) forest in Scania, which showed an 44.5% increase in the area of beech forest between 1976 and 1986 (Brunet and Berlin 2004). Results from the Swedish National Forest Inventory show an increase in the area of hardwood forest (from 21,000 ha to 73,000 ha) in Scania between 1962 and 1999 (Brunet and Berlin 2004). Data from Statistics Sweden (http://www.statistikdatabasen.scb.se) show that the total built-up area in Scania increased by 8.7% between 1980 and 2000. In the present study, however, the total built-up area was lower in the contemporary landscape than in the historical landscape, indicating that the approach we used to investigate changes in built-up areas may not have been optimal. Previous studies have shown that the investigation of land-cover and land-cover change using historical satellite data with low spatial resolution may be problematic in built-up areas that contain a diversity of small features (e.g. rooftops, asphalt, parks, playgrounds and ponds) (Getman et al. 2008). In our study, we re-sampled the pixel size of the ETM+ data to match the 60 m spatial resolution of the MSS data. Our results suggest that the relatively low radiometric and spectral resolution of the historical Landsat MSS data, compared with that of the contemporary Landsat ETM+ data, may have limited the accuracy of the comparison between the proportions of built-up areas in the two landscapes. In addition, the contemporary ETM+ data (with a higher radiometric and spectral resolution than the MSS data) may have a better ability than the MSS data to characterize the contribution of relict patches of semi-natural habitats ("small biotopes") to the within-square heterogeneity in land-cover. The fact that the radiometric and spectral resolution of the Landsat MSS data may have been too coarse to fully characterize the contribution of small biotopes to the heterogeneity in the historical landscape may have contributed to the increase in landscape heterogeneity between the historical and the contemporary landscapes shown in our study.

## 4.3 Species richness

Our results agree with those of many ecological studies which show that the intensification of agriculture, and the associated conversion of semi-natural habitats into arable land, has had a negative effect on levels of species diversity in the present-day landscape in Europe (e.g. Benton et al. 2003; Tscharntke et al. 2005; Billeter et al. 2008). For example, in Sweden, approximately a third of all red-listed species and 60 % of the red-listed vascular plant species are found within the agricultural landscape (Artdatabanken).

The majority of the grasslands in the present-day agricultural landscape represent young grasslands on grazed fields which represent the earliest stages in the succession from arable cultivation towards species-rich old grasslands which have had a long continuity of grazing management (cf. Cousins et al. 2009). The transformation of arable land into grazed grassland (Table S5) is expected to offer new possibilities for the establishment of species-rich and diverse grassland vegetation (Ewers and Didham 2005; Plieninger and Gaertner 2011; Römermann et al. 2005). Our study showed that the squares with a high proportion of grazed grasslands within both the historical and contemporary landscapes tended to be associated with a higher level of within-square species richness (particularly for grassland specialists) than those squares that are characterized by a low proportion of grazed grassland (Fig. 2, Table S8). Earlier studies have shown that previously arable fields that have been grazed for more than 30 years can harbour relatively high number of species of value for conservation (Lindborg 2006). Non-crop habitats (such as grazed ex-arable fields, field margins, wetlands and deciduous forests) may provide refugia and resources for a wide range of organisms, and enhance the dispersal of species through the landscape (e.g. Baudry et al. 2000; Marshall and Moonen 2002; Delattre et al. 2013). In intensively managed landscapes that have undergone a progressive loss of semi-natural habitats, wetlands and deciduous forests are recognized as being among the remaining habitats that are most important for maintaining biodiversity (cf. Brinson and Malvárez 2002; Gilliam 2007). Our study suggests that an increase in the within-square proportion of deciduous forest between the mid-1970s and the turn of the millennium has a positive impact on the levels of present-day species richness (Figs. 4, 5, Table S10). High within-square proportions of wetlands in both the historical and contemporary landscapes were positively associated with the present-day levels of species richness.

A substantial proportion of the coniferous forest within our study area – particularly within the Woodland district – is intensively managed. Managed coniferous forests are typically characterized by a low level of species richness which reflects the negative effects of drainage and fertilization as well as habitat damage by forest machinery (cf. Hartley 2002). In the present study, an increase in the proportions of coniferous forest in the Central and Woodland districts between the mid-1970s and the turn of the millennium appears to have had a negative impact on the within-square richness of red-listed species and, in the whole of Scania and the Central district, a negative impact on the number grassland species (Figs. 4, 5 Table S10). The hot squares (the 30 out of 473 squares with the highest plant species richness) for grassland specialists were characterized by a reduction in the proportion of coniferous forest, whereas the cold squares (the 30 out of 473 squares with the lowest plant species richness) were associated with an increase in the proportion of coniferous forest. However, contrary to our expectations, the proportion of coniferous forests in both the historical and contemporary landscapes occasionally showed a positive association with the levels of species richness in the contemporary landscape in the whole of Scania and the Plains and Woodland districts (Table S8). In general, the proportions of the different land-cover classes and the changes in those proportions between 1975 and 2001 affected the different species-richness categories (total number of plant species, number of red listed species and number of grassland specialists) in a comparable way. However, the number of red-listed species showed a stronger association with the proportion of built-up area than was observed for the other species categories. In urban areas, habitats such as gardens, parks, and ponds may have a positive impact on species richness (Kowarik 1995).

The positive relationships between the levels of species richness and the levels of landscape heterogeneity shown in the present study (Fig. 3, Table S9) are in accordance with the expectation that a diversity of habitats or niches will promote levels of species co-existence at a range of spatial scales (Whittaker 1972; Silvertown 2004; Stein et al. 2014). For example, the hot squares showed a greater increase in patch-number and the total length of the patch edges between the historical and contemporary landscapes than the cold squares. However, the red-listed species showed a somewhat weaker relationship with the landscape heterogeneity for both the historical and contemporary landscapes than the other species categories, possibly indicating the red-listed species may have already undergone a decline before the mid-1970s. The results of previous studies suggest that red-listed species may be affected by land-cover change earlier than other species-categories (Cousins et al. 2015; Irminger-Street et al. 2015).

While our study reveals significant relationships between Landsat-based measures of landscape variables and field data on species richness, much of the observed variation in species richness remained unexplained by the satellite-based approach used in the present study. The levels of within-square plant species richness may be influenced by a range of environmental and landscape variables that were not investigated in our study. For example, the proportions of land-cover types and the landscape heterogeneity in the surrounding squares (cf. Roschewitz et al. 2005, Tscharntke et al. 2005), and the connectivity between habitats (Taylor et al. 1993; Donald and Evans 2006) are also expected to have an impact on levels of species richness.

# 5. Conclusions

An improved understanding of the way in which the distributions of different types of land-cover, and changes in the distribution of land-cover classes, are related to present-day species richness requires detailed information on both the historical and contemporary landscape. Our study shows that Landsat satellite data have the potential to deliver the necessary information. The inclusion of ancillary data, on topography and on seasonal variation in the spectral properties of the vegetation enhanced the reliability of the satellite-based land-cover classifications. The ancillary data made a particularly valuable contribution to the classification of historical land-cover classes, which was based on satellite data that have a relatively low radiometric and spectral resolution. The results from the present satellite-based study are consistent with the results from ecological studies which show that increasing proportions of cropland and coniferous forests are associated with declining levels of landscape species richness. Our study supports the conclusions, based on ecological studies, that there are positive relationships between levels of species richness and the availability of ecologically valuable habitats (such as grazed grasslands, deciduous forests and wetlands), and between species richness and levels of landscape heterogeneity.

# Acknowledgments

We would like to thank the Lund Botanical Society for granting us access to the database for the project Skånes Flora, and Oskar Löfgren and Barbara C. Schmid for valuable discussions. The study was financed by grants from The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) to Karin Hall and Honor C. Prentice.

# References

ArtDatabanken, Swedish Species Information Centre. http://www.artfakta.se/GetSpecies.aspx?SearchType=Advanced. Downloaded 2016-05-10.

Baudry, J., Bunce, R. G. H., Burel, F. (2000) Hedgerows: an international perspective on their origin, function and management. Journal of Environmental Management 60: 7-22.

Benton, T. G., Vickery, J. A., Wilson, J. D. (2003) Farmland biodiversity: is habitat heterogeneity the key? Trends in Ecology & Evolution 18: 182-188.

Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek, R., Burel, F., Cerny, M., De Blust, G., De Cock, R., Diekotter, T., Dietz, H., Dirksen, J., Dormann, C., Durka, W., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., Le Coeur, D., Maelfait, J. P., Opdam, P., Roubalova, M., Schermann, A., Schermann, N., Schmidt, T., Schweiger, O., Smulders, M. J. M., Speelmans, M., Simova, P., Verboom, J., van Wingerden, W. K. R. E., Zobel, M., Edwards, P. J. (2008) Indicators for biodiversity in agricultural landscapes: a pan-European study. Journal of Applied Ecology 45: 141-150.

Brinson, M. M., Malvarez, A. I. (2002) Temperate freshwater wetlands: types, status, and threats. Environmental Conservation 29: 115-133.

Brunet, J., Berlin, G. (2004) Skånes skogar – historia, mångfald och skydd. Kunskapsunderlag för bevarande av värdefulla skogsområden och andra trädbärande marker. Länsstyrelsen i Skåne län.

Campbell, J.B. (2002) Introduction to Remote Sensing. Taylor & Francis, London, UK.

Chander, G. Markham, B.L., Helder, D.L. (2009) Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sensing of Environment 113: 893–903.

Chang, C. C., Lin, C. J. (2011) LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2:1-27.

Chapin, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., Hooper, D. U., Lavorel, S., Sala, O., Hobbie, S. E., Mack, M. C., Diaz, S. (2000) Consequences of changing biodiversity. Nature 405: 234-242.

Cohen, J. (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurements 20: 37–46.

Comber, A., Balzter, H., Cole, B., Fisher, P., Johnson, S. C. M., Ogutu, B. (2016) Methods to Quantify Regional Differences in Land Cover Change. Remote Sensing 8:176.

Cousins, S. A. O., Lindborg, R., Mattsson, S. (2009) Land use history and site location are more important for grassland species richness than local soil properties. Nordic Journal of Botany 27: 483-489.

Cousins, S. A. O., Auffret, A. G., Lindgren, J., Trank, L. (2015) Regional-scale land-cover change during the 20th century and its consequences for biodiversity. Ambio 44: 17-27.

Cramer, V. A., Hobbs, R. J., Standish, R. J. (2008) What's new about old fields? Land abandonment and ecosystem assembly. Trends in Ecology & Evolution 23: 104-112.

Delattre, T., Vernon, P., Burel, F. (2013) An agri-environmental scheme enhances butterfly dispersal in European agricultural landscapes. Agriculture Ecosystems & Environment, 166: 102-109.

Donald, P. F., Evans, A. D. (2006) Habitat connectivity and matrix restoration: the wider implications of agri-environment schemes. Journal of Applied Ecology 43: 209-218.

Dorner, B., Lertzman, K., Fall, J. (2002) Landscape pattern in topographically complex landscapes: issues and techniques for analysis. Landscape Ecology 17: 729-743.

CORINE Land Cover project. http://www.eea.europa.eu/publications/COR0-landcover. Downloaded 2016-05-10.

Ekstam, U., Forshed, N. (1992) Om hävden upphör: kärlväxter som indikatorarter i ängs- och hagmarker. (If Grassland Management Ceases: Vascular plants as Indicator Species in Meadows and Pastures). Naturvårdsverket, Stockholm, SE (in Swedish with English summary).

Ewers, R. M., Didham, R. K. (2005) Confounding factors in the detection of species response to habitat fragmentation. Biological Reviews 81:117–142.

Foody, G. M. (2002) Status of land cover classification accuracy assessment. Remote Sensing of Environment 80: 185-201.

Foody, G. M., Mathur, A. (2004) Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. Remote Sensing of Environment 93: 107-117.

Gerard, F., Petit, S., Smith, G., Thomson, A., Brown, N., Manchester, S., Wadsworth, R., Bugar, G., Halada, L., Bezak, P., Boltiziar, M., De Badts, E., Halabuk, A., Mojses, M., Petrovic, F., Gregor, M., Hazeu, G., Mucher, C. A., Wachowicz, M., Huitu, H., Tuominen, S., Kohler, R., Olschofsky, K., Ziese, H., Kolar, J., Sustera, J., Luque, S., Pino, J., Pons, X., Roda, F., Roscher, M., Feranec, J. (2010) Land cover change in

Europe between 1950 and 2000 determined employing aerial photography. Progress in Physical Geography 34: 183-205.

Germundsson, T., Schlyter, P. (Eds.) (1999) Atlas över Skåne. Sveriges Nationatlas förlag. Stockholm, SE.

Getman, D. J., Harbor, J. M, Johannsen, C. J., Engel, B. A., Shao, G. (2008) Improving the accuracy of Panchromatic Data. The Journal of Terrestrial Observation 1: 70-87.

Gilliam, F. S. (2007) The ecological significance of the herbaceous layer in temperate forest ecosystems. Bioscience 57(10): 845-858.

Goetz, S. J., Bunn, A. G., Fiske, G. J., Houghton, R. A. (2005) Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance. Proceedings of the National Academy of Sciences of the United States of America 102: 13521-13525.

Haines-Young, R. (2009) Land use and biodiversity relationships. Land Use Policy 26: 178-186.

Hartley, M. J. (2002) Rationale and methods for conserving biodiversity in plantation forests. Forest Ecology and Management 155: 81-95.

Helm, A., Hanski, I., Partel, M. (2006) Slow response of plant species richness to habitat loss and fragmentation. Ecology Letters 9: 72-77.

Hooftman, D.A.P., Bullock, J. M. (2012) Mapping to inform conservation: a case study of changes in semi-natural habitats and their connectivity over 70 years. Biological Conservation 145: 30–38.

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 species diversity. Landscape and Urban Planning 63: 241-250.

Huang, C., Davis, L. S., Townshend, J. R. G. (2002) An assessment of support vector machines for land cover classification. International Journal of Remote Sensing, 23, 725–749.

Ihse, M. (1995) Swedish agricultural landscapes - patterns and changes during the last 50 years, studied by aerial photos. Landscape and Urban Planning 31: 21-37.

Irminger Street, T., Prentice, H. C., Hall, K., Smith, H. G., Olsson, O. (2015) Removal of woody vegetation from uncultivated field margins is insufficient to promote non-woody vascular plant diversity. Agriculture Ecosystems & Environment 201: 1-10.

Kowarik, I. (1995) On the role of alien species in urban flora and vegetation. Plant invasions: General Aspects and Special Problems (ed. by P. Pysek, K. Prach, M. Rejmanek, and M. Wade (1995), pp. 85-103. SPB Academic Publishing, Amsterdam.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., Coomes, O. T., Dirzo, R., Fischer, G., Folke, C., George, P. S., Homewood, K., Imbernon, J., Leemans, R., Li, X. B., Moran, E. F., Mortimore, M., Ramakrishnan, P. S., Richards, J. F., Skanes, H., Steffen, W., Stone, G. D., Svedin, U., Veldkamp, T. A., Vogel, C., Xu, J. C. (2001) The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change 11: 261-269.

Li, H., Reynolds, J. F. (1995) On definition and quantification of heterogeneity. Oikos 73: 280-284.

Lindborg, R. (2006) Recreating grasslands in Swedish rural landscapes - effects of seed sowing and management history. Biodiversity and Conservation 15: 957-969.

MacDonald, D., Crabtree, J. R., Wiesinger, G., Dax, T., Stamou, N., Fleury, P., Lazpita, J. G., Gibon, A. (2000) Agricultural abandonment in mountain areas of Europe: Environmental consequences and policy response. Journal of Environmental Management 59: 47-69.

Marshall, E. J. P., Moonen, A. C. (2002) Field margins in northern Europe: their functions and interactions with agriculture. Agriculture Ecosystems & Environment 89: 5-21.

McGarigal, K., Cushman, S. A, Ene, E. (2012) FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. http://www.umass.edu/landeco/research/fragstats/fragstats.html. Downloaded 2016-05-10.

Müller, H., Rufin, P., Griffiths, P., Siqueira, A. J. B., Hostert, P. (2015) Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. Remote Sensing of Environment 156: 490-499.

Pal, M., Mather, P. M. (2004) Assessment of the effectiveness of support vector machines for hyperspectral data. Future Generation Computer Systems 20: 1215-1225.

Parviainen, M., Luoto, M., Heikkinen, R. K. (2010) NDVI-based productivity and heterogeneity as indicators of plant-species richness in boreal landscapes. Boreal Environment Research 15: 301-318.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., Stenseth, N. C. (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology & Evolution 20: 503-510.

Plexida, S. G., Sfougaris, A. I., Ispikoudis, I. P., Papanastasis, V. P. (2014) Selecting landscape metrics as indicators of spatial heterogeneity − a comparison among Greek landscapes. International Journal of Applied Earth Observation and Geoinformation 26: 26-35.

Plieninger, T., Gaertner, M. (2011) Harnessing degraded lands for biodiversity conservation. Journal for Nature Conservation, 19: 18–23.

Poschlod, P., WallisDeVries, M. F. (2002) The historical and socioeconomic perspective of calcareous grasslands - lessons from the distant and recent past. Biological Conservation 104: 361-376.

Pullin, A. S., Baldi, A., Can, O. E., Dieterich, M., Kati, V., Livoreil, B., Lovei, G., Mihok, B., Nevin, O., Selva, N., Sousa-Pinto, I. (2009) Conservation Focus on Europe: Major Conservation Policy Issues That Need to Be Informed by Conservation Science. Conservation Biology 23: 818-824.

Reitalu, T., Purschke, O., Johansson, L. J., Hall, K., Sykes, M. T., Prentice, H. C. (2012) Responses of grassland species richness to local and landscape factors depend on spatial scale and habitat specialization. Journal of Vegetation Science 23: 41-51.

Robinson, R. A., Sutherland, W. J. (2002) Post-war changes in arable farming and biodiversity in Great Britain. Journal of Applied Ecology 39: 157-176.

Roschewitz, I., Gabriel, D., Tscharntke, T., Thies, C. (2005) The effects of landscape complexity on arable weed species diversity in organic and conventional farming. Journal of Applied Ecology 42: 873-882.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W. (1973) Monitoring vegetation systems in the great plains with ERTS. In: Proceedings of the Third ERTS Symposium NASA SP-351, 1, U.S. Government Printing Office Washington DC, 309–317.

Römermann, C., Dutoit, T., Poschlod, P., Buisson, E. (2005) Influence of former cultivation on the unique Mediterranean steppe of France and consequences for conservation management. Biological Conservation 121: 21–33.

Schuster, C., Schmidt, T., Conrad, C., Kleinschmit, B., Forster, M. (2015) Grassland habitat mapping by intra-annual time series analysis - Comparison of RapidEye and TerraSAR-X satellite data. International Journal of Applied Earth Observation and Geoinformation 34: 25-34.

Silvertown, J. (2004) Plant coexistence and the niche. Trends in Ecology & Evolution 19: 605–611.

Sloan, S. (2012) Historical tropical successional forest cover mapped with Landsat MSS imagery. International Journal of Remote Sensing 33: 7902-7935.

Statistics Sweden. Statistical database. http://www.statistikdatabasen.scb.se. Downloaded 2016-05-10.

Stein, A., Gerstner, K., Kreft, H. (2014) Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. Ecology Letters 17: 866-880.

Storkey, J., Meyer, S., Still, K. S., Leuschner, C. (2012) The impact of agricultural intensification and land-use change on the European arable flora. Proceedings of Biological Sciences 279: 1421-1429.

Svenska Marktäckedata (SMD). http://gis-services.metria.se/nvfeed/atom/annex2.xml. Downloaded 2016-05-10.

Tarantino, C., Adamo, M., Lucas, R., Blonda, P. (2016) Detection of changes in semi-natural grasslands by cross correlation analysis with WorldView-2 images and new Landsat 8 data. Remote Sensing of Environment 175: 65-72.

Taylor, P.D., Fahrig, L., Henein, K., Merriam, G. (1993) Connectivity is a vital element of landscape structure. Oikos 68: 571–573.

Tscharntke, T., Klein, A. M., Kruess, A., Steffan-Dewenter, I., Thies, C. (2005) Landscape perspectives on agricultural intensification and biodiversity − ecosystem service management. Ecology Letters 8: 857-874.

Tucker, C.J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8:127–150.

Tyler, T., Olsson, K.-A., Johansson, H. & Sonesson, M. (2007) Floran i Skåne. Arterna och deras utbredning. Lunds Botaniska Förening, Lund.

Török, P., Miglecz, T., Valko, O., Kelemen, A., Deak, B., Lengyel, S., Tothmeresz, B. (2012) Recovery of native grass biodiversity by sowing on former croplands: is weed suppression a feasible goal for grassland restoration? Journal for Nature Conservation 20: 41-48.

Vapnik, V.N. (1995) The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Vellend, M., Verheyen, K., Jacquemyn, H., Kolb, A., Van Calster, H., Peterken, G., Hermy, M. (2006) Extinction debt of forest plants persists for more than a century following habitat fragmentation. Ecology 87: 542-548.

Vogelmann, J. E., Sohl, T., Howard, S.M. (1998) Regional characterization of land cover using multiple sources of data. Photogrammetric Engineering and Remote Sensing 64: 45-57.

Whittaker, R.H. (1972) Evolution and measurement of species diversity. Taxon 21: 213–251.

Willis, K. S. (2015) Remote sensing change detection for ecological monitoring in United States protected areas. Biological Conservation 182: 233-242.

Wrbka, T., Erb, K. H., Schulz, N. B., Peterseil, J., Hahn, C., Haberl, H. (2004) Linking pattern and process in cultural landscapes. An empirical study based on spatially explicit indicators. Land Use Policy 21: 289-306.

Xu, B., Gong, P. (2007) Land-use/land-cover classification with multispectral and hyperspectral E0-1 data. Photogrammetric Engineering and Remote Sensing 73: 955-965.

Zoungrana, B. J. B., Conrad, C., Amekudzi, L. K., Thiel, M., Da, E. D., Forkuor, G., Low, F. (2015) Multi-temporal Landsat images and ancillary data for land use/cover change (LULCC) detection in the Southwest of Burkina Faso, West Africa. Remote Sensing 7: 12076-12102.

# Supplementary information

**Table S1**. ID-codes and class names for the original 51 land-cover classes in the SMD data, and the class names after reclassifying into the eight classes used in the present study.

| SMD ID | SMD class name | Classes in the present study | SMD ID | SMD class name | Classes in the present study |
|---|---|---|---|---|---|
| 141 | Green urban area | Built-up | 3131 | Mixed woodland | Deciduous |
| 1422 | Airport | Built-up | 3132 | Mixed woodland mire | Deciduous |
| 1424 | Golf course | Built-up | 3133 | Mixed woodland rock | Deciduous |
| 1425 | Nonurban park | Built-up | 3241 | Scrub | Deciduous |
| 1426 | Camping holiday homes | Built-up | 211 | Arable land | Cropland |
| 111 | City center | Built-up | 222 | Permanent crop | Cropland |
| 121 | Industrial area | Built-up | 231 | Grazed grassland | Grazed grassland |
| 122 | Road railroad | Built-up | 321 | Natural grassland | Grazed grassland |
| 123 | Harbour | Built-up | 322 | Moorland | Grazed grassland |
| 124 | Airport | Built-up | 511 | Watercourse | Water |
| 132 | Dumps | Built-up | 521 | Coastal lagoon | Water |
| 133 | Construction site | Built-up | 522 | Estuary | Water |
| 1122 | Village | Built-up | 5121 | Open lake pond | Water |
| 1123 | Rural settlement | Built-up | 5122 | Covered lake pond | Water |
| 1421 | Sport leisure facility | Built-up | 5231 | Open sea | Water |
| 11211 | City urban | Built-up | 5232 | Covered sea | Water |
| 11212 | City suburb | Built-up | 331 | Beaches dunes sand planes | Barren |
| 3122 | Coniferous woodland mire | Conifers | 332 | Bare rock | Barren |
| 3123 | Coniferous woodland rock | Conifers | 1311 | Sand gravel extraction | Barren |
| 3242 | Clearing | Conifers | 1312 | Mineral extraction | Barren |
| 3243 | Young forest | Conifers | 411 | Marshland | Wetlands |
| 312121 | Coniferous woodland 5 - 15y | Conifers | 421 | Saltmarsh | Wetlands |
| 312122 | Coniferous woodland >15y | Conifers | 4121 | Wet mire | Wetlands |
| 3111 | Deciduous woodland | Deciduous | 4122 | Mire | Wetlands |
| 3112 | Deciduous woodland mire | Deciduous | 4123 | Peat extraction site | Wetlands |
| 3113 | Deciduous woodland rock | Deciduous | | | |

**Table S2**. Pearson's correlation coefficients between each of three vascular plant species richness categories; total number of plant species (total no. spp.), number of red-listed species (no. red-listed spp.) and number of grassland specialists (no. grassland specialist spp.), within 2.5 km × 2.5 km grid squares representing the whole province of Scania, and the subsets of squares representing the Plains, Central, and Woodland districts. All the correlations are statistically significant (p ≤ 0.05).

| | **Number of squares** | **total no. spp./ no. red-listed spp.** | **total no. spp./ no. grassland specialist spp.** | **no. red-listed spp./ no.grassland specialist spp.** |
|---|---|---|---|---|
| Scania | 473 | 0.70 | 0.93 | 0.63 |
| Plains | 193 | 0.66 | 0.92 | 0.52 |
| Central | 226 | 0.75 | 0.94 | 0.77 |
| Woodland | 54 | 0.73 | 0.93 | 0.69 |

**Table S3**. Error matrix for the SVM classification for the historical landscape, based on 4 bands from a Landsat 2 MSS scene from 3 July 1975. The diagonal indicating the correct class-assignments is shaded.

| | **Historical landscape** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Land-cover classes identified from aerial photos** | **Land-cover classes identified from a SVM-classified 4-band Landsat 2 MSS scene** | | | | | | | | | |
| | Cropland | Gr. grassl. | Deciduous | Conifers | Water | Wetlands | Barren | Built-up | Total | User's acc. |
| Cropland | 69 | 16 | 0 | 0 | 0 | 8 | 2 | 5 | 100 | 69.0% |
| Grazed grassland | 27 | 55 | 1 | 2 | 0 | 6 | 1 | 8 | 100 | 55.0% |
| Deciduous | 1 | 1 | 87 | 2 | 0 | 9 | 0 | 0 | 100 | 87.0% |
| Conifers | 1 | 0 | 0 | 96 | 0 | 3 | 0 | 0 | 100 | 96.0% |
| Water | 0 | 0 | 0 | 4 | 95 | 0 | 1 | 0 | 100 | 95.0% |
| Wetlands | 13 | 5 | 15 | 14 | 0 | 53 | 0 | 0 | 100 | 53.0% |
| Barren | 3 | 0 | 0 | 0 | 1 | 0 | 78 | 18 | 100 | 78.0% |
| Built-up | 1 | 9 | 0 | 0 | 0 | 0 | 1 | 89 | 100 | 89.0% |
| Total | 115 | 86 | 103 | 118 | 96 | 79 | 83 | 120 | 800 | |
| Producer's acc. | 60.0% | 64.0% | 84.5% | 81.4% | 99.0% | 67.1% | 94.0% | 74.2% | | |

Overall accuracy = 77.8%
Kappa = 0.746

**Table S4**. Error matrix for the SVM classification for the contemporary landscape, using 6 bands from a Landsat 7 ETM+ scene from 1 July 2001. The diagonal indicating the correct class-assignments is shaded.

| | **Contemporary landscape** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Land-cover classes identified from aerial photos** | **Land-cover classes identified from a SVM-classified 6-band Landsat 7 ETM+ scene** | | | | | | | | | |
| | Cropland | Gr. grassl. | Deciduous | Conifers | Water | Wetlands | Barren | Built-up | Total | User's acc. |
| Cropland | 90 | 7 | 2 | 0 | 0 | 0 | 0 | 1 | 100 | 90.0% |
| Grazed grassland | 2 | 88 | 2 | 0 | 0 | 7 | 0 | 1 | 100 | 88.0% |
| Deciduous | 1 | 1 | 95 | 0 | 0 | 3 | 0 | 0 | 100 | 95.0% |
| Conifer | 6 | 0 | 1 | 85 | 0 | 8 | 0 | 0 | 100 | 85.0% |
| Water | 0 | 0 | 0 | 1 | 98 | 1 | 0 | 0 | 100 | 98.0% |
| Wetlands | 4 | 11 | 4 | 5 | 0 | 75 | 0 | 1 | 100 | 75.0% |
| Barren | 0 | 1 | 0 | 0 | 0 | 0 | 94 | 5 | 100 | 94.0% |
| Built-up | 2 | 2 | 0 | 0 | 0 | 1 | 6 | 89 | 100 | 89.0% |
| Total | 105 | 110 | 104 | 91 | 98 | 95 | 100 | 97 | 800 | |
| Producer's acc. | 85.7% | 80.0% | 91.3% | 93.4% | 100.0% | 78.9% | 94.0% | 91.8% | | |

Overall accuracy = 89.3%
Kappa = 0.877

**Table S5**. Land cover transition matrix. The diagonal (shaded) shows the percentage of the landscape that persists between the historical (year 1975) and contemporary (year 2001) landscapes for each class of land-cover. The rows show the percentage of the landscape that a land-cover class (row) has lost to another land-cover class (column) during the transition from the historical landscape to the contemporary landscape. The columns show the percentage of the landscape that a land-cover class (column) has gained from another land-cover class (row) during the transition from the historical to the contemporary landscape.

**Land-cover transition matrix between the historical (year 1975) and contemporary (year 2001) landscapes**

| | Cropland | Gr. grassl. | Deciduous | Conifers | Water | Wetland | Barren | Built-up | Total 1975 | Loss |
|---|---|---|---|---|---|---|---|---|---|---|
| Cropland | 44.52% | 6.06% | 0.80% | 0.11% | 0.31% | 1.33% | 0.52% | 1.67% | 55.32% | 10.80% |
| Gr. grassl. | 3.42% | 8.75% | 1.20% | 0.32% | 0.21% | 1.34% | 0.11% | 0.57% | 15.92% | 7.17% |
| Deciduous | 0.16% | 0.21% | 5.37% | 0.43% | 0.01% | 0.45% | 0.01% | 0.01% | 6.65% | 1.29% |
| Conifers | 0.20% | 0.44% | 0.98% | 2.65% | 0.12% | 0.81% | 0.06% | 0.35% | 5.60% | 2.95% |
| Water | 0.01% | 0.02% | 0.01% | 0.00% | 1.21% | 0.01% | 0.02% | 0.05% | 1.33% | 0.12% |
| Wetland | 1.05% | 0.54% | 1.53% | 0.90% | 0.12% | 0.99% | 0.02% | 0.05% | 5.19% | 4.20% |
| Barren | 0.16% | 0.21% | 0.01% | 0.02% | 0.03% | 0.01% | 0.14% | 0.31% | 0.90% | 0.75% |
| Built-up | 1.80% | 1.73% | 0.08% | 0.14% | 0.15% | 0.18% | 0.16% | 4.84% | 9.09% | 4.25% |
| Total 2001 | 51.31% | 17.96% | 9.98% | 4.58% | 2.16% | 5.13% | 1.05% | 7.83% | 100.00% | 31.53% |
| Gain | 6.79% | 9.21% | 4.62% | 1.93% | 0.95% | 4.14% | 0.91% | 2.99% | 31.53% | |

**Table S6**. Pearson's correlation coefficients between measures of land-cover heterogeneity, within the historical and contemporary landscapes, respectively, in Scania. Results are based on the SVM classifications using the Landsat 2 MSS scene from 3 July 1975, and the Landsat 7 ETM+ scene from 1 July 2001, with supporting data on seasonal changes in the NDVI, and data on topography. NP = number of patches, TE = total edge (m), PR = patch richness, SHDI = Shannon's diversity index, SHEI = Shannon's evenness index, SD = standard deviation. All correlations are statistically significant (p ≤ 0.05).

| | **Historical landscape** | | | | |
|---|---|---|---|---|---|
| | NP | TE | PR | SHDI | SHEI |
| NP | 1 | 0.95 | 0.45 | 0.77 | 0.71 |
| TE | 0.95 | 1 | 0.43 | 0.79 | 0.75 |
| PR | 0.45 | 0.43 | 1 | 0.57 | 0.33 |
| SHDI | 0.77 | 0.79 | 0.57 | 1 | 0.95 |
| SHEI | 0.71 | 0.75 | 0.33 | 0.95 | 1 |

| | **Contemporary landscape** | | | | |
|---|---|---|---|---|---|
| | NP | TE | PR | SHDI | SHEI |
| NP | 1 | 0.95 | 0.61 | 0.8 | 0.72 |
| TE | 0.95 | 1 | 0.56 | 0.81 | 0.75 |
| PR | 0.61 | 0.56 | 1 | 0.64 | 0.44 |
| SHDI | 0.8 | 0.81 | 0.64 | 1 | 0.96 |
| SHEI | 0.72 | 0.75 | 0.44 | 0.96 | 1 |

**Table S7.** Measures of landscape heterogeneity for the historical and contemporary landscapes, respectively. The heterogeneity measures were calculated from 8 land-cover classes (Table S1), classified with the help of Landsat data, together with supporting data on seasonal changes in the NDVI, and data on topography. NP = number of patches, TE = total edge (m), PR = patch richness, SHDI = Shannon's diversity index, SHEI = Shannon's evenness index, SD = standard deviation.

| | **Historical landscape** | | | **Contemporary landscape** | | |
|---|---|---|---|---|---|---|
| **Scania** | **Mean** | **SD** | **Range** | **Mean** | **SD** | **Range** |
| NP | 36.1 | 21.2 | 105 | 45.0 | 27.1 | 130 |
| TE | 21275 | 11378 | 49620 | 23185 | 12894 | 51540 |
| PR | 5.9 | 1.24 | 8 | 6.6 | 1.3 | 8 |
| SHDI | 0.97 | 0.43 | 1.82 | 1.01 | 0.45 | 1.88 |
| SHEI | 0.54 | 0.23 | 0.97 | 0.53 | 0.22 | 0.95 |
| **Plains** | | | | | | |
| NP | 24.2 | 12.1 | 73 | 32.2 | 20.2 | 102 |
| TE | 14643 | 7460 | 38520 | 16105 | 9903 | 49140 |
| PR | 5.8 | 1.32 | 8 | 6.3 | 1.4 | 8 |
| SHDI | 0.77 | 0.39 | 1.76 | 0.81 | 0.44 | 1.83 |
| SHEI | 0.43 | 0.2 | 0.95 | 0.43 | 0.21 | 0.88 |
| **Central** | | | | | | |
| NP | 41.0 | 21.4 | 104 | 50.8 | 28 | 129 |
| TE | 24020 | 11163 | 47040 | 26014 | 12404 | 51480 |
| PR | 6.2 | 1.16 | 7 | 6.8 | 1.3 | 7 |
| SHDI | 1.07 | 0.42 | 1.79 | 1.11 | 0.43 | 1.77 |
| SHEI | 0.59 | 0.21 | 0.95 | 0.58 | 0.2 | 0.88 |
| **Woodland** | | | | | | |
| NP | 57.9 | 20.7 | 94 | 66.1 | 23.1 | 107 |
| TE | 33490 | 8826 | 43500 | 36651 | 8412 | 41880 |
| PR | 5.4 | 1.0 | 7 | 6.4 | 1.1 | 6 |
| SHDI | 1.24 | 0.33 | 1.49 | 1.27 | 0.27 | 1.35 |
| SHEI | 0.74 | 0.18 | 0.77 | 0.69 | 0.12 | 0.56 |

**Table S8**. Pearson's correlation coefficients between the proportion of land-cover classes within the historical and contemporary landscapes (based on Landsat classifications with supporting data on seasonal changes in the NDVI, and data on topography), and the contemporary within-square species richness within each of three categories of vascular plants (total number of plant species, number of red listed species, and number of grassland specialists), respectively. The shaded information is presented graphically in Fig. 2. All correlations that are not marked "ns" are statistically significant (p ≤ 0.05).

| | **Historical landscape** | | | | **Contemporary landscape** | | | |
|---|---|---|---|---|---|---|---|---|
| **Total number of plant species** | Scania | Plains | Central | Woodland | Scania | Plains | Central | Woodland |
| Cropland | -0.33 | -0.34 | -0.21 | ns | -0.35 | -0.35 | -0.23 | -0.39 |
| Gr. grassl. | 0.24 | ns | 0.31 | ns | 0.22 | ns | 0.22 | ns |
| Deciduous | 0.24 | 0.22 | ns | ns | 0.33 | 0.31 | 0.26 | ns |
| Conifers | 0.17 | ns | ns | ns | 0.15 | ns | ns | 0.39 |
| Water | ns | 0.21 | ns | ns | ns | 0.23 | ns | ns |
| Wetland | 0.35 | 0.27 | 0.27 | ns | 0.36 | 0.31 | 0.30 | ns |
| Barren | -0.18 | ns | -0.29 | ns | -0.14 | ns | -0.26 | ns |
| Built-up | ns | 0.25 | ns | ns | ns | ns | ns | ns |
| **Number of red-listed species** | | | | | | | | |
| Cropland | -0.33 | -0.43 | -0.25 | ns | -0.32 | -0.40 | -0.26 | ns |
| Gr. grassl. | 0.21 | ns | 0.33 | ns | 0.19 | ns | 0.25 | ns |
| Deciduous | ns | ns | ns | ns | 0.14 | ns | 0.24 | ns |
| Conifers | ns | ns | ns | 0.42 | ns | ns | ns | 0.37 |
| Water | ns | 0.29 | ns | ns | ns | 0.22 | ns | ns |
| Wetland | 0.13 | ns | ns | ns | 0.15 | ns | 0.19 | ns |
| Barren | ns | ns | ns | ns | ns | ns | ns | ns |
| Built-up | 0.23 | 0.43 | ns | 0.41 | 0.18 | 0.35 | ns | 0.50 |
| **Number of grassland specialist species** | | | | | | | | |
| Cropland | -0.36 | -0.35 | -0.21 | ns | -0.39 | -0.35 | -0.25 | ns |
| Gr. grassl. | 0.37 | 0.36 | 0.33 | ns | 0.35 | 0.32 | 0.27 | ns |
| Deciduous | 0.31 | 0.27 | ns | ns | 0.41 | 0.36 | 0.27 | ns |
| Conifers | 0.25 | ns | ns | ns | 0.20 | 0.22 | ns | ns |
| Water | ns | ns | ns | ns | ns | 0.30 | ns | ns |
| Wetland | 0.38 | 0.30 | 0.24 | ns | 0.46 | 0.44 | 0.33 | ns |
| Barren | -0.19 | ns | -0.26 | ns | -0.13 | ns | -0.23 | ns |
| Built-up | -0.15 | ns | ns | ns | -0.19 | ns | ns | ns |

ns = non-significant

**Table S9.** Pearson's correlation coefficients between the measures of landscape heterogeneity within the historical and contemporary landscapes (based on Landsat classifications with supporting data on seasonal changes in the NDVI, and data on topography), and the contemporary within-square species richness within each of three categories of vascular plants (total number of plant species, number of red listed species, and number of grassland specialists), respectively. The shaded information is presented graphically in Fig. 3. NP = number of patches, TE = total edge (m), PR = patch richness, SHDI = Shannon's diversity index, SHEI = Shannon's evenness index, SD = standard deviation. All correlations that are not marked "ns" are statistically significant (p ≤ 0.05).

| | **Historical landscape** | | | | **Contemporary landscape** | | | |
|---|---|---|---|---|---|---|---|---|
| **Total number of plant species** | Scania | Plains | Central | Woodland | Scania | Plains | Central | Woodland |
| NP | 0.46 | 0.47 | 0.39 | ns | 0.46 | 0.47 | 0.35 | 0.42 |
| TE | 0.50 | 0.51 | 0.43 | ns | 0.50 | 0.47 | 0.41 | ns |
| PR | 0.45 | 0.47 | 0.48 | ns | 0.46 | 0.46 | 0.43 | 0.52 |
| SHDI | 0.49 | 0.48 | 0.38 | 0.40 | 0.51 | 0.52 | 0.39 | 0.48 |
| SHEI | 0.39 | 0.38 | 0.26 | ns | 0.43 | 0.50 | 0.25 | ns |
| **Number of red-listed species** | | | | | | | | |
| NP | 0.23 | 0.30 | 0.25 | ns | 0.24 | 0.29 | 0.21 | ns |
| TE | 0.23 | 0.27 | 0.26 | ns | 0.22 | 0.23 | 0.23 | ns |
| PR | 0.37 | 0.32 | 0.39 | ns | 0.30 | 0.24 | 0.31 | 0.44 |
| SHDI | 0.33 | 0.36 | 0.32 | ns | 0.33 | 0.34 | 0.31 | 0.39 |
| SHEI | 0.26 | 0.31 | 0.24 | ns | 0.28 | 0.33 | 0.24 | ns |
| **Number of grassland specialist species** | | | | | | | | |
| NP | 0.52 | 0.46 | 0.36 | ns | 0.49 | 0.48 | 0.29 | ns |
| TE | 0.56 | 0.50 | 0.38 | ns | 0.54 | 0.48 | 0.35 | ns |
| PR | 0.38 | 0.45 | 0.40 | ns | 0.40 | 0.43 | 0.34 | 0.54 |
| SHDI | 0.52 | 0.52 | 0.32 | 0.41 | 0.54 | 0.58 | 0.32 | 0.49 |
| SHEI | 0.46 | 0.44 | 0.23 | ns | 0.49 | 0.56 | 0.22 | 0.35 |

ns = non-significant

**Table S10**. Pearson's correlation coefficients between the change in the proportions of cropland, deciduous forest and coniferous forest (based on Landsat classifications with support from data on seasonal changes in the NDVI, and data on topography) between the historical and contemporary landscapes, and the contemporary within-square species richness within each of three categories of vascular plants (total number of plant species, number of red listed species, and number of grassland specialists) for the whole study area in Scania and each of the districts in Scania. The shaded information is graphically presented in Fig. 4. All correlations that are not marked "ns" are statistically significant (p ≤ 0.05).

| | **Total number of plant species** | | | **Number of red-listed species** | | | **Number of grassland specialists** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Cropland | Deciduous | Conifers | Cropland | Deciduous | Conifers | Cropland | Deciduous | Conifers |
| Scania | ns | 0.32 | ns | ns | 0.23 | ns | -0.15 | 0.4 | -0.15 |
| Plains | ns | 0.35 | ns | ns | ns | ns | ns | 0.4 | ns |
| Central | ns | 0.22 | ns | ns | 0.32 | -0.22 | ns | 0.29 | -0.22 |
| Woodland | ns | 0.37 | ns | ns | 0.29 | -0.35 | ns | ns | ns |

ns = non-significant

# Paper III

# Small fragments of non-crop habitat make a significant contribution to landscape-scale vascular plant richness in agricultural landscapes

Therese Irminger Street, Jonas Dalmayne, Ola Olsson, Honor C. Prentice, Henrik G. Smith, Torbjörn T. Tyler, Karin Hall

## Abstract

The positive influence of non-crop habitat on farmland biodiversity has been widely recognized but the relative contribution of different types of non-crop habitat towards farmland biodiversity is less well known. In the present study, we used an information-theoretic approach with multimodel inference, to examine the extent to which vascular plant species richness at the landscape-scale (2.5×2.5 km) was influenced by land use and land cover, including small fragments of non-crop habitat such as road verges, field margins and field islets. Three regions, which differed in their degree of structural complexity, were included in the study. Our results show that the most important non-crop habitat types differed between regions: in the most simplified landscapes ("Plain districts") only the most widely distributed habitat types were important for landscape scale species richness of vascular plants whereas the extent of non-crop fragments contributed to both total and grassland species richness within the more complex landscapes ("Central districts" and "Forest districts"). However, it is possible that small biotopes influence the abundance of plant species within all three regions, in which case their importance for maintaining the farmland flora may well have been underestimated, especially in the most simplified landscapes.

**KEYWORDS:** landscape-scale species richness, vascular plants, grassland species, landscape complexity, non-crop habitat, small biotopes, multi-model inference

# 1. Introduction

Increases in agricultural intensity over the past half-century have been accompanied by major changes in management practices (Krebs et al., 1999; Tscharntke et al., 2005). There has been a general shift from small-scale mixed farming to large farming units within simplified, regionally specialized, agricultural landscapes, and an abandonment of farmland within marginal areas (Stoate et al., 2009). These changes have been accompanied by a widespread decline in farmland biodiversity (Benton et al., 2003; de Heer et al., 2005; Kleijn et al., 2009). Although agri-environment schemes have been directed, with varying degrees of success, towards mitigating the negative effects of intensification (Kleijn and Sutherland, 2003; Berendse et al., 2004; Kleijn et al., 2006), only 7 % of the habitat types linked to agriculture within the European Union are considered as being likely to continue to prosper without any change in their management (Commission of the European Communities, 2009).

In intensively cultivated areas, small fragments of habitat, such as drainage ditches, stone walls, permanent field margins, field islets, semi-natural pastures and patches of woodland, all enhance landscape heterogeneity. These different types of non-crop elements represent different environmental conditions and different histories of land use, and can provide habitat, resources and refugia for a wide range of farmland species (Duelli and Obrist, 2003; Smart et al., 2006; Billeter et al., 2008; Lindborg et al., 2014). Even small fragments of permanent vegetation, such as midfield islets, field margins and road verges (hereafter referred to collectively as “small biotopes”), can harbour grassland specialist plants (Cousins and Lindborg, 2008; Hamre et al., 2010). Hedgerows can harbour forest species (de Blois et al., 2002; Aavik et al., 2008), ditches wetland species (Aavik et al., 2008; Herzon and Helenius, 2008) and, in traditional agricultural landscapes, Wuczynski et al. (2014) found that approximately one in five field margins harboured red-listed plants species. Small biotopes can also ameliorate the negative effects of habitat fragmentation by facilitating the dispersal of organisms through the landscape (Burel, 1996; Tikka et al., 2001; van Dijk et al., 2014). Non-crop habitats within the farmland mosaic can thus have a positive effect on both the local and the landscape-scale richness of a range of organism groups, including vascular plants (e.g. Duelli and Obrist, 2003; Billeter et al., 2008; Aavik and Liira, 2009; Ernoult et al., 2013) and the importance of maintaining non-crop habitat within the agricultural landscape is widely recognised (e.g. Benton et al., 2003; Billeter et al., 2008; Lindborg et al., 2014).

Landscape indices are metrics that can be used to describe and quantify landscape structure and patterns of land use. These indices are increasingly being used to study how landscape structure influences ecological processes (Uuemaa et al.,

2013) and landscape indices based on the amount of non-crop habitat have been proposed as potential indicators of overall farmland biodiversity (Billeter et al., 2008; Banks-Leite et al., 2011). However, because landscape indices are influenced by the accuracy, scale, grain and thematic resolution (number of classes) of the maps on which they are based (Langford et al., 2006; Bailey et al., 2007b; Buyantuyev and Wu, 2007; Corry and Lafortezza, 2007; Simova and Gdulova, 2012), observed relationships between spatial patterns and the distribution of species are dependent on the quality of the data on which the calculations are based. Careful consideration therefore needs to be given both to the choice of data used to calculate the metrics and to the choice of the landscape metrics themselves (Dramstad, 2009).

Studies that analyse the drivers of species richness within agricultural landscapes often handle non-crop habitats as a single variable or pooled into broad habitat categories (e.g. Burel et al., 2013; Duflot et al., 2014; Loos et al., 2015, but see e.g. Honnay et al., 2003). Because small biotopes are seldom included in large-scale studies, and when included are often not treated separately from larger areas, little is known about the contribution of these habitat fragments to the overall species richness at the landscape-scale. Landscape context can influence species distribution patterns at the local scale and within a particular habitat type (e.g. Aavik and Liira, 2009; Concepcion et al., 2012) and may also be expected to influence the relative importance of different non-crop habitat types for landscape-scale species richness. However, the relative contribution of non-crop habitat to the species richness within landscapes of differing complexity is poorly understood.

In the present study, we aimed to determine the relative importance of different types of non-crop habitat for landscape-scale species richness of vascular plants within agricultural landscapes. We were particularly interested in the potential influence of small biotopes on large-scale species richness and therefore developed a proxy that was used to estimate the extent of small biotopes within south-Swedish agricultural landscapes. Because we were interested in whether-or-not the most influential non-crop habitat habitats differed between landscape types, we carried out the study in three types of agricultural landscapes, with differing degrees of structural complexity.

# 2. Methods

## 2.1 Description of the study region

The study was carried out in Scania, the southernmost Swedish province, which covers an area of approximately 11,000 km$^2$ (Fig. 1). The mean temperatures in the province range from 0 to -2°C in January and from 15.5 to 17°C in July, and the mean annual precipitation ranges from 500 to 900 mm (Germundsson and Schlyter, 1999). Scania can be divided into three agricultural regions (also known as production areas), on the basis of differences in edaphic, topographical and climatic conditions: the Plain, Central and Forest districts of Götaland (Fig. 1). The bedrock is mainly sedimentary in the Plain and Central districts, whereas gneiss and granites dominate within the Forest districts (Germundsson and Schlyter, 1999).

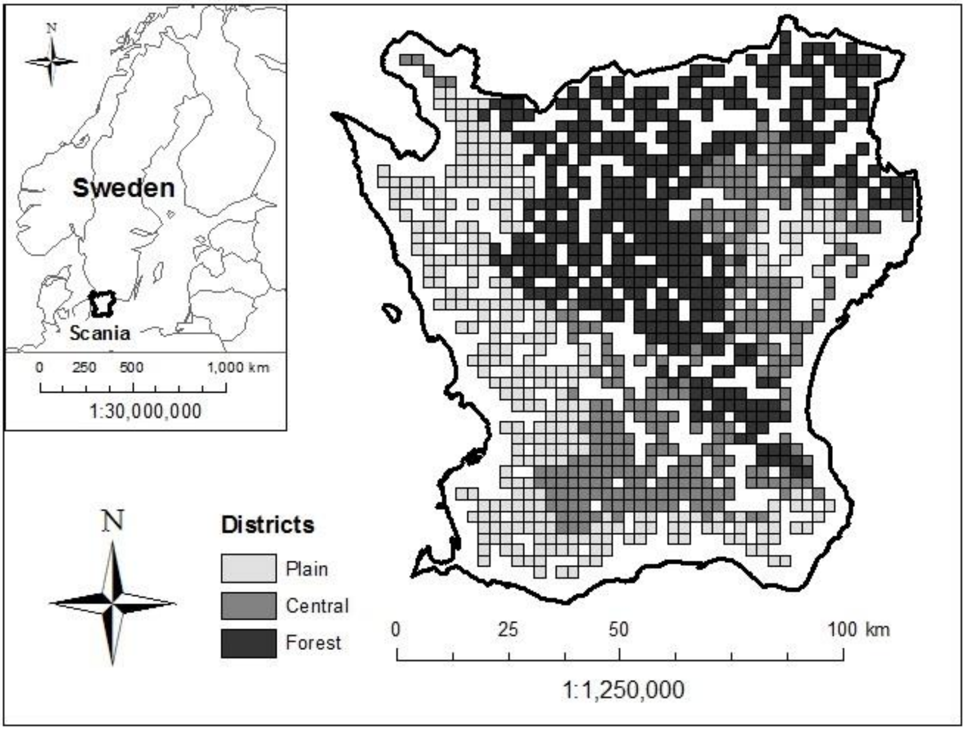


**Figure 1**: The province of Scania, Sweden showing the 1005 landscapes (2.5×2.5 km squares) included in the study.

The Plain districts contain Sweden’s most productive soils: crop yields are among the highest in the country and the landscape is dominated by large arable fields, primarily used for annual crops (StatisticsSweden, 2014). The Central districts are also dominated by farmland, but pastures and patches of woodland are more common than in the Plain districts, the landscape is more heterogeneous and the soils are less fertile (Germundsson and Schlyter, 1999). Arable fields are smaller than in the Plain districts and both cereal farming and animal husbandry are widespread, although not necessarily co-occurring within the same farming unit. A mix of deciduous and coniferous forests dominates within the Forest districts (Germundsson and Schlyter, 1999), where only a quarter of the land area is used for agricultural activities, primarily as ley fields and pastures. Fields are smaller and the structural complexity higher than in the other two districts. The three regions in Scania thus represent a gradient in landscape complexity and farming intensity, ranging from the intensively farmed and simplified landscapes of the Plain districts to the more complex agricultural landscapes of the Central districts and finally to the forest dominated landscapes of the Forest districts where the farmland is more complex than in the Central districts.

## 2.2 Vascular plant data

The species occurrence data used in the present study are derived from the project “Skånes Flora” (SF), a regional floristic survey coordinated by the Lund Botanical Society, which collected presence/absence data for all vascular plant taxa within Scania (Table 1). The survey was based on a grid with 1983 (2.5×2.5 km) cells (Tyler et al., 2007a), and was carried out by volunteer recorders between 1987 and 2005. The recorders were asked to report every taxon found in each cell at least once, and 1523 of the 1983 grid-cells were fully surveyed (at least 80 % of the taxa that could be expected to be present, on the basis of the number of taxa reported in neighbouring cells, had been recorded within the cell). Voucher specimens were required for critical taxa and these were verified by specialists. Taxa known to reproduce spontaneously in the province were considered to be resident.

For each of the 1523 grid-cells (hereafter referred to as landscapes) in the present study, we extracted information on (i) the number of regionally resident taxa (excluding only ill-defined or insufficiently documented/recorded infraspecific taxa) and (ii) the number of vascular plant taxa characteristic for grasslands (according to Ekstam and Forshed (1992)), giving us data on landscape-scale (i) richness of taxa, hereafter referred to as “species richness” (SR) and (ii) richness of grassland taxa, (GR) for 77 % of Scania. The list of regionally resident taxa was obtained from Tyler et al. (2007a) which follow the criteria for residency set up by

Jonsell (2004) and also used in the national Swedish check-list (Karlsson and Agestam).

**Table 1**: Landscape descriptors. The variables used in the present study were extracted from, or calculated on the basis of, information derived from six different data sources

| Name of data set | Source | Description | Reclassification |
|---|---|---|---|
| Project Skånes Flora (SF) | The Lund Botanical Society | Presence/absence data for all vascular plants within 2.5×2.5 km squares in a grid covering Scania. | No, but some pooling of taxa was carried out |
| Svensk Marktäckedata (SMD) | Lantmäteriet, the National Land Survey of Sweden | CORINE-based, digital land cover map with 51 classes present within Scania.<br>Spatial resolution of 25 m. | Yes, 8 broader classes defined. |
| Blockdatabasen (IACS) | Swedish Board of Agriculture | Data on land use for all registered farmland with 71 classes. | Yes, 6 broader classes defined. |
| Bedrock map (BR) | Geological Survey of Sweden | Digital bedrock map with 69 classes.<br>Spatial resolution of 50 m. | Yes, 4 broader classes defined. |
| National road database (NVDB) | Swedish transport administration | Database over all roads in Sweden. | No. |
| Digital elevation map (DEM) | Lantmäteriet, the National Land Survey of Sweden | Digital elevation map.<br>Spatial resolution of 50 m. | No. |

## 2.3 Environmental variables

### *2.3.1 Land cover and land use data from national databases*

The land cover categories used in the *Swedish Land Cover Data* (SMD) database (Table 1) do not necessarily reflect biologically relevant habitat categories. We therefore reclassified the 51 land cover classes in the SMD into eight broad habitat classes considered relevant for the present study; cultivated land (CULT), semi-natural grassland (SNG), deciduous forest (DECID), coniferous forest (CONIF), water (WATER), wetland (WETL), sparsely vegetated areas (SPARSE) and built-up areas (BUILT) (see appendix A for reclassification details). For each landscape (2.5×2.5 km), we calculated the percentage of the total area that was covered by each of the eight land cover classes (Table 2). We also used the reclassified land cover classes to calculate Simpson’s diversity index (SIDI), an index which, at the level of thematic resolution (i.e. the number of land use classes) used in the present study, has been shown to correlate well with total plant species richness in agricultural landscapes (Bailey et al., 2007a). All metrics were calculated using Fragstats v. 4 (McGarigal et al., 2012).

The timing and intensity of disturbance regimes influences landscape-scale richness of vascular plants (Buhk et al., 2007) and differs greatly between different land use categories within agricultural landscapes, a variability which is not captured by the broad class CULT. For farmland, we therefore extracted additional information about land use from *Blockdatabasen*, an annually updated Integrated Administration and Control System (IACS) from the Swedish Board of Agriculture (Table 1). The 71 original land use classes and crop types were reclassified into 6 broader classes which were expected to reflect the main differences in terms of disturbance; annual crops (ANNU), ley (LEY), fallow (FALL), riparian buffer zones (RIP), pasture (PAST) and other (OT), which together represent the total farmland area (FARM) (Table 2; appendix B). The IACS was first launched in 1998 and has, since then, been updated yearly (Jordbruksverket, 2009). To avoid using the earliest versions of the IACS, but still match the period during which the plant survey was performed (1987-2005), we extracted information from the IACS of 2001.

After including the variables extracted from the IACS, the variable CULT became redundant and was therefore excluded from further analyses as was the variable OT, which included a broad range of crops that could not be expected to influence the flora in a uniform manner (eg. *Salix* sp., *Picea* sp., *Phalaris arundinacea*, and fruit and berry plantations). For grasslands, we instead combined information from the IACS and the SMD and created a new layer "SNG_tot", in which all pastures from the IACS were combined with semi-natural grassland areas on non-farmland from the SMD.

### *2.3.2 Developing a proxy for small biotopes*

Information on the presence of small fragments of non-crop habitat, which were the central focus of the present study, could not be extracted directly from the IACS or the SMD. We, therefore, used an indirect approach, based primarily on the structure of the data stored in the IACS, to estimate the extent of small biotopes (SB) within each landscape.

In the IACS all agricultural land is divided into “blocks”, where each block is a continuous area of uniform land use (i.e. agricultural fields, hay meadows or pastures) delimited by a permanent boundary (Fig. 2). The block-edges can therefore be used as a proxy definition for uncultivated field margins (Persson et al., 2010). In addition, we hypothesised that the number of non-agricultural polygons (i.e. "holes") within agricultural blocks could be used to estimate the number of point-elements within agricultural fields by selecting only those holes that were between 0.01 and 0.1 ha and thus fit the size requirement for small biotopes in agricultural landscapes set up by the Swedish Board of Agriculture. The reliability of this approach was verified by extracting the number of point elements within the 180 (1 km$^2$) Scanian study sites used by Persson et al. (2010) and comparing our estimate (272 point elements) with that obtained by Persson et al. (2010) through manual interpretation of aerial photographs (292 habitat fragments).

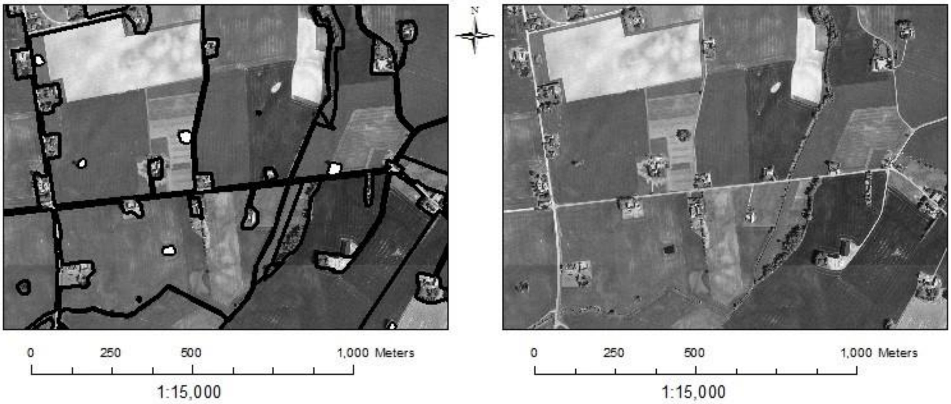


**Figure 2**: An illustration of the way in which agricultural blocks are delimited within the IACS. In the left-hand picture, block edges are shown in black and the "holes" that were extracted and referred to as point elements are shown in white.

The area of road verges within each landscape was estimated by extracting the total road length from the Swedish National Road Database (NVDB) (Table 1) and multiplying the road length by twice the width of a 1.6 m template (to account for both sides of the road). The template represented the mean width of road verges, measured at 622 locations in Scania (T. Irminger Street, unpublished data). For field margins, a mean width of 2.4 m (Persson et al., 2010) was assumed when converting boundary length to area. A mean size of 0.045 ha was assumed for the point-elements. Field margins, point elements and road verges were merged into the variable "small biotopes" (SB).

### *2.3.3 Abiotic variables*

#### *Bedrock*

Soil properties, which, in their turn, largely depend on the properties of the bedrock from which they are derived, strongly influence the distribution of vascular plants (Tyler, 2007b). In Scania, the underlying bedrock provides a reasonable representation of the properties of the overlying moraine, despite soil movements by ice, wind and water (Persson and Tyler, 2007), and the distribution of many plant species in Scania show a clear separation between areas with igneous and sedimentary bedrock (Weimarck and Andersson, 2007). Furthermore, lime influences soil pH, and hence soil properties and the distribution of species (Tyler, 2007b). To account for the main differences in substrate, we reclassified the digital bedrock map into four broad categories: igneous rocks (IG), calcareous sedimentary rocks (i.e. limestone and marl) (CALC), non-calcareous sedimentary rocks (SED) and ultrabasic bedrock types (UB) (see appendix C for reclassification details). The proportions of these bedrock categories were included in the statistical models to account for the expected influence of the bedrock on the richness of vascular plants.

#### *Topography*

In mountainous regions, the landscape-scale diversity of vascular plants is closely linked to topographical variability (Hofer et al., 2008) but topography also influences the flora in flatter regions such as Scania (where altitude ranges from 0 to 212 m a.s.l.) (Moeslund et al., 2013). We therefore calculated the standard deviation of the elevation (in metres), a commonly used measure of topographical roughness (see e.g. Luoto, 2000; Hofer et al., 2008; Aguilar-Santelises and del Castillo, 2013). The calculations were based on a digital elevation model of Scania with a spatial resolution of 50 m.

Operations and calculations were done in the software Matlab R2010a (Mathworks Inc.), ArcGIS 10.1 (ESRI) and Fragstats v. 4 (McGarigal et al., 2012).

## 2.4 Data analyses

### *2.4.1 Data preparation*

We excluded all landscapes in which the total cover of built-up areas and/or open water exceeded 40 % as well as landscapes that were within 2.5 km from the coastline (to reduce a potential influence from the sea), leaving 1006 landscapes for which plant data were available (Fig 1). Of these landscapes, 312 were located in the Plain districts, 279 in the Central districts and 415 in the Forest districts. Preliminary analyses revealed that one of the landscapes in the Forest districts, in which the extent of calcareous bedrock was five times greater than in any of the other Forest district landscapes, had a high leverage (Cook's distance close to 0.5) and that landscape was therefore excluded from further analyses. All variables that represented a proportion of the landscape were arcsine-square-root transformed prior to analyses.

### *2.4.2 Statistical models*

The environmental variables (Table 2) were fitted to each of the two response variables (SR and GR), within each of the three regions, using linear models and an information-theoretic approach to account for model uncertainty when analysing the consequences of many potential explanatory variables (Burnham and Anderson, 2002). The interdependence between all potential explanatory variables was first explored using correlation matrices and variance inflation factors (VIF). The VIF measures the degree to which the variance of the regression coefficients is inflated because of correlation with other variables, and VIF > 5 indicates potential problems with multicollinearity (Rogerson, 2001) (Appendix Correlation a, b and c). The variables CONIF, FARM and SED showed a high level of interdependence with at least one other variable within all three regions (Pearson's correlation coefficient ≥ 0.7) and were excluded from the full models for all the regions. The variable LEY was excluded from the models for the Forest districts, where it was highly correlated with the variable SB. The global (full) models for the Plain and Central districts thus contained 16 explanatory variables, whereas the global models for the Forest districts contained 15 explanatory variables (Table 2).

Linear models were fitted for each of the global models using the gls-function (R package nlme; Pinheiro et al., 2011). Models were fitted without any correction for spatial autocorrelation and with five commonly used correlation structures (exponential, Gaussian, linear, rational and spherical) (Zuur et al., 2009) and the models were ranked on the basis of their $AIC_c$-values (AIC with a second-order correction for small sample). We found only small differences between the models fitted with an exponential, Gaussian, rational or spherical correlation structure whereas the models fitted with a linear correlation structure performed relatively poorly ($\Delta AIC_c$ +1.5 to +59.6 compared with the "best" model), as did the uncorrected models ($\Delta AIC_c$ +15.5 to +82.4). Because none of the correlation structures consistently performed best, we arbitrarily chose to use the global models fitted with an exponential correlation structure (which at most had a $\Delta AIC_c$ of +3.3 compared to the "best" model) as a baseline for further model selection.

We used the dredge function (MuMIn package; Bartón, 2014) to run all possible combinations of the variables defined in each of the global models. Restrictions were added so that highly correlated (≥ 0.7) variables that remained within a global model (DECID, ANNU and SIDI in the Plain districts, ANNU and SIDI in the Central districts and ANNU and FALL in the Forest districts) were prevented from being present in the same model. However, SNG_tot and SIDI were allowed to co-occur when modelling species richness and grassland richness in the Plain districts, despite a correlation of 0.71, since the VIF-value remained relatively low (VIF = 4.2). The maximum number of terms to be tested within a single model was set to seven in order to obtain a manageable level of complexity. With these restrictions, all VIF-values were below five.

**Table 2**. The range (min-max) and mean values for all variables within the 1005 (2.5×2.5 km) landscapes representing the three types of agricultural regions in the study (the number of landscapes in each type of district is shown in parentheses). Variables that were included in the statistical analyses are shown in bold.

| **Data set** | **Variable** | **Explanation** | **Min-max (mean)** | | |
|---|---|---|---|---|---|
| | | | **Plain Districts (312)** | **Central districts (279)** | **Forest Districts (414)** |
| SF | **SR** | No. of regionally resident vascular plant taxa | 178-653 (374) | 282-587 (425) | 210-560 (372) |
| | **GR** | No. of grassland specialists[1] | 35-186 (98) | 71-178 (122) | 66-160 (112) |
| SMD | **BUILT***| Built-up areas (%) | 0-39 (6.1) | 0-37 (3.6) | 0-33 (2.4) |
| | CONIF* | Coniferous forest (%) | 0-70 (6.1) | 0-86 (17) | 0.74-89 (38) |
| | CULT* | Cultivated land (%) | 0-100 (72) | 0.48-98 (49) | 0-97 (17) |
| | **DECID*** | Deciduous forest (%) | 0-59 (7.4) | 0-59 (16) | 0.17-65 (29) |
| | SNG | Semi-natural grassland (%) | 0-40 (6.8) | 0.02-86 (12) | 0-45 (9.1) |
| | **SPARSE*** | Sparsely vegetated areas (%) | 0-9.0 (0.20) | 0-7.6 (0.12) | 0-8.3 (0.13) |
| | **WATER*** | Water (%) | 0-30 (1.0) | 0-38 (2.3) | 0-40 (2.1) |
| | **WETL*** | Wetland (%) | 0-15 (0.31) | 0-15 (0.53) | 0-46 (2.1) |
| | **SIDI** | Simpson Diversity Index | 0-0.81 (0.36) | 0.041-0.80 (0.56) | 0.063-0.80 (0.63) |
| IACS | **ANNU*** | Annual crops (%) | 0-92 (55) | 0-78 (28) | 0-56 (5.1) |
| | **FALL*** | Fallow (%) | 0-18 (4.9) | 0-20 (4.0) | 0-13 (0.76) |
| | **LEY**[2]* | Ley (%) | 0-31 (6.4) | 0-56 (13) | 0-35 (9.6) |
| | OT* | Other (%) | 0-9.3 (0.59) | 0-20 (0.86) | 0-4.4 (0.17) |
| | PAST* | Pasture (%) | 0-29 (3.9) | 0-59 (8.5) | 0-45 (6.8) |
| | **RIP*** | Riparian buffer zones (%) | 0-1.0 (0.22) | 0-1.2 (0.11) | 0-0.83 (0.030) |
| | FARM* | Total farmland (%) | 0.60-96 (71) | 2.8-95 (54) | 0-92 (23) |
| SMD + IACS | **SNG_tot*** | Semi-natural grassland (%) | 0-44 (6.6) | 0.92-60 (13) | 0.26-53 (11) |
| IACS + NVDB | **SB*** | Small biotopes (%) | 1.1-4.4 (2.5) | 0.88-4.0 (2.4) | 0.67-4.3 (2.0) |
| BR | **CALC*** | Calcareous sedimentary bedrock (%) | 0-100 (34) | 0-100 (29) | 0-2.6 (0.016) |
| | **IG*** | Igneous bedrock (%) | 0-100 (15) | 0-100 (33) | 0-100 (87) |
| | SED* | Non-calcareous sedimentary bedrock (%) | 0-100 (50) | 0-100 (37) | 0-100 (9.3) |
| | **UB*** | Ultrabasic bedrock (%) | 0-22 (1.1) | 0-6.8 (0.63) | 0-100 (3.9) |
| DEM | **TOPO** | Topographical roughness | 0.88-45 (8.1) | 1.2-42 (11) | 2.5-55 (10) |

* The variable was arcsine square root-transformed prior to analyses. Untransformed values are presented here

[1] According to Ekstam and Forshed(1992)

[2] Not included in the global model for the Forest district because of high correlation with the variable SB.

The influence of the explanatory variables was examined using a multimodel inference approach in which we used Akaike model weights to estimate the relative importance of the different variables within each data set (Burnham and Anderson, 2002). The Akaike model weight ($w_i$) represents the likelihood that a given model is the best fitted model, amongst all considered models, and is calculated by dividing the relative likelihood of model i by the sum of the relative likelihood values of all R models within the model set.

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta AICc_i\right)}{\sum_{r=1}^{R}\exp\left(-\frac{1}{2}\Delta AICc_r\right)}$$

By summarizing the $w_i$ of all models within $\Delta$AIC$_c$ 6 of the best-fitted model, we obtained an estimate of the relative importance of each variable for each data set. Because highly correlated variables were prevented from occurring together during the model selection procedure, the model sets were not balanced. Two highly correlated variables that were prevented from occurring together would, in a situation where one of the variables was always present and both variables occurred equally often, both obtain a relative variable weight of approximately 0.5. Both variables would in this case be considered to be un-influential. If instead only one of the variables had been included in the global model, and that variable was included in all models, then the relative variable importance would be 1 and the variable would instead be considered highly influential. Therefore, when interpreting the results, we also considered the summed importance of variables that had been prevented from occurring together.

To ensure that the explanatory variables (CONIF, FARM, SED and, in the Forest districts, LEY) that had been excluded a priori from the models, were not more influential than the included variables with which they were highly correlated, we created a new model for each data set with all variables that had a variable weight of at least 0.5. The “included variables” were replaced, one-by-one, by the “excluded variables” and the AIC$_c$-values of the original and the modified models were compared. As an example, in the models for the Forest districts, the variable SB was first replaced by FARM and then by LEY, and the AIC$_c$-values of the two models were compared to the AIC$_c$-value of the model with SB. All the model analyses were carried out in R, version 3.1.0 (R Core Team, 2014). To visualize the effects of the main explanatory factors in the generalized linear mixed models we used effects plots (package effects; Fox, 2003).

# 3. Results

A total of 1440 regionally resident vascular plant species were recorded within the 1005 landscapes included in the present study. Most of the species were recorded within all three regions.

## 3.1 Total species richness

The relative importance of the different variables within each data set was calculated on the basis of 243 models for species richness within the Plain districts, 319 models for species richness within the Central districts and 83 models for species richness within the Forest districts. The relative variable importance was generally higher for variables related to land cover than for variables related to crop type or to environmental heterogeneity, although the extent of fallow (Plain districts) and topographical roughness (Central and Forest districts) were also important (Table 3). The most important land cover types were: semi-natural grasslands, built-up areas and deciduous forests within the Plain districts; semi-natural grasslands, small biotopes and water within the Central districts and finally; built-up areas, small biotopes and water within the Forest districts.

**Table 3**: Estimates, adjusted standard error (Adj SE) and relative variable importance (RVI) for species richness within each region. Estimates are based on model-averaged coefficients with shrinkage (i.e. the variable estimator was set to zero when the variable was not included in the model). Variables considered to be at least moderately influential (RVI≥0.7) are shown in bold text. For definition of variables see Table 2.

| | Plain Estimate | Adj SE | RVI | Central Estimate | Adj SE | RVI | Forest Estimate | Adj SE | RVI |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | 232.8 | 28.3 | | 270.1 | 50.9 | | 196.9 | 31.2 | |
| SNG_tot* | **186.3** | **33.1** | **1.00** | **99.9** | **40.5** | **0.98** | 4.77 | 17.8 | 0.15 |
| TOPO | 0.17 | 0.48 | 0.21 | **1.00** | **0.72** | **0.80** | **1.43** | **0.49** | **1.00** |
| BUILT* | **130.2** | **24.3** | **1.00** | 36.6 | 38.2 | 0.59 | **174.2** | **22.5** | **1.00** |
| SB* | 41.3 | 144.4 | 0.17 | **460.3** | **338.9** | **0.77** | **584.7** | **196.9** | **0.99** |
| DECID* | **139.6** | **30.8** | **1.00**[1] | 15.9 | 30.2 | 0.32 | 55.6 | 27.2 | 0.52 |
| WATER* | 21.6 | 40.1 | 0.33 | **65.4** | **36.6** | **0.90** | **71.0** | **23. 5** | **1.00** |
| SIDI | 3.3e-04 | 14.9 | 0.10[1] | 42.9 | 42.7 | 0.61[1] | **57.9** | **44.5** | **0.74** |
| FALL* | **134.8** | **48.7** | **1.00** | 0.84 | 13.5 | 0.08 | 85.4 | 49. 5 | 0.34[1] |
| UB* | -8.79 | 32.4 | 0.17 | -46.1 | 71.0 | 0.41 | 0.41 | 6.07 | 0.07 |
| WETL* | 28.9 | 61.8 | 0.28 | 16.6 | 43.1 | 0.21 | -10.0 | 23.5 | 0.23 |
| IG* | 3.96 | 10.3 | 0.23 | 0.06 | 2.74 | 0.07 | -0.71 | 3.81 | 0.10 |
| SPARSE* | 68.7 | 89.7 | 0.48 | 22.7 | 65.2 | 0.19 | 13.7 | 40.8 | 0.18 |
| LEY* | 9.12 | 24.1 | 0.22 | -16.5 | 30.5 | 0.32 | NA | NA | NA |
| CALC* | -0.38 | 3.19 | 0.12 | 0.28 | 2.54 | 0.08 | 10.0 | 49.7 | 0.11 |
| RIP* | 8.01 | 54.4 | 0.12 | -12.6 | 65.2 | 0.11 | -34.8 | 102.5 | 0.17 |
| ANNU* | 0.87 | 10.0 | 0.11[1] | 0.50 | 7.05 | 0.05[1] | 1.75 | 9.81 | 0.08[1] |

* Arcsine square root-transformed variable

[1] Highly correlated variables that were prevented from occurring together during the model selection procedure

Replacing highly correlated variables with their excluded counterpart(s) did not lead to an improvement for any of the models. Within the Plain districts, DECID appeared to be a better predictor of total species richness than any of the variables with which it was highly correlated: $AIC_c$ increased by 12.37 when DECID was replaced by FARM and by 15.14 when replaced by CONIF. Within the Central districts, $AIC_c$ decreased by 0.20 when SIDI was replaced by FARM and increased by 2.33 when SIDI was replaced by CONIF. Within the Forest districts, SB was a better predictor of total species richness than the variables with which it was highly correlated ($AIC_c$ increased by 9.99 when replaced by LEY and by 6.44 when replaced by FARM). Adding CONIF, FARM and SED where appropriate did not improve any of the models.

## 3.2 The richness of grassland species

The relative importance of the different variables within each data set was calculated on the basis of 530 models for grassland richness within the Plain districts, 993 models for grassland richness within the Central districts and 152 models for grassland richness within the Forest districts. The relative importance of individual variables did not always follow the same pattern as for species richness, but the general trend was the same for both the response variables, with variables related to land cover being more important than variables related to crop type or environmental heterogeneity (Table 4). The most important land cover types were: semi-natural grasslands and deciduous forests within the Plain districts (Fig 3); semi-natural grasslands and small biotopes within the Central districts and finally; semi-natural grasslands, built-up areas, small biotopes and water within the Forest districts.

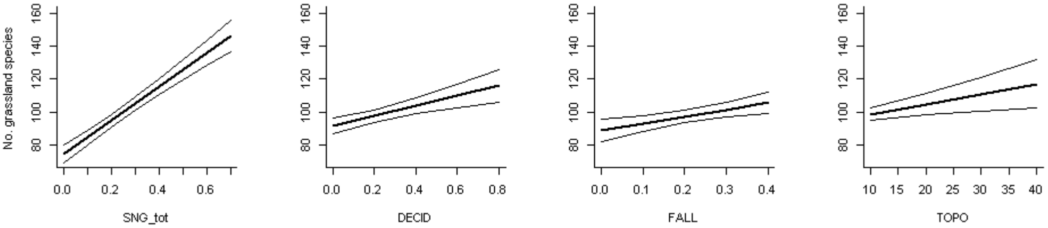
**Table 4**: Estimates, adjusted standard error (Adj SE) and relative variable importance (RVI) for the richness of grassland species within each region. Estimates are based on model-averaged coefficients with shrinkage. Variables considered to be at least moderately influential (RVI≥0.7) are shown in bold.

| | **Plain** | | | **Central** | | | **Forest** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Estimate** | **Adj SE** | **RVI** | **Estimate** | **Adj SE** | **RVI** | **Estimate** | **Adj SE** | **RVI** |
| Intercept | 54.3 | 8.83 | | 72.9 | 14.6 | | 69.2 | 8.66 | |
| SNG_tot* | **93.7** | **12.3** | **1.00** | **64.5** | **10.4** | **1.00** | **21.0** | **13.4** | **0.83** |
| TOPO | **0.57** | **0.28** | **0.93** | 0.096 | 0.16 | 0.38 | **0.34** | **0.13** | **0.98** |
| BUILT* | 0.60 | 5.08 | 0.21 | -3.19 | 7.27 | 0.28 | **23.1** | **6.65** | **1.00** |
| SB* | 5.90 | 34.4 | 0.14 | **123.2** | **90.6** | **0.81** | **129.9** | **60.8** | **0.94** |
| DECID* | **19.5** | **14.5** | **0.75** | 13.0 | 12.0 | 0.68 | 1.98 | 5.21 | 0.20 |
| WATER* | 2.13 | 7.41 | 0.17 | 0.58 | 3.59 | 0.13 | **10.3** | **8.61** | **0.71** |
| SIDI | 8.39 | 11.9 | 0.43 | 5.20 | 9.36 | 0.35 | **17.1** | **14.0** | **0.71** |
| FALL* | **42.7** | **15.1** | **1.00** | -0.15 | 4.77 | 0.11 | 1.13 | 5.42 | 0.11 |
| UB* | -3.66 | 11.7 | 0.19 | **-33.7** | **26.8** | **0.76** | -0.04 | 1.54 | 0.07 |
| WETL* | 2.05 | 10.1 | 0.13 | 8.76 | 16.1 | 0.35 | -2.84 | 6.44 | 0.25 |
| IG* | 4.05 | 5.24 | 0.49 | 1.54 | 2.82 | 0.34 | 0.005 | 0.72 | 0.07 |
| SPARSE* | 6.31 | 16.6 | 0.22 | 2.03 | 12.3 | 0.13 | 0.91 | 6.26 | 0.09 |
| LEY* | 4.49 | 9.16 | 0.29 | -3.14 | 7.37 | 0.26 | NA | NA | NA |
| CALC* | -0.88 | 1.97 | 0.27 | -0.35 | 1.34 | 0.17 | 25.7 | 52.8 | 0.28 |
| RIP* | -1.25 | 15.2 | 0.11 | -2.59 | 19.0 | 0.12 | -15.9 | 35.9 | 0.25 |
| ANNU* | -1.01 | 4.29 | 0.14 | 0.35 | 2.75 | 0.10 | -1.11 | 3.80 | 0.15 |

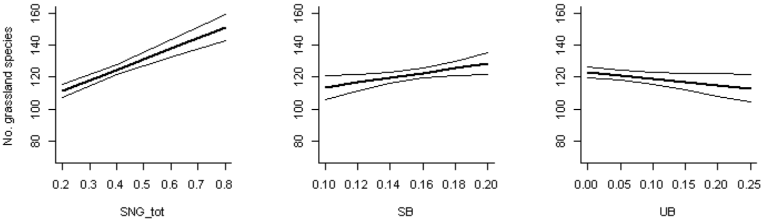
* Arcsine square root-transformed variable

[1] Highly correlated variables were prevented from occurring together during the model selection procedure

**a) Plain districts**



**b) Central districts**



**c) Forest districts**

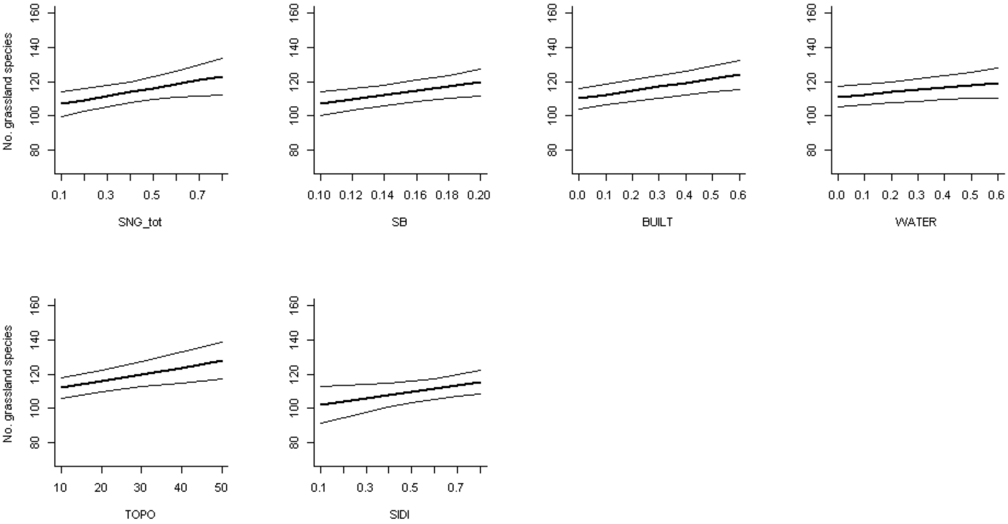


**Fig 3**: The relationship between the richness of grassland species and the variables that had a relative variable importance of at least 0.7 in each of the three districts. Only variables that had a weight of 0.5, or more, were included in the model.

Replacing highly correlated variables with their excluded counterpart(s) did not improve any of the models: within the Plain districts, the AICc-value only changed marginally when DECID was replaced by CONIF, whereas it increased by 4.16 when DECID was replaced by FARM. In the Central districts, AICc-values increased by 3.65 when DECID was replaced by FARM and by 4.49 when replaced by CONIF. And finally, within the Forest districts, SB was a better predictor for grassland richness than both LEY (ΔAICc+5.65) and FARM (ΔAICc +6.80) were. None of the models were substantially improved by the addition of CONIF, FARM or SED, where appropriate.

To estimate model uncertainties, we examined all models on which the relative variable importance had been based, and found that many of the models contained uninformative parameters (Δ AICC< 2 when compared to a more parsimonious model) (Arnold, 2010). However, because the uninformative parameters were only included in models sporadically, their relative importance was low, whereas the variables within the more parsimonious model were included more frequently and hence the ones that were interpreted as influential.

# 4. Discussion

We show that the relative importance of different non-crop habitat types depends on landscape context and that even small fragments of non-crop habitat, such as small biotopes, can increase landscape-scale species richness of vascular plants. Although previous studies have found a positive relationship between land use heterogeneity and diversity (e.g. Stein et al., 2014), we found that the extent of a number of different non-crop habitat categories was generally a better descriptor of species richness patterns than land use heterogeneity per se was.

## 4.1 Non-crop habitat and land use

### *4.1.1 The Plain districts*

The positive effect of non-crop habitat types on total and grassland species richness within the highly simplified Plain districts was expected and in accordance with previous studies (e.g. Deutschewitz et al., 2003; Billeter et al., 2008). However, the influence on species richness was limited to the most widespread non-crop habitat types (i.e. semi-natural grasslands, built-up areas and deciduous forests for species richness and semi-natural grasslands and forests for grassland richness), whereas small biotopes, wetlands, and other habitat categories that covered less than a few percent of the landscape, did not appear to influence landscape-scale species richness (Tables 2, 3 and 4). These results are, at least in part, in accordance with Edvarsen et al. (2010) who found that woodlands and pastures contributed towards overall species richness, whereas midfield islets and several border habitat types were dominated by species typical of intensively used agricultural land and thus contributed little to overall species richness.

Not only the most widespread non-crop habitat types, but also the extent of fallow had a positive influence on total and grassland species richness within the Plain districts. The richness of arable fields has been shown to decrease exponentially with land use intensity (Kleijn et al., 2009). Fallows are less intensively managed than cultivated fields, and may consequently be considerably richer than other field types (Koellner and Scholz, 2008). Field margins and hedgerows adjacent to fallows may also be richer than margins adjacent to other types of arable fields (de Blois et al., 2002). We therefore interpret the positive influence of fallow on species richness and grassland richness within the present study as an effect of the agricultural land being less intensively used. Either, because the overall land use intensity decreases within a landscape as the extent of fallow in the crop rotation increases, or because fallows are more likely to be found within the less intensively farmed parts of the Plain districts.

### *4.1.2 The Central districts*

The extent of non-crop habitats also had a positive effect on species richness within the Central districts, but the influence on species richness was not limited to the most widespread habitat types. Semi-natural grasslands, water and small biotopes were the most important habitat types for total species richness, whereas semi-natural grasslands, small biotopes and, to some extent, deciduous forests were the most important habitat types for grassland species richness (Tables 3 and 4). Hence, in these more complex agricultural landscapes, small-sized non-crop habitats also had a positive effect on species richness and the association between size and influence was less strong than in the Plain districts. Several studies have shown that small biotopes may make an important contribution to local species richness and to the richness of a particular habitat type (e.g. Gabriel et al., 2005; Aavik et al., 2008; Janisova et al., 2014); here we show that they can also be of importance for the richness of species at the landscape-scale.

The top-ranking models for the Central districts were more variable than within the other two regions, which caused an overall reduction in the relative variable importance of the most influential variables within this region. We suggest that this is, at least partly, explained by a greater variation between landscapes within the Central districts as compared with the other two regions. If different processes govern species richness patterns within the same region, then one model will not be able to capture the “full” picture.

### *4.1.3 The Forest districts*

As within the Central districts, both total and grassland species richness was increased by non-crop habitat types with a limited distribution (Tables 2, 3 and 4). However, within the Forest districts, total species richness was mainly explained by land use complexity (SIDI) and by built-up areas, water and small biotopes, none of which covered more than a low percent of the landscapes. More widely distributed land use classes, such as forest and semi-natural grasslands, on the other hand, were less influential although, and not surprisingly, the extent of semi-natural grasslands was important for the richness of grassland species.

## 4.2 Small biotopes

In our study, the positive influence of small biotopes increases with landscape complexity and hence, the marginal importance of small biotopes does not appear to decrease as the extent of non-crop habitat within the surrounding landscape increases. On the contrary, a sufficient amount of other non-crop habitat types appear to be necessary for small biotopes to be of importance for landscape-scale richness of vascular plants. Likewise, Aavik and Liira (2009) found that small fragments of non-crop habitat alone were insufficient to influence plant diversity at the landscape-scale. However, contrary to the suggestion by Aavik et al. (2009), we found that much of the positive influence of non-crop habitats within our study sites could be attributed to the extent of small biotopes and not just to the availability of stable conditions within larger habitat patches. These results hold true for both total species richness and for the richness of grasslands specialists and are hence likely not to simply reflect an increase in the number of generalist plant species. Although we cannot separate the effect of small biotopes alone, from that of small biotopes in combination with other habitat types, we believe that there are synergistic effects between large non-crop patches and small fragments of non-crop habitat so that the influence of large and small patches may be mutually reinforcing, as found by Grashof-Bokdam et al. (2009) in a study of forest species within large woodland patches and small woody fragments.

Small biotopes vary considerably in their local habitat conditions – from dry to moist, from shaded to sunny and from highly disturbed to relatively stable. Different types of small biotope are thus likely to support species with different life history traits (Baudry et al., 2000; Aavik et al., 2008). In the Plain districts of Scania, the small biotopes are likely to be more influenced by farming activities than within the other two regions: in part because of a higher overall intensity of agricultural activities but also because the small biotopes are mainly located between two arable fields, and not between an arable field and another habitat type. Hence small biotopes in the Plain districts are likely to have become impoverished to a larger extent, and to contain more agro-tolerant species (Cousins, 2006; Aavik and Liira, 2009; Ma et al., 2013), than the small biotopes within the Central and Forest districts which could, in part, explain the lack of influence of small biotopes on landscape-scale species richness within the Plain districts. Small biotopes in the Plain districts are also, on average, more isolated from other non-crop habitats types than within the other two regions and differences in edaphic conditions between arable fields and small biotopes are smaller within the (highly fertile) Plain districts than within the (less fertile) Central districts and (nutrient poor) Forest districts. Furthermore, road verges often have a higher pH than the surrounding landscape within the Forest districts, which enables species that are not found outside the road network to persist there.

## 4.3 The influence of environmental variability within the three regions

Total species richness within the Central and Forest districts, and grassland species richness within the Plain and Forest districts all increased with increasing topographical roughness (Tables 3 and 4). Topography influences habitat conditions in terms of soil moisture, solar radiation and wind exposure (Moeslund et al., 2013) and thus influences the micro-climatic variability within the landscapes. With a greater topographical roughness, local site conditions become more variable, thus enabling more species to co-occur. But topography also reflects (past and present) land use (Persson and Tyler, 2007): within Scania, flat areas are likely to be/have been cultivated whereas areas with a more rugged terrain are likely to have had a history as pasture or forest. We believe that the influence exerted by topographical roughness on the flora within our study sites reflects both of these aspects.

Although many distribution maps for species within Scania (Tyler et al., 2007a) show a clear separation between areas with igneous and sedimentary bedrock, the underlying bedrock generally had very little influence on species richness patterns within the present study. The overlying soil may locally differ from that of the underlying bedrock, which could partly explain the lack of influence, but a more probable explanation is that the main differences are between, rather than within, regions and will hence not be captured when the three regions are analysed separately. However, the difference in species richness between the three regions could in part be a reflection of differences in the substrate.

# 5. Conclusions

Detailed, large-scale inventories of vascular plants are expensive both in terms of time and resources. Nevertheless, information from such surveys may be necessary when choosing where management should be applied and how limited (financial) resources for the conservation of biodiversity should be used. Finding alternative approaches to the prediction of patterns of species richness is therefore of interest. In the present study we take one step towards such an understanding by disentangling the relative importance of different habitat types for total and grassland species richness within agricultural landscapes of varying complexity.

The different non-crop land cover types were generally more important predictors of landscape-scale species richness than land cover heterogeneity (SIDI) was, and the habitat types with a large contrast in relation to the dominant land use were generally the most influential for landscape-scale species richness. The greater importance of small biotopes within the more complex landscapes, as compared to in the Plain districts, suggests that, from a plant richness perspective, small biotopes are mainly important for maintaining landscape-level species richness within landscapes where a sufficient amount of non-crop habitat still remains. However, we only look at the richness of plant species within each landscape, not at how abundant the different species are. Had data on abundance been available, the importance of small biotopes for vascular plants may well have been higher within all three regions.

# Acknowledgments

We thank the Lund Botanical Society for granting us access to the database from the project Skånes Flora, Anna Persson for giving us access to her aerial photo interpretations and Martin Stjernman (SAPES) for assistance with the IACS. The research was funded by grants from the Swedish Research Council FORMAS through the projects “More Biodiversity at Less Cost” (O. Osson), “Sustainable Agriculture for the Production of Ecosystem Services” (SAPES) (H.G. Smith) and “High-spatial resolution imagery for inventorying grasslands” (K. Hall and H.C. Prentice).

# References

Aavik, T., Augenstein, I., Bailey, D., Herzog, F., Zobel, M. & Liira, J. (2008) *What is the role of local landscape structure in the vegetation composition of field boundaries? Applied Vegetation Science,* **11,** 375-386.

Aavik, T. & Liira, J. (2009) *Agrotolerant and high nature-value species-Plant biodiversity indicator groups in agroecosystems. Ecological Indicators,* **9,** 892-901.

Aguilar-Santelises, R. & del Castillo, R. F. (2013) *Factors affecting woody plant species diversity of fragmented seasonally dry oak forests in the Mixteca Alta, Oaxaca, Mexico. Revista Mexicana De Biodiversidad,* **84,** 575-590.

Arnold, T. W. (2010) *Uninformative Parameters and Model Selection Using Akaike's Information Criterion. Journal of Wildlife Management,* **74,** 1175-1178.

Bailey, D., Billeter, R., Aviron, S., Schweiger, O. & Herzog, F. (2007a) *The influence of thematic resolution on metric selection for biodiversity monitoring in agricultural landscapes. Landscape Ecology,* **22,** 461-473.

Bailey, D., Herzog, F., Augenstein, I., Aviron, S., Billeter, R., Szerencsits, E. & Baudry, J. (2007b) *Thematic resolution matters: Indicators of landscape pattern for European agro-ecosystems. Ecological Indicators,* **7,** 692-709.

Banks-Leite, C., Ewers, R. M., Kapos, V., Martensen, A. C. & Metzger, J. P. (2011) *Comparing species and measures of landscape structure as indicators of conservation importance. Journal of Applied Ecology,* **48,** 706-714.

Bartón, K. (2014) MuMIn: Multi-model inference. R package version 1.10.0. http://CRAN.R-project.org/package=MuMIn.

Baudry, J., Bunce, R. G. H. & Burel, F. (2000) *Hedgerows: An international perspective on their origin, function and management. Journal of Environmental Management,* **60,** 7-22.

Benton, T. G., Vickery, J. A. & Wilson, J. D. (2003) *Farmland biodiversity: is habitat heterogeneity the key? Trends in Ecology & Evolution,* **18,** 182-188.

Berendse, F., Chamberlain, D., Kleijn, D. & Schekkerman, H. (2004) *Declining biodiversity in agricultural landscapes and the effectiveness of agri-environment schemes. Ambio,* **33,** 499-502.

Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek, R., Burel, F., Cerny, M., De Blust, G., De Cock, R., Diekoetter, T., Dietz, H., Dirksen, J., Dormann, C., Durka, W., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., Le Coeur, D., Maelfait, J. P., Opdam, P., Roubalova, M., Schermann, A., Schermann, N., Schmidt, T., Schweiger, O., Smulders, M. J. M., Speelmans, M., Simova, P., Verboom, J., van Wingerden, W., Zobel, M. & Edwards, P. J. (2008) *Indicators for biodiversity in agricultural landscapes: a pan-European study. Journal of Applied Ecology,* **45,** 141-150.

Buhk, C., Retzer, V., Beierkuhnlein, C. & Jentsch, A. (2007) *Predicting plant species richness and vegetation patterns in cultural landscapes using disturbance parameters. Agriculture Ecosystems & Environment,* **122,** 446-452.

Burel, F. (1996) *Hedgerows and their role in agricultural landscapes. Critical Reviews in Plant Sciences,* **15,** 169-190.

Burel, F., Lavigne, C., Marshall, E. J. P., Moonen, A. C., Ouin, A. & Poggio, S. L. (2013) *Landscape ecology and biodiversity in agricultural landscapes*. *Agriculture Ecosystems & Environment,* **166,** 1-2.

Burnham, K. P. & Anderson, D. R. (2002) *Model selection and multimodel inference. A practical information-theoretic approach* Springer, New York.

Buyantuyev, A. & Wu, J. (2007) *Effects of thematic resolution on landscape pattern analysis*. *Landscape Ecology,* **22,** 7-13.

Commission of the European Communities. (2009) Composite Report on the Conservation Status of Habitat Types and Species as required under Article 17 of the Habitats Directive.

Concepcion, E. D., Diaz, M., Kleijn, D., Baldi, A., Batary, P., Clough, Y., Gabriel, D., Herzog, F., Holzschuh, A., Knop, E., Marshall, E. J. P., Tscharntke, T. & Verhulst, J. (2012) *Interactive effects of landscape context constrain the effectiveness of local agri-environmental management*. *Journal of Applied Ecology,* **49,** 695-705.

Corry, R. C. & Lafortezza, R. (2007) *Sensitivity of landscape measurements to changing grain size for fine-scale design and management*. *Landscape and Ecological Engineering,* **3,** 47-53.

Cousins, S. A. O. (2006) *Plant species richness in midfield islets and road verges - The effect of landscape fragmentation*. *Biological Conservation,* **127,** 500-509.

Cousins, S. A. O. & Lindborg, R. (2008) *Remnant grassland habitats as source communities for plant diversification in agricultural landscapes*. *Biological Conservation,* **141,** 233-240.

de Blois, S., Domon, G. & Bouchard, A. (2002) *Factors affecting plant species distribution in hedgerows of southern Quebec*. *Biological Conservation,* **105,** 355-367.

de Heer, M., Kapos, V. & ten Brink, B. J. E. (2005) *Biodiversity trends in Europe: development and testing of a species trend indicator for evaluating progress towards the 2010 target*. *Philosophical Transactions of the Royal Society B-Biological Sciences,* **360,** 297-308.

Deutschewitz, K., Lausch, A., Kuhn, I. & Klotz, S. (2003) *Native and alien plant species richness in relation to spatial heterogeneity on a regional scale in Germany*. *Global Ecology and Biogeography,* **12,** 299-311.

Dramstad, W. E. (2009) *Spatial metrics - useful indicators for society or mainly fun tools for landscape ecologists? Norsk Geografisk Tidsskrift-Norwegian Journal of Geography,* **63,** 246-254.

Duelli, P. & Obrist, M. K. (2003) *Regional biodiversity in an agricultural landscape: the contribution of seminatural habitat islands*. *Basic and Applied Ecology,* **4,** 129-138.

Duflot, R., Georges, R., Ernoult, A., Aviron, S. & Burel, F. (2014) *Landscape heterogeneity as an ecological filter of species traits*. *Acta Oecologica-International Journal of Ecology,* **56,** 19-26.

Edvardsen, A., Halvorsen, R., Norderhaug, A., Pedersen, O. & Rydgren, K. (2010) *Habitat specificity of patches in modern agricultural landscapes*. *Landscape Ecology,* **25,** 1071-1083.

Ekstam, U. & Forshed, N. (1992) *If Grassland Management Ceases: Vascular Plants as Indicator Species in Meadows and Pastures*. Naturvårdsverket, Solna, SE.

Ernoult, A., Vialatte, A., Butet, A., Michel, N., Rantier, Y., Jambon, O. & Burel, F. (2013) *Grassy strips in their landscape context, their role as new habitat for biodiversity*. *Agriculture Ecosystems & Environment,* **166,** 15-27.

Fox, J. (2003) *Effect Displays in R for Generalised Linear Models*. *Journal of Statistical Software,* **8,** 1-27.

Gabriel, D., Thies, C. & Tscharntke, T. (2005) *Local diversity of arable weeds increases with landscape complexity*. *Perspect. PlantEcol. Evol. Syst.,* **7,** 85-93.

Germundsson, T. & Schlyter, P. (Eds.) (1999) Atlas över Skåne. Sveriges nationalatlas förlag, Stockholm, SE.

Grashof-Bokdam, C. J., Chardon, J. P., Vos, C. C., Foppen, R. P. B., WallisDeVries, M., van der Veen, M. & Meeuwsen, H. A. M. (2009) *The synergistic effect of combining woodlands and green veining for biodiversity*. *Landscape Ecology,* **24,** 1105-1121.

Hamre, L. N., Halvorsen, R., Edvardsen, A. & Rydgren, K. (2010) *Plant species richness, composition and habitat specificity in a Norwegian agricultural landscape*. *Agriculture Ecosystems & Environment,* **138,** 189-196.

Herzon, I. & Helenius, J. (2008) *Agricultural drainage ditches, their biological importance and functioning*. *Biological Conservation,* **141,** 1171-1183.

Hofer, G., Wagner, H. H., Herzog, F. & Edwards, P. J. (2008) *Effects of topographic variability on the scaling of plant species richness in gradient dominated landscapes*. *Ecography,* **31,** 131-139.

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 species diversity*. *Landscape and Urban Planning,* **63,** 241-250.

Janisova, M., Michalcova, D., Bacaro, G. & Ghisla, A. (2014) *Landscape effects on diversity of semi-natural grasslands*. *Agriculture Ecosystems & Environment,* **182,** 47-58.

Jonsell, B. (2004) *Flora Nordica. General Volume*. The Royal Swedish Academy of Sciences, Stockholm.

Jordbruksverket (2009) Uppdatering av blockdatabasen med stöd av satellitdata. Rapport 2009:3.

Karlsson, T. & Agestam, M. Checklista över Nordens kärlväxter. Access date: 2014-07-05. http://www.euphrasia.nu/checklista

Kleijn, D., Baquero, R. A., Clough, Y., Diaz, M., De Esteban, J., Fernandez, F., Gabriel, D., Herzog, F., Holzschuh, A., Johl, R., Knop, E., Kruess, A., Marshall, E. J. P., Steffan-Dewenter, I., Tscharntke, T., Verhulst, J., West, T. M. & Yela, J. L. (2006) *Mixed biodiversity benefits of agri-environment schemes in five European countries*. *Ecology Letters,* **9,** 243-254.

Kleijn, D., Kohler, F., Baldi, A., Batary, P., Concepcion, E. D., Clough, Y., Diaz, M., Gabriel, D., Holzschuh, A., Knop, E., Kovacs, A., Marshall, E. J. P., Tscharntke, T. & Verhulst, J. (2009) *On the relationship between farmland biodiversity and land-use intensity in Europe*. *Proceedings of the Royal Society B-Biological Sciences,* **276,** 903-909.

Kleijn, D. & Sutherland, W. J. (2003) *How effective are European agri-environment schemes in conserving and promoting biodiversity? Journal of Applied Ecology,* **40,** 947-969.

Koellner, T. & Scholz, R. W. (2008) *Assessment of land use impacts on the natural environment - Part 2: Generic characterization factors for local species diversity in central Europe*. *International Journal of Life Cycle Assessment,* **13,** 32-48.

Krebs, J. R., Wilson, J. D., Bradbury, R. B. & Siriwardena, G. M. (1999) *The second silent spring? Nature,* **400,** 611-612.

Langford, W. T., Gergel, S. E., Dietterich, T. G. & Cohen, W. (2006) *Map misclassification can cause large errors in landscape pattern indices: Examples from habitat fragmentation*. *Ecosystems,* **9,** 474-488.

Lindborg, R., Plue, J., Andersson, K. & Cousins, S. A. O. (2014) *Function of small habitat elements for enhancing plant diversity in different agricultural landscapes*. *Biological Conservation,* **169,** 206-213.

Loos, J., Turtureanu, P. D., von Wehrden, H., Hanspach, J., Dorresteijn, M., Frink, J. P. & Fischer, J. (2015) *Plant diversity in a changing agricultural landscape mosaic in Southern Transylvania (Romania)*. *Agriculture Ecosystems & Environment,* **199,** 350-357.

Luoto, M. (2000) *Modelling of rare plant species richness by landscape variables in an agriculture area in Finland*. *Plant Ecology,* **149,** 157-168.

Ma, M., Hietala, R., Kuussaari, M. & Helenius, J. (2013) *Impacts of edge density of field patches on plant species richness and community turnover among margin habitats in agricultural landscapes*. *Ecological Indicators,* **31,** 25-34.

Mathworks http://www.mathworks.se/.

McGarigal, K., Cushman, S. A. & Ene, E. (2012) FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. http://www.umass.edu/landeco/research/fragstats/fragstats.html.

Moeslund, J. E., Arge, L., Bocher, P. K., Dalgaard, T. & Svenning, J.-C. (2013) *Topography as a driver of local terrestrial vascular plant diversity patterns*. *Nordic Journal of Botany,* **31,** 129-144.

Persson, A. S., Olsson, O., Rundlof, M. & Smith, H. G. (2010) *Land use intensity and landscape complexity-Analysis of landscape characteristics in an agricultural region in Southern Sweden*. *Agriculture Ecosystems & Environment,* **136,** 169-176.

Persson, H. & Tyler, T. (2007) Den skånska florans geologiska förutsättningar. In *Floran i Skåne. Arterna och deras utbredning* (Eds Olsson. K.-A. Tyler T., Johansson H., Sonesson M). Lunds Botaniska Förening, Lund.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & Team, R. D. C. (2011) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-98, http://CRAN.R-project.org/package=nlme.

. (ed^(eds.

Rogerson, P. A. (2001) *Statistical methods for geography*. SAGE Publications, London.

Simova, P. & Gdulova, K. (2012) *Landscape indices behavior: A review of scale effects*. *Applied Geography,* **34,** 385-394.

Smart, S. M., Marrs, R. H., Le Duc, M. G., Thompson, K., Bunce, R. G. H., Firbank, L. G. & Rossall, M. J. (2006) *Spatial relationships between intensive land cover and residual plant species diversity in temperate farmed landscapes*. *Journal of Applied Ecology,* **43,** 1128-1137.

StatisticsSweden (2014) *Yearbook of agricultural statistics 2014*. SCB-Tryck, Örebro.

Stein, A., Gerstner, K. & Kreft, H. (2014) *Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales*. *Ecology Letters,* **17,** 866-880.

Stoate, C., Baldi, A., Beja, P., Boatman, N. D., Herzon, I., van Doorn, A., de Snoo, G. R., Rakosy, L. & Ramwell, C. (2009) *Ecological impacts of early 21st century agricultural change in Europe - A review*. *Journal of Environmental Management,* **91,** 22-46.

Team, R. C. (2014) R: A Language and Environment for Statistical Computing. (ed^(eds. R Foundation for Statistical Computing, Vienna, Austria.

Tikka, P. M., Hogmander, H. & Koski, P. S. (2001) *Road and railway verges serve as dispersal corridors for grassland plants*. *Landscape Ecology,* **16,** 659-666.

Tscharntke, T., Klein, A. M., Kruess, A., Steffan-Dewenter, I. & Thies, C. (2005) *Landscape perspectives on agricultural intensification and biodiversity - ecosystem service management*. *Ecology Letters,* **8,** 857-874.

Tyler, G. (2007b) Markegenskapernas betydelse för arternas fördelning. In *Floran i Skåne. Arterna och deras utbredning* (eds Tyler T., Olsson, K.-A., Johansson H., Sonesson M). Lunds Botaniska Förening, Lund, Sweden.

Tyler, T., Olsson, K.-A., Johansson, H. & Sonesson, M. (Eds) (2007a) *Floran i Skåne. Arterna och deras utbredning.* Lunds Botaniska Förening, Lund.

Uuemaa, E., Mander, U. & Marja, R. (2013) *Trends in the use of landscape spatial metrics as landscape indicators: A review*. *Ecological Indicators,* **28,** 100-106.

van Dijk, W. F. A., van Ruijven, J., Berendse, F. & de Snoo, G. R. (2014) *The effectiveness of ditch banks as dispersal corridor for plants in agricultural landscapes depends on species' dispersal traits*. *Biological Conservation,* **171,** 91-98.

Weimarck, G. & Andersson, P.-A. (2007) Skånes floristiska växtgeografi. In *Skånes Flora. Arterna och deras utbredning* (Eds Tyler T., Olsson. K.-A., Johansson H., Sonesson M). Lunds Botaniska Förning, Lund.

Wuczynski, A., Dajdok, Z., Wierzcholska, S. & Kujawa, K. (2014) *Applying red lists to the evaluation of agricultural habitat: regular occurrence of threatened birds, vascular plants, and bryophytes in field margins of Poland*. *Biodiversity and Conservation,* **23,** 999-1017.

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

# Appendix A: Original land cover classes and reclassification details

| Original class | After reclassification | Original class | After reclassification |
|---|---|---|---|
| 141 | Built-up areas | 3111 | Deciduous forests |
| 1422 | Built-up areas | 3112 | Deciduous forests |
| 1424 | Built-up areas | 3113 | Deciduous forests |
| 1425 | Built-up areas | 3131 | Deciduous forests |
| 1426 | Built-up areas | 3132 | Deciduous forests |
| 111 | Built-up areas | 3133 | Deciduous forests |
| 121 | Built-up areas | 3241 | Deciduous forests |
| 122 | Built-up areas | 211 | Cultivated |
| 123 | Built-up areas | 222 | Cultivated |
| 124 | Built-up areas | 231 | Semi-natural grasslands |
| 132 | Built-up areas | 321 | Semi-natural grasslands |
| 133 | Built-up areas | 322 | Semi-natural grasslands |
| 1122 | Built-up areas | 511 | Water bodies |
| 1123 | Built-up areas | 521 | Water bodies |
| 1421 | Built-up areas | 522 | Water bodies |
| 11211 | Built-up areas | 5121 | Water bodies |
| 11212 | Built-up areas | 5122 | Water bodies |
| 3122 | Coniferous forests | 5231 | Water bodies |
| 3123 | Coniferous forests | 5232 | Water bodies |
| 3242 | Coniferous forests | 331 | Sparsely vegetated areas |
| 3243 | Coniferous forests | 332 | Sparsely vegetated areas |
| 312121 | Coniferous forests | 1311 | Sparsely vegetated areas |
| 312122 | Coniferous forests | 1312 | Sparsely vegetated areas |
| | | 411 | Wetlands |
| | | 421 | Wetlands |
| | | 4121 | Wetlands |
| | | 4122 | Wetlands |
| | | 4123 | Wetlands |

# Appendix B: Original code in the IACS and reclassification details for all farmland

| **Code** | ANNU | PAST | LEY | LIM | FALL | OT |
|---|---|---|---|---|---|---|
| 1 | 1 | | | | | |
| 2 | 1 | | | | | |
| 3 | 1 | | | | | |
| 4 | 1 | | | | | |
| 5 | 1 | | | | | |
| 7 | 1 | | | | | |
| 8 | 1 | | | | | |
| 9 | 1 | | | | | |
| 11 | 1 | | | | | |
| 12 | 1 | | | | | |
| 13 | 1 | | | | | |
| 14 | 1 | | | | | |
| 15 | 1 | | | | | |
| 16 | 1 | | | | | |
| 20 | 1 | | | | | |
| 21 | 1 | | | | | |
| 22 | 1 | | | | | |
| 23 | 1 | | | | | |
| 24 | 1 | | | | | |
| 25 | 1 | | | | | |
| 26 | 1 | | | | | |
| 27 | 1 | | | | | |
| 30 | 1 | | | | | |
| 31 | 1 | | | | | |
| 32 | 1 | | | | | |
| 33 | 1 | | | | | |
| 36 | 1 | | | | | |
| 37 | 1 | | | | | |
| 40 | 1 | | | | | |
| 41 | 1 | | | | | |
| 45 | 1 | | | | | |
| 46 | 1 | | | | | |
| 47 | 1 | | | | | |
| 48 | 1 | | | | | |
| 49 | | | 1 | | | |
| 50 | | | 1 | | | |
| 51 | | | 1 | | | |
| 52 | | 1 | | | | |
| 53 | | 1 | | | | |
| 54 | | 1 | | | | |
| 60 | | | | | 1 | |
| 64 | | | | | | 1 |
| 65 | | | | | | 1 |
| 67 | | | | | | 1 |
| 68 | | | | | | 1 |
| 70 | | | | | | 1 |
| 71 | | | | | | 1 |
| 72 | | | | | | 1 |
| 73 | 1 | | | | | |
| 74 | 1 | | | | | |
| 75 | | | | 1 | | |
| 76 | | | | 1 | | |
| 77 | | | | 1 | | |
| 80 | 1 | | | | | |
| 81 | 1 | | | | | |
| 83 | | | | | | 1 |
| 84 | | | | | | 1 |
| 89 | | 1 | | | | |
| 99 | | | | | | 1 |

# Appendix C: The original bedrock categories with reclassification details.

| Class | UB | IG | SED | CALC |
|---|---|---|---|---|
| Basisk bergart | 1 | | | |
| Gabbro | 1 | | | |
| Kvartsmonzodiorit | 1 | | | |
| Amfibolit | 1 | | | |
| Grönsten | 1 | | | |
| Amfibolit | 1 | | | |
| Grönsten | 1 | | | |
| Ultramafit | 1 | | | |
| Basalt | 1 | | | |
| Amfibolit | 1 | | | |
| Grönsten | 1 | | | |
| Metaandesit | 1 | | | |
| Diabas | 1 | | | |
| Gnejs | | 1 | | |
| Kvartsdominerad hydrot. gång/segr. | | 1 | | |
| Silicifierad bergart | | 1 | | |
| SEDIMENTÄR BERGART | | | 1 | |
| Konglomerat | | | 1 | |
| Sandsten | | | 1 | |
| Kvartsarenit | | 1 | | |
| Arkosisk arenit | | | 1 | |
| Kalkförande sandsten | | | | 1 |
| Siltsten | | | 1 | |
| Slamsten | | | 1 | |
| Lerskiffer | | | 1 | |
| Kalksten | | | | 1 |
| Alunskiffer | | | 1 | |
| Märgelsten | | | | 1 |
| Sand | | | 1 | |
| Silt | | | 1 | |
| Lersten | | | 1 | |
| Lera | | | 1 | |
| Märgel | | | | 1 |
| METASEDIMENTÄR BERGART | | | 1 | |
| Metaarenit | | | 1 | |
| Kvartsit | | 1 | | |
| Glimmerkvartsit | | 1 | | |
| Paragnejs | | 1 | | |
| Kvartsit | | 1 | | |
| GNEJS | | 1 | | |
| INTRUSIV BERGART | | 1 | | |
| Syenitoid | | 1 | | |
| Monzonit/kvartsmonzonit | | 1 | | |
| Monzonit | | 1 | | |
| Kvartsmonzonit | | 1 | | |
| Syenit | | 1 | | |
| Tonalit | | 1 | | |
| Granodiorit | | 1 | | |
| Granit | | 1 | | |
| Pegmatit | | 1 | | |
| Aplit | | 1 | | |
| Gnejsgranodiorit | | 1 | | |
| Gnejsgranit | | 1 | | |
| Metamonzonit | | 1 | | |
| Ortognejs | | 1 | | |
| Felsisk bergart | | 1 | | |
| Dacit/ryolit | | 1 | | |
| Metadacit | | 1 | | |
| Metaryolit | | 1 | | |
| Metaandesit/-dacit | | 1 | | |
| Silt | | | 1 | |
| Lera till silt | | | 1 | |
| Lera till sand | | | 1 | |
| Lera | | | 1 | |
| Felsisk bergart till amfibolit | | 1 | | |
| Lera till siltsten | | | 1 | |
| Märgel till märgelsten | | | | 1 |
| Märgel till sand | | | | 1 |

# Appendix D: Pearson's correlation coefficients between the different variables within each of the three regions. Variables which were excluded from the global models are shaded in grey and all variables with a correlation of 0.7 or more are bold.

**Plain districts**

| | GR | SR | SNG_tot | DECID | WATER | WETL | SPARSE | BUILT | CONIF | SIDI | ANNU | LEY | RIP | FALL | SB | FARM | IG | CALC | SED | UB | TOPO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **GR** | 1,00 | 0,89 | 0,68 | 0,51 | 0,23 | 0,20 | 0,25 | 0,05 | 0,43 | 0,60 | -0,55 | 0,33 | -0,27 | -0,04 | -0,06 | -0,44 | 0,44 | -0,24 | -0,05 | 0,23 | 0,38 |
| **SR** | 0,89 | 1,00 | 0,58 | 0,46 | 0,26 | 0,18 | 0,26 | 0,24 | 0,35 | 0,60 | -0,52 | 0,24 | -0,20 | -0,03 | 0,04 | -0,46 | 0,39 | -0,14 | -0,12 | 0,16 | 0,31 |
| **SNG_tot** | 0,68 | 0,58 | 1,00 | 0,46 | 0,27 | 0,32 | 0,19 | 0,09 | 0,39 | **0,71** | -0,61 | 0,36 | -0,25 | -0,09 | -0,01 | -0,44 | 0,39 | -0,15 | -0,12 | 0,24 | 0,26 |
| **DECID** | 0,51 | 0,46 | 0,46 | 1,00 | 0,29 | 0,30 | 0,13 | -0,13 | **0,80** | **0,77** | **-0,85** | 0,28 | -0,41 | -0,28 | -0,38 | **-0,83** | 0,62 | -0,24 | -0,20 | 0,50 | 0,52 |
| **WATER** | 0,23 | 0,26 | 0,27 | 0,29 | 1,00 | 0,38 | -0,03 | -0,06 | 0,22 | 0,39 | -0,36 | 0,06 | -0,21 | -0,22 | -0,34 | -0,36 | 0,15 | 0,13 | -0,22 | 0,03 | 0,04 |
| **WETL** | 0,20 | 0,18 | 0,32 | 0,30 | 0,38 | 1,00 | -0,07 | -0,13 | 0,30 | 0,34 | -0,37 | 0,03 | -0,21 | -0,16 | -0,28 | -0,34 | 0,18 | 0,11 | -0,24 | 0,12 | -0,04 |
| **SPARSE** | 0,25 | 0,26 | 0,19 | 0,13 | -0,03 | -0,07 | 1,00 | 0,09 | 0,10 | 0,22 | -0,18 | 0,09 | -0,01 | -0,04 | 0,11 | -0,17 | 0,26 | -0,15 | -0,01 | 0,09 | 0,22 |
| **BUILT** | 0,05 | 0,24 | 0,09 | -0,13 | -0,06 | -0,13 | 0,09 | 1,00 | -0,20 | 0,30 | -0,10 | -0,11 | -0,03 | 0,05 | 0,54 | -0,19 | -0,05 | 0,03 | 0,01 | -0,10 | 0,00 |
| **CONIF** | 0,43 | 0,35 | 0,39 | **0,80** | 0,22 | 0,30 | 0,10 | -0,20 | 1,00 | **0,71** | **-0,83** | 0,20 | -0,43 | -0,31 | -0,44 | **-0,84** | 0,62 | -0,28 | -0,18 | 0,58 | 0,33 |
| **SIDI** | 0,60 | 0,60 | **0,71** | **0,77** | 0,39 | 0,34 | 0,22 | 0,30 | **0,71** | 1,00 | **-0,89** | 0,31 | -0,40 | -0,21 | -0,09 | **-0,86** | 0,53 | -0,20 | -0,18 | 0,40 | 0,36 |
| **ANNU** | -0,55 | -0,52 | -0,61 | **-0,85** | -0,36 | -0,37 | -0,18 | -0,10 | **-0,83** | **-0,89** | 1,00 | -0,39 | 0,40 | 0,27 | 0,21 | **0,94** | -0,67 | 0,24 | 0,25 | -0,54 | -0,43 |
| **LEY** | 0,33 | 0,24 | 0,36 | 0,28 | 0,06 | 0,03 | 0,09 | -0,11 | 0,20 | 0,31 | -0,39 | 1,00 | 0,00 | 0,04 | 0,11 | -0,15 | 0,26 | -0,24 | 0,05 | 0,25 | 0,20 |
| **RIP** | -0,27 | -0,20 | -0,25 | -0,41 | -0,21 | -0,21 | -0,01 | -0,03 | -0,43 | -0,40 | 0,40 | 0,00 | 1,00 | 0,30 | 0,29 | 0,44 | -0,34 | 0,18 | 0,07 | -0,23 | -0,29 |
| **FALL** | -0,04 | -0,03 | -0,09 | -0,28 | -0,22 | -0,16 | -0,04 | 0,05 | -0,31 | -0,21 | 0,27 | 0,04 | 0,30 | 1,00 | 0,33 | 0,37 | -0,36 | 0,00 | 0,26 | -0,32 | -0,19 |
| **SB** | -0,06 | 0,04 | -0,01 | -0,38 | -0,34 | -0,28 | 0,11 | 0,54 | -0,44 | -0,09 | 0,21 | 0,11 | 0,29 | 0,33 | 1,00 | 0,24 | -0,16 | 0,05 | 0,08 | -0,18 | -0,07 |
| **FARM** | -0,44 | -0,46 | -0,44 | -0,83 | -0,36 | -0,34 | -0,17 | -0,19 | **-0,84** | **-0,86** | **0,94** | -0,15 | 0,44 | 0,37 | 0,24 | 1,00 | -0,64 | 0,18 | 0,28 | -0,52 | -0,41 |
| **IG** | 0,44 | 0,39 | 0,39 | 0,62 | 0,15 | 0,18 | 0,26 | -0,05 | 0,62 | 0,53 | -0,67 | 0,26 | -0,34 | -0,36 | -0,16 | -0,64 | 1,00 | -0,34 | -0,36 | 0,67 | 0,50 |
| **CALC** | -0,24 | -0,14 | -0,15 | -0,24 | 0,13 | 0,11 | -0,15 | 0,03 | -0,28 | -0,20 | 0,24 | -0,24 | 0,18 | 0,00 | 0,05 | 0,18 | -0,34 | 1,00 | **-0,75** | -0,34 | -0,26 |
| **SED** | -0,05 | -0,12 | -0,12 | -0,20 | -0,22 | -0,24 | -0,01 | 0,01 | -0,18 | -0,18 | 0,25 | 0,05 | 0,07 | 0,26 | 0,08 | 0,28 | -0,36 | **-0,75** | 1,00 | -0,21 | -0,08 |
| **UB** | 0,23 | 0,16 | 0,24 | 0,50 | 0,03 | 0,12 | 0,09 | -0,10 | 0,58 | 0,40 | -0,54 | 0,25 | -0,23 | -0,32 | -0,18 | -0,52 | 0,67 | -0,34 | -0,21 | 1,00 | 0,38 |
| **TOPO** | 0,38 | 0,31 | 0,26 | 0,52 | 0,04 | -0,04 | 0,22 | 0,00 | 0,33 | 0,36 | -0,43 | 0,20 | -0,29 | -0,19 | -0,07 | -0,41 | 0,50 | -0,26 | -0,08 | 0,38 | 1,00 |

**Central districts**

| | GR | SR | SNG_tot | DECID | WATER | WETL | SPARSE | BUILT | CONIF | SIDI | ANNU | LEY | RIP | FALL | SB | FARM | IG | CALC | SED | UB | TOPO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **GR** | 1,00 | 0,80 | 0,46 | 0,12 | -0,02 | 0,10 | 0,07 | -0,02 | 0,10 | 0,26 | -0,20 | 0,04 | -0,11 | -0,10 | 0,07 | -0,08 | 0,13 | -0,02 | -0,10 | 0,02 | 0,07 |
| **SR** | 0,80 | 1,00 | 0,27 | 0,14 | 0,13 | 0,09 | 0,14 | 0,15 | 0,13 | 0,31 | -0,21 | -0,06 | -0,14 | -0,09 | 0,00 | -0,19 | 0,08 | 0,06 | -0,12 | 0,01 | 0,14 |
| **SNG_tot** | 0,46 | 0,27 | 1,00 | 0,01 | -0,01 | 0,17 | 0,03 | 0,00 | 0,01 | 0,35 | -0,33 | 0,22 | -0,09 | -0,16 | 0,08 | 0,02 | 0,05 | 0,10 | -0,13 | 0,00 | -0,12 |
| **DECID** | 0,12 | 0,14 | 0,01 | 1,00 | 0,11 | 0,14 | 0,05 | -0,26 | 0,56 | 0,62 | -0,69 | -0,29 | -0,37 | -0,39 | -0,48 | **-0,75** | 0,55 | -0,23 | -0,33 | 0,48 | 0,34 |
| **WATER** | -0,02 | 0,13 | -0,01 | 0,11 | 1,00 | 0,34 | -0,06 | -0,03 | 0,02 | 0,28 | -0,24 | -0,23 | -0,21 | -0,17 | -0,37 | -0,30 | -0,12 | 0,14 | -0,01 | -0,11 | -0,11 |
| **WETL** | 0,10 | 0,09 | 0,17 | 0,14 | 0,34 | 1,00 | -0,10 | -0,08 | 0,12 | 0,21 | -0,26 | -0,15 | -0,14 | -0,22 | -0,27 | -0,21 | -0,06 | 0,23 | -0,17 | -0,07 | -0,20 |
| **SPARSE** | 0,07 | 0,14 | 0,03 | 0,05 | -0,06 | -0,10 | 1,00 | -0,03 | 0,04 | 0,10 | -0,04 | -0,01 | -0,06 | 0,06 | 0,00 | -0,04 | 0,05 | -0,04 | 0,00 | 0,06 | 0,28 |
| **BUILT** | -0,02 | 0,15 | 0,00 | -0,26 | -0,03 | -0,08 | -0,03 | 1,00 | -0,12 | 0,15 | 0,03 | -0,07 | 0,01 | 0,06 | 0,27 | -0,04 | -0,10 | 0,16 | -0,04 | -0,05 | -0,07 |
| **CONIF** | 0,10 | 0,13 | 0,01 | 0,56 | 0,02 | 0,12 | 0,04 | -0,12 | 1,00 | 0,62 | **-0,74** | -0,44 | -0,30 | -0,39 | -0,59 | **-0,83** | 0,39 | -0,01 | -0,38 | 0,28 | 0,16 |
| **SIDI** | 0,26 | 0,31 | 0,35 | 0,62 | 0,28 | 0,21 | 0,10 | 0,15 | 0,62 | 1,00 | **-0,70** | -0,27 | -0,30 | -0,27 | -0,37 | **-0,73** | 0,37 | 0,09 | -0,44 | 0,29 | 0,17 |
| **ANNU** | -0,20 | -0,21 | -0,33 | -0,69 | -0,24 | -0,26 | -0,04 | 0,03 | **-0,74** | **-0,70** | 1,00 | 0,14 | 0,45 | 0,48 | 0,50 | **0,85** | -0,44 | 0,07 | 0,38 | -0,34 | -0,16 |
| **LEY** | 0,04 | -0,06 | 0,22 | -0,29 | -0,23 | -0,15 | -0,01 | -0,07 | -0,44 | -0,27 | 0,14 | 1,00 | 0,14 | 0,09 | 0,53 | 0,53 | -0,07 | -0,18 | 0,23 | -0,02 | -0,15 |
| **RIP** | -0,11 | -0,14 | -0,09 | -0,37 | -0,21 | -0,14 | -0,06 | 0,01 | -0,30 | -0,30 | 0,45 | 0,14 | 1,00 | 0,27 | 0,25 | 0,40 | -0,20 | 0,05 | 0,14 | -0,14 | -0,10 |
| **FALL** | -0,10 | -0,09 | -0,16 | -0,39 | -0,17 | -0,22 | 0,06 | 0,06 | -0,39 | -0,27 | 0,48 | 0,09 | 0,27 | 1,00 | 0,32 | 0,47 | -0,35 | 0,21 | 0,15 | -0,33 | 0,02 |
| **SB** | 0,07 | 0,00 | 0,08 | -0,48 | -0,37 | -0,27 | 0,00 | 0,27 | -0,59 | -0,37 | 0,50 | 0,53 | 0,25 | 0,32 | 1,00 | 0,62 | -0,09 | -0,15 | 0,24 | -0,05 | -0,04 |
| **FARM** | -0,08 | -0,19 | 0,02 | **-0,75** | -0,30 | -0,21 | -0,04 | -0,04 | **-0,83** | **-0,73** | **0,85** | 0,53 | 0,40 | 0,47 | 0,62 | 1,00 | -0,43 | 0,02 | 0,41 | -0,36 | -0,23 |
| **IG** | 0,13 | 0,08 | 0,05 | 0,55 | -0,12 | -0,06 | 0,05 | -0,10 | 0,39 | 0,37 | -0,44 | -0,07 | -0,20 | -0,35 | -0,09 | -0,43 | 1,00 | -0,48 | -0,53 | 0,62 | 0,37 |
| **CALC** | -0,02 | 0,06 | 0,10 | -0,23 | 0,14 | 0,23 | -0,04 | 0,16 | -0,01 | 0,09 | 0,07 | -0,18 | 0,05 | 0,21 | -0,15 | 0,02 | -0,48 | 1,00 | -0,48 | -0,34 | -0,23 |
| **SED** | -0,10 | -0,12 | -0,13 | -0,33 | -0,01 | -0,17 | 0,00 | -0,04 | -0,38 | -0,44 | 0,38 | 0,23 | 0,14 | 0,15 | 0,24 | 0,41 | -0,53 | -0,48 | 1,00 | -0,32 | -0,13 |
| **UB** | 0,02 | 0,01 | 0,00 | 0,48 | -0,11 | -0,07 | 0,06 | -0,05 | 0,28 | 0,29 | -0,34 | -0,02 | -0,14 | -0,33 | -0,05 | -0,36 | 0,62 | -0,34 | -0,32 | 1,00 | 0,25 |
| **TOPO** | 0,07 | 0,14 | -0,12 | 0,34 | -0,11 | -0,20 | 0,28 | -0,07 | 0,16 | 0,17 | -0,16 | -0,15 | -0,10 | 0,02 | -0,04 | -0,23 | 0,37 | -0,23 | -0,13 | 0,25 | 1,00 |

**Forest districts**

| | GR | SR | SNG_tot | DECID | WATER | WETL | SPARSE | BUILT | CONIF | SIDI | ANNU | LEY | RIP | FALL | SB | FARM | IG | CALC | SED | UB | TOPO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **GR** | 1,00 | 0,90 | 0,41 | 0,18 | -0,08 | -0,34 | 0,08 | 0,28 | -0,39 | 0,35 | 0,24 | 0,29 | 0,11 | 0,29 | 0,42 | 0,33 | -0,15 | 0,10 | 0,13 | 0,09 | 0,19 |
| **SR** | 0,90 | 1,00 | 0,35 | 0,24 | -0,05 | -0,38 | 0,09 | 0,43 | -0,49 | 0,41 | 0,30 | 0,28 | 0,13 | 0,33 | 0,46 | 0,33 | -0,21 | 0,07 | 0,16 | 0,16 | 0,25 |
| **SNG_tot** | 0,41 | 0,35 | 1,00 | 0,04 | -0,28 | -0,26 | -0,01 | 0,08 | -0,54 | 0,47 | 0,40 | 0,60 | 0,15 | 0,38 | 0,58 | 0,68 | -0,20 | -0,06 | 0,22 | 0,00 | -0,13 |
| **DECID** | 0,18 | 0,24 | 0,04 | 1,00 | -0,06 | -0,30 | -0,01 | -0,05 | -0,36 | 0,41 | -0,15 | -0,07 | -0,10 | -0,09 | -0,07 | -0,14 | -0,03 | 0,10 | -0,14 | 0,46 | 0,45 |
| **WATER** | -0,08 | -0,05 | -0,28 | -0,06 | 1,00 | 0,18 | -0,04 | -0,05 | 0,11 | 0,07 | -0,26 | -0,30 | -0,10 | -0,22 | -0,27 | -0,31 | 0,16 | 0,08 | -0,16 | -0,02 | 0,02 |
| **WETL** | -0,34 | -0,38 | -0,26 | -0,30 | 0,18 | 1,00 | -0,05 | -0,18 | 0,35 | -0,07 | -0,34 | -0,30 | -0,18 | -0,32 | -0,48 | -0,35 | 0,26 | -0,06 | -0,23 | -0,12 | -0,30 |
| **SPARSE** | 0,08 | 0,09 | -0,01 | -0,01 | -0,04 | -0,05 | 1,00 | 0,03 | 0,03 | 0,01 | -0,01 | -0,05 | 0,00 | -0,01 | 0,04 | -0,05 | 0,05 | 0,13 | -0,08 | 0,05 | 0,12 |
| **BUILT** | 0,28 | 0,43 | 0,08 | -0,05 | -0,05 | -0,18 | 0,03 | 1,00 | -0,23 | 0,31 | 0,03 | 0,01 | -0,01 | 0,03 | 0,37 | 0,02 | -0,02 | -0,03 | 0,01 | 0,08 | 0,04 |
| **CONIF** | -0,39 | -0,49 | -0,54 | -0,36 | 0,11 | 0,35 | 0,03 | -0,23 | 1,00 | -0,52 | **-0,71** | -0,64 | -0,35 | -0,58 | -0,61 | **-0,75** | 0,49 | -0,07 | -0,42 | -0,23 | -0,19 |
| **SIDI** | 0,35 | 0,41 | 0,47 | 0,41 | 0,07 | -0,07 | 0,01 | 0,31 | -0,52 | 1,00 | 0,12 | 0,39 | 0,02 | 0,14 | 0,38 | 0,27 | -0,05 | 0,03 | -0,03 | 0,26 | 0,11 |
| **ANNU** | 0,24 | 0,30 | 0,40 | -0,15 | -0,26 | -0,34 | -0,01 | 0,03 | **-0,71** | 0,12 | 1,00 | 0,63 | 0,46 | **0,76** | 0,60 | **0,87** | -0,58 | 0,01 | 0,62 | -0,03 | 0,05 |
| **LEY** | 0,29 | 0,28 | 0,60 | -0,07 | -0,30 | -0,30 | -0,05 | 0,01 | -0,64 | 0,39 | 0,63 | 1,00 | 0,39 | 0,48 | **0,76** | **0,88** | -0,32 | -0,08 | 0,31 | 0,05 | -0,13 |
| **RIP** | 0,11 | 0,13 | 0,15 | -0,10 | -0,10 | -0,18 | 0,00 | -0,01 | -0,35 | 0,02 | 0,46 | 0,39 | 1,00 | 0,31 | 0,34 | 0,45 | -0,28 | 0,06 | 0,26 | 0,05 | 0,03 |
| **FALL** | 0,29 | 0,33 | 0,38 | -0,09 | -0,22 | -0,32 | -0,01 | 0,03 | -0,58 | 0,14 | **0,76** | 0,48 | 0,31 | 1,00 | 0,46 | **0,71** | -0,54 | 0,04 | 0,56 | -0,01 | 0,04 |
| **SB** | 0,42 | 0,46 | 0,58 | -0,07 | -0,27 | -0,48 | 0,04 | 0,37 | -0,61 | 0,38 | 0,60 | **0,76** | 0,34 | 0,46 | 1,00 | **0,75** | -0,30 | -0,05 | 0,30 | 0,05 | 0,02 |
| **FARM** | 0,33 | 0,33 | 0,68 | -0,14 | -0,31 | -0,35 | -0,05 | 0,02 | **-0,75** | 0,27 | **0,87** | **0,88** | 0,45 | **0,71** | **0,75** | 1,00 | -0,52 | -0,03 | 0,54 | -0,02 | -0,07 |
| **IG** | -0,15 | -0,21 | -0,20 | -0,03 | 0,16 | 0,26 | 0,05 | -0,02 | 0,49 | -0,05 | -0,58 | -0,32 | -0,28 | -0,54 | -0,30 | -0,52 | 1,00 | 0,00 | **-0,92** | -0,27 | -0,06 |
| **CALC** | 0,10 | 0,07 | -0,06 | 0,10 | 0,08 | -0,06 | 0,13 | -0,03 | -0,07 | 0,03 | 0,01 | -0,08 | 0,06 | 0,04 | -0,05 | -0,03 | 0,00 | 1,00 | -0,03 | -0,07 | 0,26 |
| **SED** | 0,13 | 0,16 | 0,22 | -0,14 | -0,16 | -0,23 | -0,08 | 0,01 | -0,42 | -0,03 | 0,62 | 0,31 | 0,26 | 0,56 | 0,30 | 0,54 | **-0,92** | -0,03 | 1,00 | -0,12 | 0,00 |
| **UB** | 0,09 | 0,16 | 0,00 | 0,46 | -0,02 | -0,12 | 0,05 | 0,08 | -0,23 | 0,26 | -0,03 | 0,05 | 0,05 | -0,01 | 0,05 | -0,02 | -0,27 | -0,07 | -0,12 | 1,00 | 0,18 |
| **TOPO** | 0,19 | 0,25 | -0,13 | 0,45 | 0,02 | -0,30 | 0,12 | 0,04 | -0,19 | 0,11 | 0,05 | -0,13 | 0,03 | 0,04 | 0,02 | -0,07 | -0,06 | 0,26 | 0,00 | 0,18 | 1,00 |

# Paper IV

Ecological Informatics 18 (2013) 1–9

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

# Assessment of fine-scale plant species beta diversity using WorldView-2 satellite spectral dissimilarity

Jonas Dalmayne [a], Thomas Möckel [a], Honor C. Prentice [b], Barbara C. Schmid [b], Karin Hall [a],*

[a] *Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden*
[b] *Department of Biology, Lund University, Sölvegatan 37, SE-223 62 Lund, Sweden*

ARTICLE INFO

*Article history:*
Received 25 March 2013
Received in revised form 10 May 2013
Accepted 13 May 2013
Available online 22 May 2013

*Keywords:*
High-spatial resolution
Spectral distance
Habitat heterogeneity
Species turnover
Plot size

ABSTRACT

Plant species beta diversity is influenced by spatial heterogeneity in the environment. This heterogeneity can potentially be characterised with the help of remote sensing. We used WorldView-2 satellite data acquired over semi-natural grasslands on The Baltic island of Öland (Sweden) to examine whether dissimilarities in remote sensing response were related to fine-scale, between-plot dissimilarity (beta diversity) in non-woody vascular plant species composition within the grasslands. Fieldwork, including the on-site description of a set of 30 2 m × 2 m plots and a set of 30 4 m × 4 m plots, was performed to record the species dissimilarity between pairs of same-sized plots. Spectral data were extracted by associating each plot with a suite of differently sized pixel windows, and spectral dissimilarity was calculated between pairs of same-sized pixel windows. Relationships between spectral dissimilarity and beta diversity were analysed using univariate regression and partial least squares regression. The study revealed significant positive relationships between spectral dissimilarity and fine-scale (2 m × 2 m and 4 m × 4 m) between-plot species dissimilarity. The correlation between the predicted and the observed species dissimilarity was stronger for the set of large plots (4 m × 4 m) than for the set of small plots (2 m × 2 m), and the association between spectral and species data at both plot scales decreased when pixel windows larger than 3 × 3 pixels were used. We suggest that the significant relationship between spectral dissimilarity and species dissimilarity is a reflection of between-plot environmental heterogeneity caused by differences in grazing intensity (which result in between-plot differences in field-layer height, and amounts of biomass and litter). This heterogeneity is reflected in dissimilarities in both the species composition and the spectral response of the grassland plots. Between-plot dissimilarities in both spectral response and species composition may also be caused by between-plot variations in edaphic conditions. Our results indicate that high spatial resolution satellite data may potentially be able to complement field-based recording in surveys of fine-scale species diversity in semi-natural grasslands.

## 1. Introduction

The maintenance of diversity in plant communities has been an important question in ecosystem studies for many years (e.g. Connell, 1978; Hutchinson, 1961). Semi-natural grasslands, once widespread throughout Europe (Poschlod and WallisDeVries, 2002), show high levels of species richness, particularly at smaller spatial scales (e.g. Kull and Zobel, 1991). The area of semi-natural grassland in Europe has declined dramatically over the last century with the increasing intensification of agriculture (Poschlod and WallisDeVries, 2002). The remaining fragments of old grassland are of great importance for the overall species richness in agricultural landscapes (Cousins and Lindborg, 2008) and therefore also for the preservation of ecosystem services (Tscharntke et al., 2005). Semi-natural grasslands are identified as a priority for conservation throughout Europe (cf. Janišová et al., 2011).

Levels of species diversity at the scale of individual grassland fragments may be influenced by a range of spatial, historical and ecological factors, and by interactions among these factors (e.g. Luoto et al., 2003; Raatikainen et al., 2009; Reitalu et al., 2010). In particular, fine scale habitat heterogeneity may have an influence on species diversity (e.g. Bruun, 2000; Gazol et al., 2012). Heterogeneous habitats are expected to support a greater diversity of potential niches for species than habitats with more homogenous conditions (Silvertown, 2004; Whittaker, 1972) and environmental heterogeneity has been shown to have a positive effect on species diversity in studies of grassland communities (e.g. Bruun, 2000; Öster et al., 2007). When different species are favoured by different environmental conditions, high environmental heterogeneity may lead to high dissimilarity (turnover) in species composition between plots (Kneitel and Chase, 2004; Nekola and White, 1999). Levels of turnover between sampled plots may be related to the spatial size of the plots (Keil et al., 2012; Nekola and White, 1999), and

---

* Corresponding author. Tel.: +46 46 222 9579; fax: +46 46 222 0321.
*E-mail address:* Karin.Hall@nateko.lu.se (K. Hall).

1574-9541/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.05.004

the use of smaller plots is generally assumed to result in higher estimates of between-plot turnover than the use of larger plots (Nekola and White, 1999). The amount of between-plot turnover in species composition is characterised by beta diversity (Anderson et al., 2011; Koleff et al., 2003). Levels of beta diversity offer insights into spatial and temporal patterns of biodiversity (Whittaker, 1972) and provide essential information for conservation management (e.g. McKnight et al., 2007).

Field-based inventories of species diversity for management planning may be time consuming and it has been suggested that remote sensing may have the potential to support and supplement direct inventories of species diversity (Gillespie et al., 2008; Rocchini et al., 2010a; Turner et al., 2003). For example, spectral information acquired with the help of Earth Observation satellite data have shown to be related to local-scale plant species richness (number of species per unit area) in a newly burned area in Spain (Viedma et al., 2012). The WorldView-2 satellite is a recently launched Earth Observation satellite that provides broadband spectral data (data based on optical sensors with broad spectral bands) in eight wavebands and with high spatial resolution (2 m).

There has been considerable recent progress in the development of methods for handling and analysing remote sensing data (cf. Boyd and Foody, 2011), including improvements in the methodology for examining variability within spectral responses (cf. Rocchini et al., 2010a), and improved statistical methods for estimating habitat properties (e.g. Feilhauer and Schmidtlein, 2011). Partial least squares regression (PLSR; Wold et al., 2001) is increasingly used in remote sensing and has been shown to be a useful technique for studying grassland vegetation (e.g. Chen et al., 2009; Darvishzadeh et al., 2008; Feilhauer and Schmidtlein, 2011) and in the monitoring of species diversity (Feilhauer and Schmidtlein, 2009). PLSR has commonly been used in studies based on hyperspectral sensors (sensors that collect data in many narrow and contiguous spectral bands) and recent studies demonstrate that the technique also has a good potential for use with broadband satellite data (Feilhauer et al., 2012).

In remote sensing, vegetation properties such as the leaf area index, the fraction of intercepted photosynthetic active radiation, absorbed photosynthetically active radiation, green biomass and primary productivity, and leaf chlorophyll content can be interpreted in terms of vegetation indices (VIs) (mathematical transformations of vegetation reflectance into dimensionless indicators) (e.g. Boschetti et al., 2007; Cristiano et al., 2010; Vescovo and Gianelle, 2008). A routinely used index for the estimation of vegetation properties is the normalised difference vegetation index (NDVI) (Rouse et al., 1974), which is based on the contrast between absorption in the red band by chlorophyll pigments and reflectance in the near infrared (NIR) band caused by internal scattering within leaves. NDVIs calculated from different WorldView-2 satellite bands have been used for the estimation of above-ground biomass (Mutanga et al., 2012).

Differences in environmental conditions between plots may be expected to result in between-plot dissimilarities in the remotely sensed spectral response (cf. Rocchini et al., 2010a). Because levels of turnover in species composition are often related to environmental heterogeneity (Kneitel and Chase, 2004), spectral dissimilarities between plots or sites have been shown to function as a useful predictor of plant species turnover and plant species beta diversity in a variety of vegetation types (e.g. Rocchini et al., 2009, 2010b). For example, Rocchini et al. (2010b) used hyperspectral data to analyse the relation between spectral variation and beta diversity at local scales in highland savannahs, and satellite data have also been used for estimating plant species turnover in semi-natural grassland sites (Hall et al., 2011). Landsat data have been used to describe environmental heterogeneity when modelling the beta diversity of forest trees in a fragmented landscape (Cayuela et al., 2006). Spectral distances based on the MODIS NDVI have shown to be related to beta diversity at the county level in the USA (He et al., 2009), and the MODIS NDVI has been used as a surrogate for productivity in studies of beta diversity on a global scale (He and Zhang, 2009).

Strategic conservation planning requires information on species distributions at different spatial levels (Auestad et al., 2008; Turner et al., 2003) and there is still a need for studies on the potential of Earth Observation satellite-based sensors for collecting fine-scale information on plant species beta diversity. In the present paper, we examine whether dissimilarities in spectral response are related to vascular plant species beta diversity between semi-natural grassland plots. The study has a focus on the relationships between WorldView-2 satellite spectral dissimilarity and fine-scale species dissimilarity, using different sizes of plot (2 m × 2 m and 4 m × 4 m) and satellite pixel window (between 1 × 1 pixel and 11 × 11 pixels). Univariate regression and regression modelling of VIs, and PLSR modelling using the full set of WorldView-2 satellite bands are used to analyse the relationships between satellite and field data. We ask the following questions: (1) is WorldView-2 satellite spectral dissimilarity related to fine-scale species dissimilarity? (2) are the possible associations between spectral dissimilarity and fine-scale species dissimilarity affected by the spatial extent of (a) the grassland plots and (b) the satellite pixel windows?

## 2. Material and methods

### 2.1. Study area

The study area (centred on 56°40′49″N, 16°33′58″E) is located on the Baltic island of Öland in SE Sweden (Fig. 1a) and covers approximately 22.5 km$^2$. The bedrock of the area consists of Cambro-Silurian limestone and the overall topography is flat, with a few low ridges and deposits of glaciofluvial material. The mean temperature is 7 °C (July mean = 16 °C; January mean = −1 °C) and the mean annual precipitation is 468 mm (Forslund, 2001). In the early 19th century, approximately 80% of the study area was covered by grassland (Johansson et al., 2008). The present-day landscape consists of a mosaic of arable fields, villages, forest, and grasslands. The majority of the remaining grasslands (approximately 10% of the landscape (Johansson et al., 2008)) are grazed, with varying intensity, mainly by cattle.

### 2.2. Vegetation sampling

With the help of a land-use map of the study area (Johansson et al., 2008), interpretation of aerial photographs, and field inventories, we identified a total of 77 dry grassland sites that were larger than 0.25 ha, and which had been managed by grazing for more than 50 years. The grassland sites were separated from each other and the surrounding landscape by walls or fences. Within each dry grassland site, two coordinate points were randomly positioned in open (not covered by shrubs or trees) grassland vegetation, with the constraints that they were at least 25 m apart, at least 13.5 m from the site boundary, and at least 13.5 m from shrubs or trees. We also used a bioassay approach, based on the presence/absence of indicator species (cf. Prentice and Cramer, 1990; Reitalu et al., 2009) to exclude vegetation with high levels of eutrophication. The 15 grassland sites that could accommodate these constraints were used for the vegetation sampling (Fig. 1b).

Plots of two sizes: (i) 2 m × 2 m plots, nested within (ii) 4 m × 4 m plots were centred over each of the two coordinate points within each of the 15 sites. A total of 30 2 m × 2 m plots and 30 4 m × 4 m plots were inventoried. The presence of all non-woody vascular plants was recorded within each individual 2 m × 2 m and 4 m × 4 m plot. The 2 m × 2 m and 4 m × 4 m plots were treated as two separate datasets. Each dataset was divided into a training and a validation subset by randomly assigning the two plots from each site to one or other of the two subsets (Fig. 1c).

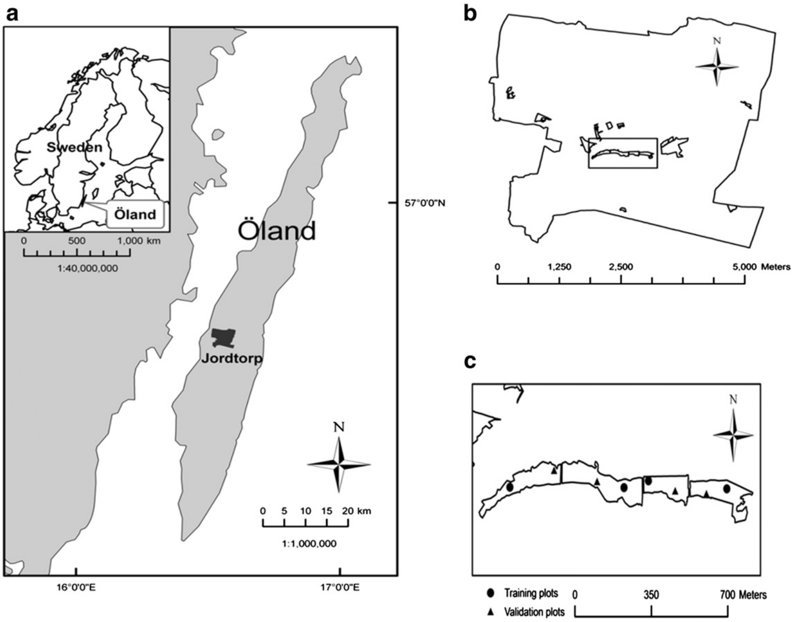


**Fig. 1.** (a) The Jordtorp study area on the Baltic island of Öland, Sweden. (b) The distribution of the 15 semi-natural grassland sites included in the present study. (c) The distribution of field sample plots within four grassland sites.

A hand-held differential global positioning system (DGPS) receiver (Topcon GRS-1 GNSS, equipped with a PG-A1 external antenna (Topcon Corporation, Japan)) connected to a real-time positioning service (SWEPOS) was used to log (with an accuracy of ~1 cm) the ground coordinates of the plot centres.

The field-work was carried out between 15 May and 15 July 2011.

### 2.3. Species beta diversity

The abundance-based Bray–Curtis dissimilarity index (Clarke et al., 2006) was used to quantify dissimilarity in species composition between pairs of plots. Assuming $x[ij]$ equals the quantity $x$ of a given species $i$ in plot $j$, and $x[ik]$ equals the quantity $x$ of a given species $i$ in plot $k$, the dissimilarity index was calculated using the vegan package (Oksanen et al., 2012) in the R programming environment (R Development Core Team, 2011) as $\sum(abs(x[ij] - x[ik]))/\sum(x[ij] + x[ik])$.

A pairwise beta diversity dissimilarity matrix was computed for each of the four subsets (training and validation subsets representing the two plot sizes 2 m × 2 m and 4 m × 4 m, respectively). Within each subset, the Bray–Curtis dissimilarity index was calculated for the 105 pair-wise relations between 15 plots. The Bray–Curtis dissimilarity index ranges between 0 and 1. When the species composition between pairs of plots is similar, the Bray–Curtis dissimilarity index will be low, and the index value will increase as the between-plot dissimilarity in species composition increases.

### 2.4. Satellite data

We used remote sensing data acquired on 21 May 2011 by the WorldView-2 satellite (DigitalGlobe). WorldView-2 provides eight-band multispectral imagery 400–450 nm (costal), 450–510 nm (blue), 510–581 nm (green), 585–625 nm (yellow), 630–690 nm (red), 705–745 nm (red edge), 770–895 nm (NIR1), and 860–1040 nm (NIR2) with a spatial resolution of 2 m. The imagery was orthorectified and geometrically corrected by the satellite data providers. The pixel digital numbers (DNs) were converted to top-of-atmosphere band-averaged reflectance according to Updike and Comp (2010).

### 2.5. Extracting spectral data

GPS readings of plot-centre coordinates were used to identify the plot pixel (i.e. the 2 × 2 m pixel whose coordinates were closest to the centre of the plot) in the satellite image. Spectral data extracted from six different sizes of pixel windows, were retrieved by overlaying the plot pixels with windows ranging in size from 1 × 1 pixel to 11 × 11 pixels. The mean spectral values of WorldView-2 bands, calculated for pixels falling within the individual pixel windows were used to compute three normalised difference-based VIs (NDVI = (NIR1 − red) / (NIR1 + red), NDRE = (NIR1 − red edge) / (NIR1 + red edge), NDVI2 = (NIR2 − red) / (NIR2 + red)).

Finally, the Euclidean spectral differences (for each individual band and the VIs) between all pairs of same-sized pixel windows were calculated. A pair-wise spectral dissimilarity matrix was computed; for each size of pixel window, each band, and each VI, and for each of the four subsets (training and validation subsets representing the two plot sizes 2 m × 2 m and 4 m × 4 m, respectively).

### 2.6. Statistical analyses

For the training subsets, the relationship between spectral dissimilarity and field-observed species dissimilarity was investigated using (i) univariate regression (including VIs) and (ii) "leave-one-out" cross-validated partial least squares regression (PLSR) (including the full set of WorldView-2 satellite bands).

The validation subsets were used to evaluate the regression models for the training subsets (i) with the help of the reduced major axis (RMA) regression-modelling approach (Cohen et al., 2003; Curran and Hay, 1986; Heiskanen, 2006) and (ii) by fitting the final cross-validated PLSR models of the training subsets to the validation subsets. Following Piñeiro et al. (2008), the regression models for the training subsets were evaluated by regression of the observed species dissimilarity values of the validation subsets (on the y-axis) versus the predicted dissimilarity values of the validation subsets (on the x-axis). To test the significance of the relationship between spectral dissimilarity and species dissimilarity, a Mantel test with 999 permutations was performed, using the vegan package (Oksanen et al., 2012) in the R programming environment (R Development Core Team, 2011). The spectral dissimilarity variables were log-transformed before the statistical analyses.

#### 2.6.1. Univariate regression and regression-modelling using vegetation indices

The coefficient of determination ($R^2$), the root mean square error (RMSE), and the relative RMSE (rRMSE, %) of the linear relationship between the spectral dissimilarity (calculated from VIs (NDVI, NDRE, NDVI2)) and the Bray–Curtis dissimilarity index were computed for the training subsets. With the help of the RMA regression-modelling approach (Cohen et al., 2003; Curran and Hay, 1986; Heiskanen, 2006), the regression models for the training subsets were applied to the validation subsets and used to predict the Bray–Curtis index from the spectral dissimilarity for the validation subsets. The quality of the predictions was assessed with the help of Pearson correlation coefficient ($r$) values and the root mean square error of the prediction ($RMSE_P$) of the relationship between the beta diversity predicted from VI-based spectral dissimilarities and the observed beta diversity.

#### 2.6.2. PLSR using the full set of WorldView-2 satellite bands

PLSR is a multivariate analysis method used for modelling the relationship between two matrices consisting, respectively, of a set of several predictor variables (X) (in this case the set of spectral dissimilarities calculated from individual WorldView-2 satellite bands) and one or more dependent variable(s) (Y) (in this case dissimilarities in species composition computed by the Bray–Curtis dissimilarity index). PLSR allows statistical analysis of data sets where the predictor variables are not only numerous, but also strongly correlated and subject to noise (Wold et al., 2001). PLSR is built on the assumption that there are only a few variables (Latent Variables (LVs)) that influence the process under study. The PLSR-model attempts to model the LVs by constructing X-scores (or components) by creating linear combinations of the original X-variables, weighted so as to maximise the covariance of X and Y. X-scores are, therefore, predictors of Y but, at the same time, are also modelling X. The relative importance of individual predictor variables in a PLSR model can be described by the variable importance in projection (VIP) (Eriksson et al., 1999). The higher the VIP value of an X-variable, the greater its contribution to the model. The most influential predictor variables in a model are those that have VIP-values greater than 1. Using the PLS package (Mevik et al., 2011) in the R programming environment (R Development Core Team, 2011), two methods were used to quantify the predictive capabilities of PLSR models:

(i) “Leave-one-out” cross validation was employed on the training subsets to measure the predictive fits (measured as $R^2_{CV}$) and the predictive errors (measured as root mean square error ($RMSE_{CV}$) and relative $RMSE_{CV}$ ($rRMSE_{CV}$, %)). The appropriate number of model components was identified by adopting the rule that the $RMSE_{CV}$ must be reduced by >2% by the addition of another component (e.g. Chen et al., 2009; Cho et al., 2007).

(ii) The final cross-validated models of the training subsets were fitted to the validation subsets and Y was predicted on the basis of new X-variables. The predictive capabilities of the models were measured by Pearson correlation coefficient ($r$) values, and the root mean square error of the prediction ($RMSE_P$) of the relationship between the beta diversity predicted from WorldView-2 satellite data and the observed beta diversity.

## 3. Results

Table 1 presents summary statistics for the species richness and the Bray–Curtis dissimilarity index values within the training and validation subsets for the 2 m × 2 m and 4 m × 4 m plots, respectively. The five most abundant species recorded within the grasslands were *Plantago lanceolata*, *Galium verum*, *Ranunculus bulbosus*, *Poa pratensis*, and *Achillea millefolium*.

The WorldView-2 satellite visible bands (coastal, blue, green, yellow and red) were correlated with each other within each of the different sizes of pixel window ($0.60 \le r \le 0.88$, $p < 0.001$), and the red-edge band and the near infrared bands were also intercorrelated ($0.70 \le r \le 0.83$, $p < 0.001$) with each other within each of the different sizes of pixel window.

### 3.1. Training subsets

#### 3.1.1. Univariate regression using vegetation indices

For each set of plots (2 m × 2 m and 4 m × 4 m), the results of the Mantel permutation tests showed that, for each of the different sizes of pixel window, the spectral dissimilarity calculated from each of the three normalised difference-based VIs (NDVI, NDRE, and NDVI2) was significantly ($p < 0.002$) and positively correlated with the Bray–Curtis dissimilarity index. Our results indicate that an increase in the species dissimilarity between plots was accompanied by an increase in the spectral dissimilarity between the plots. The NDVI showed slightly stronger relationships between spectral dissimilarity and species dissimilarity than the other VIs.

The strength of the relationship between spectral dissimilarity and species dissimilarity for the two sets of plots (2 m × 2 m and 4 m × 4 m) was comparable. Using the two smallest pixel windows 1 × 1 pixel (2 m × 2 m) and 3 × 3 pixels (6 m × 6 m), the NDVI-based spectral dissimilarity explained approximately 30% of the variation in the species dissimilarity for each individual set of plots. For example, using the 2 m × 2 m plot set for the 1 × 1 pixel window resulted in a $R^2$ value of 0.32 ($p < 0.001$), a RMSE of 0.105, and a rRMSE of 18.4% (Fig. 2a). Using the 4 m × 4 m plot set for the 3 × 3 pixel window produced a $R^2$ value of 0.33 ($p < 0.001$) and RMSE and rRMSE values of 0.106 and 17.4%, respectively (Fig. 2b).

**Table 1**
Summary statistics for non-woody vascular plant species richness and the Bray–Curtis dissimilarity index between grassland plots within the Jordtorp study area on the Baltic island of Öland, Sweden. The Bray–Curtis dissimilarity index is calculated between pairs of same-sized plots (2 m × 2 m and 4 m × 4 m) within the training and validation subsets, respectively. In each of the four subsets (representing the two plot sizes in both the training and validation subsets), n = 15 for species richness and n = 105 for the Bray–Curtis dissimilarity index.

| | Training subsets | | Validation subsets | |
|---|---|---|---|---|
| | 2 × 2 m | 4 × 4 m | 2 × 2 m | 4 × 4 m |
| *Species richness* | | | | |
| Total | 105 | 125 | 107 | 128 |
| Mean | 37.4 | 47.7 | 37.7 | 48.9 |
| Minimum | 21 | 32 | 26 | 37 |
| Maximum | 48 | 60 | 45 | 61 |
| *Bray–Curtis index* | | | | |
| Minimum | 0.23 | 0.20 | 0.21 | 0.19 |
| Maximum | 0.80 | 0.81 | 0.83 | 0.78 |

The association between spectral dissimilarity and species dissimilarity became progressively poorer when windows larger than 3 × 3 pixels were used (Fig. 3a,b), implying that the spectral response captured from increasingly larger pixel windows had a successively lower ability to characterise the species dissimilarity at fine-scales compared with the spectral response acquired from smaller windows. The $R^2$ values decreased from above 0.30 ($p < 0.001$) using the two smallest pixel windows to approximately 0.26 ($p < 0.001$), using the largest window of 11 × 11 pixels (22 m × 22 m) (Fig. 3a,b).

### 3.1.2. PLSR using the full set of WorldView-2 satellite bands

The $RMSE_{CV}$ values were used to determine how many components should be used for modelling the species dissimilarity. Adopting the rule that the $RMSE_{CV}$ must be reduced by >2% by the addition of another component (Chen et al., 2009; Cho et al., 2007), gave the inclusion of two components in all PLSR models. The red, NIR1 and NIR2 bands were associated with VIP values higher than 1 in all PLSR models and the red and NIR1 bands had higher VIP values than the NIR2 band in all PLSR models.

The relationships between the predicted and the observed species dissimilarity were similar in both the 2 m × 2 m and 4 m × 4 m sets of plots. For example, $R^2_{CV}$ values equal to or higher than 0.35, were obtained for each individual set of plots using the two smallest pixel windows (1 × 1 pixel (2 m × 2 m) and 3 × 3 pixels (6 m × 6 m)). Using the 1 × 1 pixel window for the 2 m × 2 m plot set gave a $R^2_{CV}$ of 0.38, a $RMSE_{CV}$ of 0.098 and a $rRMSE_{CV}$ of 17.2% (Fig. 4a). Using the 3 × 3 pixel window for the 4 m × 4 m plot set resulted in $R^2_{CV} = 0.35$, $RMSE_{CV} = 0.104$ and $rRMSE_{CV} = 17.1\%$ (Fig. 4b). $R^2_{CV}$ values decreased and $RMSE_{CV}$/$rRMSE_{CV}$ values increased when windows larger than 3 × 3 pixels were used (Fig. 5a,b). For example, $R^2_{CV}$ values decreased from equal to or higher than 0.35 using the two smallest windows to approximately 0.28 using the largest window of 11 × 11 pixels (Fig. 5a,b).

### 3.2. Validation subsets

Both the RMA regression-modelling and the PLSR modelling resulted in correlation coefficients between the field-observed and the predicted species dissimilarity that were higher for the 4 m × 4 m plot set compared with the 2 m × 2 m plot set (Table 2), indicating that the spatial size of the plots influenced the association between the satellite-based spectral dissimilarity and the beta diversity. For example, for the 2 m × 2 m plot set, both the RMA and the PLSR approaches gave $r$ values that were between 0.32 and 0.33 ($p < 0.001$) using the two smallest pixel windows (1 × 1 pixel and 3 × 3 pixels), whereas for the 4 m × 4 m plot set, these window sizes resulted in $r$ values that were between 0.37 and 0.39 ($p < 0.001$) (Table 2). For both the 2 m × 2 m and 4 m × 4 m plot sets, the relationships between the predicted and the observed species dissimilarity decreased when larger pixel windows were used and, for the largest analysed window (11 × 11 pixels), the $r$ values were lower than 0.30 ($p < 0.01$) (Table 2).

## 4. Discussion

Positive associations between spectral dissimilarity and fine-scale (2 m × 2 m and 4 m × 4 m) species dissimilarity were found in the training subsets of plots. The correlation between the predicted and the observed species dissimilarity for the validation subsets was stronger for the set of large plots (4 m × 4 m) than for the set of small plots (2 m × 2 m). Spectral data acquired over windows of 1 × 1 pixel (2 m × 2 m) and 3 × 3 pixels (6 m × 6 m) performed better in the assessment of fine-scale (2 m × 2 m and 4 m × 4 m) species dissimilarity than the spectral response extracted over larger pixel windows. Other remote sensing studies have shown that PLSR models may represent a valuable complement to univariate statistical methods (e.g. Darvishzadeh et al., 2008; Fava et al., 2010), and PLSR models were found to be associated with higher $R^2$ values than models based on linear and nonlinear regression of VIs in the estimation of biomass in grasslands (Chen et al., 2009). In our study, the models based on PLSR also had higher $R^2$ values than those based on univariate, VI-based models. Spectral information may be lost when only two wavelengths are used for VIs and the use of several spectral bands in PLSR models is expected to improve the relationships between response and explanatory variables (cf. Darvishzadeh et al., 2008; Fava et al., 2010).

### 4.1. The relationship between spectral dissimilarity and species dissimilarity

The reflectance in the red wavelength rises over areas with low vegetation cover, while the reflectance in the NIR wavelengths rises as the vegetation cover and density increase — because of increased multiple scattering within the canopy (e.g. Thenkabail et al., 2012). In the present study, the spectral dissimilarities based on the NDVI, which uses the contrast between the red and the NIR1 bands, showed slightly stronger relationships with the species dissimilarity than the other VIs. The red and the NIR1 bands were also the most influential predictor variables in the PLSR models. Previous studies have shown that information provided by spectral data may be useful for the estimation of field-layer height and above-ground biomass in grasslands

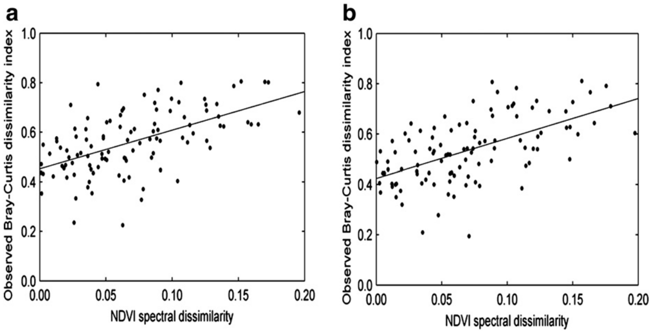


**Fig. 2.** Linear relationships between the NDVI-based spectral dissimilarity and the field-observed Bray–Curtis dissimilarity index for the training subsets. (a) Plot size 2 m × 2 m, pixel window 1 × 1 pixel (2 m × 2 m): $R^2 = 0.32$ ($p < 0.001$), intercept = 0.45 ($p < 0.001$), n = 105. (b) Plot size 4 m × 4 m, pixel window 3 × 3 pixels (6 m × 6 m): $R^2 = 0.33$ ($p < 0.001$), intercept = 0.41 ($p < 0.001$), n = 105.

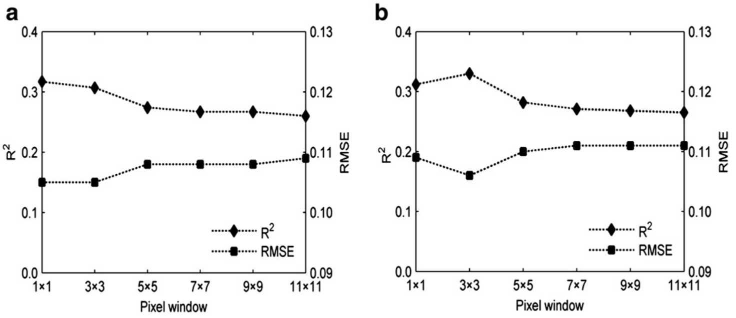


**Fig. 3.** The coefficient of determination ($R^2$) and the root mean square error (RMSE) of the linear relation between the NDVI-based spectral dissimilarity and the field-observed Bray–Curtis dissimilarity index for the training subsets using six pixel windows, ranging between 1 × 1 pixel (2 m × 2 m) and 11 × 11 pixels (22 m × 22 m). (a) Plot size 2 m × 2 m: n = 105. (b) Plot size 4 m × 4 m: n = 105. All the $R^2$ values are significant ($p < 0.001$).

(e.g. Boschetti et al., 2007; Poças et al., 2012). It has also been shown that the amount of dead above-ground biomass (litter) may influence the spectral response from vegetation, particularly in the red and NIR bands (Zhang and Guo, 2008).

The productivity–diversity relationship (Al-Mufti et al., 1977) has received much attention in ecology but the nature of the processes that underlie the relationship is still under debate (e.g. Adler et al., 2011; Gillman and Wright, 2006). Biomass may be well-correlated with productivity and it has often been used as a proxy for productivity (e.g. Ni et al., 2007). The NDVI has been used as a surrogate for productivity in studies of species diversity relationships (e.g. He and Zhang, 2009; Psomas et al., 2011). For example, Psomas et al. (2011) used estimates of biomass based on spectral data as a proxy for productivity in studies of the relationship between biomass and plant species richness in grassland habitats. However, it has been pointed out that biomass may not necessarily be a reliable surrogate for productivity (Gillman and Wright, 2006). The relationship between biomass and productivity can, for example, vary between different plant species and may be influenced by the level of disturbance in the habitat and the management regime (cf. Gillman and Wright, 2006; Guo, 2007).

In semi-natural grasslands, a reduction in management (grazing or mowing) intensity is followed by an increase in field-layer height (and amount of above-ground biomass) and changes in the light conditions in the vegetation canopy (e.g. Werger et al., 2002). These changes, in their turn, have an effect on the species composition of grassland habitats (Kull and Zobel, 1991; Luoto et al., 2003; Marion et al., 2010) by allowing more competitive, and often generalist, species to become dominant. Decreasing grazing intensity may also increase the accumulation of litter within grasslands (cf. Jensen and Gutekunst, 2003), which in its turn, influences plant community composition and species diversity (Galvánek and Leps, 2012). We suggest that the positive associations between spectral dissimilarity and beta diversity in the present study may be related to differences in grazing intensity between plots. Increased between-plot differences in grazing intensity may be accompanied by increased between-plot differentiation in both spectral response and species composition — and an increase in beta diversity that is reflected in the spectral dissimilarity.

Spectral data may also be influenced by abiotic factors such as soil properties (Huete, 1988). Edaphic conditions may have a significant influence on the species composition within semi-natural grasslands (e.g. Pärtel and Helm, 2007) and between-plot dissimilarities in soil properties may also contribute to the positive associations between spectral dissimilarity and species dissimilarity in the present study.

### 4.2. The spatial extent of plots and pixel windows

The effect of the spatial size of the sampling unit on the relationship between spectral variability and beta diversity has previously

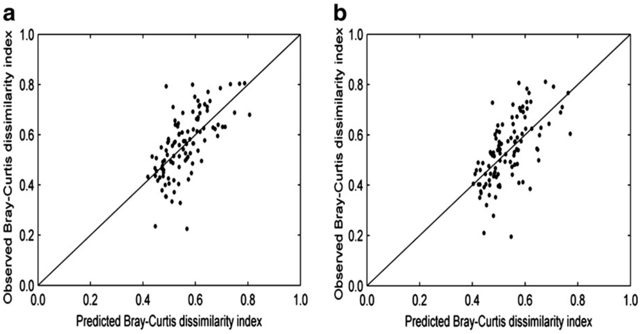


**Fig. 4.** The field-observed versus the predicted Bray–Curtis dissimilarity index using partial least squares regression (PLSR) analysis for the training subsets. (a) Plot size 2 m × 2 m, pixel window 1 × 1 pixel (2 m × 2 m): $R^2_{CV} = 0.38$, $RMSE_{CV} = 0.098$, $rRMSE_{CV} = 17.2\%$, n = 105. (b) Plot size 4 m × 4 m, pixel window 3 × 3 pixels (6 m × 6 m): $R^2_{CV} = 0.35$, $RMSE_{CV} = 0.104$, $rRMSE_{CV} = 17.1\%$, n = 105.

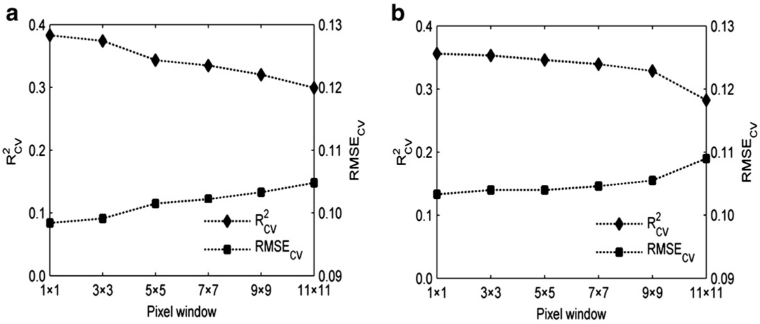


**Fig. 5.** The predictive fit ($R^2_{CV}$) and the root mean square error ($RMSE_{CV}$) for the relation between the predicted and the field-observed Bray–Curtis dissimilarity index using partial least squares regression (PLSR) analysis for the training subsets for six pixel windows, ranging between 1 × 1 pixel (2 m × 2 m) and 11 × 11 pixels (22 m × 22 m). (a) Plot size 2 m × 2 m: n = 105. (b) Plot size 4 m × 4 m: n = 105.

been studied in African highland savannahs using plot sizes of 10 m × 10 m and 20 m × 50 m (Rocchini et al., 2010b). Whereas Rocchini et al. (2010b) showed that correlations between spectral variation and species turnover were weaker for small plots than for large plots, the strength of the relationship between NDVI-based spectral dissimilarity and species dissimilarity was similar for both the training subsets (2 m × 2 m and 4 m × 4 m) in our study. And, in contrast to the results of Rocchini et al. (2010b), the PLSR analysis showed a slightly stronger relationship between spectral dissimilarity and species dissimilarity for the set of small plots (2 m × 2 m) than for the set of large plots (4 m × 4 m) within the training subset. However, in agreement with the results of Rocchini et al. (2010b), the validation subset of small plots was characterised by poorer correlations between predicted and field-observed beta diversity than the validation subset of large plots (Table 2). Rocchini et al. (2010b) concluded that the relationship between spectral variability and beta diversity was comparatively stronger for large plots than for small plots, because species data recorded in small plots are likely to be more affected by random disturbances than data recorded in large plots.

The species pool hypothesis (Hodgson, 1987; Zobel, 1997), proposes that the local species diversity will be related to the size of the regional pool of species that are potentially capable of coexisting within a particular community. The role of species pools in determining plant species diversity has been studied on a variety of scales. For instance, Franzén and Eriksson (2001) showed that the species diversity at the 1 dm × 1 dm scale could be explained by the size of the species pool at the 2 m × 2 m scale within Swedish semi-natural grasslands. The influence of the species pool on local species diversity may be driven by migration and dispersal processes as well as by environmental “filtering” (Zobel, 1997). Because the size of the species pool at larger scales affects the availability of species at smaller scales, measures of species diversity at different spatial scales within semi-natural grasslands are usually correlated with each other (Franzén and Eriksson, 2001; Öster et al., 2007; Pärtel and Zobel, 1999). The present study focuses only on fine-scale variation and the spatial extent of the plots and the size difference between the plots were small. The lack of distinctly different relationships between spectral dissimilarity and species dissimilarity for the two studied grain dimensions in our study is likely to reflect the fact that the species composition within the 2 m × 2 m plots represents a subset of the species pool in the 4 m × 4 m plots.

**Table 2**
Correlation coefficients (*r*) and root mean square error (RMSE_P) values between predicted and field-observed fine-scale beta diversity derived from (a) reduced major axis (RMA) regression modelling (using the normalised difference based vegetation index NDVI (NIR − red) / (NIR + red)) and (b) partial least square regression (PLSR) of WorldView-2 satellite data for the validation subsets (n = 105 at the 2 m × 2 m and 4 m × 4 m plot size, respectively). All the *r* values are significant (p < 0.01).

| Pixel window | 2 × 2 m plot | | 4 × 4 m plot | |
|---|---|---|---|---|
| | *r* | RMSE_P | *r* | RMSE_P |
| *a. RMA modelling* | | | | |
| 1 × 1 | 0.33 | 0.12 | 0.37 | 0.12 |
| 3 × 3 | 0.32 | 0.12 | 0.37 | 0.12 |
| 5 × 5 | 0.32 | 0.12 | 0.35 | 0.12 |
| 7 × 7 | 0.29 | 0.13 | 0.32 | 0.12 |
| 9 × 9 | 0.26 | 0.13 | 0.31 | 0.12 |
| 11 × 11 | 0.25 | 0.13 | 0.28 | 0.13 |
| *b. PLSR* | | | | |
| 1 × 1 | 0.32 | 0.13 | 0.37 | 0.12 |
| 3 × 3 | 0.33 | 0.13 | 0.39 | 0.12 |
| 5 × 5 | 0.33 | 0.12 | 0.37 | 0.12 |
| 7 × 7 | 0.30 | 0.12 | 0.33 | 0.12 |
| 9 × 9 | 0.26 | 0.13 | 0.32 | 0.12 |
| 11 × 11 | 0.26 | 0.14 | 0.29 | 0.13 |

The relationships between spectral dissimilarity and species dissimilarity became progressively somewhat poorer when pixel windows larger than 3 × 3 pixels (6 m × 6 m) were used (Figs. 3 and 5). Semi-natural grasslands are often characterised by a small-scale environmental heterogeneity, which may influence the fine-scale spatial composition and distribution of species within grassland sites (Bruun, 2000; Ekstam and Forshed, 1992; Gazol et al., 2012). Our results suggest that the spectral response captured with the help of pixel windows that corresponded approximately to the size of the plots was better at characterising the environmental conditions (e.g. field layer-height, amount of biomass and litter, and soil properties) that shape the pattern of species composition at fine scales (2 m × 2 m and 4 m × 4 m) than the spectral response captured by larger pixel windows. It is likely that the spectral information extracted over increasingly larger pixel windows was influenced by a progressive increase in environmental heterogeneity that lead to a gradual attenuation of the relationship between spectral dissimilarity and fine-scale (2 m × 2 m and 4 m × 4 m) beta diversity.

### *4.3. Conclusion*

In the present study we used WorldView-2 satellite spectral dissimilarity to infer environmental heterogeneity in dry semi-natural grasslands. Our results revealed a significant positive association between spectral dissimilarity and fine-scale (2 m × 2 m and 4 m × 4 m) plant species beta diversity, suggesting that the satellite-based measurement of environmental variation may provide a promising approach to the detection of variability in species composition at detailed spatial scales. We conclude that very high spatial resolution data, such as the WorldView-2 satellite data may potentially contribute to the development of improved methods for use in basic ecological research and in the monitoring of plant species beta diversity in semi-natural grasslands at fine scales. However, our study had a focus on a specific type of species-rich grassland vegetation, and further research in other types of semi-natural grassland habitats is needed before it will be possible to fully assess the potential of satellite data as an operational tool for monitoring grassland plant species beta diversity at fine scales.

## Acknowledgements

The study was financed by grants from The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) to KH and HCP. We thank Sofia Pallander, Andreas Press, and Johan Rydlöv for their field assistance and Jonas Ardö and two anonymous reviewers for their helpful comments on an earlier version of this paper. The "Station Linné" research station at Ölands Skogsby was used as a base for the fieldwork.

## References

Adler, P., Seabloom, E., Borer, E., Hillebrand, H., Hautier, Y., Hector, A., et al., 2011. Productivity is a poor predictor of plant species richness. Science 333, 1750–1753.

Al-Mufti, M., Sydes, C., Furness, S., Grime, J., Band, S., 1977. A quantitative analysis of shoot phenology and dominance in herbaceous vegetation. Journal of Ecology 65, 759–791.

Anderson, M., Crist, T., Chase, J., Vellend, M., Inouye, B., Freestone, A., Swenson, N., 2011. Navigating the multiple meanings of $\beta$ diversity: a roadmap for the practicing ecologist. Ecology Letters 14, 19–28.

Auestad, I., Rydgren, K., Økland, R., 2008. Scale-dependence of vegetation-environment relationships in semi-natural grasslands. Journal of Vegetation Science 19, 139–148.

Boschetti, M., Bocchi, S., Brivio, P., 2007. Assessment of pasture production in the Italian Alps using spectrometric and remote sensing information. Agriculture, Ecosystems & Environment 118, 267–272.

Boyd, D.S., Foody, G.M., 2011. An overview of recent remote sensing and GIS based research in ecological informatics. Ecological Informatics 6, 25–36.

Bruun, H.H., 2000. Patterns of species richness in dry grassland patches in an agricultural landscape. Ecography 23, 641–650.

Cayuela, L., Rey Benayas, J.M., Justel, A., Salas-Rey, J., 2006. Modelling tree diversity in a highly fragmented tropical montane landscape. Global Ecology and Biogeography 15, 602–613.

Chen, J., Song, G., Shen, M., Tang, Y., Matsushita, B., 2009. Estimating aboveground biomass of grassland having a high canopy cover: an exploratory analysis of in situ hyperspectral data. International Journal of Remote Sensing 30, 6497–6517.

Cho, M., Skidmore, A., Corsi, F., van Wieren, S., Sobhan, I., 2007. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. International Journal of Applied Earth Observation and Geoinformation 9, 414–424.

Clarke, R., Somerfield, P., Chapman, G., 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and zero-adjusted Bray–Curtis coefficient for denuded assemblages. Journal of Experimental Marine Biology and Ecology 330, 55–80.

Cohen, W., Maiersperger, T., Gower, S., Turner, D., 2003. An improved strategy for regression of biophysical variables and Landsat ETM + data. Remote Sensing of Environment 84, 561–571.

Connell, J.H., 1978. Diversity in tropical rain forest and coral reefs. Science 1999, 1302–1310.

Cousins, S., Lindborg, R., 2008. Remnant grassland habitats as source communities for plant diversification in agricultural landscapes. Biological Conservation 141, 233–240.

Cristiano, P., Posse, G., Di Bella, C., Jaimes, F., 2010. Uncertainties in fPAR estimation of grass canopies under different stress situations and differences in architecture. International Journal of Remote Sensing 15, 4095–4109.

Curran, P., Hay, A., 1986. The importance of measurements error for certain procedures in remote sensing at optical wavelengths. Photogrammetric Engineering and Remote Sensing 52, 229–241.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., Cho, M., 2008. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. ISPRS Journal of Photogrammetry and Remote Sensing 63, 409–426.

Ekstam, U., Forshed, N., 1992. Om hävden upphör: kärlväxter som indikatorarter i ängs- och hagmarker (If Grassland Management Ceases: Vascular plants as Indicator Species in Meadows and Pastures). Naturvårdsverket, Stockholm (in Swedish with English summary).

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 1999. Multi- and Megavariate Data Analysis Using Projections Methods (PCA&PLS). Umetrics, Umeå, Sweden.

Fava, F., Parolo, G., Colombo, R., Gusmeroli, F., Della Maianna, G., Monteiro, A., Bocchi, S., 2010. Fine-scale assessment of hay meadow productivity and plant diversity in the European Alps using field spectrometric data. Agriculture, Ecosystems & Environment 137, 151–157.

Feilhauer, H., Schmidtlein, S., 2009. Mapping continuous fields of forest alpha and beta diversity. Applied Vegetation Science 12, 429–439.

Feilhauer, H., Schmidtlein, S., 2011. On variable relations between vegetation patterns and canopy reflectance. Ecological Informatics 6, 83–92.

Feilhauer, H., He, K.S., Rocchini, D., 2012. Modeling species distribution using niche-based proxies derived from composite bioclimatic variables and MODIS NDVI. Remote Sensing 4, 2057–2075.

Forslund, F., 2001. Natur och kultur på Öland. Länstyrelsen i Kalmar län, Kalmar (in Swedish).

Franzén, D., Eriksson, O., 2001. Small-scale patterns of species richness in Swedish semi-natural grasslands: the effects of community species pools. Ecography 24, 505–510.

Galvánek, D., Leps, J., 2012. The effect of management on productivity, litter accumulation and seedling recruitment in a Carpathian mountain grassland. Plant Ecology 213, 523–533.

Gazol, A., Tamme, R., Takkis, K., Kasari, L., Saar, L., Helm, A., Pärtel, M., 2012. Landscape- and small-scale determinants of grassland species diversity: direct and indirect influences. Ecography 35, 944–951.

Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., Saatchi, S., 2008. Measuring and modeling biodiversity from space. Progress in Physical Geography 32, 203–221.

Gillman, L., Wright, S., 2006. The influence of productivity on the species richness of plants: a critical assessment. Ecology 87, 1234–1243.

Guo, Q., 2007. The diversity–biomass–productivity relationships in grassland management and restoration. Basic and Applied Ecology 8, 199–208.

Hall, K., Reitalu, T., Sykes, M.T., Prentice, H.C., 2011. Spectral heterogeneity of QuickBird satellite data is related to fine-scale plant species spatial turnover in semi-natural grasslands. Applied Vegetation Science 15, 145–157.

He, K., Zhang, J., 2009. Testing the correlation between beta diversity and differences in productivity among global ecoregions, biomes, and biogeographical realms. Ecological Informatics 4, 93–98.

He, K., Zhang, J., Zhang, Q., 2009. Linking variability in species composition and MODIS NDVI based on beta diversity measurements. Acta Oecologica 35, 14–21.

Heiskanen, J., 2006. Estimating aboveground tree biomass and leaf area index in mountain birch forest using ASTER satellite data. International Journal of Remote Sensing 27, 1135–1158.

Hodgson, J.G., 1987. Why do so few plant species exploit productive habitats? An investigation into cytology, plant strategies and abundance within a local flora. Functional Ecology 1 (243), 250.

Huete, A.R., 1988. A soil adjusted vegetation index (SAVI). Remote Sensing of Environment 25, 295–309.

Hutchinson, G.E., 1961. The paradox of the plankton. The American Naturalist 95, 137–145.

Janišová, M., Bartha, K., Kiehl, K., Dengler, J., 2011. Advances in the conservation of dry grasslands: introduction to contributions from the seventh European Dry Grassland meeting. Plant Biosystems 3, 507–513.

Jensen, K., Gutekunst, K., 2003. Effects of litter on establishment of grassland species: the role of seed size and successional status. Basic and Applied Ecology 4, 579–587.

Johansson, L., Hall, K., Prentice, H.C., Ihse, M., Reitalu, T., Sykes, M.T., Kindström, M., 2008. Semi-natural grassland continuity, long-term land-use change and plant species richness in an agricultural landscape on Öland, Sweden. Landscape and Urban Planning 84, 200–211.

Keil, P., Schweiger, O., Kühn, I., Kunin, W., Kuussaari, M., Settele, J., Storch, D., 2012. Patterns of beta diversity in Europe: the role of climate, land cover and distance across scales. Journal of Biogeography 39, 1473–1486.

Kneitel, J., Chase, J., 2004. Trade-offs in community ecology: linking spatial scales and species coexistence. Ecology Letters 7, 69–80.

Koleff, P., Gaston, K., Lennon, J., 2003. Beta diversity for presence–absence data. Journal of Animal Ecology 72, 367–382.

Kull, K., Zobel, M., 1991. High species richness in an Estonian wooded meadow. Journal of Vegetation Science 2, 715–718.

Luoto, M., Rekolainen, S., Aakkula, J., Pykälä, J., 2003. Loss of plant species richness and habitat connectivity in grasslands associated with agricultural change in Finland. Ambio 32, 447–452.

Marion, B., Bonis, A., Bouzillé, J.-B., 2010. How much does grazing-induced heterogeneity impact plant diversity in wet grasslands? Ecoscience 17, 229–239.

McKnight, M., White, P., McDonald, R., Lamoreux, J., Sechrest, W., Ridgely, R., Stuart, S., 2007. Putting beta-diversity on the map: broad-scale congruence and coincidence in the extremes. PLoS Biology 5, 2424–2432.

Mevik, B.-H., Wehrens, R., Hovde Liland, K., 2011. PLS: partial least squares and principal component regression. R package version 2.3-0. http://cran.r-project.org/package=pls.

Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. International Journal of Applied Earth Observation and Geoinformation 18, 399–406.

Nekola, J., White, P., 1999. The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26, 867–878.

Ni, J., Wang, G., Bai, Y., Li, X., 2007. Scale-dependent relationships between plant diversity and above-ground biomass in temperate grasslands, south-eastern Mongolia. Journal of Arid Environments 68, 132–142.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Simpson, G., Solymos, P., Henry, M., Stevens, M., Wagner, H., 2012. vegan: community ecology package. R package version 2.0-3. http://cran.r-project.org/package=vegan.

Öster, M., Cousins, S.A.O., Eriksson, O., 2007. Size and heterogeneity rather than landscape context determine plant species richness in semi-natural grasslands. Journal of Vegetation Science 18, 859–868.

Pärtel, M., Helm, A., 2007. Invasion of woody species into temperate grasslands: relationships with abiotic and biotic soil resource heterogeneity. Journal of Vegetation Science 18, 63–70.

Pärtel, M., Zobel, M., 1999. Small-scale plant species richness in calcareous grasslands determined by the species pool, community age and shoot density. Ecography 22, 153–159.

Piñeiro, G., Perelman, S., Guerschman, J., Paruelo, J., 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? Ecological Modelling 216, 316–322.

Poças, I., Cunha, M., Pereira, L., 2012. Dynamics of mountain semi-natural grassland meadows inferred from SPOT–VEGETATION and field spectroradiometer data. International Journal of Remote Sensing 33, 4334–4355.

Poschlod, P., WallisDeVries, M., 2002. The historical and socioeconomic perspective of calcareous grasslands — lessons from the distant and recent past. Biological Conservation 104, 361–376.

Prentice, H.C., Cramer, W., 1990. The plant community as a niche bioassay: environmental correlates of local variation in *Gypsophila fastigiata*. Journal of Ecology 14, 313–325.

Psomas, A., Kneubühler, M., Huber, S., Itten, K., Zimmermann, N., 2011. Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. International Journal of Remote Sensing 32, 9007–9031.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria3-900051-07-0 (URL http://www.R-project.org/).

Raatikainen, K., Heikkinen, R., Luoto, M., 2009. Relative importance of habitat area, connectivity, management and local factors for vascular plants: spring ephemerals in boreal semi-natural grasslands. Biodiversity and Conservation 18, 1067–1085.

Reitalu, T., Sykes, M.T., Johansson, L., Lönn, M., Hall, K., Vandewalle, M., Prentice, H.C., 2009. Small-scale plant species richness and evenness in semi-natural grasslands respond differently to habitat fragmentation. Biological Conservation 142, 899–908.

Reitalu, T., Johansson, L., Sykes, M.T., Hall, K., Prentice, H.C., 2010. History matters: village distances, grazing and grassland species diversity. Journal of Applied Ecology 47, 1216–1224.

Rocchini, D., He, K., Zhang, J., 2009. Is spectral distance a proxy of beta diversity at different taxonomic ranks? A test using quantile regression. Ecological Informatics 4, 254–259.

Rocchini, D., Balkenhol, N., Carter, G., Foody, G.M., Gillespie, T., He, K., Neteler, M., 2010a. Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. Ecological Informatics 5, 318–329.

Rocchini, D., He, K., Oldeland, J., Wesuls, D., Neteler, 2010b. Spectral variation versus species β-diversity at different spatial scales: a test in African highland savannas. Journal of Environmental Monitoring 12, 825–831.

Rouse, J., Haas, R., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the Great Plains with ERTS-1. Proceedings, 3rd Earth Resources Technology Satellite Symposium, 1, pp. 309–317.

Silvertown, J., 2004. Plant coexistence and the niche. Trends in Ecology & Evolution 19, 605–611.

Thenkabail, P., Lyon, J., Huete, A., 2012. Hyperspectral Remote Sensing of Vegetation. Taylor and Francis Group, LLC, Boca Raton.

Tscharntke, T., Klein, A., Kruess, K., Steffan-Dewenter, I., Thies, C., 2005. Landscape perspectives on agricultural intensification and biodiversity — ecosystem service management. Ecology Letters 8, 857–874.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. Trends in Ecology & Evolution 18, 306–314.

Updike, T., Comp, C., 2010. Radiometric Use of WorldView-2 Imagery. Technical Note, 2010-11-01. DigitalGlobe, Inc., Longmont, Colorado, USA.

Vescovo, L., Gianelle, D., 2008. Using the MIR bands in vegetation indices for the estimation of grassland biophysical parameters from satellite remote sensing in the Alps region of Trentino (Italy). Advances in Space Research 41, 1764–1772.

Viedma, O., Torres, I., Pérez, B., Moreno, J., 2012. Modeling plant species richness using reflectance and texture data derived from QuickBird in a recently burned area of Central Spain. Remote Sensing of Environment 199, 208–221.

Werger, M., Hirose, T., During, H., Heil, G., Hikosaka, K., Ito, T., Nachinshonhor, U., Nagamatsu, D., Shibasaki, K., Takatsuki, S., van Rheenen, J., Anten, N., 2002. Light partitioning among species and species replacement in early successional grasslands. Journal of Vegetation Science 13, 615–626.

Whittaker, R.H., 1972. Evolution and measurement of species diversity. Taxon 21, 213–251.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58, 109–130.

Zhang, C., Guo, X., 2008. Monitoring northern mixed prairie health using broadband satellite imagery. International Journal of Remote Sensing 29, 2257–2271.

Zobel, M., 1997. The relative role of species pools in determining plant species richness: an alternative explanation of species coexistence? Trends in Ecology & Evolution 12, 266–269.

# Paper V

remote sensing

MDPI

Article

# Airborne Hyperspectral Data Predict Fine-Scale Plant Species Diversity in Grazed Dry Grasslands

**Thomas Möckel [1,2,3,*], Jonas Dalmayne [2], Barbara C. Schmid [2,3], Honor C. Prentice [3] and Karin Hall [2]**

[1] Department of Grassland Science and Renewable Plant Resources, University of Kassel, Steinstrasse 19, 37213 Witzenhausen, Germany

[2] Department of Physical Geography and Ecosystem Sciences, Lund University, Sölvegatan 12, 223 62 Lund, Sweden; jonas.dalmayne@nateko.lu.se (J.D.); Barbara.Schmid@ekol.lu.se (B.C.S.); karin.hall@nateko.lu.se (K.H.)

[3] Department of Biology, Lund University, Sölvegatan 37, 223 62 Lund, Sweden; Honor_C.Prentice@biol.lu.se

* Correspondence: thmoeck@uni-kassel.de; Tel.: +49-561-804-1337; Fax: +49-561-804-1230

Academic Editors: Susan L. Ustin, Parth Sarathi Roy and Prasad S. Thenkabail
Received: 16 September 2015; Accepted: 25 January 2016; Published: 8 February 2016

**Abstract:** Semi-natural grasslands with grazing management are characterized by high fine-scale species richness and have a high conservation value. The fact that fine-scale surveys of grassland plant communities are time-consuming may limit the spatial extent of ground-based diversity surveys. Remote sensing tools have the potential to support field-based sampling and, if remote sensing data are able to identify grassland sites that are likely to support relatively higher or lower levels of species diversity, then field sampling efforts could be directed towards sites that are of potential conservation interest. In the present study, we examined whether aerial hyperspectral (414–2501 nm) remote sensing can be used to predict fine-scale plant species diversity (characterized as species richness and Simpson's diversity) in dry grazed grasslands. Vascular plant species were recorded within 104 (4 m × 4 m) plots on the island of Öland (Sweden) and each plot was characterized by a 245-waveband hyperspectral data set. We used two different modeling approaches to evaluate the ability of the airborne spectral measurements to predict within-plot species diversity: (1) a spectral response approach, based on reflectance information from (i) all wavebands, and (ii) a subset of wavebands, analyzed with a partial least squares regression model, and (2) a spectral heterogeneity approach, based on the mean distance to the spectral centroid in an ordinary least squares regression model. Species diversity was successfully predicted by the spectral response approach (with an error of *ca.* 20%) but not by the spectral heterogeneity approach. When using the spectral response approach, iterative selection of important wavebands for the prediction of the diversity measures simplified the model but did not improve its predictive quality (prediction error). Wavebands sensitive to plant pigment content (400–700 nm) and to vegetation structural properties, such as above-ground biomass (700–1300 nm), were identified as being the most important predictors of plant species diversity. We conclude that hyperspectral remote sensing technology is able to identify fine-scale variation in grassland diversity and has a potential use as a tool in surveys of grassland plant diversity.

**Keywords:** arable-to-grassland succession; northern Europe; species richness; inverse Simpson's diversity index; HySpex spectrometer; partial least squares regression

## 1. Introduction

The threats to biodiversity from habitat loss, fragmentation and climate change continue to escalate [1], and the mapping of habitats and the investigation of the processes that determine local patterns of biodiversity have become increasingly important tasks [2]. Extensively managed, semi-natural grasslands are among the most diverse ecosystems in Europe, and both agricultural intensification and the abandonment of grazing management have led to a decrease in the plant species diversity in grassland habitats (*cf.* [1]). The conservation and monitoring of grazed semi-natural grasslands has become a high priority within the European Union [3] and target areas for habitat conservation need to be identified and prioritized in order to maintain and enhance biodiversity [4]. In the future, the conservation of species diversity in modern agricultural landscapes will require the development of techniques for monitoring and predicting patterns of grassland species diversity: the need for tools that are applicable at detailed spatial scales and over large areas has been identified as a central problem [5].

While a range of edaphic, topographic, historical and stochastic processes may act as drivers of species diversity within grazed semi-natural grasslands (e.g., [6,7]), many studies show that local plant species richness is influenced by present-day variation in grazing intensity [8] and by the historical continuity of grazing management (e.g., [9]). The activity of grazing animals influences the availability of essential resources, such as light and soil nutrients (the resource availability hypothesis) [10]. The activity of grazers may also lead to a greater spatial heterogeneity of resources, as a result of trampling or patchy removal of above-ground biomass (the spatial heterogeneity hypothesis) [10]. Heterogeneous habitats are expected to contain a greater diversity of potential niches for species rather than habitats with more homogeneous conditions [11], and environmental heterogeneity has been shown to promote fine-scale species diversity in grassland communities (e.g., [3,6]). Plant species richness (SR) is regarded as an important ecosystem characteristic [2] and may also provide an indication of ecosystem health and resilience [12]. Whereas data on the numbers of species (SR) recorded within a particular sample or habitat are important in conservation planning, diversity indices that account for both the number of species present and the abundance of each species (e.g., the inverse Simpson's diversity index, iSDI) are often preferred in ecological studies because it is assumed that the most dominant species are likely to contribute most to processes within local communities [13]. Species diversity indices, such as SR and iSDI, are usually estimated on the basis of standardized field sampling or ground surveys, and the fact that detailed field inventories are time-consuming may limit the spatial extent of diversity surveys. Remote sensing techniques have the potential to play a valuable supporting role in the mapping of plant species diversity, and in the identification of habitat patches that may be of conservation interest [14] if, for example, spectral data correlate with species diversity or with vegetation properties that are associated with species diversity (*cf.* [15]).

Nagendra [16] identified three categories of methods for the assessment of species diversity using remotely sensed data: (1) mapping individual organisms or communities; (2) mapping habitat characteristics that are expected to be associated with species diversity; and (3) modeling-based methods by which species diversity is predicted from the direct relationship between spectral data and field-based measures of species diversity. Modeling-based approaches have been shown to be successful in the prediction of fine-scale plant species diversity using remote sensing data acquired with the help of hyperspectral sensors (sensors that collect data in many narrow and contiguous spectral bands) within a range of different grassland habitats and geographic regions [17–19]. The direct relationship between hyperspectral data and species diversity has also been examined using measures of the spatial variation of remotely sensed data (hereafter referred to as spectral heterogeneity). The spectral heterogeneity is expected to be associated with the environmental heterogeneity (the spectral variation hypothesis (SVH); [20]), and can, thus, be used as a proxy for species diversity (*cf.* [21]). Hyperspectral data have also been used, in combination with topographic data, for predicting plant distributions in French and Swiss alpine grasslands [22]. To our knowledge, no studies have modeled the direct relationship between hyperspectral data and plant species diversity in northern European grasslands.

In the present study, we explore the ability of hyperspectral remote sensing technology to characterize fine-scale plant species diversity in dry, grazed grassland habitats in an agricultural landscape on the Baltic island of Öland (Sweden). We compare the performance of two modeling-based approaches to the prediction of species diversity in 4 m × 4 m plots, using data from airborne HySpex hyperspectral imagers (415–2345 nm). We ask the following questions: can hyperspectral data be used to predict the SR and iSDI in dry grazed grasslands via the direct relationship between reflectance data and field-based measures of plant species diversity using (1) an analysis of reflectance, based on information from (i) all wavebands; and (ii) a subset of wavebands, analyzed with a partial least squares regression model (hereafter referred to as the spectral response approach); and (2) an analysis of spectral heterogeneity, based on the mean distance to the spectral centroid in an ordinary least squares regression model (hereafter referred to as the spectral heterogeneity approach)? We also investigate whether the possible relationship between hyperspectral data and species diversity is influenced by environmental conditions (grazing continuity, nutrient and moisture status, field-layer height, and soil- and litter-cover fractions).

## 2. Materials and Methods

### *2.1. Study Area and Site Selection*

The study area is located in the southeast of Sweden on the Baltic island of Öland (Figure 1a; centered on 56°40′49′′N, 16°33′58′′E) and covers approximately 22.5 km$^2$. The bedrock consists mainly of Ordovician limestone, and the area is characterized by a generally flat topography (*cf.* [9]). The climate on the island is maritime (mean annual temperature: 7 °C, mean annual precipitation: 468 mm) (*cf.* [9]). The present-day landscape consists of a mosaic of arable fields, deciduous forest, villages and grazed grasslands; the majority of the grasslands are grazed by cattle at varying intensities. The grassland sites in the landscape represent different stages of succession—ranging from young grasslands on recently abandoned arable fields to grassland sites with a history of grazing of more than 280 years.

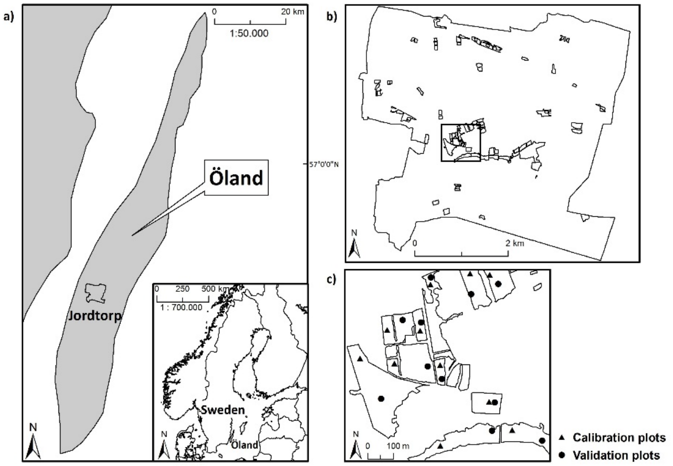


**Figure 1.** (**a**) The location of the study area on the Baltic Island of Öland, Sweden; (**b**) the distribution of grassland sites included in the present study ($n$ = 52); (**c**) an example of the distribution of field plots within some of the grassland sites.

A total of 299 grassland sites were identified within the study area, with the help of the most recent (2005) land-use map (B.C. Schmid, unpublished data) and field visits. An overlay analysis of land-use maps from different time periods was used to assign each of the grassland sites into one of three age classes within the arable-to-grassland succession (young grasslands: 5–14 years, intermediate-aged grasslands: 15–49 years, and old grasslands: ⩾50 years of grazing management). The overlay analysis was based on land-use maps from 2005 (B.C. Schmid, unpublished data), 1994, 1959 and 1730 (*cf.* [9]). The land-use map from 2005 was developed from a digital color infrared aerial orthophoto (0.5 m pixel size) geometrically corrected with the help of a digital terrain model. The land-use maps from 1994 to 1959 were developed from aerial photos (1994: color infrared, 1959: black and white) at the scale of 1: 30,000. The land-use map from 1730 was based on a set of large-scale (1: 4000), high-quality survey maps from the early 18th century (*cf.* [9]).

Within each of the 299 sites, we randomly positioned two coordinate points in open (*i.e.*, not covered by shrubs or trees) grassland vegetation with the constraints that they had to be at least 25 m apart, at least 13.5 m from the site boundary (to minimize edge effects in the vegetation), and at least 13.5 m from shrubs or trees that were higher than 50 cm (to minimize shading effects in the vegetation). A total of 239 out of the 299 grassland sites could accommodate these constraints. A hand-held differential global positioning system (GPS) receiver (Topcon GRS-1 GNSS, equipped with a PG-A1 external antenna; Topcon Corporation, Japan) was used to log the ground coordinates of the points.

Sixty sites (20 young, 20 intermediate, and 20 old) were randomly selected from the 239 sites. Within these 60 sites, a bioassay approach (*cf.* [23]) based on indicator species, such as *Sesleria caerulea* and *Molinia caerulea*, was used to identify sites with "dry" grassland vegetation, and to exclude moist grassland vegetation. A total of 52 sites (17 young, 18 intermediate, and 17 old) out of the 60 sites were characterized as dry grassland vegetation; these sites were used for the field-based vegetation and remote sensing sampling (Figure 1b).

### *2.2. Field Sampling*

#### 2.2.1. Vascular Plant Species Richness and Diversity

The fieldwork was carried out between 15 May and 15 July 2011. A 4 m × 4 m plot (divided into a grid of 16 sub-plots, each 1 m × 1 m), was centered over each of the two coordinate points, within each of the 52 chosen sites (Figure 1c). The presence of all non-woody vascular plants was recorded within each 1 m × 1 m sub-plot. The within-plot species diversity was characterized: (a) in terms of species richness (SR), calculated as the total number of vascular plant species present in a 4 m × 4 m plot and (b) in terms of the inverse Simpson diversity index (iSDI)

$$iSDI = 1/\sum_{i=1}^{S} p_i^2 \tag{1}$$

where S is the number of species in the plot and $p_i$ is the proportion of the $i$th species in a plot [24]. The inverse Simpson diversity index assigns a lower weight to rare species, thus emphasizing the most abundant species in the vegetation canopy [25]. Because the inverse Simpson diversity index characterizes the dominance structure of plant communities (which, in its turn, contributes to the spectral signal of vegetation canopies), it may be particularly informative in remote sensing studies [17]. From a statistical point of view, the number of species within a specific area represents discontinuous data, and thus has a Poisson or negative-binomial distribution. Because the partial least squares regression (PLSR) and ordinary least squares regression (OLSR) analyses used in the present study (Section 2.4) assume a normal distribution, we ln-transformed the SR values before analysis, so that the ln(SR) data approximate to a normal distribution. Ln(SR) and iSDI were calculated using the *vegan* package [26] in the R programming environment [27].

#### 2.2.2. Environmental Variables

Each of the 4 m × 4 m plots was assigned values for the following environmental variables: grazing continuity, nutrient and moisture status, field-layer height, and soil- and litter-cover fractions (Table A1). Soil nutrient and moisture status were assessed indirectly, with the help of Ellenberg indicator values for nutrient (Ellenberg N) and moisture (Ellenberg M) availability [28]. For each of the sampled plant species, values for Ellenberg N and M were extracted from the JUICE database [29]. A community-weighted mean value (CWM) [30] was calculated for N and M within each plot:

$$CWM\,(x) = \sum_i p_i \times x_i \tag{2}$$

where $p_i$ is the relative frequency of the $i$th species and $x_i$ is the Ellenberg indicator value of the $i$th species. The CWM for Ellenberg N (Ellenberg mN) and Ellenberg M (Ellenberg mM) were calculated using the *FD* package [31] in the R programming environment [27]. Field-layer height (cm) was estimated as the mean vegetation height at 100 points in the 1 m × 1 m sub-plot in the south-west corner of each 4 m × 4 m plot. Mean percentage covers of soil and litter fractions, respectively, for each of the 4 m × 4 m plots were based on the cover values (visually estimated by experienced field ecologists) within each of the 16 1 m × 1 m sub-plots.

### *2.3. Remote Sensing Data*

#### 2.3.1. Spectral Data Collection

Hyperspectral data were acquired by the company Terratec AS, Lysaker, at around solar noon on 9 July 2011. Two airborne HySpex hyperspectral imagers (Norsk Elektro Optikk, Lörenskog, Norway), VNIR-1600 and SWIR-320m-e, were used in the push broom scanning mode (Table A2). The flight altitude was approximately 1500 m and the weather conditions were cloud-free. A total of 25 flight lines (conducted either from north to south or from south to north, to minimize illumination effects) were recorded.

#### 2.3.2. Preparation of Spectral Data

Wavebands between 962–985 nm, 1322–1496 nm, 1803–2050 nm, and 2351–2501 nm were deleted from the hyperspectral data set because of strong atmospheric interference or detector overlap, leaving 245 wavebands that were used for further analysis. ATCOR-4 software [32], which is based on the radiative transfer model MODTRAN 5 [33], was used for atmospheric and topographic corrections of the hyperspectral data. The atmospheric correction was carried out using the settings for rugged terrain, the desert aerosol model, a water vapor column of 1.0 $\text{g}\cdot\text{m}^{-2}$, visibility of 28.4 km, and an ozone concentration of 330 Dopson units. The radiance was converted into reflectance using the Fontenla-2011 solar irradiance spectrum [34]. The images were orthorectified with an accuracy of approximately 0.3 m, by the data providers, using the PARGE software [35]. To match the spatial resolution of the two HySpex sensors, the spectral data originating from the VNIR-1600 spectrometer were resampled to a spatial resolution of 1 m, using a triangulated nearest neighborhood method, and a spectral resolution of 6 nm, using locally weighted scatterplot smoothing (LOESS) interpolation [36]. High frequency noise in the spectral data was reduced by using a cubic Savitzky–Golay filter [37] with a kernel size of 21 nm. The resampling was done using the *signal* [38] and *raster* packages [39] and the filtering was done using *hyperSpec* package [40] in the R statistical environment [27].

#### 2.3.3. Calculating Mean Spectral Reflectance and Spectral Heterogeneity

Vector polygons of the grassland sites were overlaid onto the hyperspectral imagery, and each individual 4 m × 4 m field plot was located on the HySpex image using the GPS coordinates taken during the field work. Although we used a GPS receiver with a high accuracy, GPS errors may exist. To account for possible positional uncertainties, a pixel window of 8 × 8 pixels (8 m × 8 m) was centered on each of the 4 m × 4 m field plots and the reflectance of each pixel ($n$ = 64) within each pixel window from the 245 individual wavebands was extracted.

Prior to the spectral response analyses, the reflectance was $\log_{10}(1/R_\lambda)$ transformed (where $R_\lambda$ is the reflectance at each waveband). The mean transformed reflectance of each pixel window was obtained by calculating the mean spectral value of the pixels ($n$ = 64) for each of the 245 individual HySpex wavebands. $\text{Log}_{10}(1/R_\lambda)$ transformed reflectance shows a near-linear relationship with the concentration of absorbing land surface components [41], and may provide important information on the environmental and ecological processes underlying any potential associations between hyperspectral data and plant species diversity [42].

Several different methods have been used to compute spectral heterogeneity in remotely sensed data. For example, Viedma [43] used methods based on spectral texture data, while Heumann [44] applied an approach using measures of statistical dispersion to represent spectral diversity. Warren [45] compared two categories of spectral heterogeneity metrics—one category calculated with the help of principle component analysis (PCA) and one category developed from semivariogram descriptors—and found that both types of metrics performed equally well as predictors of species diversity. Oldeland [17] and Rocchini [46] also used spectral heterogeneity calculated with the help of PCA to model species diversity. Following Oldeland [17] and Rocchini [46], we applied a PCA-based approach to calculate spectral heterogeneity. The PCA was conducted on the spectral data set ($n$ = 245) within each pixel window consisting of 64 pixels, using untransformed reflectance data. The spectral heterogeneity was calculated as the mean of the Euclidean distances between each of the 64 pixels and the centroid of the pixel-cloud within the PCA space using the first five principal components, which summarized at least 97% of the total spectral variation. We predicted that increasing within-plot environmental heterogeneity should be accompanied by an increasing mean distance to the spectral centroid (see [17,46]).

### 2.4. Data Analysis

The 104 plots were divided into a calibration subset and a validation subset by randomly assigning the two plots from each of the 52 grassland sites to one or other of the two subsets (Figure 1c). Two plots (one each from the calibration and validation subset) that had an unusually low reflectance in the spectral range 900–1300 nm were excluded from further analyses. The exclusion of the two plots resulted in a total of 102 plots (51 plots in each subset) used for the data analysis (Figure 2).

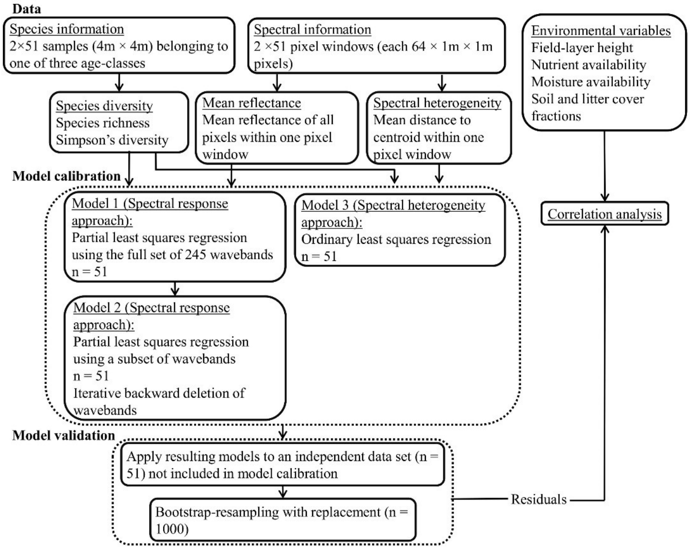


**Figure 2.** Schematic overview of the workflow used in the present study.

The Pearson correlation between the species diversity indices (ln(SR) and iSDI) and the mean spectral reflectance of each pixel window for individual wavebands were analyzed for all plots ($n = 102$). In the spectral response approach, the relationships between hyperspectral data and field-observed ln(SR) and iSDI were investigated for the plots in the calibration subset, using "leave-one-out" cross-validated PLSR of (i) the full set of 245 (Model 1) and (ii) a subset (Model 2) of HySpex wavebands (Figure 2). Earlier studies showed that the exclusion of wavebands that provide little information related to the response variable improves the PLSR-based prediction of vegetation variables [47]. We used iterative variable deletion to identify the wavebands used in Model 2 (*i.e.*, the bands most important for prediction of species diversity). Marten's uncertainty test [48] and the variable importance in projection (VIP) values [49] were calculated for each waveband to identify the least important wavebands (variables) within each iterative step. Variables that were non-significant and had a VIP value lower than 0.8 were deleted before recalibrating the model. This procedure was repeated until none of the remaining variables could be deleted. The relationships between hyperspectral data and field-observed species diversity were also examined using spectral heterogeneity and OLSR analysis (Model 3) (Figure 2). The validation subset was used to evaluate the regression models of the calibration subset.

### 2.4.1. Partial Least Squares Regression Analysis (PLSR)—Models 1 and 2 (Spectral Response Approach)

A PLSR analysis [49] between each of the two diversity indices (dependent variables) and the average spectral reflectances (explanatory variables) was carried out. PLSR is a method of multivariate analysis that is suitable for the analysis of data sets that include a larger number of (highly correlated) explanatory variables than samples [50]—which is often the case in remote sensing–based species diversity studies. Several studies have shown that the PLSR method outperforms other methods (e.g., OLSR) when analyzing highly co-linear hyperspectral remote sensing data sets [51]. The PLSR algorithm attempts to find latent variables (LVs) that summarize the variation in the explanatory matrix and, at the same time, maximize the covariance with the dependent variable (see [52]).

The optimal number of LVs needs to be identified in order to avoid model over-fitting in PLSR analyses, but there is, at present, no consensus about the best method to use [53]. The number of LVs is usually determined by a cross-validation procedure, which is used to find the lowest cross-validated root mean square error ($\mathrm{RMSE_{CV}}$) of the PLSR model. It has been shown that the use of the global minimum of the $\mathrm{RMSE_{CV}}$ can lead to erroneous and over-fitted prediction models [53]. To avoid over-fitting, the first local minimum of the $\mathrm{RMSE_{CV}}$ is usually used [54]. However, this approach may lead to model under-fitting, if the minimum error results in a negative coefficient of determination for the cross-validated predicted dependent variables ($R^2_{\mathrm{CV}}$). Negative values of $R^2_{\mathrm{CV}}$ indicate that the model residuals exceed those obtained from the mean observation as predictors. We used the number of LVs which resulted in the first local minimum absolute $\mathrm{RMSE_{CV}}$ ($\mathrm{aRMSE_{CV}}$) and a positive $R^2_{\mathrm{CV}}$ value. To allow comparison between Models 1 and 2, we normalized the $\mathrm{aRMSE_{CV}}$ ($\mathrm{nRMSE_{CV}}$, %) values by the range of the field-observed values for ln(SR) and iSDI of the calibration subset.

The PLSR was run in two ways: (i) using all wavebands from the full reflectance spectrum ($n = 245$; Model 1); and (ii) using the subset of spectral wavebands that were most important for the prediction of species diversity (Model 2). Cross-validation was carried out using "leave-one-out" cross-validation, with each plot being excluded in turn, and the calibration model based on the remaining plots used to predict the excluded plot. Both the models (Model 1 and Model 2) for each of the two species diversity indices were validated with the plots from the independent validation subset. The root mean square error of the predicted values (absolute value: $\mathrm{aRMSE_P}$) and the squared correlation coefficient ($R^2_{\mathrm{P}}$) between the field-observed and predicted ln(SR) and iSDI for the validation subset were used to evaluate the predictive qualities of the models. The $\mathrm{aRMSE_P}$ was normalized ($\mathrm{nRMSE_P}$, %) by the range of the field-observed ln(SR) and iSDI values of the validation subset. A good validation result is characterized by low values for both $\mathrm{aRMSE_P}$ and $\mathrm{nRMSE_P}$ and high values for $R^2_{\mathrm{P}}$, indicating that the hyperspectral data have a good ability to predict the species diversity indices. To assess the robustness of the predictive performance of the calibrated PLSR models, the validation subset was bootstrapped 1000 times (with replacement), building 1000 "new" validation subsets. The calibration model was applied to each of the 1000 validation subsets and the $\mathrm{nRMSE_P}$ (%) and $R^2{}_\mathrm{P}$ between the field-observed and predicted ln(SR) and iSDI were calculated for each validation subset. Mean $\mathrm{nRMSE_P}$ (%) and $R^2{}_\mathrm{P}$ values, and 95% confidence levels were then calculated for the 1000 validation subsets.

### 2.4.2. Residual Analysis of Models 1 and 2 (Spectral Response Approach)

The residuals of Models 1 and 2 were separately correlated with the environmental variables Ellenberg mN and mM, field-layer height, bare ground and litter-cover fractions (Table A1), to examine whether the relationships between hyperspectral data and species diversity was influenced by the environmental variables characterizing the plots. The significance of the correlations was tested using a two-sided Student's *t*-test. Because multiple tests may result in an increased risk of Type I error, the significance values of the correlations were assessed after Bonferroni correction. To investigate whether the relationships between hyperspectral data and species diversity were influenced by the age-class of plots (young grasslands: 5–14 years, intermediate-aged grasslands: 15–49 years, and old grasslands with $\geqslant$50 years of grazing continuity), we examined if there were significant differences between the residuals associated with each of the three grassland age-classes. Because the residuals were not normally distributed we used the Kruskal-Wallis test [55]. All analyses were conducted in the R statistical environment [27], using the *pls* package [54].

### 2.4.3. Ordinary Least Squares Regression Analysis (OLSR) and Reduced Major Axis Regression—Model 3 (Spectral Heterogeneity Approach)

Previous studies successfully applied OLSR to examine the relationship between spectral heterogeneity and plant species diversity [17]. In the present study, separate OLSR analyses of the relationships between each of the species diversity indices (dependent variables) and the spectral heterogeneity (explanatory variable) were carried out on the calibration subset (Model 3) (Figure 2). Model 3 was tested on the validation subset using a reduced major axis (RMA) regression. Curran [56] showed that RMA is an appropriate method for the remote sensing–based prediction of grassland variables, in cases where there is no available information on measurement error. The $\mathrm{nRMSE_P}$ and $R^2{}_\mathrm{P}$ were used to evaluate the performance of the validation.

## 3. Results

The summary statistics for the dependent variables for the plots within each grassland age-class (young, intermediate-aged, and old grasslands) are presented in in Table A3. The Pearson's correlation coefficients between the ln(SR) and the iSDI were significant for both the calibration ($r = 0.98$, $p < 0.001$) and validation subsets ($r = 0.97$, $p < 0.001$). There were significant negative correlations ($p < 0.05$) between the reflectance associated with wavebands in the near-infrared (758–1316 nm) (NIR) part of the electromagnetic spectrum and the ln(SR) (Figure 3). There were positive but non-significant correlations between the reflectance at wavebands in the blue (415–499 nm) and red (602–752 nm) parts of the spectrum and the dependent variables, and (non-significant) negative correlations between the reflectance at wavebands in the green (505–595 nm) and SWIR (1502–2345 nm) parts of the spectrum and the dependent variables (Figure 3).

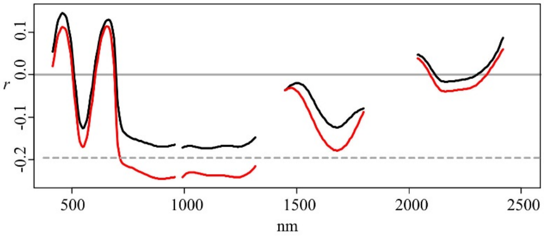


**Figure 3.** Pearson's correlation coefficients ($r$) between single wavebands and the species richness ln(SR) (red), and the inverse Simpson's diversity index iSDI (black) for the whole data set ($n$ = 102). Correlations below the dotted line are significant ($p < 0.05$).

## *3.1. Spectral Reflectance—Models 1 and 2 (Spectral Response Approach)*

### 3.1.1. PLSR Using the Full Set of 245 HySpex Wavebands—Model 1

The inclusion of seven LVs gave the first local minimum absolute aRMSE$_{CV}$ in the PLSR model developed from the calibration subset (Model 1; Figure 2), for both the ln(SR) (aRMSE$_{CV}$ = 0.34) and the iSDI (aRMSE$_{CV}$ = 8.87) (Table 1).

**Table 1.** Summary of the ability of PLSR models, based on spectral reflectance using the full set of wavebands (Model 1) or a subset of wavebands (Model 2), to predict the species richness (ln(SR)) and the inverse Simpson's diversity index (iSDI). The cross-validated error of the calibration models ($n$ = 51) is indicated by the absolute (aRMSE$_{CV}$) and normalized RMSE$_{CV}$ (nRMSE$_{CV}$, %). LV indicates the number of latent variables used in the PLSR models. The absolute and normalized prediction errors (aRMSE$_P$, nRMSE$_P$ (%)) indicate the ability of the model to predict the observed species diversity measure. The squared correlation ($R^2_P$) indicates the fit between the predicted and observed diversity values from the validation subset ($n$ = 51).

| | | **aRMSE$_{CV}$** | **nRMSE$_{CV}$** | **LV** | **aRMSE$_P$** | **nRMSE$_P$** | **$R^2_P$** | **No. of Wavebands** |
|---|---|---|---|---|---|---|---|---|
| **ln(SR)** | Model 1 | 0.34 | 21% | 7 | 0.29 | 19% | 0.43 | 245 |
| | Model 2 | 0.37 | 23% | 5 | 0.34 | 22% | 0.19 | 25 |
| **iSDI** | Model 1 | 8.87 | 23% | 7 | 6.77 | 20% | 0.45 | 245 |
| | Model 2 | 9.29 | 25% | 4 | 7.07 | 21% | 0.40 | 35 |

The correlations between the field-observed and predicted measures of species diversity were significant for both the ln(SR) ($R^2_P$ = 0.43, $p < 0.001$) and the iSDI ($R^2_P$ = 0.45, $p < 0.001$) (Table 1, Figure 4a,b). The nRMSE$_P$ values were approximately 20% for both the ln(SR) (nRMSE$_P$ = 19%) and the iSDI (nRMSE$_P$ = 20%) (Table 1, Figure 4a,b). Out of the 245 wavebands used in Model 1, 25 bands were most important for the prediction of the ln(SR) (Figure 5a, Table A4), while 35 bands were most important for the prediction of the iSDI (Figure 5b, Table A4). The relationships between the residuals associated with the prediction of both dependent variables (using Model 1) and the values for individual environmental variables (Ellenberg mN and mM, field-layer height, and soil- and litter-cover fractions) were non-significant (Figure 6a,b). There were significant ($p < 0.05$), positive associations between the residuals (in the prediction of both the ln(SR) and the iSDI) and the grassland age: the shorter the grazing continuity of the grassland, the more the values for ln(SR) and iSDI were overestimated (negative residuals) (Figure 6c,d).

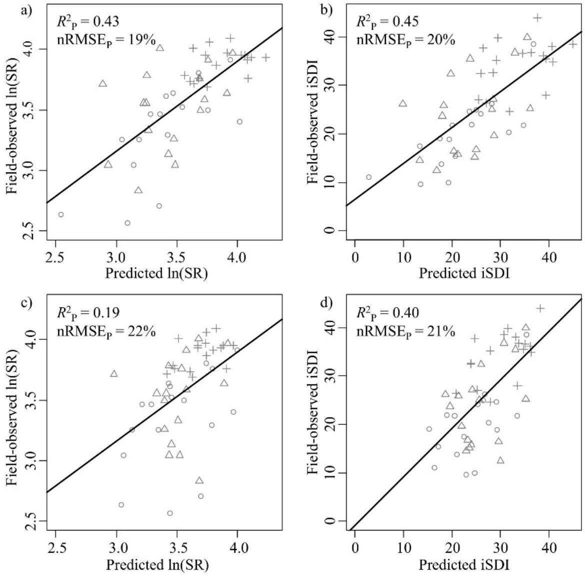


**Figure 4.** Correlations between field-observed and predicted (**left column**) species richness (ln(SR)) and (**right column**) inverse Simpson's diversity (iSDI) for the validation subset ($n$ = 51). (**a**,**b**) show the field-observed *versus* the predicted correlations for the PLSR model based on the full set of wavebands (Model 1) ($n$ = 245); (**c**,**d**) show the field-observed *versus* the predicted correlations for the model based on a subset of wavebands (Model 2) ($n$ = 25 (for ln(SR)) or 35 (for iSDI)). The normalized prediction error (nRMSE$_P$, %) indicates the quality of the model in predicting the observed species diversity measure, and the squared correlation ($R^2{}_P$) indicates the fit between the predicted and observed diversity value. The age-class of the grassland plots is also displayed (key: ○ young, Δ intermediate, and + old). Black lines indicate the relationship between the predicted and the measured values.

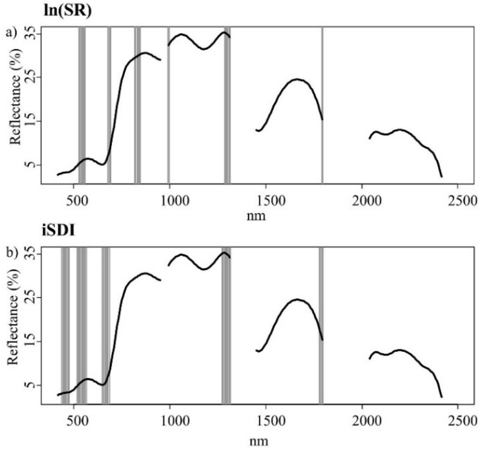


**Figure 5.** Important wavebands (grey bars) selected with the help of an iterative variable deletion procedure, for estimating (**a**) the species richness (ln(SR)) and (**b**) the inverse Simpson's diversity index (iSDI) in grassland plots using the calibration subset ($n$ = 51). The black line represents the mean spectral reflectance curve for grassland plots in the whole data set ($n$ = 102).

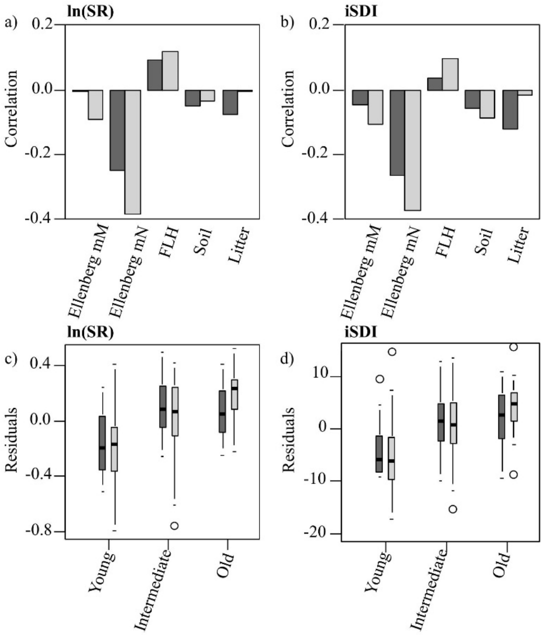


**Figure 6.** (**a**,**b**) Pearson's correlation coefficients of the residuals of the PLSR models' (Model 1 = dark; Model 2 = light) predictions of (**a**) the species richness (ln(SR)) and (**b**) the inverse Simpson's diversity index (iSDI) with different environmental variables (moisture availability, Ellenberg mM; nutrient availability, Ellenberg mN; field-layer height, FLH; cover of bare ground, Bare ground; and cover of litter, Litter); (**c**,**d**) Distribution of the residuals of (**c**) the species richness (ln(SR)) and (**d**) the inverse Simpson's diversity index (iSDI), within the three grassland age-classes, predicted by Model 1 (dark) and Model 2 (light).

### 3.1.2. PLSR Using the Subset of HySpex Wavebands—Model 2

Using the subsets of HySpex wavebands (Model 2; Figure 2), the inclusion of five LVs gave the first local minimum $\text{aRMSE}_{\text{CV}}$ in the PLSR model for ln(SR) ($\text{aRMSE}_{\text{CV}} = 0.37$) (Table 1). For the prediction of iSDI, the inclusion of four LVs gave the first minimum $\text{aRMSE}_{\text{CV}}$ ($\text{aRMSE}_{\text{CV}} = 9.29$) (Table 1). The correlations between the field-observed and predicted measures of plant diversity for the validation subset were significant for ln(SR) ($R^2_{\text{P}} = 0.19$, $p < 0.001$) and iSDI ($R^2_{\text{P}} = 0.40$, $p < 0.001$) (Table 1, Figure 4c,d). The $\text{nRMSE}_{\text{P}}$ values were above 20% for ln(SR) ($\text{nRMSE}_{\text{P}} = 22\%$) and iSDI ($\text{nRMSE}_{\text{P}} = 21\%$) (Table 1, Figure 4c,d). In the Model 2 approach, there were significant negative correlations ($p < 0.05$) between the residuals (associated with the prediction of ln(SR) and iSDI) and the Ellenberg mN: the higher the Ellenberg mN, the more the values for ln(SR) and iSDI were overestimated (negative residuals) (Figure 6a,b). There were also significant ($p < 0.05$) associations between the residuals (in the prediction of ln(SR) and iSDI) and the grassland age: the shorter the grazing continuity of the grassland, the more the values for ln(SR) and iSDI were overestimated (negative residuals) (Figure 6c,d).

### 3.1.3. The Robustness of the Prediction Models (Models 1 and 2)

The bootstrapping procedure revealed that the mean $R^2_{\text{P}}$ were higher and the mean $\text{nRMSE}_{\text{P}}$ were lower when using the PLSR-based models developed from the full set of bands (Model 1) than when using the models developed from the subset of bands (Model 2) to predict both the dependent variables (Table 2).

**Table 2.** Bootstrap results showing the ability of the PLSR models based on the spectral reflectance in the full set of wavebands (Model 1) and the subset of wavebands (Model 2) to predict the species richness (ln(SR)) and the inverse Simpson's diversity index (iSDI). Mean $R^2_P$ and mean $\text{nRMSE}_P$ (%) are the average squared correlation coefficients and normalized prediction errors for the validation subset, based on 1000 bootstraps. The 95% confidence limit indicates the upper and lower confidence intervals of the mean values.

| | | **Mean $R^2_P$** | **95% Confidence Limit** | **Mean $\text{nRMSE}_P$** | **95% Confidence Limit** |
|---|---|---|---|---|---|
| **ln(SR)** | Model 1 | 0.39 | ±0.010 | 20% | ±0.1% |
| | Model 2 | 0.17 | ±0.009 | 23% | ±0.1% |
| **iSDI** | Model 1 | 0.43 | ±0.008 | 21% | ±0.1% |
| | Model 2 | 0.38 | ±0.007 | 22% | ±0.1% |

### *3.2. Linear Regression Based on Spectral Heterogeneity—Model 3 (Spectral Heterogeneity Approach)*

When the regression models (Model 3; Figure 2) developed from the calibration subset using the spectral heterogeneity approach were applied to the validation subsets, the results showed non-significant relationships between both the field-observed and predicted SR ($R^2 = 0.06$, $p > 0.05$) and the field-observed and predicted iSDI ($R^2 = 0.04$, $p > 0.05$). The $\text{nRMSE}_P$ values were above 30% for both ln(SR) ($\text{nRMSE}_P = 31\%$) and iSDI ($\text{nRMSE}_P = 35\%$).

## 4. Discussion

### *4.1. The Relationship between Hyperspectral Reflectance Measurements and Plant Species Diversity (Spectral Response Approach—Models 1 and 2)*

The spectral response approach, using both Model 1 and Model 2, resulted in good predictions, with a relative error ($\text{nRMSE}_P$) of approximately 20% of both ln(SR) and iSDI (Figure 4, Tables 1 and 2). PLSR-based approaches have been used in the past to predict plant species diversity in grassland ecosystems with the help of hyperspectral remote sensing [18,19]. The study by Fava [18] used ground-based hyperspectral measurements and PLSR to assess plant species diversity (species richness and Shannon diversity) in alpine meadow systems and showed a somewhat lower average error (<15%) for the plant diversity estimates than was found in the present study. The relatively higher error values in the present study may reflect the fact that airborne spectrometric measurements are affected by noise caused by, for example, effects of atmospheric scattering. Carter [19] used airborne spectrometric measurements (400–2500 nm) to estimate species richness in mesic grasslands, and revealed somewhat weaker relationships between the spectral data and species diversity than those in the present study. The analyses in Carter [19] were based on simple linear regression using individual wavebands and band ratios, whereas our analyses used a PLSR-based approach involving multiple bands. The use of a limited number of wavebands may result in poorer relationships between remote sensing data and vegetation parameters compared with the use of many spectral bands. Although much information in a hyperspectral data set may be redundant, important spectral information may nevertheless be lost when only a small number of wavebands are used to predict vegetation variables [57].

### *4.2. The Most Important Hyperspectral Wavebands for Predicting Plant Species Diversity Using the Spectral Response Approach (Models 1 and 2)*

Previous studies revealed that the soils of young grasslands with a short history of grazing management often contain high levels of nitrogen and phosphorus as a result of their recent use as fertilized arable fields [58]. Over time, the continuous removal of above-ground biomass by grazing animals leads to a progressive decrease in soil nutrient levels [58]. Nutrient-poor conditions promote the maintenance and establishment of plant species with a low competitive ability. Grasslands with a long grazing continuity, such as the old grasslands in our study (Table A3), are often associated with a lower above-ground biomass and a higher plant species diversity than young grasslands [59,60]. An increase in the amount of above-ground biomass is often accompanied by an increase in the reflectance in the NIR region of the spectrum because of increased multiple scattering within the canopy (e.g., [61]). Relationships between species richness and above-ground biomass have been extensively studied (e.g., [12]) and have increasingly been used to interpret the relationship between reflectance and species diversity measures in grasslands (e.g., [18,62]. For example, [18] explained the relationship between reflectance in the NIR wavebands and species richness in terms of a negative relationship between biomass and plant diversity. In the present study, the identification of key wavebands for predicting species diversity in the PLSR models 1 and 2 showed that several NIR wavebands were important for the assessment of species diversity (Figure 5). Significant negative correlations between reflectance and diversity in the NIR spectral region (Figure 3) indicate that the species diversity increased as the above-ground biomass decreased. Other studies have shown that the visible wavebands may also contain important information for the assessment of plant species richness (e.g., [18]). The relationships we identified between wavebands in the visible and SWIR parts of the spectrum and the species diversity are in line with those findings (Figure 5). The positive associations that we observed between reflectance and diversity in the chlorophyll absorption regions of the spectrum (blue and red) and the negative associations with the green and the water absorption parts (SWIR bands) of the spectrum (Figure 3) suggest that the species-poor plots were characterized by higher levels of plant chlorophyll content and vegetation water content than the species-rich plots.

### *4.3. Residual Analysis of Models 1 and 2 (Spectral Response Approach)*

The lack of significant relationships between the residuals of Model 1 and the environmental variables indicates that none of these variables had an effect on the species diversity that was not predicted by the full set of hyperspectral wavebands (Figure 6a,b). In contrast, the significant relationships between the residuals for Model 2 and nutrient status suggest that the unexplained variance in the models based on a subset of wavebands may be related to within-plot soil nutrient availability (Figure 6a,b). In the present study, the use of a subset of wavebands (Model 2) may have resulted in the loss of spectral information on the variation in nutrient status, which may, at least partly, explain the poorer prediction of species diversity in Model 2 compared with that for Model 1 (Figure 4, Tables 1 and 2).

For both Model 1 and Model 2, the residuals differed significantly between the age-classes, indicating that the unexplained variance was related to the grazing continuity within the plots (Figure 6c,d). Thus, in Model 1 and Model 2, the grazing continuity had effects on the species diversity that were not predicted from the spectral data. Although young grasslands are often associated with higher amounts of above-ground biomass and lower species diversity than old grasslands [60], the current grazing intensity is expected to influence the amount of biomass within the grasslands. Some of the young grassland sites may be subject to relatively high grazing intensity, and thus be characterized by a relatively low amount of above-ground biomass, which may explain the overestimation of the predicted species diversity in some of those plots (Figure 4). The overestimation of ln(SR) in low ln(SR) plots tended to be more accentuated in the Model 2 approach than in the Model 1 approach (Figure 4a,c). The full set of wavebands may include individual bands that were important for the prediction of particular (heavily grazed) low ln(SR) plots—and these bands were excluded in the backward deletion procedure which may explain the poorer $R^2$ value in Model 2 than in Model 1. A higher number of wavebands were retained in Model 2 for the iSDI response variable than for the ln(SR) variable, and there was little difference between the Model 1 and Model 2 results for iSDI (Figure 4b,d). Although the number of LVs may not always be a good measure of model complexity [63], the lower number of LVs in the Model 2 analyses (Table 1) indicate that the Model 2 analyses were more parsimonious than the Model 1 analyses.

### *4.4. The Relationship between Spectral Heterogeneity and Plant Species Diversity (Spectral Heterogeneity Approach—Model 3)*

Relationships between spectral heterogeneity and measures of species diversity have been investigated in recent studies [17,46]. However, these studies were carried out at relatively coarse spatial scales (sampling areas $\geqslant$100 m$^2$) [17,46]. Large sampling areas may be characterized by a greater variability in the environmental conditions than smaller areas (*cf.* [64]), and the strength of the relationship between spectral heterogeneity and plant species diversity has been shown to increase with the size of the spectral and vegetation sampling areas [43]. In the present study, spectral heterogeneity failed to predict species diversity within our comparatively small (4 m × 4 m) vegetation plots. Whereas our study was based on a pixel size of 1 m, previous studies of patterns of plant species co-existence within the same study area revealed heterogeneity in the plant community composition at a scale of decimeters rather than meters [65]. The failure to detect a significant relationship between spectral variability and species diversity in our study is, therefore, likely to reflect the fact that the scale (dictated by the pixel size) at which we worked was too large to detect ecologically relevant heterogeneity in the grassland community composition.

### *4.5. Limitations*

One possible explanation for the discrepancy between the field-observed and predicted species diversity, in both the spectral response approach and the spectral heterogeneity approach, may be the mismatch between the size of the pixel windows (8 m × 8 m) and the size of the vegetation plots (4 m × 4 m). Although the paired spectral and vegetation plots belong to the same grassland age categories, and there is no variation in grazing intensity between the spectral plot and its corresponding vegetation plot, we cannot exclude the possibility that the relationship may have been influenced by environmental conditions in the 2 m zone surrounding the vegetation plots. However, species diversity levels at different spatial scales often show a positive relationship within semi-natural grasslands (e.g., [66,67]). For example, within Swedish semi-natural grassland the species diversity at the 1 dm × 1 dm scale was explained by the size of the species pool at the 2 m × 2 m scale [66].

## 5. Conclusions

The monitoring of biodiversity is regarded as a central task for nature conservation, and hyperspectral remote sensing has recently been identified as a method that has the potential to make a substantial contribution to the mapping of habitat and species diversity at local to regional scales [68]. The present study presents a novel methodology for the assessment of fine-scale (4 m × 4 m) vascular plant species diversity in dry grasslands based on hyperspectral data obtained with the help of airborne spectrometers covering 414 to 2501 nm. We used two different approaches to evaluate the ability of hyperspectral measurements to predict fine-scale grassland species diversity (characterized with the help of the species richness (SR) and the inverse Simpson's diversity index (iSDI)). The spectral response approach included information on reflectance based on (i) all wavebands (Model 1), and (ii) a subset of wavebands (Model 2), input into a partial least squares regression (PLSR) model. The spectral heterogeneity approach was based on the spectral variation hypothesis, and included an analysis of spectral variation, based on the mean distance to the spectral centroid, in an ordinary least squares regression model (Model 3).

Our study demonstrates that a spectral response approach using airborne hyperspectral data can be used to predict fine-scale species diversity in dry grasslands. The relationships between the field-observed and predicted measures of plant species diversity were significant for both the SR and the iSDI with a normalized root mean square error of approximately 20% for the predicted values of both the diversity indices. The PLSR-based approach allows a large number of hyperspectral wavebands to be compressed into a few latent variables (LVs) while decreasing the risk of model overfitting. Although the average prediction quality for both SR and iSDI was poorer for the Model 2 procedure than for the Model 1 procedure, the lower number of LVs in the Model 2 analyses indicated that the Model 2 analyses were more parsimonious. The prediction quality of the PLSR algorithm is dependent on the optimal selection of LVs used in the final prediction model. Although there are different ways of selecting LVs, there has been no systematic comparison of the performance of the different approaches. We suggest, therefore, that alternatives to the first-local-minimum rule—such as the total minimum cross-validated error or an overall F-test of the loss function [53]—should be evaluated further in future studies. There was a negative correlation between the reflectance in the NIR spectral region and species diversity, indicating that the species diversity increased as the above-ground biomass decreased. Although the prediction errors of the two PLSR models derived from the spectral response approach are low for both the species diversity indices, a certain amount of variation within the predicted diversity indices remained unexplained in our study. We suggest that the unexplained variance in the predicted species diversity may, at least in part, result from between-site variation in grazing intensity (particularly in the younger grasslands on recently abandoned arable fields) that results in between-site differences in the amount of biomass.

The spectral heterogeneity approach, using spectral variability as a proxy for habitat heterogeneity, was unable to predict species diversity. Our results, together with results from earlier ecological studies [65], suggest that the relevant scale for the investigation of the relationships between environmental heterogeneity and fine-scale grassland species diversity in our study system may be smaller than the 1 m × 1 m pixels used in the study. We suggest that future studies should examine a wide range of pixel sizes to identify the scale, or scales, at which a relationship between environmental heterogeneity and species diversity can be identified.

In the present study, we used remotely sensed data acquired at a single time-point in July. If leaf senescence in response to summer drought is associated with lowered levels of spectral variation within and between the grassland plots, then a multi-temporal approach might improve the ability to predict grassland species diversity with the help of remotely sensed data. The use of unmanned aerial vehicles (UAVs), which can provide high levels of both spatial and temporal resolution, is attracting increasing attention within the field of fine-scale remote sensing (e.g., [44,68]). Future studies should examine the potential use of UAVs to deliver improved spectral data that can be used in the assessment of grassland species diversity.

**Acknowledgments:** We thank Sofia Pallander, Andreas Press and Johan Rydlöv for field assistance. We are grateful to Oskar Löfgren, Heather Reese, Duccio Rocchini and Susanne Thulin for their helpful comments on an earlier version of the manuscript. The research station "Station Linné" at Ölands Skogsby was used as a base for the fieldwork. We are grateful to the anonymous reviewers for their valuable comments on an earlier version of the manuscript.The study is a contribution to the Lund University Strategic Research Area "Biodiversity and Ecosystem Services in a Changing Climate" (BECC). The study was financed by grants from The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) to Karin Hall and Honor C. Prentice.

**Author Contributions:** T.M., J.D., B.C.S., H.C.P. and K.H. planned and participated in the field work. T.M., J.D. and K.H. planned the collection of remotely sensed data. B.C.S. processed the land-use maps into a time-classified land-use map. T.M. processed and analyzed the data. All authors were involved in the research design, the interpretation of results, and the preparation of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Mean values and standard deviations (stdev) for the environmental variables, measured in 4 m × 4 m plots belonging to the validation subset ($n$ = 52).

| Environmental Variables | Mean | Stdev |
|---|---|---|
| Ellenberg M | 3.55 | 0.53 |
| Ellenberg N | 3.57 | 0.84 |
| Field-layer height (cm) | 3.56 | 4.00 |
| Soil-cover (%) | 6.13 | 8.18 |
| Litter-cover (%) | 7.29 | 17.01 |

**Table A2.** Technical characteristics of spectral sensors used in the study.

| | VNIR Sensor | SWIR Sensor |
|---|---|---|
| **Sensor name** | VNIR-1600 | SWIR-320m-e |
| **Spectral coverage (nm)** | 415–992 | 967–2501 |
| **Spectral sampling interval (nm)** | 3.7 | 6.0 |
| **Spectral bands** | 160 | 256 |
| **Field of view (FOV)** | 17 | 13.5 |
| **Ground sampling distance (m)** | 0.5 | 1.0 |
| **Radiometric resolution (bit)** | 12 | 14 |

**Table A3.** Summary statistics for the ln-transformed species richness ln(SR) and the inverse Simpson's diversity index (iSDI) values within the calibration and validation subsets for the 4 m × 4 m plots within each of the three grassland age-classes (young, intermediate-aged and old grasslands). The five most abundant species recorded within each grassland age-class are also presented.

| | | ln(SR) | | | | iSDI | | | | Five Most Abundant Species |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **Mean** | **Stdev** | **Min** | **Max** | **Mean** | **Stdev** | **Min** | **Max** | |
| Young | Calibration (*n* = 16) | 3.28 | 0.41 | 2.64 | 4.23 | 19.41 | 9.66 | 10.10 | 47.97 | *Taraxacum* agg., *Dactylis glomerata, Poa pratensis, Lolium perenne, Convolvulus arvensis* |
| | Validation (*n* = 17) | 3.34 | 0.40 | 2.56 | 3.91 | 20.00 | 7.13 | 9.68 | 38.60 | |
| Intermediate | Calibration (*n* = 18) | 3.59 | 0.33 | 2.89 | 4.17 | 25.70 | 7.73 | 13.49 | 44.14 | *Poa pratensis, Dactylis glomerata, Taraxacum* agg., *Festuca rubra, Ranunculus bulbosus* |
| | Validation (*n* = 17) | 3.51 | 0.35 | 2.83 | 4.01 | 24.03 | 8.44 | 12.42 | 39.95 | |
| Old | Calibration (*n* = 17) | 3.87 | 0.18 | 3.47 | 4.11 | 34.49 | 6.05 | 20.63 | 43.21 | *Plantago lanceolata, Galium verum, Achillea millefolium, Ranunculus bulbosus, Poa pratensis* |
| | Validation (*n* = 17) | 3.89 | 0.12 | 3.69 | 4.09 | 34.38 | 5.24 | 24.62 | 43.97 | |

**Table A4.** Number of important wavebands selected with the help of an iterative variable deletion procedure, for predicting the ln-transformed species richness ln(SR) and the inverse Simpson's diversity index (iSDI) in grassland plots using the calibration subset (*n* = 51).

| ln(SR) | | iSDI | |
|---|---|---|---|
| **Number of Wavebands** | **Part of the Electromagnetic Spectrum** | **Number of Wavebands** | **Part of the Electromagnetic Spectrum** |
| 6 | Green (529–559 nm) | 7 | Blue (439–475 nm) |
| 3 | Red (680–692 nm) | 9 | Green (517–565 nm) |
| 14 | NIR (818–1316 nm) | 7 | Red (650–686 nm) |
| 2 | SWIR (1791, 1797 nm) | 8 | NIR (1274–1316 nm) |
| | | 4 | SWIR (1779–1797 nm) |

## References

1. Habel, J.; Dengler, J.; Janišová, M.; Török, P.; Wellstein, C.; Wiezik, M. European grassland ecosystems: Threatened hotspots of biodiversity. *Biodivers. Conserv.* **2013**, *22*, 2131–2138. [CrossRef]
2. Gould, W. Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecol. Appl.* **2000**, *10*, 1861–1870. [CrossRef]
3. Council of the European Union. Council Directive 92/43/EEC of 21 May 1992 on the Conservation of Natural Habitats and of Wild Fauna and Flora. Available online: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31992L0043 (accessed on 2 February 2016).

4. Myers, N.; Mittermeier, R.A.; Mittermeier, C.G.; da Fonseca, G.A.B.; Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **2000**, *403*, 853–858. [CrossRef] [PubMed]
5. Sutherland, W.J.; Aveling, R.; Brooks, T.M.; Clout, M.; Dicks, L.V.; Fellman, L.; Fleishman, E.; Gibbons, D.W.; Keim, B.; Lickorish, F.; *et al*. A horizon scan of global conservation issues for 2014. *Trends Ecol. Evol.* **2014**, *29*, 15–22. [CrossRef] [PubMed]
6. Löbel, S.; Dengler, J.; Hobohm, C. Species richness of vascular plants, bryophytes and lichens in dry grasslands: The effects of environment, landscape structure and competition. *Folia Geobot.* **2006**, *41*, 377–393. [CrossRef]
7. Moeslund, J.E.; Arge, L.; Bøcher, P.K.; Dalgaard, T.; Ejrnæs, R.; Odgaard, M.V.; Svenning, J.C. Topographically controlled soil moisture drives plant diversity patterns within grasslands. *Biodivers. Conserv.* **2013**, *22*, 2151–2166. [CrossRef]
8. Klimek, S.; Richter gen. Kemmermann, A.; Hofmann, M.; Isselstein, J. Plant species richness and composition in managed grasslands: The relative importance of field management and environmental factors. *Biol. Conserv.* **2007**, *134*, 559–570. [CrossRef]
9. Johansson, L.J.; Hall, K.; Prentice, H.C.; Ihse, M.; Reitalu, T.; Sykes, M.T.; Kindström, M. Semi-natural grassland continuity, long-term land-use change and plant species richness in an agricultural landscape on Öland, Sweden. *Landsc. Urban Plan.* **2008**, *84*, 200–211. [CrossRef]
10. Bakker, C.; Blair, J.M.; Knapp, A.K. Does resource availability, resource heterogeneity or species turnover mediate changes in plant species richness in grazed grasslands? *Oecologia* **2003**, *137*, 385–391. [CrossRef] [PubMed]
11. Ricklefs, R.E. Environmental heterogeneity and plant species diversity: A hypothesis. *Am. Nat.* **1977**, *111*, 376–381. [CrossRef]
12. Tilman, D. Biodiversity: Population *versus* ecosystem stability. *Ecology* **1996**, *77*, 350–363. [CrossRef]
13. Grime, J.P. Benefits of plant diversity to ecosystems: Immediate, filter and founder effects. *J. Ecol.* **1998**, *86*, 902–910. [CrossRef]
14. Rose, R.A.; Byler, D.; Eastman, J.R.; Fleishman, E.; Geller, G.; Goetz, S.; Guild, L.; Hamilton, H.; Hansen, M.; Headley, R.; *et al*. Ten ways remote sensing can contribute to conservation. *Conserv. Biol.* **2014**, *29*, 350–359. [CrossRef] [PubMed]
15. Nagendra, H.; Lucas, R.; Honrado, J.P.; Jongman, R.H.G.; Tarantino, C.; Adamo, M.; Mairota, P. Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecol. Indic.* **2013**, *33*, 45–59. [CrossRef]
16. Nagendra, H. Using remote sensing to assess biodiversity. *Int. J. Remote Sens.* **2001**, *22*, 2377–2400. [CrossRef]
17. Oldeland, J.; Wesuls, D.; Rocchini, D.; Schmidt, M.; Jürgens, N. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? *Ecol. Indic.* **2010**, *10*, 390–396. [CrossRef]
18. Fava, F.; Parolo, G.; Colombo, R.; Gusmeroli, F.; Della Marianna, G.; Monteiro, A.T.; Bocchi, S. Fine-scale assessment of hay meadow productivity and plant diversity in the European Alps using field spectrometric data. *Agric. Ecosyst. Environ.* **2010**, *137*, 151–157. [CrossRef]
19. Carter, G.A.; Knapp, A.K.; Anderson, J.E.; Hoch, G.A.; Smith, M.D. Indicators of plant species richness in AVIRIS spectra of a mesic grassland. *Remote Sens. Environ.* **2005**, *98*, 304–316. [CrossRef]
20. Palmer, M.W.; Earls, P.G.; Hoagland, B.W.; White, P.S.; Wohlgemuth, T. Quantitative tools for perfecting species lists. *Environmetrics* **2002**, *13*, 121–137. [CrossRef]
21. Rocchini, D.; Balkenhol, N.; Carter, G.A.; Foody, G.M.; Gillespie, T.W.; He, K.S.; Kark, S.; Levin, N.; Lucas, K.; Luoto, M.; *et al*. Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecol. Inform.* **2010**, *5*, 318–329. [CrossRef]
22. Pottier, J.; Malenovský, Z.; Psomas, A.; Homolová, L.; Schaepman, M.E.; Choler, P.; Thuiller, W.; Guisan, A.; Zimmermann, N.E. Modelling plant species distribution in alpine grasslands using airborne imaging spectroscopy. *Biol. Lett.* **2014**, *10*, 20140347. [CrossRef] [PubMed]
23. Prentice, H.C.; Cramer, W. The plant community as a niche bioassay: Environmental correlates of local variation in *Gypsophila fastigata*. *J. Ecol.* **1990**, *78*, 313–325. [CrossRef]
24. Simpson, E.H. Measurement of diversity. *Nature* **1949**, *163*, 688. [CrossRef]
25. Nagendra, H. Opposite trends in response for the Shannon and Simpson indices of landscape diversity. *Appl. Geogr.* **2002**, *22*, 175–186. [CrossRef]

26. Oksanen, J.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Henry, M.; Stevens, H.; *et al*. Vegan: Community Ecology Package. Available online: https://cran.r-project.org/web/packages/vegan/index.html (accessed on 2 February 2016).
27. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
28. Ellenberg, H.; Weber, H.E.; Düll, R.; Wirth, V.; Werner, W.; Paulissen, D. *Zeigerwerte von Pflanzen in Mitteleuropa*; GOLTZE: Göttingen, Germany, 1991.
29. Tichý, L. JUICE, software for vegetation classification. *J. Veg. Sci.* **2002**, *13*, 451–453. [CrossRef]
30. Garnier, E.; Cortez, J.; Billés, G.; Navas, M.L.; Roumet, C.; Debussche, M.; Laurent, G.; Blanchard, A.; Aubry, D.; Bellmann, A.; *et al*. Plant functional markers capture ecosystem properties during secondary succession. *Ecology* **2004**, *85*, 2630–2637. [CrossRef]
31. Laliberté, E.; Legendre, P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **2010**, *91*, 299–305. [CrossRef] [PubMed]
32. Richter, R.; Schläpfer, D. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction. *Int. J. Remote Sens.* **2002**, *23*, 2631–2649. [CrossRef]
33. Richter, R.; Schläpfer, D. Geo-atmospheric processing of wide FOV airborne imaging spectrometry data. In *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology*; Ehlers, M., Ed.; SPIE: Bellingham, WA, USA, 2002; pp. 264–273.
34. Fontenla, J.M.; Harder, J.; Livingston, W.; Snow, M.; Woods, T. High-resolution solar spectral irradiance from extreme ultraviolet to far infrared. *J. Geophys. Res. Atmos.* **2011**, *116*, D20108. [CrossRef]
35. Schläpfer, D.; Richter, R. Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: Parametric orthorectification. *Int. J. Remote Sens.* **2002**, *23*, 2609–2630. [CrossRef]
36. Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836. [CrossRef]
37. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least square procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
38. Signal, Signal Processing. 2013. Available online: http://r-forge.r-project.org/projects/signal/ (accessed on 2 February 2016).
39. Hijmans, R.J. Raster: Geographic Data Analysis and Modeling. Available online: https://cran.r-project.org/web/packages/raster/index.html (accessed on 2 February 2016).
40. Beleites, C.; Sergo, V. HyperSpec: A Package to Handle Hyperspectral Data Sets in R, 2015. Available online: http://hyperspec.r-forge.r-project.org (accessed on 2 February 2016).
41. Kumar, L.; Schmidt, K.; Dury, S.; Skidmore, A. Imaging spectrometry and vegetation science. In *Imaging Spectrometry*; Meer, F.D., Jong, S., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2001; pp. 111–155.
42. Schmidtlein, S. Imaging spectroscopy as a tool for mapping Ellenberg indicator values. *J. Appl. Ecol.* **2005**, *42*, 966–974. [CrossRef]
43. Viedma, O.; Torres, I.; Pérez, B.; Moreno, J.M. Modeling plant species richness using reflectance and texture data derived from QuickBird in a recently burned area of Central Spain. *Remote Sens. Environ.* **2012**, *119*, 208–221. [CrossRef]
44. Heumann, B.W.; Hackett, R.A.; Monfils, A.K. Testing the spectral diversity hypothesis using spectroscopy data in a simulated wetland community. *Ecol. Inform.* **2015**, *25*, 29–34. [CrossRef]
45. Warren, S.D.; Alt, M.; Olson, K.D.; Irl, S.D.H.; Steinbauer, M.J.; Jentsch, A. The relationship between the spectral diversity of satellite imagery, habitat heterogeneity, and plant species richness. *Ecol. Inform.* **2014**, *24*, 160–168. [CrossRef]
46. Rocchini, D. Effects of spatial and spectral resolution in estimating ecosystem alpha-diversity by satellite imagery. *Remote Sens. Environ.* **2007**, *111*, 423–434. [CrossRef]
47. Möckel, T.; Dalmayne, J.; Prentice, H.; Eklundh, L.; Purschke, O.; Schmidtlein, S.; Hall, K. Classification of grassland successional stages using airborne hyperspectral imagery. *Remote Sens.* **2014**, *6*, 7732–7761. [CrossRef]
48. Martens, H.; Martens, M. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Preference* **2000**, *11*, 5–16. [CrossRef]

49. Wold, S.; Johansson, E.; Cocchi, M. PLS—Partial least squares projections to latent structures. In *3D QSAR in Drug Design: Volume 1: Theory Methods and Applications*; Kubinyi, H., Ed.; Springer: Drodrecht, The Netherlands, 1994; Volume 1, pp. 523–550.
50. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
51. Hansen, P.M.; Schjoerring, J.K. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* **2003**, *86*, 542–553. [CrossRef]
52. Li, B.; Morris, J.; Martin, E.B. Model selection for partial least squares regression. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 79–89. [CrossRef]
53. Osten, D.W. Selection of optimal regression models via cross-validation. *J. Chemom.* **1988**, *2*, 39–48. [CrossRef]
54. Mevik, B.H.; Wehrens, R. The pls package: Principal component and partial least squares regression in R. *J. Stat. Softw.* **2007**, *18*, 1–24. [CrossRef]
55. Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. [CrossRef]
56. Curran, P.J.; Hay, A.M. The importance of measurement error for certain procedures in remote sensing at optical wavelength. *Photogramm. Eng. Remote Sens.* **1986**, *52*, 229–241.
57. de Jong, S.M.; Pebesma, E.J.; Lacaze, B. Above-ground biomass assessment of Mediterranean forests using airborne imaging spectrometry: The DAIS Peyne experiment. *Int. J. Remote Sens.* **2003**, *24*, 1505–1520. [CrossRef]
58. Carbajo, V.; den Braber, B.; van der Putten, W.H.; de Deyn, G.B. Enhancement of late successional plants on ex-arable land by soil inoculations. *PLoS ONE* **2011**, *6*, e21943. [CrossRef] [PubMed]
59. Werger, M.J.A.; Hirose, T.; During, H.J.; Heil, G.W.; Hikosaka, K.; Ito, T.; Nachinshonhor, U.G.; Nagamatsu, D.; Shibasaki, K.; Takatsuki, S.; *et al.* Light partitioning among species and species replacement in early successional grasslands. *J. Veg. Sci.* **2002**, *13*, 615–626. [CrossRef]
60. Purschke, O.; Schmid, B.C.; Sykes, M.T.; Poschlod, P.; Michalski, S.G.; Durka, W.; Kühn, I.; Winter, M.; Prentice, H.C. Contrasting changes in taxonomic, phylogenetic and functional diversity during a long-term succession: Insights into assembly processes. *J. Ecol.* **2013**, *101*, 857–866. [CrossRef]
61. Tucker, C.J. Spectral estimation of grass canopy variables. *Remote Sens. Environ.* **1977**, *6*, 11–26. [CrossRef]
62. Psomas, A.; Kneubühler, M.; Huber, S.; Itten, K.; Zimmermann, N.E. Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. *Int. J. Remote Sens.* **2011**, *32*, 9007–9031. [CrossRef]
63. Ginzburg, L.R.; Jensen, C.X.J. Rules of thumb for judging ecological theories. *Trends Ecol. Evol.* **2004**, *19*, 121–126. [CrossRef] [PubMed]
64. Öster, M.; Cousins, S.A.; Eriksson, O. Size and heterogeneity rather than landscape context determine plant species richness in semi-natural grasslands. *J. Veg. Sci.* **2007**, *18*, 859–868. [CrossRef]
65. Reitalu, T.; Prentice, H.C.; Sykes, M.T.; Lonn, M.; Johansson, L.J.; Hall, K. Plant species segregation on different spatial scales in semi-natural grasslands. *J. Veg. Sci.* **2008**, *19*, 407–416. [CrossRef]
66. Franzen, D.; Eriksson, O. Small-scale patterns of species richness in Swedish semi-natural grasslands: The effects of community species pools. *Ecography* **2001**, *24*, 505–510. [CrossRef]
67. Pärtel, M.; Zobel, M. Small-scale plant species richness in calcareous grasslands determined by the species pool, community age and shoot density. *Ecography* **1999**, *22*, 153–159. [CrossRef]
68. Von Bueren, S.K.; Burkart, A.; Hueni, A.; Rascher, U.; Tuohy, M.P.; Yule, I.J. Deploying four optical UAV-based sensors over grassland: Challenges and limitations. *Biogeosciences* **2015**, *12*, 163–175. [CrossRef]

# Paper VI

*Remote Sens.* **2014**, *6*, 7732-7761; doi:10.3390/rs6087732

OPEN ACCESS

*remote sensing*

**ISSN 2072-4292**

www.mdpi.com/journal/remotesensing

*Article*

# Classification of Grassland Successional Stages Using Airborne Hyperspectral Imagery

**Thomas Möckel [1,2,]*, Jonas Dalmayne [1], Honor C. Prentice [2], Lars Eklundh [1], Oliver Purschke [3,4], Sebastian Schmidtlein [5] and Karin Hall [1]**

[1] Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden; E-Mails: jonas.dalmayne@nateko.lu.se (J.D.); lars.eklundh@nateko.lu.se (L.E.); karin.hall@nateko.lu.se (K.H.)

[2] Department of Biology, Lund University, Sölvegatan 37, SE-223 62 Lund, Sweden; E-Mail: Honor_C.Prentice@biol.lu.se

[3] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany; E-Mail: oliver.purschke@idiv.de

[4] Institute of Biology/Geobotany and Botanical Garden, Martin Luther University of Halle Wittenberg, Am Kirchtor 1, 06108 Halle (Saale), Germany

[5] Karlsruhe Institute of Technology (KIT), Institute of Geography and Geoecology, 76131 Karlsruhe, Germany; E-Mail: schmidtlein@kit.edu

* Author to whom correspondence should be addressed; E-Mail: thomas.mockel@nateko.lu.se; Tel: +46-46-222-4887; Fax: +46-46-222-0321.

*Received: 8 April 2014; in revised form: 30 July 2014 / Accepted: 30 July 2014 / Published: 20 August 2014*

---

**Abstract:** Plant communities differ in their species composition, and, thus, also in their functional trait composition, at different stages in the succession from arable fields to grazed grassland. We examine whether aerial hyperspectral (414–2501 nm) remote sensing can be used to discriminate between grazed vegetation belonging to different grassland successional stages. Vascular plant species were recorded in 104.1 $m^2$ plots on the island of Öland (Sweden) and the functional properties of the plant species recorded in the plots were characterized in terms of the ground-cover of grasses, specific leaf area and Ellenberg indicator values. Plots were assigned to three different grassland age-classes, representing 5–15, 16–50 and >50 years of grazing management. Partial least squares discriminant analysis models were used to compare classifications based on aerial hyperspectral data with the age-class classification. The remote sensing data successfully classified the plots into age-classes: the overall classification accuracy was higher for a model based on a pre-selected set of wavebands (85%, Kappa statistic value = 0.77) than one using the full set of wavebands (77%, Kappa statistic value = 0.65). Our results show that nutrient availability and grass cover differences between grassland age-classes are detectable by spectral imaging. These techniques may potentially be used for mapping the spatial distribution of grassland habitats at different successional stages.

**Keywords:** arable-to-grassland succession; Ellenberg indicator values; HySpex spectrometer; imaging spectroscopy; partial least square discriminant analysis

---

## Abbreviations

| | |
|---|---|
| ATCOR | Atmospheric and Topographic Correction |
| CWM | Community-Weighted Mean |
| DEM | Digital Elevation Model |
| DGPS | Differential Global Positioning System |
| GIS | Geographic Information System |
| LDMC | Leaf Dry Matter Content |
| LiDAR | Light Detection and Ranging |
| LMA | Leaf Mass per Area |
| LOESS | Local Polynomial Regression |
| LV | Latent Variable |
| MICE | Multivariate Imputation by Chained Equations |
| MODTRAN | Moderate Resolution Atmospheric Transmission |
| NEO | Norsk Elektro Optikk |
| NIR | Near-Infrared |
| PARGE | Parametric Geocoding |
| PLS | Partial Least Squares |
| PLS-DA | Partial-Least-Squares Discriminant Analysis |
| SLA | Specific Leaf Area |
| SWIR | Short-Wave Infrared |
| VIP | Variable Importance in Projection |
| VNIR | Visible and Near Infrared |

## 1. Introduction

The rationalization of European agricultural landscapes during the last century has resulted in the fragmentation and loss of species-rich semi-natural habitats, leading to a dramatic decrease in farmland biodiversity across Europe [1]. The remaining old fragments of grazed semi-natural grasslands are among the most species-rich habitats within the agricultural landscape [2], and they are of great importance for the overall species richness of agricultural landscapes [3]. As well as agricultural intensification in many areas, there has also been a successive increase in the area of abandoned cultivated land in several parts of Europe over the past 50 years [4]. In many European regions, abandoned arable fields are gradually being transformed into grazed grasslands [5,6]. The majority of the grasslands in the present-day agricultural landscape are grazed fields that represent early stages in the succession from arable cultivation towards species-rich old grasslands [7].

The transformation of abandoned arable land into grazed grassland is expected to offer new possibilities for the establishment of species-rich and diverse grassland vegetation and to mitigate biodiversity loss in agricultural landscapes [8,9]. If species diversity in cultivated landscapes is to be maintained and enriched in the future, species will need to be able to disperse from old, species-rich grassland fragments into younger grasslands [10]. Optimization of the spatial distribution of grassland fragments will require information that discriminates between land cover belonging to different stages in the succession from arable fields to old semi-natural grasslands. It may be difficult to collect explicit spatial information on grassland age and habitat characteristics over wide areas solely from field-based assessments [11]. Field-based inventories of grasslands are time consuming and are therefore often based on plot-scale sampling within spatially restricted areas (*cf.* [12]).

Remote sensing technology has the ability to support and supplement field-based habitat inventories [13,14], and the potential of remote sensing data as a source of information both within vegetation science [15] and as a tool within conservation biology has recently been highlighted [16]. The development of remote sensing-based methods that can be used for the mapping of habitats at detailed scales is considered to be particularly important [16]. Aerial photographs and broadband satellite-based spectral data have been used to map and monitor grassland properties. For example, Waldhardt and Otte [17] showed that grassland vegetation of different ages could be discriminated with the help of the colour tonal values from false-colour infrared aerial photographs. Kawamura *et al.* [18] used spectral information acquired from satellite data to assess grazing intensity in grasslands. However, the low spectral resolution of aerial photographs and broad-band sensors limits the collection of detailed information on vegetation properties [19,20].

Hyperspectral remote sensing (imaging spectroscopy) is a particularly good method for assessing and monitoring vegetation characteristics [21–23]. Spectral measurements acquired by hyperspectral sensors provide detailed information on the structural and biochemical properties of vegetation [24,25]. For example, plant functional traits and properties, such as leaf nitrogen content, leaf chlorophyll content, leaf water content, and leaf area index, have been successfully estimated with the help of hyperspectral data [26–29]. Ecological indicators, such as Ellenberg values [30], are commonly used to describe relationships between vegetation and environment [31]. Schmidtlein [12] mapped gradients of community-weighted mean Ellenberg indicator values for nutrient and moisture availability in montane pastures, and Klaus *et al.* [32] predicted mean Ellenberg indicator values for nutrient and moisture availability in agricultural grasslands using spectroscopy data.

Previous studies have shown that grassland plant communities representing different stages in the arable-to-grassland succession are characterized by different habitat conditions and plant community characteristics [33,34]. Old grasslands have lower community-weighted mean values for Ellenberg indicators for nutrient and moisture availability than young grasslands [34]. In addition, old grasslands are typically characterized by lower community-weighted mean values for specific leaf area (SLA), canopy height, and leaf size, and by higher mean values for leaf dry matter content (LDMC) than young grasslands [33]. Because SLA is associated with the assemblage of leaf chemicals that control photosynthesis, positive relationships between SLA, nutrient availability, chlorophyll content and leaf water content can be expected [33,35–39].

There has been considerable recent progress in the development of methods for both handling and analysing spectral data (e.g., [40,41]) and characterizing and discriminating between habitats with the help of spectral data [42]. Partial least squares (PLS) [43] regression is commonly used in hyperspectral remote sensing and has shown to be a powerful technique for studying grassland vegetation [44–46]. In addition, partial least squares discriminant analysis (PLS-DA), using pre-selected wavebands, is increasingly used in remote sensing-based classification of plant communities (e.g., [47]).

Here, we examine whether a combination of airborne hyperspectral data and PLS-DA can be used to discriminate between grazed, dry grasslands belonging to different age-classes in an agricultural landscape on the Baltic island of Öland (Sweden). We used data from HySpex hyperspectral spectrometers (414–2501 nm) to classify grassland age at a spatial resolution of 3 m × 3 m. We compare the classification accuracies using two different PLS-DA models: Model 1 based on the full set of HySpex wavebands and Model 2 based on a pre-selected subset of wavebands. We explore the dissimilarities in plant community and spectral characteristics between grasslands representing different age-classes and ask the following questions: (1) can grassland age-classes be classified with the help of hyperspectral HySpex data and PLS-DA? In addition, (2) can the classification accuracies of grassland age-classes be improved by pre-selecting the hyperspectral wavebands that are used in the PLS-DA model?

## 2. Materials and Methods

### *2.1. Study Area*

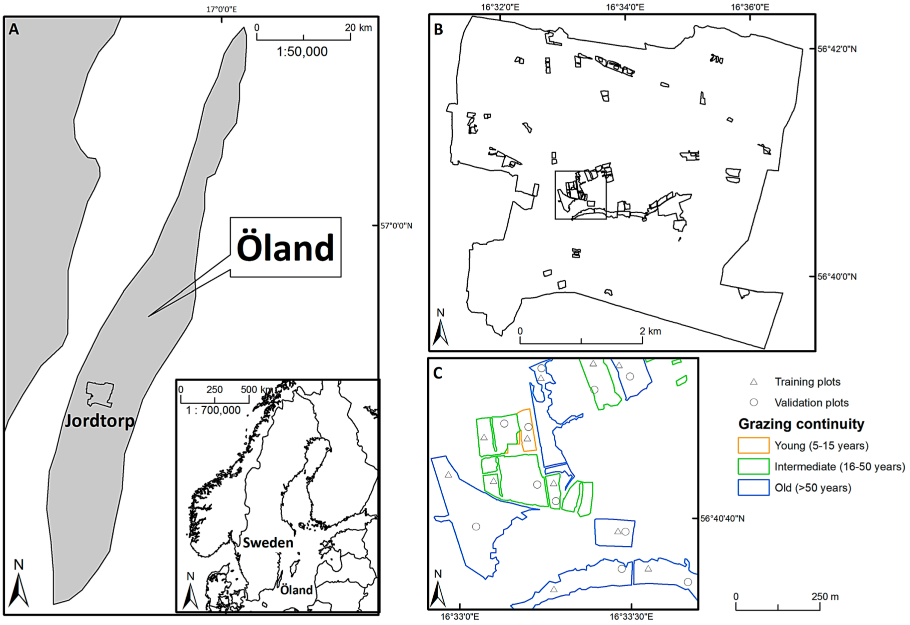
The study area (centred on 56°40′49″N, 16°33′58″E) covers approximately 22.5 $km^2$ and is located on the Baltic island of Öland in SE Sweden (Figure 1A). The bedrock consists of Cambro-Silurian limestone, the average elevation is approximately 36 m above sea level, and the overall topography is flat [48]. The area is crossed by a few low ridges of glaciofluvial deposits. The mean annual temperature is 7 °C and the mean annual precipitation is 468 mm [48]. The present-day landscape consists of a mosaic of arable fields, villages, forests, and grasslands. The majority of the grasslands are grazed, with varying intensity, mainly by cattle.

Detailed information on the historical land use within the study area is available [49]. Whereas the study area was dominated by grasslands during the 18th century [49], most (>80%) of the ancient grasslands have been transformed to arable land during the last 300 years [49]. Approximately 15% of the current grasslands have developed during the last 50 years on previously arable land [49].

### *2.2. Grassland Sites and Grassland Age-Classes*

Using historical and present-day land-use maps, aerial photographs [49] and field inventories, we identified 299 grazed grassland sites that were separated from each other and the surrounding landscape by walls or fences. With the help of geographic information system (GIS) overlay analysis of the land-use maps and aerial photographs, the sites were categorized according to their grazing continuity (grassland age) and assigned to one of three age-classes within the arable-to-grassland succession; 5–15 (young grasslands), 16–50 (intermediate-aged grasslands), and >50 (old grasslands) years of grassland continuity (Figure A1).

**Figure 1.** (**A**) The Jordtorp study area on the Baltic Island of Öland, Sweden; (**B**) The distribution of grassland sites included in the present study; (**C**) An example of the distribution of field sample plots within grassland sites.



Within each site, two sampling points were randomly positioned in open (not covered by shrubs or trees) grassland vegetation, with the constraints that they should be at least 25 m apart (to minimize effects of spatial autocorrelation in the vegetation), at least 13.5 m from the site boundary (to minimize edge-effects in the vegetation [50]), and at least 13.5 m from shrubs or trees higher than 50 cm (to minimize shading-effects). In total 239 grassland sites (89 young, 73 intermediate, and 77 old) out of the 299 sites could accommodate these constraints. From these 239 grassland sites, we randomly selected 60 sites (20 young, 20 intermediate, and 20 old). Within these 60 sites, a bioassay approach (*cf.* [51,52]) based on indicator species (such as *Sesleria caerulea* and *Molinia caerulea*) was used to define sites with "dry" grassland vegetation, and exclude moister grassland vegetation. A total of 52 sites (17 young, 18 intermediate, and 17 old) out of the 60 sites were characterized by dry grassland vegetation and used for the vegetation and remote sensing sampling (Figure 1B). Of the 17 "old" sites, 13 sites had a management continuity of >280 years [49]: the vegetation in these sites falls within the Natura 2000 habitat type "Fennoscandian lowland species-rich dry to mesic grasslands".

A hand-held differential global positioning system (DGPS) receiver (Topcon GRS-1 GNSS, equipped with a PG-A1 external antenna (Topcon Corporation, Japan)) connected to a real-time positioning service (SWEPOS) was used to log (with an accuracy of ~1 cm) the ground coordinates of the two sampling points within each of the 52 grassland sites. The sampling points were divided into two data sets (a training and a validation data set) by randomly assigning the two sampling points from each site to one or other of the two data sets (Figure 1C).

### *2.3. Plant Community Characteristics*

Vegetation sampling was carried out between 15 May and 15 July 2011. A 1 m × 1 m plot was centred over each of the two sampling points within each of the 52 sites. Each of the 104 (1 m × 1 m) plots was divided into a grid of 100 (10 cm × 10 cm) sub-plots, and the presence of all non-woody vascular plants was recorded within each sub-plot (Table A1). For each 1 m × 1 m plot, the frequency of each species was calculated as the number of sub-plots in which the species was present.

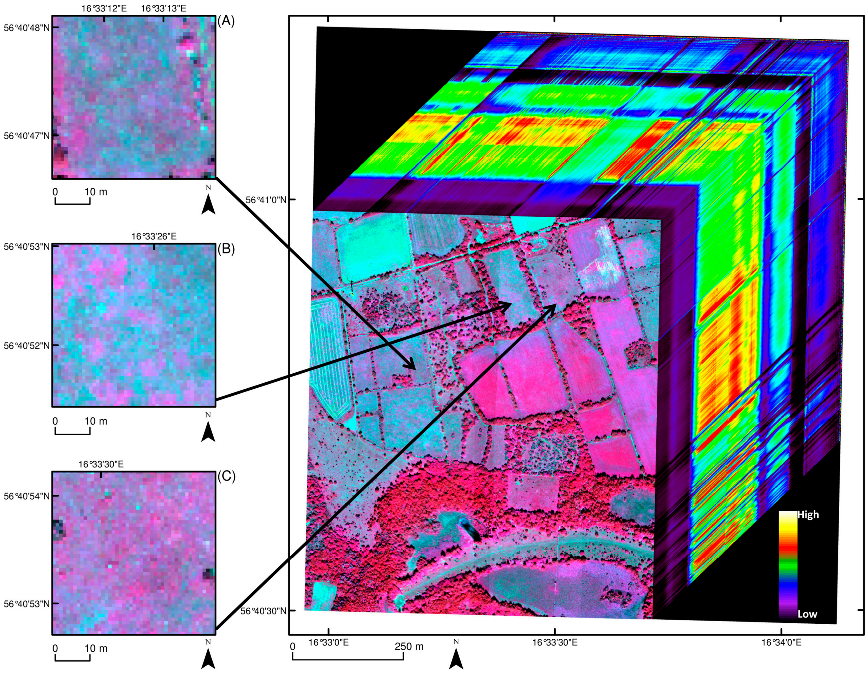
The plant species recorded in the 1 m × 1 m plots were assigned values for SLA and two Ellenberg indicator values (for nutrients and moisture availability). The trait information was compiled from the LEDA trait database [53], which provides species-level SLA values that represent aggregated trait values. The LEDA values are based on multiple individuals per species, measured under different environmental conditions. Trait data were not available for the full set of species ($n$ = 185) in the present study. For SLA, previous grassland studies have shown that imputation methods based on multiple imputation by chained equations may be used for filling gaps in functional trait databases [54]. Following Taugourdeau *et al.* [54], estimates for missing values (approximately 9% of the species) were obtained using the multivariate imputation by chained equation (MICE) method [33,55,56]. Ellenberg indicator values were extracted from the JUICE database [57]: values for the four species that were not found in the database were extracted from Ellenberg *et al.* [58]. A frequency-weighted mean value (CWM) [35] was calculated, for SLA and each of the Ellenberg values, for each plant community: $CWM(x) = \sum_i p_i \times x_i$, where $p_i$ is the relative frequency of the $i$th species and $x_i$ is the trait or Ellenberg index value of the $i$th species.

The percentage ground cover of grasses was estimated within four 50 cm × 50 cm sub-plots within each 1 m × 1 m plot, and a mean within-plot value for grass cover (excluding dead litter) was calculated for each plot.

### *2.4. Remote Sensing Data*

Remote sensing data over the study area were acquired on 9 July 2011. Twenty-five flight lines were recorded at around solar noon, using two airborne HySpex hyperspectral spectrometers (Norsk Elektro Optikk AS (NEO), Lörenskog, Norway), and a push-broom scanning mode at a flight altitude of approximately 1500 m. The flight was carried out by Terratec AS, Lysaker, Norway. All flight lines were conducted either from north to south or from south to north to minimize illumination effects. The output from the two HySpex spectrometers (VNIR-1600 operating over the 414 to 991 nm range and SWIR-320m-e operating over the 966 to 2501 nm range) consisted of 416 wavebands with spectral resolutions of 3.7 nm (VNIR-1600) and 6.0 nm (SWIR-320m-e) (Figure 2). The spatial resolution (pixel size) of the image captured from the VNIR-1600 spectrometer was 0.5 m × 0.5 m and the spatial resolution of the SWIR-320m-e image was 1.0 m × 1.0 m. A digital elevation model (DEM) was created with the help of LiDAR data recorded during the flight.

**Figure 2.** Hyperspectral data cube showing examples of data associated with (**A**) young; (**B**) intermediate-aged; and (**C**) old grassland sites within a subarea of 0.71 km$^2$ within the Jordtorp study area (see Figure 1B). The colour-composite on the top was obtained using three hyperspectral wavebands (861 nm, 651 nm, and 549 nm).



### 2.4.1. Pre-Processing of HySpex Data

Physically based atmospheric correction of the HySpex data was carried out with the help of the ATCOR-4 software [59], which is based on the radiative transfer model MODTRAN 5 [60]. In accordance with the standard procedure in the ATCOR-4 software, the following atmospheric parameters were used during the correction process: desert aerosol model, water vapour column of 1.0 g·m$^{-2}$, visibility of 28.4 km, and an ozone concentration of 330 Dopson units. The conversion of radiance into reflectance was based on the Fontenla-2011 solar irradiance spectrum [61,62]. While the overall topography within the study area is flat, there are local differences in elevation. To account for surface elevation, slope, and orientation, the correction for topographic illumination effects was carried out in the "rugged terrain" model of the ATCOR-4 software using the DEM created with the help of the LiDAR data recorded during the flight. The spectral data acquired with the help of the VNIR-1600 spectrometer were resampled to a spatial resolution of 1 m × 1 m (with the help of a triangulated nearest neighbourhood method) and to a spectral resolution of 6.0 nm (by LOESS interpolation) to match the data collected from the SWIR-320m-e spectrometer. A Savitzky-Golay smoothing filter [63], with a degree of 3 and a width of 9 was used to reduce the effect of random noise in the spectral data. Following [64], brightness normalization was carried out to dampen differences in the brightness of the spectra caused by subpixel shade. The interpolation was done using the hyperSpec package [65] and the filtering was done using the signal package [66] in the R statistical environment [67]. The images were orthorectified by the HySpex data providers (Terratec AS, Lysaker, Norway) using the PARGE software [60] to a spatial accuracy of approximately 0.3 m for both the VNIR-1600 and SWIR-320m-e spectral bands.

#### 2.4.2. Extracting HySpex Data

Spectral bands that (i) overlapped between the two spectrometers (962–985 nm) and (ii) were in spectral domains that are strongly influenced by atmospheric water vapour absorption (1321–1443 nm, 1803–2032 nm and 2420–2501 nm) were deleted (*cf.* [68,69]). A pixel window of 3 × 3 pixels (3 m × 3 m) for the remaining 269 bands (414–2417 nm) was centred on each of the 52 coordinate points associated with the training and validation data sets. The mean spectral value for each of the 269 HySpex bands was calculated for the pixels falling within each of the 104 pixel windows.

### *2.5. Statistical Analyses*

#### 2.5.1. Partial Least Squares Discriminant Analysis

PLS regression [43] allows statistical analysis of data sets where the explanatory variables are strongly correlated and where the number of explanatory variables is similar to or higher than the number of samples [70,71]. Whereas the use of a high number of inter-correlated explanatory variables may influence random noise, which, in its turn, may lead to model overfitting and reduced model accuracy, PLS regression builds on the assumption that only a few variables influence the process that is under study. By combining the information in a large number of inter-correlated explanatory variables into a few latent components, the risk of model overfitting is reduced in PLS [72]. The latent variables (LVs) are identified by finding the loading weights for each explanatory variable that maximize the covariance between the explanatory variables and the dependent variables. In the case of binary dependent variables, the PLS algorithm [43] can be used for discriminatory purposes [73].

A PLS-DA based on hyperspectral data and vegetation data was carried out to assess whether (i) the three grassland age-classes can be classified with the help of HySpex data, and (ii) it is possible to identify a subset of HySpex wavebands that can be used for the classification of the grassland age-classes. Two PLS-DA models—Model 1 developed from the full set of HySpex bands and Model 2 from a subset of HySpex bands—were generated and the capabilities of the two models for classifying grassland age-classes were compared.

The two models were developed using a similar PLS-DA-based procedure. Initially, a binary membership vector was built for each grassland age-class individually. For each of the three individual grassland age-class membership vectors, each individual plot was defined as 1 (belonging to a particular age-class) or 0 (not belonging to a particular age-class). The three membership vectors were used to build the dependent matrix. The explanatory matrix was generated from the HySpex bands representing the plots associated with the training data set ($n$ = 52). An increasing number of latent variables will normally improve the predictive capability of a PLS-DA model because several variables can carry more information than a few [74]. However, because too many latent variables can overfit the final model, the optimal number of latent variables needs to be identified (*cf.* [75]). In the present study, the number of latent variables that gave the lowest misclassification rate was used in the final model.

The training data sets associated with the two models were used to quantify their accuracies for classifying grassland age-classes with the help of tenfold cross-validated discriminant analysis. The validation data sets associated with the models were used to evaluate them for the training data sets by fitting the final cross-validated PLS-DA models of the training data sets to the validation data sets. The calculations were implemented in the R statistical environment [67] using the *pls* package [76].

### 2.5.2. Explanatory Matrices Used in the Two PLS-DA Models

The explanatory matrix used for Model 1 included 269 spectral bands. The relative importance of each individual predictor variable (*i.e.*, each HySpex band) in Model 1 was described by the variable importance in projection (VIP) value [77]. The HySpex bands associated with VIP values greater than 0.8 (*cf.* [78]) in Model 1 were assigned to a subset of spectral bands. The explanatory matrix used for Model 2 was then based on the subset of HySpex bands.

### 2.5.3. Accuracies of the PLS-DA Models Used for the Classification of Grassland Age-Classes

Two approaches were used to quantify the ability of each of the two PLS-DA models to classify grassland age-classes:

(i) Tenfold cross-validation was used on the training data set and a confusion matrix was created to assess the accuracy of the HySpex based classification of the three grassland age-classes. With the help of the confusion matrix, the producer's accuracy and the user's accuracy were calculated for each age-class. The producer's accuracy refers to the probability that a plot associated with a specific grassland age-class on the ground will be assigned to the correct age-class on the basis of the grassland spectral response acquired from the plot. The user's accuracy represents the probability that a grassland plot classified (with the help of the grassland spectral response acquired from the plot) as belonging to a specific grassland age-class is associated with this class on the ground. The overall prediction accuracy and the Kappa statistic value [79,80], which assesses the interclassifier agreement, were also calculated from the confusion matrix.

(ii) The final cross-validated model of the training data set was fitted to the validation data set and used to classify the age-class of each plot associated with the validation data set. The predictive capability of the PLS-DA model was investigated by calculating the producer's and user's accuracy for each age-class associated with the validation data set. The overall prediction accuracy and Kappa statistic value [79,80] for the classification results based on the validation data set were also calculated.

## 3. Results

### *3.1. Classification Based on the Full Set of 269 HySpex Wavebands* (*Model 1*)

The misclassification rate decreased progressively with the number of latent variables (from 57.7% with one latent variable), reaching a minimum (30.8%; Figure 3) with the first four latent variables before increasing again. We used the first four latent variables in the final model.

The producer's and user's classification accuracies for the individual grassland age-classes varied between 47% and 100% for the training data set, and between 65% and 94% for the validation data set (Table 1).

The overall classification accuracy and the Kappa statistic value were 77% and 0.65, respectively, for both the training and the validation data set (Table 1).

**Figure 3.** The relationship between the number of LVs and the misclassification rate in Model 1 (the PLS-DA model developed from the full set of HySpex hyperspectral wavebands) (*n* = 269). Tenfold cross-validation was used to determine the lowest misclassification rate on the training data set. The arrow indicates the optimal (giving the lowest misclassification rate) number of LVs.

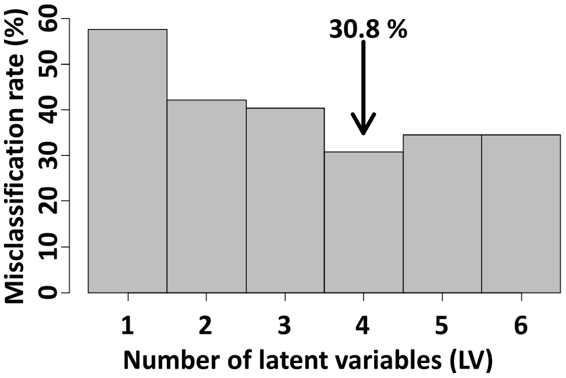


**Table 1.** Confusion matrix produced from Model 1 (the partial least squares discriminant analysis (PLS-DA) model based on the full set of HySpex hyperspectral wavebands) (*n* = 269). The producer's and user's classification accuracies are shown for the training and the validation data for each of three grassland successional stages, represented by young (*n* = 17), intermediate-aged (*n* = 18), and old (*n* = 17) grassland plots. Values in bold represent the number of correctly classified grassland plots.

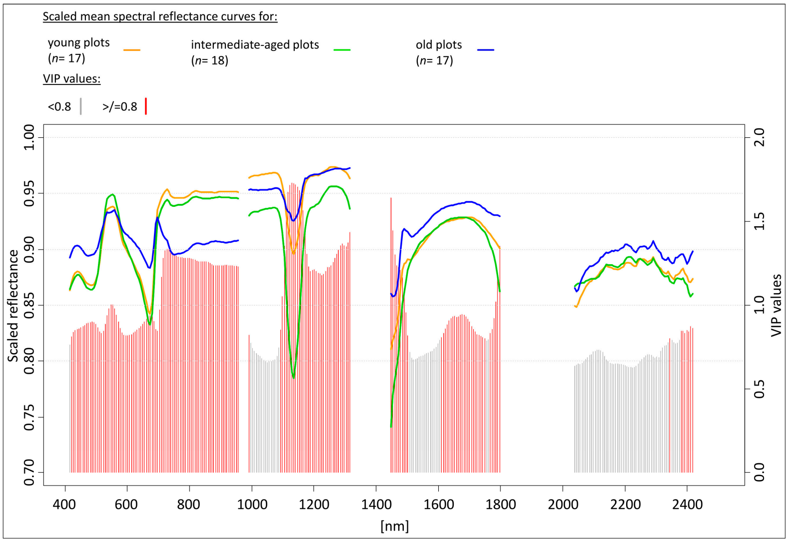
| **Grassland Age-Class Classified from HySpex Data** | **Grassland Age-Class as Derived from Land-Use Maps, Field Inventories, and Aerial Photographs** | | | |
|---|---|---|---|---|
| **Training Data Set** | **Young** | **Intermediate** | **Old** | **User's Accuracy (%)** |
| Young | 8 | 0 | 3 | 73 |
| Intermediate | 6 | 18 | 0 | 75 |
| Old | 3 | 0 | 14 | 82 |
| Producer's accuracy (%) | 47 | 100 | 82 | |
| Validation data set | | | | |
| Young | 11 | 0 | 5 | 69 |
| Intermediate | 4 | 17 | 0 | 81 |
| Old | 2 | 1 | 12 | 80 |
| Producer's accuracy (%) | 65 | 94 | 71 | |

For the training data set: overall classification accuracy = 77%, Kappa statistic value = 0.65. For the validation data set: overall classification accuracy = 77%, Kappa statistic value = 0.65.

### *3.2. The HySpex Wavebands That are Most Important for the Classification of Grassland Age-Classes*

The VIP value is an indicator of the relative influence of each predictor in a PLS model. Out of the 269 HySpex bands used as explanatory variables in the PLS-DA model, 177 bands had VIP values greater than 0.8, indicating that these bands were the most important predictor variables in the remote sensing-based classification of grassland age-classes (Figure 4). In total, 14 wavebands in the blue region (414–499 nm), 16 wavebands in the green region (505–596 nm), 20 wavebands in the red region (602–716 nm), 17 wavebands in the red-edge portion (722–818 nm), 23 wavebands in the NIR1 part (824–969 nm), 39 wavebands in NIR2 part (975–1394 nm), and 48 wavebands in the SWIR portion (1448–2417 nm) of the electromagnetic spectrum were associated with VIP scores greater than 0.8 (Figure 4).

**Figure 4.** VIP values as a function of the wavelengths used in Model 1 (the PLS-DA model developed from the full set of HySpex hyperspectral wavebands) (*n* = 269). The higher the VIP value of a waveband, the greater its contribution to the model. The most influential wavebands in the PLS-DA model (with VIP values >0.8) are marked in red. The orange, green, and blue lines represent the scaled mean spectral reflectance curves obtained using the training subsets for young, intermediate-aged, and old grassland plots, respectively.



### *3.3. Classification Based on a Reduced Set of HySpex Wavebands (Model 2)*

The subset of HySpex bands used to develop Model 2 included the 177 spectral bands that were associated with VIP values greater than 0.8 in the model based on the full set of 269 HySpex bands. The inclusion of four latent variables gave the lowest misclassification rate in the second PLS-DA model.

The producer's and user's classification accuracies for the individual grassland age-classes for Model 2 varied between 65% and 94% for the training data set (Table 2). The classification accuracies for the individual grassland age-classes for the validation data set ranged between 74% and 94% (Table 2). The overall classification accuracy and the Kappa statistic value were 81% and 0.71, respectively, for the training data set and 85% and 0.77, respectively, for the validation data set (Table 2).

For the validation data sets, the overall classification accuracy and the Kappa statistic value were higher for Model 2 than for Model 1 (Tables 1 and 2). The differences in classification accuracies between the two models were 8% for the overall accuracy and 0.12 for the Kappa statistic value (Tables 1 and 2).

**Table 2.** Confusion matrix produced from Model 2 (the PLS-DA model based on a subset of HySpex hyperspectral wavebands) (*n* = 177). The producer's and user's classification accuracies are shown for the training and the validation data for each of three grassland successional stages, represented by young (*n* = 17), intermediate-aged (*n* = 18), and old (*n* = 17) grassland plots. Values in bold represent the number of correctly classified grassland plots.

| Grassland Age-Class Classified from HySpex-Data | Grassland Age-Class as Derived from Land-Use Maps, Field Inventories, and Aerial Photographs | | | |
|---|---|---|---|---|
| | Young | Intermediate | Old | User's accuracy (%) |
| | Training data set | | | |
| Young | **11** | 1 | 2 | 79 |
| Intermediate | 3 | **17** | 1 | 81 |
| Old | 3 | 0 | **14** | 82 |
| Producer's accuracy (%) | 65 | 94 | 82 | |
| | Validation data set | | | |
| Young | **14** | 2 | 3 | 74 |
| Intermediate | 1 | **16** | 0 | 94 |
| Old | 2 | 0 | **14** | 88 |
| Producer's accuracy (%) | 82 | 89 | 82 | |

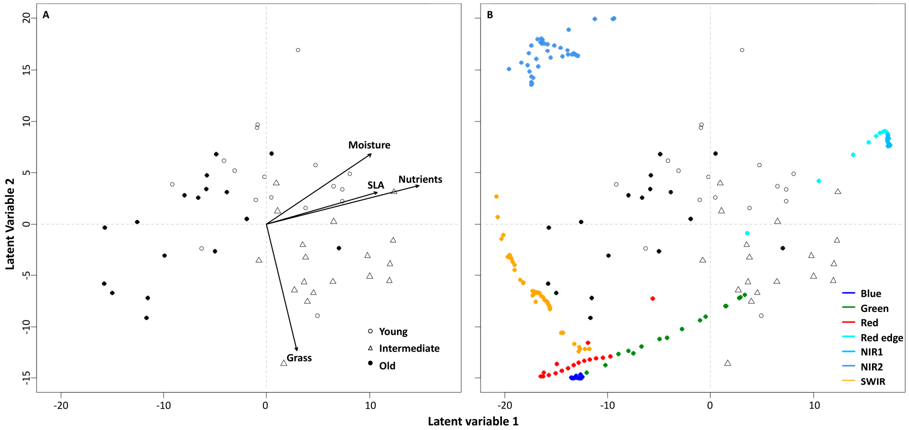
For the training data set: overall classification accuracy = 81%, Kappa statistic value = 0.71. For the validation data set: overall classification accuracy = 85%, Kappa statistic value = 0.77.

### *3.4. Plant Community and Spectral Characteristics of Grasslands Representing Different Age-Classes*

The first four latent variables (LV1–LV4) in Model 2 explained 34%, 41%, 12%, and 10% of the variation in the explanatory matrix (spectral data), respectively. Old (negative scores) and intermediate-aged (positive scores) grassland plots were separated on the first latent variable (LV1), whereas LV2 discriminated between young (positive scores) and intermediate-aged (negative scores) plots (Figure 5A).

Correlation coefficients between the scores for grassland plots on the first two latent variables, and the ecological variables (SLA, Ellenberg indicator value for nutrients and moisture availability, and the percentage cover of grasses) were calculated with the help of multiple regressions. Significance values for the correlation coefficients were based on 9999 permutations. The correlation coefficients are represented as arrows in Figure 5A, with the direction and length of the arrow indicating, respectively, the sign and the strength of the correlation coefficient. The scores for the grassland plots were significantly associated with the mean within-plot Ellenberg indicator values for nutrient availability (*p* = 0.010), and with the within-plot cover of grasses (*p* = 0.045). The young plots were characterized by the highest and the old plots by the lowest indicator values for nutrients, whereas the intermediate-aged plots were characterized by the highest cover of grasses (Figure 5A). The mean within-plot SLA and Ellenberg indicator value for moisture were not significantly associated with the plot scores in the PLS-DA.

**Figure 5.** Plots of the first two latent variables from Model 2 (the PLS-DA model based on 177 HySpex wavebands) using the training data set. (**A**) Arrows representing the correlation coefficients between the score values for grassland plots and the values for four ecological variables (Ellenberg indicator value for nutrient (Nutrients) (*p* = 0.010) and moisture availability (Moisture) (*p* = 0.053), specific leaf area (SLA) (*p* = 0.087), and cover of grasses (Grasses) (*p* = 0.045); (**B**) Loadings for individual HySpex wavebands coloured into seven spectral regions (blue (414–499 nm), green (505–596 nm), red (602–716 nm), red-edge (722–818 nm), NIR1 (824–969 nm), NIR2 (975–1394 nm), and SWIR (1448–2417 nm)).



The old plots were associated with negative loadings for the blue, red, and SWIR regions on the first latent variable, and the intermediate-aged plots were associated with positive loadings for the red-edge and NIR1 regions (Figure 5B). The old plots, which are characterized by lower indicator values for nutrients than the intermediate-aged plots (Figure 5A), had higher reflectance values than the intermediate-aged plots in the blue, red, and SWIR regions. The intermediate-aged plots, which are characterized by higher cover of grasses than the old plots (Figure 5A), had higher reflectance values than the old plots in the green, red-edge, and NIR1 regions (Figure 5B).

On LV2, the young plots were associated with positive loadings for the red-edge, NIR1, and NIR2 regions, while the intermediate-aged plots were associated with negative loading for the green region (Figure 5B). The young plots, which are characterized by higher indicator values for nutrients than the intermediate-aged plots (Figure 5A), had higher reflectance values than the intermediate-aged plots in the red-edge, NIR1, and NIR2 regions (Figure 5B). The intermediate-aged plots, which are characterized by higher cover of grasses than the young plots (Figure 5A), had higher reflectance than the young plots in the green region (Figure 5B).

Species lists for each of the three grassland age-classes are provided in Table A1 (Supporting information). Figure A1 (Supporting information) shows a field photograph of a sample plot representing each of the three grassland age-classes.

## 4. Discussion

### *4.1. PLS-DA Based Classification of Grassland Age-Classes*

The three age-classes in our study were successfully classified using a pre-selected set of hyperspectral wavebands and PLS-DA, indicating that fine-scale hyperspectral measurements are able to accurately identify grassland successional stages within a local landscape. Our results are consistent with those of earlier studies of other types of vegetation that also show that the exclusion of wavebands that provide little information related to the response variable improves the PLS-DA based classification of plant communities [47,81]. For example, using airborne AISA Eagle hyperspectral imagery, Peerbhay *et al.* [47] showed that a PLS-DA model based on the 78 wavebands that were most relevant for the classification of forest species gave 8.17% higher overall classification accuracy than a model utilising all 230 AISA Eagle wavebands. Our results are also consistent with the results of Peerbhay *et al.* [47] in that they show that a high number of hyperspectral wavebands can be compressed into a few latent variables with the help of PLS, reducing the risk of model overfitting and, at the same time providing a successful classification of different vegetation types.

### *4.2. Spectral Dissimilarities between Grasslands Belonging to Different Age-Classes*

The majority of the present-day grazed grasslands in Sweden originate from abandoned arable fields that have been cultivated as leys [82] and seeded with grasses, such as *Lolium perenne, Festuca pratensis*, *Poa pratensis*, and *Phleum pratense* (*cf.* [83]). The effects of the addition of nutrients during arable cultivation may persist during the early stages of the arable-to-grassland succession [84,85]. Recently cultivated fields, such as the young grasslands in our study (Figure 5A), are generally characterized by a high availability of nutrients that allows nitrophilous and ruderal species (fast growing and disturbance-tolerant species) to invade and restricts the establishment of less competitive species during early succession (*cf.* [33]). Early-successional grasslands are typically characterized by species that are adapted to nutrient-rich environments (e.g., *Dactylis glomerata*, *Poa pratensis*, and *Geranium molle* [86]) and nutrient-rich vegetation commonly contains plants with a relatively high SLA [87]. An increased availability of nutrients may result in greater above-ground production of biomass (*cf.* [85]), which, in its turn, influences the light conditions within the vegetation canopy: an increased leaf area represents a common adaptation to low light environments [38,88].

Residual effects of fertilization during arable cultivation may still persist within the intermediate-aged grasslands in our study (*cf.* [84,85]). However, the fact that continuous grazing management contributes to the removal of nutrients explains the reduced abundance of nitrophilous species during the second successional time step. Previous investigations within the study area have shown that grasslands with a grazing continuity of 16–50 years (corresponding to the intermediate-aged grasslands in the present study) are characterized by significantly lower community mean values for SLA than grasslands that have been grazed for 5–15 years (corresponding to the young grasslands in the present study) [33].

During the third successional time step, a long continuity of grazing management has contributed to a substantial reduction of nutrient levels, allowing species with a low competitive ability to establish in the old grasslands. Purschke *et al.* [33] showed that, within the study area, values of community-level SLA are lower in grassland sites that have been grazed for more than 51 years (corresponding to the old grasslands in the present study) than in sites with a shorter grazing continuity. Long grazing continuity is essential for the maintenance and establishment of species-rich ancient grassland vegetation characterized by, for example, *Festuca ovina* and *Helianthemum nummularium* [89].

Changes in habitat conditions during the arable-to-grassland succession have been shown to be accompanied by changes in plant functional composition and vegetation properties within the same study area [33,90]. In addition, changes in, for example, leaf nitrogen content, SLA, leaf chlorophyll content, leaf water content, and above-ground biomass have shown to be followed by changes in the vegetation spectral response within a range of different habitats [26,91–93]. For example, previous remote sensing studies have revealed that differences in leaf nitrogen content have strong influence on the reflectance from vegetation across the electromagnetic spectrum (visible (400–700 nm) [94], NIR (700–1300 nm) [95], SWIR (1300–2500 nm) [96]). The red-edge region has, been used to estimate the nitrogen concentration in ryegrass (*Lolium* spp.) [97], and SLA has previously been shown to influence reflectance in several wavebands. For example, Lymburner *et al.* [98] estimated SLA at the landscape scale with the help of broadband satellite data, and Asner *et al.* [99] showed that hyperspectral measurements were related to the SLA of tropical forest leaves. The SWIR wavelengths have been shown to be most important for estimation of leaf mass per area (the inverse of SLA [100]).

In the present study, differences in ecological variables (availability of nutrients and cover of grasses) between grassland age-classes are associated with differences in the spectral response from the grasslands (Figure 5). The low blue (414–499 nm) as well as the red (602–716 nm) and the high red-edge (722–818 nm) reflectance for the young plots (Figure 5B) suggest that these plots have a higher chlorophyll content in the vegetation than the old plots. Earlier remote sensing studies have shown that the chlorophyll content in grasslands can be assessed at both fine and landscape scales [46,101], and Darvishzadeh *et al.* [46] used the red-edge region to map canopy chlorophyll content in heterogeneous grasslands.

The higher cover of grasses (excluding dead litter) in the intermediate-aged plots compared with the young and old plots can be expected to lead to higher green (505–596 nm) reflectance in the intermediate-aged plots. Spectral measurements near the green reflectance peak at 550 nm have previously been shown to be important for the classification of vegetation communities (e.g., [47]).

The high red-edge (722–818 nm) and NIR1 (824–969 nm) reflectance for the young and intermediate-aged plots suggest that these plots were associated with larger above-ground biomass than the old plots (*cf.* [45]). The low NIR2 (975–1394 nm) reflectance for the intermediate-aged and the low SWIR (1448–2417 nm) reflectance for the young plots indicate that these were characterized by higher canopy moisture than the old plots. The SWIR wavelengths are strongly affected by water absorption [100]; in particular, the moisture-sensitive bands around 1200 nm (located in the NIR2 region in the present study) have been shown to be sensitive to the water content of vegetation in mixed woodland [102] and in grasslands [68].

Trampling by grazing animals is likely to lead to differences in the cover of bare soil between the different age-classes. Spectral dissimilarities between the grasslands belonging to different successional stages may also be explained by between-site edaphic dissimilarity (*cf.* [103]). The red part of the visible electromagnetic spectrum has been shown to be highly influenced by variation in soil properties [104]. Furthermore, increased levels of shading are expected within the vegetation cover as the above-ground biomass increases. Increased amounts of shadow are accompanied by a decrease in reflectance in all wavebands [105]. The presence of shadows may also have contributed to between-stage spectral dissimilarities in the present study. Monocotyledonous plants, such as grasses, have a compact leaf mesophyll structure, and a lower reflectance in the NIR than plants with a more porous mesophyll structure [106]. The high cover of grasses within the intermediate-aged grasslands in our study may have contributed to between-grassland stage spectral differences in the NIR. Grazing management decreases the accumulation of dead above-ground biomass [107] and the amount of litter has shown to affect the spectral response of grassland vegetation [108]. Our different-aged grassland plots—characterized by different continuity of grazing—may be expected to be associated with different amount of litter, which, in its turn, may have contributed to the spectral dissimilarities between the plots.

The results of our study are consistent with the results from previous ecological studies, which show that plant community characteristics such as the cover of grasses and community-weighted mean Ellenberg indicator values for nutrient availability, change during grassland succession [34,109]. The results of the present study also agree with those of previous interdisciplinary studies within remote sensing and ecology. Earlier studies have shown, for example, that hyperspectral reflectance measurements in grasslands are related to community-weighted mean Ellenberg indicator values for nutrient availability and for plant functional traits, such as SLA [12,32,93].

Although the accuracy of the predictions we obtained using models 1 and 2 were high, some grassland plots in our sample study were not classified correctly by either model. There are a number of possible reasons for this discrepancy. First, we conducted the spectral measurements in early July, when leaf senescence in response to summer drought is likely to have reduced the spectral differences between grassland age-classes. Second, hyperspectral images are known to be affected by radiometric noise [110]. While we used the commonly applied Savitzky-Golay filter [63], other denoising methods, such as the Minimum Noise Fraction (MNF) transform [111], might have improved the noise-reduction in our data. Thirdly, PLS-DA may not be the most optimal method for classifying grassland age-classes. Other classification methods, such as artificial neural networks or support-vector machines (*cf.* [112]), which are able to take into account non-linear relationships between dependent and explanatory variables, might provide better discrimination between grassland successional stages.

### *4.3. Future Directions*

Future research targeting the improvement of hyperspectral methods for classifying grassland successional stages may include the development of vegetation spectral indices (e.g., [93]), which could potentially provide enhanced information on the plant community characteristics associated with different grassland successional stages. There is also a need for studies that focus on the development of classification methods for discriminating between age-classes at the scale of entire grassland sites. In addition, in order to be able to develop the possible use of hyperspectral data as a source of information within ecological research, we need a better understanding of the ways in which different plant community variables (characterizing different grassland successional stages) influence the hyperspectral response. The spectral response of vegetation canopies changes over the vegetation period as a result of temporal variation in the biochemical and structural properties that influence the reflectance of the vegetation canopy [44]. Additional studies will need to explore seasonal variation in the relationships between hyperspectral data and plant community variables.

## 5. Conclusions

In the present study we demonstrate that remote sensing data, acquired with the help of two airborne HySpex hyperspectral spectrometers (together covering 414 to 2501 nm), can be used to discriminate between dry grassland vegetation associated with different stages of arable-to-grassland succession within a local agricultural landscape in Sweden. Differences between the spectral responses of different grassland successional stages were associated with differences in the Ellenberg indicator value for nutrient availability and in the ground-cover of grasses.

We analysed the hyperspectral data using partial least squares discriminant analysis—a recently introduced method for the remote sensing-based classification of vegetation (e.g., [47])—and successfully classified grasslands into three different grassland age-classes, representing 5–15, 16–50, and >50 years of grazing management, respectively. We used an independent validation dataset to evaluate the classification-accuracy of our method.

The study shows that the variable importance in projection method [77] can be used to identify the wavebands that are the most important predictor variables in the hyperspectral classification of grassland age-classes. The accuracy of a partial least squares classification based on a subset of 177 wavebands, identified with the help of the variable importance in projection approach as those that were most important for discriminating between successional stages, was 85% (8% higher than for a classification based on the full set of 269 bands). Among the 177 hyperspectral wavebands that gave the most efficient discrimination between grassland age-classes, 50 wavebands were located in the visible region (414–716 nm), 79 wavebands in the red-edge to near-infrared regions (722–1394 nm), and 48 wavebands in the shortwave infrared region (1448–2417 nm) of the electromagnetic spectrum. The fact that the best wavebands for discriminating between grassland age-classes fell within the operating range of both the HySpex VNIR-1600 spectrometer (414 to 991 nm) and the HySpex SWIR-320m-e spectrometer (966 to 2501 nm) suggests that data from specific wavebands covering the full 400–2500 nm spectral range are likely to provide the best classification of grassland successional stages. Our results also show that the partial least squares-based classification procedure is a suitable method for the classification of grasslands successional stages, allowing a large number of hyperspectral wavebands to be compressed into a few latent variables while decreasing the risk of model overfitting. In our study, the first four latent variables explained approximately 97% of the variation in the spectral data.

A recent horizon-scan review [16] identifies 15 issues, or emerging topics, that are expected to have increasingly important implications for the global conservation of biological diversity, and which require wider consideration. The potential use of remote sensing-based techniques for monitoring land cover change is recognized as one of the 15 topics, and the review identifies the need to develop remote sensing methods for monitoring land cover change at detailed spatial scales and over large areas. Native grasslands are specifically mentioned as a type of land cover that is currently relatively difficult to monitor with the help of remote sensing techniques [16]. The results of the present study demonstrate that airborne hyperspectral data are capable of capturing detailed-scale information that discriminates between grassland plant communities representing different stages of an arable-to-grassland succession within a local landscape—suggesting that a similar approach may hold promise for the remote sensing-based mapping of grasslands belonging to different successional stages over larger areas.

There are a number of possible ways in which the methodology that we used might be improved in future studies. For example, whereas the present study was based on remote sensing data from a single time-point, a multi-temporal approach is likely to improve the ability to discriminate successfully between grasslands belonging to different successional stages. Future studies should use spectral data collected at different time-points during the vegetation season. In our study, we only used one method (the Savitzky-Golay filter) to reduce radiometric noise in the hyperspectral data: subsequent studies should explore the relative efficiencies of different denoising methods. Finally, because aerial remote sensing systems have a smaller coverage area than satellite-based remote sensing systems, future studies should investigate the possibility of using hyperspectral satellite data, for example data acquired with the help of the Hyperion satellite [113] and the planned EnMAP satellite [114], for monitoring grassland successional stages over wider areas in agricultural landscapes.

## Acknowledgments

The study was financed by grants from The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) to Karin Hall and Honor C. Prentice. We thank Sofia Pallander, Andreas Press, and Johan Rydlöv for field assistance. The “Station Linné” research station at Ölands Skogsby was used as a base for the fieldwork. We are grateful to the anonymous reviewers for their valuable comments on an earlier version of the manuscript. The study is a contribution to the Lund University Strategic Research Area “Biodiversity and Ecosystem Services in a Changing Climate” (BECC). Oliver Purschke acknowledges the support of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118).

## Author Contributions

Thomas Möckel, Jonas Dalmayne, Honor C. Prentice and Karin Hall planned and participated in the field work. Thomas Möckel, Jonas Dalmayne and Karin Hall planned the collection of remotely sensed data. Thomas Möckel processed and analysed the data. All authors were involved in the research design, the interpretation of results, and the preparation of the manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## Appendix

**Table A1.** List of recorded species (taxa). The presence (1) and absence (0) for each species within grasslands of 5–15 (Yng), 16–50 (Int) and >50 (Old) years of grazing management. Nomenclature mainly follows Mossberg & Stenberg (2010) [115].

| Species (Taxon) | Yng | Int | Old |
|---|---|---|---|
| *Achillea millefolium* | 1 | 1 | 1 |
| *Agrimonia eupatoria* | 0 | 1 | 1 |
| *Agrostis capillaris* | 1 | 1 | 1 |
| *Agrostis gigantea/stolonifera* | 1 | 1 | 1 |
| *Agrostis vinealis* | 1 | 1 | 1 |
| *Alchemilla* sp. | 1 | 1 | 1 |
| *Allium vineale/oleraceum* | 1 | 1 | 1 |
| *Alopecurus pratensis* | 0 | 1 | 0 |
| *Anagallis arvensis* | 0 | 1 | 0 |
| *Anthemis arvensis* | 1 | 1 | 1 |
| *Anthoxanthum odoratum* | 1 | 1 | 1 |
| *Anthriscus sylvestris* | 1 | 0 | 0 |
| *Anthyllis vulneraria* | 1 | 1 | 1 |
| *Aphanes arvensis* | 1 | 1 | 1 |
| *Arabidopsis thaliana* | 0 | 1 | 1 |
| *Arabis hirsuta* | 1 | 1 | 1 |
| *Arenaria serpyllifolia* | 1 | 1 | 1 |
| *Arrhenatherum elatius* | 1 | 1 | 1 |
| *Artemisia absinthium* | 1 | 1 | 1 |
| *Artemisia campestris* | 0 | 0 | 1 |
| *Asperula tinctoria* | 0 | 0 | 1 |
| *Barbarea vulgaris* | 1 | 0 | 0 |
| *Bellis perennis* | 1 | 1 | 1 |
| *Brachypodium sylvaticum* | 0 | 1 | 0 |
| *Briza media* | 1 | 1 | 1 |
| *Bromopsis erecta* | 1 | 1 | 0 |
| *Bromopsis inermis* | 1 | 1 | 0 |
| *Bromus hordeaceus* | 1 | 1 | 1 |
| *Campanula persicifolia* | 1 | 1 | 1 |
| *Campanula rotundifolia* | 0 | 0 | 1 |
| *Capsella bursa-pastoris* | 1 | 1 | 1 |
| *Cardamine hirsuta* | 0 | 0 | 1 |
| *Carex caryophyllea/ericetorum* | 0 | 1 | 1 |
| *Carex echinata* | 0 | 1 | 0 |
| *Carex flacca* | 1 | 1 | 1 |
| *Carex hirta* | 0 | 1 | 0 |
| *Carex spicata* group | 1 | 0 | 0 |
| *Carex sylvatica* | 1 | 1 | 0 |
| *Carex tomentosa* | 0 | 0 | 1 |
| *Carlina vulgaris* | 1 | 1 | 1 |

**Table A1.** *Cont.*

| Species (Taxon) | Yng | Int | Old |
|---|---|---|---|
| *Centaurea jacea* | 1 | 1 | 1 |
| *Centaurea scabiosa* | 1 | 1 | 1 |
| *Centaurium erythraea* | 0 | 1 | 0 |
| *Cerastium* sp. | 1 | 1 | 1 |
| *Cerastium arvense* | 1 | 1 | 0 |
| *Chenopodium album* | 0 | 1 | 0 |
| *Cichorium intybus* | 1 | 1 | 0 |
| *Cirsium acaule* | 1 | 1 | 1 |
| *Cirsium arvense* | 1 | 1 | 1 |
| *Convolvulus arvensis* | 1 | 1 | 1 |
| *Crepis tectorum* | 1 | 1 | 1 |
| *Cynosurus cristatus* | 1 | 1 | 1 |
| *Dactylis glomerata* | 1 | 1 | 1 |
| *Daucus carota* | 1 | 1 | 1 |
| *Deschampsia cespitosa* | 0 | 1 | 0 |
| *Dianthus deltoides* | 1 | 1 | 1 |
| *Draba muralis* | 1 | 1 | 1 |
| *Elytrigia repens* | 1 | 1 | 1 |
| *Erodium cicutarium* | 1 | 0 | 1 |
| *Erophila verna* | 0 | 0 | 1 |
| *Falcaria vulgaris* | 1 | 0 | 0 |
| *Festuca ovina* | 1 | 1 | 1 |
| *Festuca pratensis* | 1 | 1 | 1 |
| *Festuca rubra* | 1 | 1 | 1 |
| *Filipendula vulgaris* | 0 | 1 | 1 |
| *Fragaria vesca/viridis* | 1 | 1 | 1 |
| *Gagea* sp. | 0 | 1 | 0 |
| *Galium album* | 1 | 1 | 1 |
| *Galium boreale* | 0 | 1 | 1 |
| *Galium verum* | 1 | 1 | 1 |
| *Geranium columbinum* | 1 | 1 | 1 |
| *Geranium dissectum* | 1 | 0 | 0 |
| *Geranium molle* | 1 | 1 | 1 |
| *Geum urbanum* | 1 | 1 | 1 |
| *Helianthemum nummularium* | 0 | 0 | 1 |
| *Helianthemum oelandicum* | 0 | 0 | 1 |
| *Helictotrichon pratensis* | 0 | 1 | 1 |
| *Helictotrichon pubescens* | 1 | 1 | 1 |
| *Herniaria glabra* | 1 | 0 | 0 |
| *Hypericum perforatum* | 1 | 1 | 1 |
| *Hypochoeris radicata* | 1 | 1 | 0 |
| *Knautia arvensis* | 1 | 1 | 1 |
| *Lactuca serriola* | 0 | 0 | 1 |

**Table A1.** *Cont.*

| Species (Taxon) | Yng | Int | Old |
|---|---|---|---|
| *Lapsana communis* | 0 | 1 | 0 |
| *Lathyrus pratensis* | 1 | 0 | 0 |
| *Leontodon autumnalis* | 0 | 1 | 0 |
| *Leucanthemum vulgare* | 1 | 1 | 1 |
| *Linaria vulgaris* | 0 | 1 | 1 |
| *Linum catharticum* | 1 | 1 | 1 |
| *Listera ovata* | 1 | 0 | 0 |
| *Lolium perenne* | 1 | 1 | 1 |
| *Lotus corniculatus* | 1 | 1 | 1 |
| *Luzula campestris* | 1 | 1 | 1 |
| *Malva neglecta* | 1 | 0 | 0 |
| *Matricaria perforata* | 0 | 1 | 0 |
| *Medicago falcata* | 1 | 1 | 1 |
| *Medicago lupulina* | 1 | 1 | 1 |
| *Moehringia trinervia* | 0 | 1 | 0 |
| *Myosotis arvensis/ramosissima* | 0 | 0 | 1 |
| *Myosurus minimus* | 1 | 0 | 0 |
| *Odontites vulgaris* | 1 | 0 | 0 |
| *Ononis spinosa* | 1 | 1 | 1 |
| *Orchis mascula/militaris* | 0 | 0 | 1 |
| *Origanum vulgare* | 1 | 0 | 0 |
| *Ornithogalum angustifolium* | 1 | 1 | 0 |
| *Oxytropis campestris* | 0 | 0 | 1 |
| *Papaver* sp. | 1 | 0 | 0 |
| *Phleum phleoides* | 0 | 1 | 1 |
| *Phleum pratense* | 1 | 1 | 1 |
| *Pilosella officinarum* | 1 | 1 | 1 |
| *Plantago lanceolata* | 1 | 1 | 1 |
| *Plantago major* | 0 | 1 | 0 |
| *Plantago media* | 0 | 1 | 0 |
| *Platanthera* sp. | 0 | 0 | 1 |
| *Poa annua* | 1 | 1 | 1 |
| *Poa bulbosa* | 0 | 0 | 1 |
| *Poa compressa* | 1 | 1 | 1 |
| *Poa pratensis* | 1 | 1 | 1 |
| *Poa trivialis* | 1 | 0 | 0 |
| *Polygala amarella* | 0 | 1 | 0 |
| *Polygala comosa/vulgaris* | 1 | 1 | 1 |
| *Polygonum aviculare* | 1 | 1 | 1 |
| *Potentilla anserina* | 1 | 0 | 0 |
| *Potentilla argentea* | 1 | 1 | 1 |
| *Potentilla reptans* | 1 | 1 | 1 |
| *Potentilla tabernaemontani* | 1 | 1 | 1 |

**Table A1.** *Cont.*

| Species (Taxon) | Yng | Int | Old |
|---|---|---|---|
| *Primula veris* | 1 | 1 | 1 |
| *Prunella grandiflora/vulgaris* | 1 | 1 | 1 |
| *Prunus spinosa* | 1 | 1 | 1 |
| *Pulsatilla pratensis* | 0 | 0 | 1 |
| *Ranunculus acris* | 1 | 1 | 1 |
| *Ranunculus auricomus* | 0 | 1 | 0 |
| *Ranunculus bulbosus* | 1 | 1 | 1 |
| *Ranunculus ficaria* | 0 | 1 | 0 |
| *Ranunculus illyricus* | 0 | 0 | 1 |
| *Ranunculus repens* | 0 | 1 | 0 |
| *Rosa* sp. | 0 | 0 | 1 |
| *Rubus* sp. | 1 | 0 | 1 |
| *Rumex acetosa* | 1 | 1 | 1 |
| *Rumex acetosella* | 1 | 1 | 0 |
| *Rumex longifolius* | 0 | 1 | 0 |
| *Sanguisorba minor* | 1 | 1 | 0 |
| *Satureja acinos* | 1 | 1 | 1 |
| *Saxifraga granulata* | 0 | 0 | 1 |
| *Scabiosa columbaria* | 0 | 0 | 1 |
| *Scleranthus annuus* | 0 | 0 | 1 |
| *Sedum acre* | 1 | 1 | 1 |
| *Senecio jacobaea* | 1 | 1 | 1 |
| *Sesleria caerulea* | 0 | 1 | 1 |
| *Sherardia arvensis* | 1 | 1 | 0 |
| *Silene nutans* | 1 | 0 | 1 |
| *Stellaria graminea* | 0 | 1 | 1 |
| *Stellaria media* | 1 | 1 | 0 |
| *Tanacetum vulgare* | 0 | 1 | 0 |
| *Taraxacum* agg. | 1 | 1 | 1 |
| *Thalictrum flavum* | 1 | 0 | 0 |
| *Thlaspi* sp. | 1 | 1 | 0 |
| *Thymus serpyllum* | 0 | 1 | 1 |
| *Tragopogon pratensis* | 1 | 1 | 1 |
| *Trifolium arvense* | 1 | 0 | 1 |
| *Trifolium campestre/dubium* | 1 | 1 | 1 |
| *Trifolium pratense* | 1 | 1 | 1 |
| *Trifolium repens* | 1 | 1 | 1 |
| *Valeriana officinalis* | 1 | 0 | 0 |
| *Valerianella locusta* | 0 | 1 | 0 |
| *Veronica arvensis* | 1 | 1 | 1 |
| *Veronica chamaedrys* | 1 | 1 | 1 |
| *Veronica hederifolia* | 0 | 1 | 0 |
| *Veronica officinalis* | 0 | 1 | 1 |

**Table A1.** *Cont.*

| Species (Taxon) | Yng | Int | Old |
|---|---|---|---|
| *Veronica persica* | 1 | 1 | 0 |
| *Veronica serpyllifolia* | 1 | 1 | 0 |
| *Veronica spicata* | 0 | 0 | 1 |
| *Vicia angustifolia* | 1 | 1 | 0 |
| *Vicia cracca* | 0 | 1 | 0 |
| *Vicia hirsuta* | 1 | 1 | 1 |
| *Vicia lathyroides* | 1 | 0 | 0 |
| *Vicia tetrasperma* | 0 | 0 | 1 |
| *Viola arvensis* | 0 | 0 | 1 |
| *Viola hirta* | 1 | 1 | 1 |

**Figure A1.** Field photographs of sample plots representing each of the three grassland age-classes. (**a**) Young grassland with a grazing management of 5–15 years; (**b**) Intermediate-aged grassland with a grazing management of 16–50 years; (**c**) Old grassland with a grazing management of >50 years.

(**a**) (**b**) (**c**)

## References

1. Benton, T.G.; Vickery, J.A.; Wilson, J.D. Farmland biodiversity: Is habitat heterogeneity the key? *Trends Ecol. Evol.* **2003**, *18*, 182–188.
2. Eriksson, O.; Cousins, S.A.O.; Bruun, H.H. Land-use history and fragmentation of traditionally managed grasslands in Scandinavia. *J. Veg. Sci.* **2002**, *13*, 743–748.
3. Tscharntke, T.; Klein, A.M.; Kruess, A.; Steffan-Dewenter, I.; Thies, C. Landscape perspectives on agricultural intensification and biodiversity—Ecosystem service management. *Ecol. Lett.* **2005**, *8*, 857–874.
4. Cramer, V.A.; Hobbs, R.J.; Standish, R.J. What's new about old fields? Land abandonment and ecosystem assembly. *Trends Ecol. Evol.* **2008**, *23*, 104–112.
5. Reger, B.; Mattern, T.; Otte, A.; Waldhardt, R. Assessing the spatial distribution of grassland age in a marginal European landscape. *J. Environ. Manag.* **2009**, *90*, 2900–2909.
6. Cramer, V.; Hobbs, R.J. *Old Fields: Dynamics and Restoration of Abandoned Farmland*; International Island Press: Washington, DC, USA, 2007.
7. Cousins, S.A.O.; Lindborg, R.; Mattsson, S. Land use history and site location are more important for grassland species richness than local soil properties. *Nord. J. Bot.* **2009**, *27*, 483–489.

8. Lengyel, S.; Varga, K.; Kosztyi, B.; Lontay, L.; Déri, E.; Török, P.; Tóthmérész, B. Grassland restoration to conserve landscape-level biodiversity: A synthesis of early results from a large-scale project. *Appl. Veg. Sci.* **2012**, *15*, 264–276.
9. Stevenson, M.J.; Bullock, J.M.; Ward, L.K. Re-creating semi-natural communities: Effect of sowing rate on establishment of calcareous grassland. *Restor. Ecol.* **1995**, *3*, 279–289.
10. Purschke, O.; Sykes, M.T.; Reitalu, T.; Poschlod, P.; Prentice, H.C. Linking landscape history and dispersal traits in grassland plant communities. *Oecologia* **2012**, *168*, 773–783.
11. Kerr, J.T.; Ostrovsky, M. From space to species: Ecological applications for remote sensing. *Trends Ecol. Evol.* **2003**, *18*, 299–305.
12. Schmidtlein, S. Imaging spectroscopy as a tool for mapping Ellenberg indicator values. *J. Appl. Ecol.* **2005**, *42*, 966–974.
13. Rocchini, D.; Balkenhol, N.; Carter, G.A.; Foody, G.M.; Gillespie, T.W.; He, K.S.; Kark, S.; Levin, N.; Lucas, K.; Luoto, M.; *et al.* Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecol. Inform.* **2010**, *5*, 318–329.
14. Gillespie, T.W.; Foody, G.M.; Rocchini, D.; Giorgi, A.P.; Saatchi, S. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* **2008**, *32*, 203–221.
15. Ustin, S.L.; Gamon, J.A. Remote sensing of plant functional types. *New Phytol.* **2010**, *186*, 795–816.
16. Sutherland, W.J.; Aveling, R.; Brooks, T.M.; Clout, M.; Dicks, L.V.; Fellman, L.; Fleishman, E.; Gibbons, D.W.; Keim, B.; Lickorish, F.; *et al.* A horizon scan of global conservation issues for 2014. *Trends Ecol. Evol.* **2014**, *29*, 15–22.
17. Waldhardt, R.; Otte, A. Indicators of plant species and community diversity in grasslands. *Agric. Ecosyst. Environ.* **2003**, *98*, 339–351.
18. Kawamura, K.; Akiyama, T.; Yokota, H.; Tsutsumi, M.; Yasuda, T.; Watanabe, O.; Wang, S. Quantifying grazing intensities using geographic information systems and satellite remote sensing in the Xilingol steppe region, Inner Mongolia, China. *Agric. Ecosyst. Environ.* **2005**, *107*, 83–93.
19. Thenkabail, P.S.; Smith, R.B.; De Pauw, E. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 607–621.
20. Thenkabail, P.S.; Smith, R.B.; de Pauw, E. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* **2000**, *71*, 158–182.
21. Chan, J.C.-W.; Paelinckx, D. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens. Environ.* **2008**, *112*, 2999–3011.
22. Thenkabail, P.S.; Enclona, E.A.; Ashton, M.S.; van der Meer, B. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ.* **2004**, *91*, 354–376.
23. Thenkabail, P.S.; Enclona, E.A.; Ashton, M.S.; Legg, C.; de Dieu, M.J. Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests. *Remote Sens. Environ.* **2004**, *90*, 23–43.
24. Jacquemoud, S.; Ustin, S.L.; Verdebout, J.; Schmuck, G.; Andreoli, G.; Hosgood, B. Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sens. Environ.* **1996**, *56*, 194–202.

25. Curran, P.J. Remote-sensing of foliar chemistry. *Remote Sens. Environ.* **1989**, *30*, 271–278.
26. Ustin, S.L.; Gitelson, A.A.; Jacquemoud, S.; Schaepman, M.; Asner, G.P.; Gamon, J.A.; Zarco-Tejada, P. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sens. Environ.* **2009**, *113*, 67–77.
27. Blackburn, G.A. Hyperspectral remote sensing of plant pigments. *J. Exp. Bot.* **2007**, *58*, 855–867.
28. Cheng, Y.-B.; Zarco-Tejada, P.J.; Riaño, D.; Rueda, C.A.; Ustin, S.L. Estimating vegetation water content with hyperspectral data for different canopy scenarios: Relationships between AVIRIS and MODIS indexes. *Remote Sens. Environ.* **2006**, *105*, 354–366.
29. Darvishzadeh, R.; Skidmore, A.; Atzberger, C.; van Wieren, S. Estimation of vegetation LAI from hyperspectral reflectance data: Effects of soil type and plant architecture. *Int. J. Appl. Earth Obs.* **2008**, *10*, 358–373.
30. Ellenberg, H. *Zeigerwerte der Gefässpflanzen Mitteleuropas*; Goltze: Göttingen, Germany, 1974; Volume 9.
31. Diekmann, M. Species indicator values as an important tool in applied plant ecology—A review. *Basic Appl. Ecol.* **2003**, *4*, 493–506.
32. Klaus, V.H.; Kleinebecker, T.; Boch, S.; Müller, J.; Socher, S.A.; Prati, D.; Fischer, M.; Hölzel, N. NIRS meets Ellenberg's indicator values: Prediction of moisture and nitrogen values of agricultural grassland vegetation by means of near-infrared spectral characteristics. *Ecol. Indic.* **2012**, *14*, 82–86.
33. Purschke, O.; Schmid, B.C.; Sykes, M.T.; Poschlod, P.; Michalski, S.G.; Durka, W.; Kühn, I.; Winter, M.; Prentice, H.C. Contrasting changes in taxonomic, phylogenetic and functional diversity during a long-term succession: Insights into assembly processes. *J. Ecol.* **2013**, *101*, 857–866.
34. Pykälä, J.; Luoto, M.; Heikkinen, R.K.; Kontula, T. Plant species richness and persistence of rare plants in abandoned semi-natural grasslands in northern Europe. *Basic Appl. Ecol.* **2005**, *6*, 25–33.
35. Garnier, E.; Cortez, J.; Billés, G.; Navas, M.-L.; Roumet, C.; Debussche, M.; Laurent, G.; Blanchard, A.; Aubry, D.; Bellmann, A.; *et al.* Plant functional markers capture ecosystem properties during secondary succession. *Ecology* **2004**, *85*, 2630–2637.
36. Wright, I.J.; Reich, P.B.; Westoby, M.; Ackerly, D.D.; Baruch, Z.; Bongers, F.; Cavender-Bares, J.; Chapin, T.; Cornelissen, J.H.C.; Diemer, M.; *et al.* The worldwide leaf economics spectrum. *Nature* **2004**, *428*, 821–827.
37. Shipley, B.; Vu, T.-T. Dry matter content as a measure of dry matter concentration in plants and their parts. *New Phytol.* **2002**, *153*, 359–364.
38. Evans, J.R.; Poorter, H. Photosynthetic acclimation of plants to growth irradiance: The relative importance of specific leaf area and nitrogen partitioning in maximizing carbon gain. *Plant Cell Environ.* **2001**, *24*, 755–767.
39. Shipley, B. Structured interspecific determinants of specific leaf area in 34 species of herbaceous angiosperms. *Funct. Ecol.* **1995**, *9*, 312–319.
40. Viedma, O.; Torres, I.; Pérez, B.; Moreno, J.M. Modeling plant species richness using reflectance and texture data derived from QuickBird in a recently burned area of Central Spain. *Remote Sens. Environ.* **2012**, *119*, 208–221.

41. Leutner, B.F.; Reineking, B.; Müller, J.; Bachmann, M.; Beierkuhnlein, C.; Dech, S.; Wegmann, M. Modelling forest alpha-diversity and floristic composition—On the added value of LiDAR plus hyperspectral remote sensing. *Remote Sens.* **2012**, *4*, 2818–2845.
42. Pal, M.; Mather, P.M. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* **2005**, *26*, 1007–1011.
43. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
44. Feilhauer, H.; Schmidtlein, S. On variable relations between vegetation patterns and canopy reflectance. *Ecol. Inform.* **2011**, *6*, 83–92.
45. Chen, J.; Gu, S.; Shen, M.; Tang, Y.; Matsushita, B. Estimating aboveground biomass of grassland having a high canopy cover: An exploratory analysis of *in situ* hyperspectral data. *Int. J. Remote Sens.* **2009**, *30*, 6497–6517.
46. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C.; Corsi, F.; Cho, M. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 409–426.
47. Peerbhay, K.Y.; Mutanga, O.; Ismail, R. Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa. *ISPRS J. Photogramm. Remote Sens.* **2013**, *79*, 19–28.
48. Forslund, M. *Natur och Kultur på Öland*; Länsstyrelsen i Kalmar: Kalmar, Sweden, 2001.
49. Johansson, L.J.; Hall, K.; Prentice, H.C.; Ihse, M.; Reitalu, T.; Sykes, M.T.; Kindström, M. Semi-natural grassland continuity, long-term land-use change and plant species richness in an agricultural landscape on Oland, Sweden. *Landsc. Urban Plan.* **2008**, *84*, 200–211.
50. Reitalu, T.; Prentice, H.C.; Sykes, M.T.; Lönn, M.; Johansson, L.J.; Hall, K. Plant species segregation on different spatial scales in semi-natural grasslands. *J. Veg. Sci.* **2008**, *19*, 407–416.
51. Reitalu, T.; Sykes, M.T.; Johansson, L.J.; Lönn, M.; Hall, K.; Vandewalle, M.; Prentice, H.C. Small-scale plant species richness and evenness in semi-natural grasslands respond differently to habitat fragmentation. *Biol. Conserv.* **2009**, *142*, 899–908.
52. Prentice, H.C.; Cramer, W. The plant community as a niche bioassay: Environmental correlates of local variation in *Gypsophila-fastigata*. *J. Ecol.* **1990**, *78*, 313–325.
53. Kleyer, M.; Bekker, R.M.; Knevel, I.C.; Bakker, J.P.; Thompson, K.; Sonnenschein, M.; Poschlod, P.; van Groenendael, J.M.; Klimeš, L.; Klimešová, J.; *et al.* The LEDA Traitbase: A database of life-history traits of the Northwest European flora. *J. Ecol.* **2008**, *96*, 1266–1274.
54. Taugourdeau, S.; Villerd, J.; Plantureux, S.; Huguenin-Elie, O.; Amiaud, B. Filling the gap in functional trait databases: Use of ecological hypotheses to replace missing data. *Ecol. Evol.* **2014**, *4*, 944–958.
55. Van Buuren, S.; Groothuis-Oudshoorn, K. Mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67.
56. Azur, M.J.; Stuart, E.A.; Frangakis, C.; Leaf, P.J. Multiple imputation by chained equations: What is it and how does it work? *Int. J. Methods Psychiatr. Res.* **2011**, *20*, 40–49.
57. Tichý, L. JUICE, software for vegetation classification. *J. Veg. Sci.* **2002**, *13*, 451–453.

58. Ellenberg, H. Zeigerwerte von Pflanzen in Mitteleuropa. In *Zeigerwerte von Pflanzen in Mitteleuropa*; Ellenberg, H., Weber, H.E., Düll, R.; Wirth, V., Werner, W., Paulissen, D., Eds.; Goltze: Göttingen, Germany, 1991; Volume 18.
59. Richter, R.; Schläpfer, D. Geo-atmospheric processing of wide FOV airborne imaging spectrometry data. In *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology*; Ehlers, M., Ed.; Society of Photo Optical: Bellingham, WA, USA, 2002; Volume 4545, pp. 264–273.
60. Schläpfer, D.; Richter, R. Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: Parametric orthorectification. *Int. J. Remote Sens.* **2002**, *23*, 2609–2630.
61. Fontenla, J.M.; Harder, J.; Livingston, W.; Snow, M.; Woods, T. High-resolution solar spectral irradiance from extreme ultraviolet to far infrared. *J. Geophys. Res. Atmos.* **2011**, *116*, doi:10.1029/2011JD016032.
62. Fontenla, J.M.; Curdt, W.; Haberreiter, M.; Harder, J.; Tian, H. Semiempirical models of the solar atmosphere. III. Set of non-lte models for far-ultraviolet/extreme-ultraviolet irradiance computation. *Astrophys. J.* **2009**, *707*, 482–502.
63. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least square procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
64. Feilhauer, H.; Asner, G.P.; Martin, R.E.; Schmidtlein, S. Brightness-normalized partial least squares regression for hyperspectral data. *J. Quant. Spectrosc. Radiat.* **2010**, *111*, 1947–1957.
65. Beleites, C.; Sergo, V. HyperSpec: A Package to Handle Hyperspectral Data Sets in R. Available online: http://hyperspec.r-forge.r-project.org (accessed on 29 May 2014).
66. Signal; Developers Signal: Signal Processing. Available online: http://r-forge.r-project.org/projects/signal/ (accessed on 29 May 2014).
67. R Development Core Team R: A Language and Environment for Statistical Computing. Available online: http://www.R-project.org/ (accessed on 29 May 2014).
68. Psomas, A.; Kneubühler, M.; Huber, S.; Itten, K.; Zimmermann, N.E. Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. *Int. J. Remote Sens.* **2011**, *32*, 9007–9031.
69. Carter, G.A.; Knapp, A.K.; Anderson, J.E.; Hoch, G.A.; Smith, M.D. Indicators of plant species richness in AVIRIS spectra of a mesic grassland. *Remote Sens. Environ.* **2005**, *98*, 304–316.
70. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; García Marquéz, J.R.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; *et al.* Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46.
71. Schmidtlein, S.; Feilhauer, H.; Bruelheide, H. Mapping plant strategy types using remote sensing. *J. Veg. Sci.* **2011**, *23*, 395–405.
72. Carrascal, L.M.; Galván, I.; Gordo, O. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* **2009**, *118*, 681–690.
73. Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166–173.
74. Whelehan, O.P.; Earll, M.E.; Johansson, E.; Toft, M.; Eriksson, L. Detection of ovarian cancer using chemometric analysis of proteomic profiles. *Chemom. Intell. Lab. Syst.* **2006**, *84*, 82–87.

75. Feilhauer, H.; Faude, U.; Schmidtlein, S. Combining Isomap ordination and imaging spectroscopy to map continuous floristic gradients in a heterogeneous landscape. *Remote Sens. Environ.* **2011**, *115*, 2513–2524.
76. Mevik, B.H.; Wehrens, R. The pls package: Principal component and partial least squares regression in R. *J. Stat. Softw.* **2007**, *18*, 1–24.
77. Wold, S.; Johansson, E.; Cocchi, M. PLS—Partial least squares projections to latent structures. In *3D QSAR in Drug Design: Volume 1: Theory Methods and Applications*; Kubinyi, H., Ed. Springer: Berlin/Heidelberg, Germany, 1993; pp. 523–550.
78. Chong, I.-G.; Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112.
79. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46.
80. Campbell, J.B. *Introduction to Remote Sensing*; Taylor & Francis: London, UK, 2002.
81. Dale, L.M.; Thewis, A.; Boudry, C.; Rotar, I.; Păcurar, F.S.; Abbas, O.; Dardenne, P.; Baeten, V.; Pfister, J.; Fernández Pierna, J.A. Discrimination of grassland species and their classification in botanical families by laboratory scale NIR hyperspectral imaging: Preliminary results. *Talanta* **2013**, *116*, 149–154.
82. Olsson, Y. *Jordbruksmarkens Användning 2011. Sveriges Officiella Statistik*; Jordbruksverket: Stockholm, Sweden, 2012.
83. Weidow, B. *Växtodlingens Grunder*; LT: Stockholm, Sweden, 1998.
84. Carbajo, V.; den Braber, B.; van der Putten, W.H.; de Deyn, G.B. Enhancement of late successional plants on ex-arable land by soil inoculations. *PLoS One* **2011**, *6*, doi:10.1371/journal.pone.0021943.
85. Ceulemans, T.; Merckx, R.; Hens, M.; Honnay, O. A trait-based analysis of the role of phosphorus *vs.* nitrogen enrichment in plant species loss across North-West European grasslands. *J. Appl. Ecol.* **2011**, *48*, 1155–1163.
86. Pavlů, V.; Schellberg, J.; Hejcman, M. Cutting frequency *vs.* N application: Effect of a 20-year management in *Lolio-Cynosuretum* grassland. *Grass Forage Sci.* **2011**, *66*, 501–515.
87. Liu, X.; Swenson, N.G.; Wright, S.J.; Zhang, L.; Song, K.; Du, Y.; Zhang, J.; Mi, X.; Ren, H.; Ma, K. Covariation in plant functional traits and soil fertility within two species-rich forests. *PLoS One* **2012**, *7*, e34767:1–e34767:9.
88. Chapin, F.S.; Bloom, A.J.; Field, C.B.; Waring, R.H. Plant responses to multiple environmental factors. *Bioscience* **1987**, *37*, 49–57.
89. Ekstam, U. Forshed *Svenska Alvarmarker*; Naturvårdsverket Förlag: Stockholm, Sweden, 2002.
90. Purschke, O.; Sykes, M.T.; Poschlod, P.; Michalski, S.G.; Römermann, C.; Durka, W.; Kühn, I.; Prentice, H.C. Interactive effects of landscape history and current management on dispersal trait diversity in grassland plant communities. *J. Ecol.* **2014**, *102*, 437–446.
91. Schlerf, M.; Atzberger, C.; Hill, J. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sens. Environ.* **2005**, *95*, 177–194.
92. Kokaly, R.F.; Asner, G.P.; Ollinger, S.V.; Martin, M.E.; Wessman, C.A. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sens. Environ.* **2009**, *113*, 78–91.

93. Oldeland, J.; Wesuls, D.; Jürgens, N. RLQ and fourth-corner analysis of plant species traits and spectral indices derived from HyMap and CHRIS-PROBA imagery. *Int. J. Remote Sens.* **2012**, *33*, 6459–6479.
94. Serrano, L.; Peñuelas, J.; Ustin, S.L. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sens. Environ.* **2002**, *81*, 355–364.
95. Johnson, L.F.; Hlavka, C.A.; Peterson, D.L. Multivariate analysis of AVIRIS data for canopy biochemical estimation along the oregon transect. *Remote Sens. Environ.* **1994**, *47*, 216–230.
96. Kokaly, R.F.; Clark, R.N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* **1999**, *67*, 267–287.
97. Lamb, D.W.; Steyn-Ross, M.; Schaare, P.; Hanna, M.M.; Silvester, W.; Steyn-Ross, A. Estimating leaf nitrogen concentration in ryegrass (*Lolium* spp.) pasture using the chlorophyll red-edge: Theoretical modelling and experimental observations. *Int. J. Remote Sens.* **2002**, *23*, 3619–3648.
98. Lymburner, L.; Beggs, P.J.; Jacobson, C.R. Estimation of canopy-average surface-specific leaf area using Landsat TM data. *Photogramm. Eng. Remote* **2000**, *66*, 183–191.
99. Asner, G.P.; Martin, R.E.; Ford, A.J.; Metcalfe, D.J.; Liddell, M.J. Leaf chemical and spectral diversity in Australian tropical forests. *Ecol. Appl.* **2009**, *19*, 236–253.
100. Riaño, D.; Vaughan, P.; Chuvieco, E.; Zarco-Tejada, P.J.; Ustin, S.L. Estimation of fuel moisture content by inversion of radiative transfer models to simulate equivalent water thickness and dry matter content: Analysis at leaf and canopy level. *IEEE Trans. Geosci. Remote* **2005**, *43*, 819–826.
101. Wong, K.K.; He, Y. Estimating grassland chlorophyll content using remote sensing data at leaf, canopy, and landscape scales. *Can. J. Remote Sens.* **2013**, *39*, 155–166.
102. Asner, G.P. Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sens. Environ.* **1998**, *64*, 234–253.
103. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309.
104. Garrigues, S.; Allard, D.; Baret, F.; Morisette, J. Multivariate quantification of landscape spatial heterogeneity using variogram models. *Remote Sens. Environ.* **2008**, *112*, 216–230.
105. Chen, Y.; Wen, D.; Jing, L.; Shi, P. Shadow information recovery in urban areas from very high resolution satellite imagery. *Int. J. Remote Sens.* **2007**, *28*, 3249–3254.
106. Daughtry, C.S.T.; Walthall, C.L. Spectral Discrimination of *Cannabis sativa* L. leaves and canopies. *Remote Sens. Environ.* **1998**, *64*, 192–201.
107. Jensen, K.; Gutekunst, K. Effects of litter on establishment of grassland plant species: The role of seed size and successional status. *Basic Appl. Ecol.* **2003**, *4*, 579–587.
108. Zhang, C.; Guo, X. Monitoring northern mixed prairie health using broadband satellite imagery. *Int. J. Remote Sens.* **2008**, *29*, 2257–2271.
109. Hansson, M.; Fogelfors, H. Management of a semi-natural grassland; results from a 15-year-old experiment in southern Sweden. *J. Veg. Sci.* **2000**, *11*, 31–38.
110. Gómez-Chova, L.; Alonso, L.; Guanter, L.; Camps-Valls, G.; Calpe, J.; Moreno, J. Correction of systematic spatial noise in push-broom hyperspectral sensors: Application to CHRIS/PROBA images. *Appl. Opt.* **2008**, *47*, F46–F60.

111. Green, A.A.; Berman, M.; Switzer, P.; Craig, M.D. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 65–74.
112. Thissen, U.; Pepers, M.; Üstün, B.; Melssen, W.J.; Buydens, L.M.C. Comparing support vector machines to PLS for spectral regression applications. *Chemom. Intell. Lab. Syst.* **2004**, *73*, 169–179.
113. Pearlman, J.S.; Barry, P.S.; Segal, C.C.; Shepanski, J.; Beiso, D.; Carman, S.L. Hyperion, a space-based imaging spectrometer. *IEEE Trans. Geosci. Remote* **2003**, *41*, 1160–1173.
114. Stuffler, T.; Kaufmann, C.; Hofer, S.; Förster, K.P.; Schreier, G.; Mueller, A.; Eckardt, A.; Bach, H.; Penné, B.; Benz, U.; *et al.* The EnMAP hyperspectral imager—An advanced optical payload for future applications in Earth observation programmes. *Acta Astronaut.* **2007**, *61*, 115–120.
115. Mossberg, B.; Stenberg, L. *Den nya Nordiska Floran*; Wahlström & Widstrand: Stockholm, Sweden, 2010.