Master Thesis in Geographical Information Science nr 44

Relationship between tree species composition and phenology extracted from satellite data in Swedish forests




Stefan Arvidsson

2015
Department of
Physical Geography and Ecosystem Science
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Stefan Arvidsson (2015). Relationship between tree species composition and phenology extracted from satellite data in Swedish forests**.**
Master degree thesis, 30/ credits in Master in Geographical Information Sciences
Department of Physical Geography and Ecosystems Science, Lund University

Relationship between tree species composition and phenology extracted from satellite data in Swedish forests

---

Stefan Arvidsson

Master thesis, 30 credits, in Geographical Information Sciences

Supervisor:
Lars Eklundh
Department of Physical Geography and Ecosystem Science

Exam committee:
Jonas Ardö
Department of Physical Geography and Ecosystem Science

Hongxiao Jin
Department of Physical Geography and Ecosystem Science

Abstract

This study investigated the *"relationship between tree species composition and phenology extracted from satellite data in Swedish forests"*. The proposed method investigated in this study aims at mapping the fractional composition of deciduous/coniferous tree species and also the fractional composition of Norway spruce (*Picea abies*) and several pine (*Pinus sp*.). The fractions can then be used to classify a forest into forest types such as uniform deciduous/coniferous or mixed forests.

The method uses field measurements for training a regression model against satellite derived seasonality parameters. The satellite derived phenological parameters consists of a time series of normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) values where parameters such as maximum, length, start and end of growing season were extracted with the software TIMESAT. The satellite system used was the Moderate-Resolution Imaging Spectroradiometer (MODIS).

The result indicates that the EVI derived seasonality parameters correlated stronger against the fractional composition of deciduous/coniferous tree species as compared with NDVI derived seasonality. The correlation coefficient for the EVI derived seasonality was estimated to 0.88 for the best performing dataset and parameter. However, when validated against an independent dataset the accuracy proved to be low when the tested regression models were used to predict the fraction of deciduous/coniferous tree species composition. A source of error derives from differences in mapping scale between the satellite system (250x250 m) and that of the field data (either a plot size of a 7 m circle or a plot size 30x30 m). The differences in mapping scale are assumed be a major source of error. However, the correlation between satellite derived seasonality and the fractional tree species composition is strong enough to consider it worthwhile to investigate for future studies when better data will become available.

Sammanfattning

Den här studien har undersökt ”Förhållandet mellan trädslagssammansättning och fenologi extraherad ur satellitdata i svenska skogar”. Den föreslagna metoden som undersöktes i denna studie försöker att kartera andelen av lövträd/barrträd men också artsammansättningen av andelen gran (*Picea abies*) och tall (*Pinus sp.*). Andelsfördelningen kan sedan användas för att klassificera in data efter skogstyp såsom löv/barrskog eller blandskog.

Metoden använder fältmätningar för att empiriskt konstruera en regressionsmodell mot fenologiska parametrar som uppskattats utifrån satellitdata. De fenologiska parametrarna består av tidsserier med NDVI- och EVI-värden där parametrar extraherades såsom bl.a. maximum, längd, början och slutet av växtsäsongen. Det använda satellitsystemet utgjordes av Moderate-Resolution Imaging Spectroradiometer (MODIS). De fenologiska parametrarna samt brus som förekommer inom tidsserien beräknades och behandlades med programvaran TIMESAT.

Resultatet visar att de fenologiska parametrar som uppskattats utifrån EVI-värden korrelerade starkare mot andelsfördelningen av lövträd/barrträd jämfört med fenologiska parametrar extraherades från NDVI-värden. Korrelationskoefficienten för de säsongs parametrar som beräknats från EVI-värden uppskattades till 0,88 för det bäst presterande datasetet och parametern. Valideringen mot ett oberoende dataset visade dock att noggrannhet vara låg när de testade regressionsmodellerna användes för att förutsäga andelsfördelningen av lövträds-/barrträdssammansättning. En av de felkällor som bidrog mest till den låga noggrannheten härrör från skalskillnaden mellan satellitsystemet (250x250 m) och det använda fältdatasetet (provyta på 7 m cirkel eller 30x30 m). Men sambandet mellan fenologiska parametrar som uppskattats utifrån satellitdata och trädartssammansättning är så stark att det verkligen är värt att undersöka ytterligare i framtida studier när bättre data kommer att finns tillgängliga.

Acknowledgement

This master degree thesis would have looked very differently or never been possible to write without access to the expertise and skills possessed by my supervisor, Prof. Lars Eklundh. I would like to thank Dr. Cecilia Akselsson for giving me access to data and metadata from the monitoring plot program. I would also like to thank Ph.D. student Per-Ola Olsson for helping me out with the weighting of data in TIMESAT.

Also, I would like to seize the opportunity to thank Andreas Eriksson for all the time spent discussing ideas and solving problems that we encountered while studying at the iGEON-programme. These discussions have not only helped me to the point where I can graduate but more importantly it has also helped me cultivate a growing interest for GIS and remote sensing science in general.

Table of Content

1. Introduction

The change of season is a slow process that is marked by annual natural events. These natural events or phenomena can consist of the first notation of a migratory bird species in spring or the bud burst of the oak trees in the park on your way to school or work. These notations can become data of great scientific value when studying systemically. The systematic study of seasons is called phenology. Long cohesive time series of phenological events can reveal trends or patterns which may indicate changes in regional or global environment. Hence, phenological time series can be a carrier of environmental information (Linderholm 2006, Barr, Black and McCaughey 2009, Olsson 2014).

The scale of the observations is often performed for specific species or specimens but can also be done on large ecosystems. However, if phenology is to be observed on an ecosystem level, then other tools have to be adopted. Tools like remotely sensed satellite imagery arranged in time series can measure the spectral properties of different land covers or vegetation communities and therefore detect changes through time. A kind of indexed measure of vegetation spectral properties are vegetation indices such as the normalized difference vegetation index (NDVI) or the enhanced vegetation index (EVI) (Linderholm 2006, Sallaba 2011, Olsson 2014). This means that can satellite data can be used to construct long cohesive time series in which phenological events could be detected.

The changes of the spectral properties through time measured via vegetation indices can be used as a source of information not only as indications of large scales changes of regional or global environment. By recognizing unique phenological vegetation index signature for different vegetation communities, time series of satellite derived vegetation indices can be used as a source of information when mapping land cover or vegetation communities (Wang and Tenhunen 2004, Linderholm 2006, Tottrup et al. 2007, Sallaba 2011).

This study will investigate if the phenological NDVI or EVI signatures can be used to map tree species communities in the forested areas of Sweden. Sweden stretches approximately from a latitude of 55° in the southernmost part to latitude 69° in the northernmost part of the country. The climate changes with shifting latitude, and therefore also the growing conditions. The growing conditions in the southernmost part will favour broadleaf deciduous tree species communities while gradually shifting towards harsher conditions favouring evergreen needleleaf tree species communities. The fundament of this study is the underlying presumption that phenological spectral signature of two specific forest stands will deviate from each other depending on the composition between the fraction of deciduous and coniferous trees.

The phenological NDVI or EVI signature will deviate since the coniferous vegetation retains the green foliage throughout the cold season while the deciduous vegetation sheds its leaves in autumn. These differences in phenological strategies are assumed to leave quantifiable differences when measuring the NDVI or EVI signature and may be directly related to the species composition of the tree species communities. Remotely sensed satellite imagery measuring NDVI and EVI at a high temporal resolution may therefore potentially be used to quantify the fraction of e.g. deciduous/coniferous vegetation present. When the fraction of

deciduous or coniferous species composition is estimated it could then also be used as a guidance to classify the forest as either a uniform deciduous/coniferous or a mixed forest.

There are no (to the knowledge of the author) studies that attempt to model the fraction or percentage of tree species mixture using high temporal resolution satellite data.

1.1. Objectives

The overall objective is to investigate if there is a relationship between modelled phenological parameters estimated from vegetation indices extracted from multi-temporal satellite imagery, and the fraction of deciduous and coniferous trees in forest stands. It will also be investigated if the modelled phenological parameters are related with the fraction of Norway spruce (*Picea abies*) and pine (*Pinus sp.*). Statistical models will be produced and tested to investigate the possibility to predict and map e.g. the fraction of deciduous and coniferous trees on the basis of the results from the correlation analyzes.

2. Theory

2.1. Remote sensing

The sciences of observing and gathering information about physical, biological, geometrical or chemical processes or properties of the planet Earth is part of a branch within science called Earth Observation. Data obtained this way are referred to as geodata, which can be used for mapping, monitoring, modelling biophysical processes, reconstructing past historical environment or future scenarios. When observing processes of the Earth, much could be gained by taking a step back and observing things from afar away to get a wider perspective. An observation made from far without physical contact and with different instrument-based technologies is referred to as remote sensing.

The technologies used to acquire geodata can be sensors that record reflected or emitted radiation from platforms such as aircrafts or satellite systems (Tolpekin and Stein 2012). The nature of the radiation can be for example electromagnetic (EM) radiation. The recording of EM energy can be subdivided into two main categories, active or passive remote sensing. Active sensors are designed with a devise that actively emits EM radiation, the sensors then measure the backscatter of the emitted energy. The most prominent examples of active sensor within remote sensing are laser or radar systems. In everyday life, a camera with a flashlight is another example of an active sensor. If the source of radiation measured by the sensor comes from the object itself (e.g. the thermal energy of a wildfire) or another natural source (e.g. the sun), the sensor is referred to as passive. If using a camera mounted on e.g. an aircraft (*without* a flashlight), using only the EM energy emitted from the sun to illuminate an object, you have in fact a passive remote sensing system. The remotely sensed data will often be recorded as images in variable scales, reflecting the intensity of the measured radiation. What part of the EM spectrum the radiation is measured in and the scale of the measurements, is dependent on sensor and platform design. Which remote sensing system to use is dependent on the intended use, and the main three parameters to consider are spatial, temporal and spectral resolution. The user therefore have to know at what spatial extent the studied process can be observed, when and how often does observation have to be made to observe the studied phenomenon, and what spectral properties the object at hand has when choosing remote sensing system. In reality, one often has to compromise between these three criteria (Bakx et al. 2012b, Bakker et al. 2012).

2.1.1. Satellite derived spectral measurements of vegetation

Different parts of the EM spectrum will interact differently with a single leaf. How EM radiation emitted from the sun will interact with a single leaf depends on e.g. leaf pigmentation, thickness, cell structure and amount of water content. In healthy green vegetation, chlorophyll is the main chemical component that intervenes in the visual part of the EM spectrum. Chlorophyll absorbs more energy in the blue (approximately between 0.4 – 0.5 μm) and the red band (approximately between 0.6 – 0.7 μm) compared with the green band in between the blue and the red part within the visual part of the spectrum (0.38- 0.7 μm), leading the human eyes and brain to interpreted vegetation cover as green. The radiation

absorbed by the leaf in the visual part can be 70-90 % of the total incident radiation and is used as a source of energy to photosynthesise atmospheric carbon dioxide and water into carbohydrates and oxygen. A leaf reflects >40 % of the incident near infrared (NIR) energy. The dips in the reflectance curve of Figure 1 at roughly 1.45 and 1.95 µm (short wave radiation, SWIR) is mostly due to water absorption. The reflectance of the SWIR band can for example be used for example to measure how water stress affects different vegetation during e.g. droughts. Measuring the reflectance of vegetation can therefore revel much information about the properties of the pant community (Jensen 2007, Bakx et al. 2012b).

However, when formulating an idea of how the incident radiation emitted from the sun will interact with vegetation, we can not only look at the properties of a single leaf. A single leaf does not only allow the incident radiation to be either absorbed or reflected. Some of the radiance is also transmitted through the leaf. The vegetation cover seldom contains only one layer of leaves but of several layers of leaves, forming the spectral properties of canopy cover.

The internal scattering of a canopy normally allows 40-60 % of the NIR energy to be reflected while transmitting the rest of the incident NIR radiance in the first leaf layer. The visual radiance will to a large extent be absorbed in the first layer, allowing some to transmit and only small proportions (approx. 6 % of the blue, 11 % of the green and 5 of the red radiance) to reflect back to the atmosphere. The transmitted part of the visual and NIR radiance will now interact with the second leaf layer which will interact with the incident radiance in roughly the same manner. This means that the denser the canopy cover, the more energy in the NIR part of the spectrum is allowed to be reflected while the more energy in the visual part of the spectrum is absorbed. This is called *additive reflectance* and its effect is often quantified by the leaf area index (LAI). LAI is often defined as one-half of the green leaf area per unit ground surface area (m^2 / m^2). Hence, the nature of the reflectance property of a healthy vegetation canopy between the visual (especially the red part) and the NIR part of the spectrum is inversed (Jensen 2007, Bakx et al. 2012b, Alkema et al. 2012).

2.1.2. Vegetation indices

Many spectral algorithms have been developed over time, which utilize the empirical knowledge of the inverse relationship between the red and the NIR part of the EM spectrum associated with healthy vegetation. These algorithms are referred to as spectral vegetation indices and are often used to find empirical relationships between remotely sensed spectral measurements via satellite or airborne systems and biophysical parameters such as LAI, fraction of photosynthetically active radiation, biomass volume, crop yield, etc.

A vegetation index should be coupled to and maximise its sensitivity to vegetation biophysical parameters (preferably linear relationships), normalize external effects on the spectral signature (such as sun angle, sensor view angle or atmospheric interaction), and normalize internal effects on the spectral signature (such as canopy background variation, topographic variation and soil variation). Different vegetation indices will perform differently depending on e.g. what biophysical phenomena is to be monitored, atmospheric conditions, vegetation density and soil characteristics (Jensen 2007). However, the most widely known

and used example is the normalized vegetation index (NDVI), which is computed by Equation 1:

$$NDVI = \frac{(\rho_{NIR} - \rho_{Red})}{(\rho_{NIR} + \rho_{Red})} \qquad \textit{Equation 1}$$

where ρ denotes the surface directional reflectance of any part of the EM spectrum. As can be seen by Equation 1, NDVI utilizes the inverse relationship between the red and the NIR band when monitoring healthy green vegetation. NDVI will generate a value between -1 and +1, where high values indicate large amounts of green vegetation (Jensen 2007). Figure 1 illustrates how NDVI performs in healthy vegetation compared with unhealthy or autumn coloured vegetation where photosynthetic activity has decreased (NASA 2015). Some advantages of the NDVI are that this index do handle the influence of spectral noise associated with clouds/cloud shadows, topography, sun and view angle. NDVI shows in principle a high correlation to green biomass measures such as LAI. However, it is shown that NDVI becomes less responsive when LAI is high and is therefore less suitable when monitoring high biomass ecosystems such as forests. In addition, NDVI is also very sensitive to background variation when e.g. bare (especially red) soil or (especially red coloured) litter is present. In response to these restrictions, other vegetation indices have been developed such as the enhanced vegetation index (EVI) which is given by Equation 2 (Solano et al. 2010):

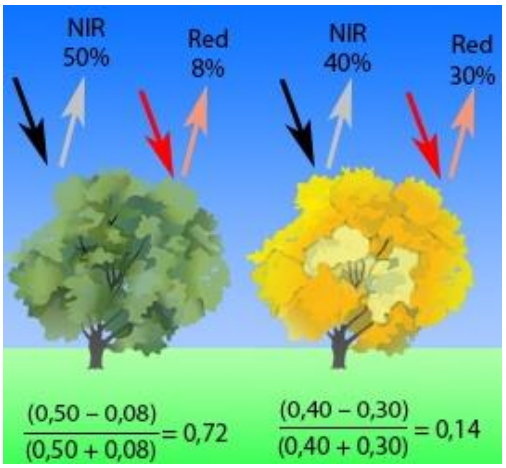

Figure 1: Illustrates how NDVI is used in healthy green vegetation contra unhealthy or autumn coloured vegetation (NASA 2015).

$$EVI = G * \frac{(\rho_{NIR} - \rho_{Red})}{(\rho_{NIR} + C_1\rho_{Red} + C_2\rho_{Blue} + L)} \qquad \textit{Equation 2}$$

where L is a canopy background adjustment term, C_1 and C_2 weigh the use of the blue channel in aerosol correction of the red channel. G is the gain or the scaling factor. EVI uses differences in the blue and the red band in a combination as an estimator of the atmospheric influence. EVI also has the advantage over NDVI regarding soil background sensitivity due to the fact that the red band becomes less dynamic when comparing EVI with NDVI. In addition, EVI has proven to be more responsive when the volume of healthy green vegetation is high. (Jensen 2007, Solano et al. 2010).

2.2. Forest phenology and remote sensing

Phenology has been defined as: "...study of timing of recurring biological events, the causes of their timing with regard to biotic and abiotic forces, and the interrelation among phases of the same or different species."(Lieth 1974). The phenological stages of plants are a function of adaption to the seasonal changes of their environment. By providing long cohesive time series of phenological observations, studying phenology can offer different kind of information, e.g. indicate environmental change or distinguish a land cover from another. For example, changes in plant phenology have been pointed out as one of the strongest responsive and observable biological phenomena for climate change (Linderholm 2006).

The spatial scale of the observations made may vary from individual plats to ecosystems. On an ecosystem level, remotely sensed data are often used, utilizing the spectral properties of the photosynthetically active green biomass. Remote sensing data has the advantage of large spatial coverage (Olsson 2014).

However, the reflectance signature through time is dependent on the land cover type. Evergreen-coniferous and deciduous-broadleaf forests has very different phenological strategies to manage the extreme seasonal temperature changes in the cold regions of mid and high latitude areas. Although sharing the same conditions during the cold season when dormancy prevails and therefore prevents growth, the seasonal cycle of LAI and photosynthesis differ when comparing evergreen-coniferous and deciduous-broadleaf vegetation. The foliage of coniferous vegetation is adapted to withstand extremely cold temperatures and minimize frost damage. Coniferous vegetation rapidly regulates down photosynthesis activities to almost zero when temperature and day length drops in autumn, fully retaining its green foliage throughout the cold season (Barr et al. 2009). When deciduous-broadleaf vegetation prepares for entering dormancy, the production of the spectrally dominating chlorophyll pigments stop and the chlorophyll slowly disappears. In this process, the photosynthesis ceases and other chemicals (e.g. anthocyanin, tannin, carotene and xanthophylls) start influencing the reflectance of the leaves. For the human eye, this makes the leaves appear as yellow or red instead of green. The leaves eventually fall off, leaving the tree bare of leaves (Jensen 2007). When peak of the growing season occurs, the difference between coniferous and deciduous forest are small. However, the differences between coniferous and deciduous forest is larger earlier on and later on in the season due to differences in the phenological processes (Atzberger et al. 2013). This effect is further illustrated by a remark made by Dave Simonett in the years of early remote sensing science: "Green is green is green" (Jensen 2007, p. 373). What Simonett meant was that the spectral signature of e.g. a tree canopy for different tree species communities will be roughly the same, which makes it hard to distinguish one species from another. Simonett's remark is today perhaps further from the truth since hyperspectral sensors has developed in a direction that allows measurements in quite narrow proportions of the EM spectrum (Jensen 2007). Even so, individual species communities of specific ecosystems, e.g. evergreen-coniferous or deciduous-broadleaf forests, can be hard to spectrally distinguish if using one satellite scene form e.g. the peak of the growing season. Therefore, it makes sense to use the information

that seasonal differences in the spectral signature may have when e.g. mapping species communities or land covers.

The idea of using metrics derived from remotely sensed phenological differences of vegetation indices for mapping is not new. Seasonality metrics such as the NDVI amplitude derived from AVHRR data was used by Wang and Tenhunen (2004) to map land use in northern China. Sallaba (2011) used MODIS derived NDVI-time series in combination with machine learning theory via a support vector machine classification for land use and cover classification of the south part of Sweden. MODIS derived NDVI-time series have also been used to map the fractional land use of non-forest, mature forest and secondary forest in southeast Asia (Tottrup et al. 2007).

2.3. Tree species communities: uniform coniferous, deciduous or mixed forests? A matter of definition

According to the *Encyclopaedia Britannica* (2015), a forest is to be regarded as "complex ecological system in which trees are the dominant life-form". When discussing tree species communities in general terms they are mainly distinguished on the basis of species composition but also on factors like tree cover density, type of soil, climatic and geological history of the region (Forest; Encyclopaedia Britannica 2015). When distinguishing, or categorizing, forest on the basis of species composition there are a number of ways in which it could be done. Bravo-Oviedo et al. (2014) reviewed a number of definitions used to define uniform and mixed forests by the National Forest Inventories (NFIs) throughout the world. Bravo-Oviedo et al. (2014) showed that the criteria used to distinguish uniform/mixed forest can be divided in to three approaches:

- No definition, only list the observed tree species and recorded measures.
- On basis of the species percentage of canopy cover.
- Definitions based on other criteria other then canopy cover, such as basal area, number of stems per hectare or volume (total volume, stem volume or commercial volume).

The problem of using different measures when defining different tree species communities of forest is that the results of the monitoring schemes are not comparable. In Austria, for example, forest is defined as mixed if the canopy cover within a plot area of 300 m^2 is shared by a minimum of two species to 20-80%. In the Swedish NFI, stems per hectare and basal area are being used when deciding if a forest is uniform or mixed (Bravo-Oviedo et al. 2014). Basal area is defined as the area outline of a tree measured from diameter at breast height (DBH = 1,3 m) (Fridman and Nilsson 2013). The used threshold for defining if a forest is uniform or mixed differs also from country to country and is sometimes even not established. This means that a forest stand regarded as uniform deciduous forest, when counting the number of stems per hectare, can also be regarded as an uniform coniferous forest if using basal area instead (Bravo-Oviedo et al. 2014). Since this study is performed in a Swedish context and also using data from the Swedish NFI, the criteria used for the correlation analysis will be both basal area and stem count per hectare.

3. Study area

Sweden is indeed a forested country. Of its 449 964 km^2, 51 % of the area consist of forest. If taking a closer look at the dominating biomes that can be seen in Figure 2, the most southward zone is called the *nemoral zone* and the natural dominating kind of forest are noble deciduous forests, although large areas are occupied by agricultural land. However, coniferous forest is today quite frequently present within the nemoral zone because of the promotion of planting coniferous trees by the forestry sector. The northernmost border of the *boreonemoral zone* coincides largely with the northernmost records of oak (*Quercus robur)* and consists of a zone where coniferous forest dominates but several noble deciduous species have the ability to grow, thrive and are frequently found. Mixed forest with Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), birch (*Betula sp.*), aspen (*Populus tremula*) are frequently found, but the coniferous species dominate due to the forestry sector's earlier mentioned favouring of these species. The *boreal zone* can be subdivided into southern, middle and northern zone and is part of the Taiga. Noble deciduous species are very rare within these zones and only a few specimens can be found in the southern boreal zone. The middle boreal zone is naturally and completely dominated by coniferous forest but the deciduous species birch and aspen are frequently found. In the most northern part of the boreal zone is the coniferous forest less dense, and at higher altitudes becomes mountain birch more frequent (Gustavsson 1996, Bergil et al. 2004).

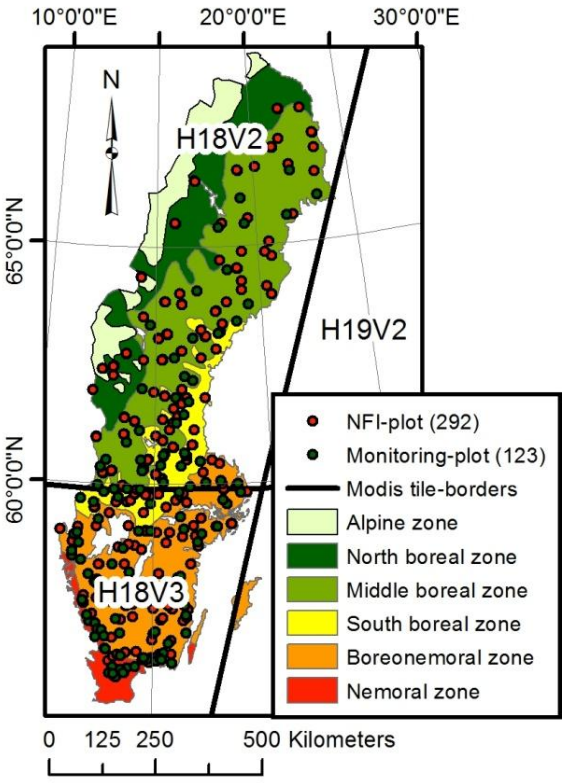

Figure 2: Map of the biomes of Sweden, spatial extent of MODIS-tiles and the field data. Source: Modification from Gustavsson (1996 p. 27)

The forest industry is important for the national economy, and as mentioned, the production practice of the forestry affects the composition of forests to a large extent. This means that the composition of an individual forest stand is often quite uniform when considering both the age and species composition, which implies that mixed forest is not common (Bergil et al. 2004).

4. Data

4.1. Field data

The field data used within this study are the relative compositions of deciduous/coniferous trees and Norway spruce and pine of the investigated forest stands regarding the basal area (m^2/ha) and number of stems (number of stems/ha).

National forest inventory plot data: Sweden has a long tradition of forest inventory. The main program started as early as 1923 by the Swedish Forest Agency (SFA), and although it changed from a program designed to only monitor timber stocks from year to year, it has developed to include other parameters that reflects several of the ecosystem services that the forests are generating. The specific sub-dataset used in this study is freely distributed via SFA:s homepage (http://www.slu.se/), visited 2015-02-17) and is supplied for the very purpose of validating remotely sensed data. This dataset is compiled from the temporary NFI plots which were collected between 2007 and 2011, but only data from 2007 will be used for this study. The data was collected in a circle with a radius of 7 m in which data from every individual tree is recorded including, i.a. species (number of stems/ha) and basal area (m^2/ha). (Fridman and Nilsson 2013, Fridman et al. 2014)

Monitoring plot data: This dataset is from an environmental monitoring program launched in 1995 and ending in 2013, managed by the SFA (Akselsson et al. 2015). The program was a development from an earlier monitoring program launched in 1984. The objectives of both the former and later program was to monitor forest degradation and dieback because of air pollution and atmospheric deposition that was reported throughout Europe in 1970_s and the early 1980_s and are therefore not developed for the prime use of training and validation within the area of remote sensing. The plot size is 30 x 30 meters in which every individual tree is recorded according to, i.a. species and DBH. When this program was developed, there was a focus on that the plots should be homogeneous and not be affected by its surroundings. This means that all plots were located at least 100 m from the nearest edge of the forest stand, at least 300 m from arable land and 4 km from a coastline, an industry or built-up areas (Akselsson et al. 2015). These characteristics are for the purpose of this study to be considered as good.

However, it was also decided that only areas where the stem count was assigned to > 70 % of the most dominant tree species was to be accepted within the program. This means that the species composition of the monitoring plot data is quite heterogeneous and few or no sample sites can be said to represent a mixed forest. This is the main reason why plots from the NFI-program were included in the study, since the NFI-data contains sample sites representing also mixed forests. Almost all measurements in this dataset used in this study were conducted during 2004-2005.

Comparison: When comparing the two datasets described above, it should be pointed out that the area of sample size differs largely. It should also be noted that none of the described field datasets was originally collected to be used for training and validation of remotely sensed data. But the monitoring plots have qualities that make them suitable to be used for moderate resolution satellite products like the MODIS-products (Jönsson et al. 2010). Hence, the

monitoring plots are a more suitable dataset compared to the NFI-plots when considering using them for training and validation of remotely sensed data. However, the data of the monitoring plots does not include any mixed forests. The NFI-plots are therefore also included.

4.2. MODIS satellite data

MODIS-satellite data (spatial resolution 250 x 250 m) was selected since it has both the temporal and spectral resolution required for this study. The specific products used was the MODIS 16 day vegetation indices MOD13Q1 (MODIS Terra) and MYD13Q1 (MODIS Aqua). When combining these two datasets the temporal resolution of the products becomes 8 days. Single point-sites (i.e. individual pixels) will be used for training and validation of the regression analysis. The quality data of pixel reliability will be used to produce weights for the seasonality modelling in the computer programme TIMESAT (see section 6.3. Phenology parameter extraction – TIMESAT). NDVI and EVI are the included vegetation indices "ready to use", but al spectral bands present in the MODIS system are also included. Note that MODIS-satellites lack a blue band at 250 meter resolution and uses therefore the blue band with 500 meter resolution when calculating EVI (Solano et al. 2010). The time series of MODIS-satellite data used in this study is collected from the first day of 2003 to the last day of 2008.

4.3. Google Earth™

A challenge for this study is to only include ground truth samples from the NFI-plots and the monitoring-plots that can be said to represent an area of 6.25 ha (i.e. the spatial resolution of the satellite data, see section below). This is a challenge because the sample has been taken for a much smaller area (NFI-plots ≈ 0.0154 ha; monitoring plots = 0.09 ha), every area was inspected in Google Earth™. The idea was to weed out sample data where the spectral signature of a MODIS pixel may be affected by other land covers within the individual sample vicinity. Land cover strongly affects the spectral signature and its relation to vegetation phenology retrieved from satellite imagery (Sallaba 2011, Bakx et al. 2012b).

The available imagery most often consists of aerial photography that Google Earth™ has bought from Lantmäteriet (National Land Survey of Sweden) or from other high-resolution imagery satellite systems e.g. WorldView-2, WorldView-3, GeoEye-1, etc. However, an individual image over a site can be from a later date compared to when the ground sample was taken. This temporal mismatch means that some samples can be excluded from the analysis since e.g. a clear-cut has been made in the time between when the sample was taken and photo was taken. This means that plots that were regarded as homogeneous throughout an area of 6.25 ha when the sample was taken, could be excluded from the analysis due to the fact that there is no way to confirm whether it was homogeneous or not since older aerial imagery is not available to confirm this. This means that a precautionary principle is applied.

5. Method

5.1. Field data

The field data acquired from the NFI plots was already in the units that are used in this study, which is basal area (m^2/ha) and stem count (stems/ha) for Scots pine, lodgepole pine (*Pinus contorta*), Norway spruce, birch and "other deciduous species". The categories of birch and "other deciduous species" were added together to form a new category, deciduous species. Scots pine and lodgepole pine were also added together to form a new category, pine. The categories of Norway spruce and pine were added together in order to calculate the basal area and stem count for the coniferous species. To calculate the relative composition in terms of basal area of e.g. deciduous species (i.e. the species fraction of sample), the basal area of the deciduous species was divided by the total basal area. The relative composition in terms of stem count was calculated in the same way, i.e. the stem count of the individual category was divided by the total stem count of the sample.

The field data acquired from the monitoring-plot data had records of DBH for individual trees of several species, including the species included in the NFI plot dataset. The individual species were added together in the same manner as for the NFI plots so that the same categories were present: Norway spruce, pine, coniferous and deciduous species. The number of individual trees was counted for every sample to determine stem count; and from DBH the basal area was calculated for each plot. Note that the DBH had been measured in both north-south and west-east direction, the mean of the two measures was used for calculating the basal area. The relative composition of the samples for all categories when regarding both basal area and stem count was then calculated as described for the NFI plots. The units of both the datasets are now comparable. See Table 1 for a summary of the extracted field data.

The dataset of the NFI plots contains > 11 000 points collected during year 2007, and the monitoring plot dataset contains 218 points collected during year 2004 and later. All these samples cannot be said to represent an area of 250 x 250 meters, i.e. an area corresponding to a MODIS satellite pixel. To select samples that represent an area of 250 meters, a stepwise exclusion of heterogeneous samples were made. The first step was to exclude every sample that did not contain any trees and samples that was classified as non productive forest (where forest production is < 1 m^3/ha/year, NFI plots).

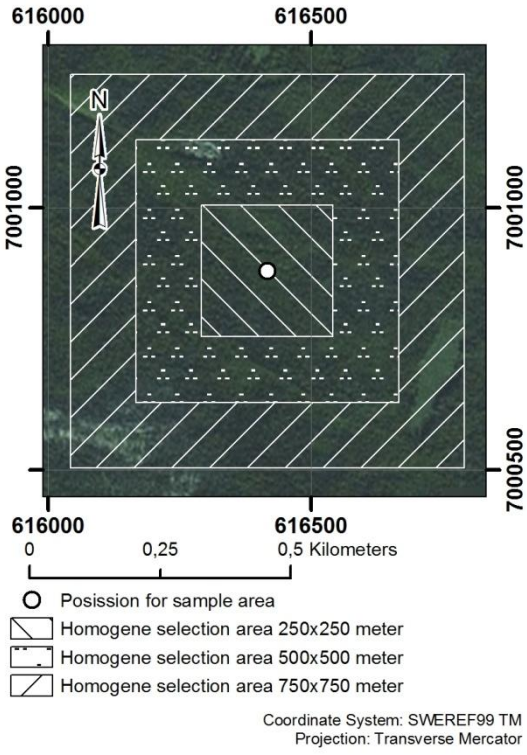

Figure 3: The three spatial selection criteria.

Table 1: Summary of the variables extracted from the field data.

Field data	Measure	Unit	Range
Relative composition of deciduous species	Basal area	Fraction	0 - 1
Relative composition of coniferous species			
Relative composition of Norway spruce (*Picea abies*)			
Relative composition of pine (*Pinus sp.*)			
Relative composition of deciduous species	Stem count	Fraction	0 - 1
Relative composition of coniferous species			
Relative composition of Norway spruce (*Picea abies*)			
Relative composition of pine (*Pinus sp.*)			

The dataset Corine Land Cover from 2006 (resolution 250 x 250 meters) was used where the classes Broad-leaved forest, Coniferous forest and Mixed forest (EEA 2006) were extracted to be used as a mask. The last step was to visually interpret the remaining areas in Google Earth™. Three spatial criteria were developed. The first spatial criterion was that the surrounding area of 250 x 250 meters had to be homogeneous (see Figure 3 for example). A total of 415 points counting from both dataset showed to fulfil this criterion. However, the literature suggests that the area of a sample size should not be smaller than 3 x 3 pixels due to several sources of errors such as low accuracy of GPS-positioning, georeferencing errors of the satellite image etc. (McCoy 2005, Congalton and Green 2009, Olofsson et al. 2014). Considering this, the sample size should correspond to a minimum of 750 x 750 meters when using MODIS satellite data. Therefore, sample points are also selected on the spatial basis of 500 x 500 meters and 750 x 750 meters to see if the results improve when the surrounding area that after visual interpretation is representing an area corresponding to the field measurements of the sample (see Figure 3).

The 250 x 250 m, 500 x 500 m and 750 x 750 meters squares around each sample form the area for which the visual interpretations was made, where created in ArcMap™ 10.1. The areas were created by first making a circular buffer zone with a diameter that corresponds to half the base of the intended area. The buffer is then enveloped with the tool "Feature Envelope To Polygon" to produce a quadrate.

5.2. Capturing and processing of satellite data

The M*D13Q1 data was downloaded from The Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) via the service "MODIS Global Subsets: Data Subsetting and Visualization" (ORNL 2014). Every site-coordinate for the samples was first converted to WGS84 datum longitude and latitude decimal degree which was used to determine the MODIS pixel to be used for every site. The time period was set to 2003-2008 for every site. This resulted in one file representing one site (one MODIS pixel) for one day for every product. The data from all the sample sites of one day were mosaicked into one image, representing the vegetation index value for every site in one day. To prepare the data for processing in the software TIMESAT, all NDVI, EVI and pixel reliability values was extracted to a point layer which attributes were converted to a ASCII-file (Eklundh and Jönsson 2015).

5.3. Phenology parameter extraction – TIMESAT

To extract seasonality parameters from time series of satellite data, the software TIMESAT 3.2 was used. This section describes which settings that were used when the seasonality was modelled and which seasonality parameters were extracted for the analysis. For a more detailed description of the mathematics for the fitting functions used, this study will refer to the software specific literature and documentation (Jönsson and Eklundh 2002, 2004, Eklundh and Jönsson 2015).

TIMESAT has a number of settings for finding the best model fit and the authors of the manual suggest that the user takes an experimental approach when finding the appropriate settings (Eklundh and Jönsson 2015). However, since the program has been available for the research community for quite some time, there is of course some guidance to be found in the literature. In this study, the seasonality extracted from the NDVI time series was tested for two modelling methods, the double logistic and Savitzky–Golay fitting methods. For the EVI time series, only double logistic was tested since the Savitzky–Golay fitting method could not simulate realistic phenology metrics due to the high VI-values that EVI has snow is present. The program settings were otherwise the same for the two time series if nothing else is mentioned.

When quality data from the producer of the remote sensed data are available, the quality data can be used to produce weights for the seasonality. There are two quality datasets provided with the M*D13Q1-products, QA-bits and pixel reliability. In this study, pixel reliability is the used dataset for assigning weights, which is a summary of the QA-bits (Solano et al. 2010). TIMSAT allows the user to define three groups of weights stretching from 0 – 1 (Eklundh and Jönsson 2015). How the weights is assigned are reported in Table 2 and was developed in consultation with Olsson (2015, personal communication). The sample sites are all located in a region at high latitude, which means that remotely sensed data can be heavily affected by long periods of snow, cloud cover and/or low zenith angles in the winter season, leading to erroneous VI-values (i.e. the scene will have low illumination and the solar emitted radiation will have a long pathway through the atmosphere). These problems are to a large extent handled by the weights, but in addition a forced minimum value of 0.1 was also used (see Figure 4), (Jönsson and Eklundh 2002, Sallaba 2011, Eklundh and Jönsson 2015).

Table 2: The used weighting scheme.

Rank Key/Pixel value	Summary QA	Description	TIMESAT-weight
-1	Fill/No Data	Not Processed	0.1
2	Snow/Ice	Target covered with snow/ice	0.1
3	Cloudy	Target not visible, covered with cloud	0.1
1	Marginal data	Useful	0.8
0	Good Data	Use with confidence	1.0

Source: (Solano et al. 2010, Olsson 2015, personal communication)

The spike detection method used was the STL-decomposition combined with the weighting scheme mentioned above. This means that the full time series are divided into a seasonal and a trend component; data values that do not fit this pattern will be assign low weights, which are then multiplied by the assigned ancillary weights. The number of seasons was set to one

since only one growing season occurs in the study area. Both the number of envelope iterations and adaption strengths was set to 2.0. A higher adaption value will make the fitted model emphasise individual high values (Eklundh and Jönsson 2015).

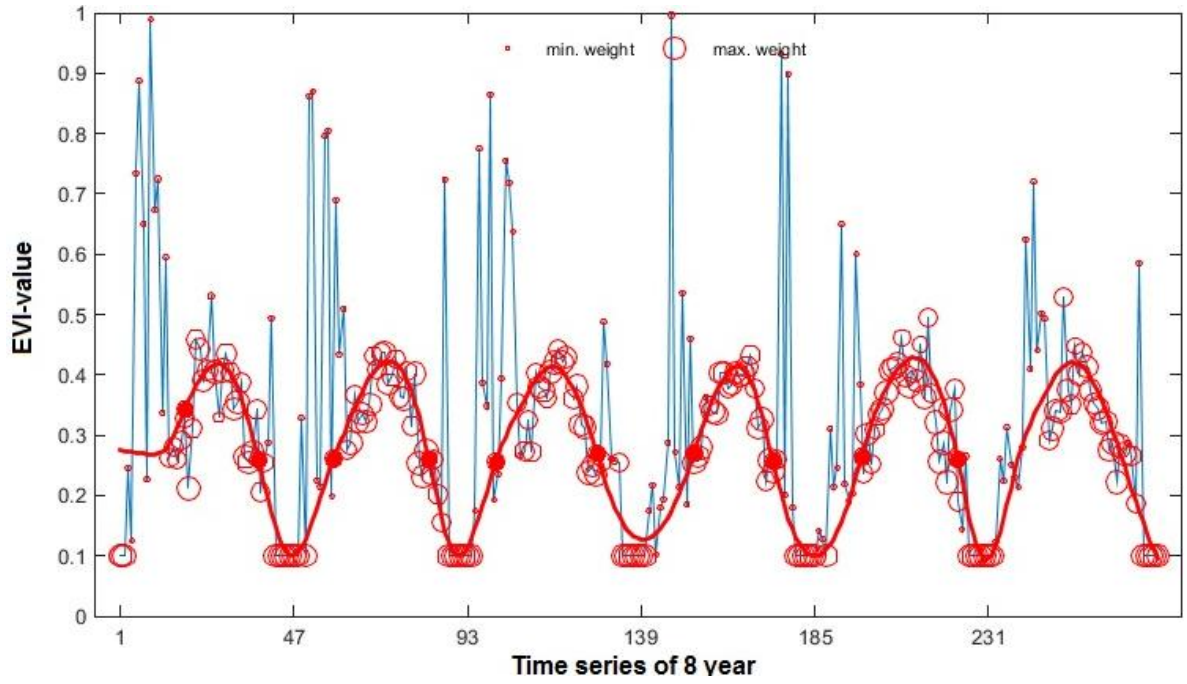

Figure 4: An example of the seasonality modelling in TIMESAT for a sample site of the EVI time series. Small circles mean that the estimated EVI value receives a low value due to poor quality when the seasonality is modelled while large circles indicate a high value. Notice the occurrence of the value 0.1 during winter season due to the force min value. Also, EVI sensitivity to snow is clearly visible by the high values during winter. The red dots singles start and end of season, here defined as 50 % of the seasonal amplitude.

Start and end of season can be defined in two different ways. The first way is to define start/end of season when a user specified proportion of the seasonal amplitude is reached, which will be referred to as the relative start/end of season. The second way is to define start/end of season when an absolute threshold value is reached. For the statistical analysis, both a relative and an absolute approach were tested. The relative start/end of season of season was set to when the seasonal amplitude reached 0.5. The absolute minimum threshold was set to 0.75 for the NDVI-time series and 0.35 for the EVI-time series. Note that when setting an absolute value, some sample sites did not reach the specified threshold and were then then be excluded from the statistical analysis.

An ASCII-file was produced from TIMESAT with the seasonal parameters. The parameters used for the statistical analysis in this study are (the letter refers to the letters in Figure 5):

a. Time for the start of the season; time for which the left edge has increased to a user defined level, which has to be multiplied by the time interval between observations in the time series (in this case 8) in order to receive the actual date. (from now referred to as S-date).
b. Time for the end of the season; time for which the right edge has decreased to a user defined level measured from the right minimum level, which has to be multiplied by the time interval between observations in the time series in order to receive the actual date (referred to as E-date).

c. Length of the season, which has to be multiplied by the time interval between observations in the time series in order to receive the actual number of days (referred to as Length.).
e. Time for the mid of the season; computed as the mean value of the times for which the left edge has increased to the 80 % level and the right edge has decreased, which has to be multiplied by the time interval between observations in the time series in order to receive the actual date (referred to as Mid-x).
f. Largest data value for the fitted function during the season, may occur at a different time compared with e (referred to as Max)*.
g. Seasonal amplitude (referred to as Amp).
h. Large seasonal integral; integral of the function describing the season from the season start to the season end. Note that the large integral has no meaning when part of the fitted function is negative (referred to as L-integ).
i. Small seasonal integral; integral of the difference between the function describing the season and the base level from season start to season end (referred to as S-integ).
j. Rate of increase at the beginning of the season; calculated as the ratio of the difference between the left 20 % and 80 % levels and the corresponding time difference (referred to as L-der).
k. Rate of decrease at the end of the season; calculated as the absolute value of the ratio of the difference between the right 20 % and 80 % levels and the corresponding time difference. The rate of decrease is thus given as a positive quantity (referred to as R-der) (Eklundh and Jönsson 2015).

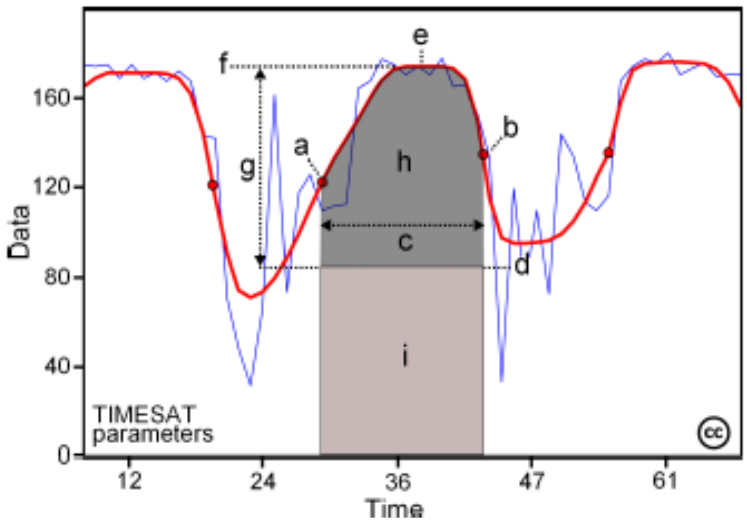

Figure 5: The solid red line indicate the modelled phenological curve by TIMESAT. The letters indicates different parameters that can be derived: a = S-date. b = E-date. c = Length. e = Mid-x. f = Max. g = Amp. h = L-integ. i = S-integ. (Eklundh and Jönsson 2015).

In addition, the VI-values for the start and end of the season were extracted. Since there was no function built into TIMESAT to automatically write out these parameters, they had to be written down by hand when visualised with the "data cursor" in the TSM_GUI. This step was very labour intensive. When including several seasons, others have taken the mean of the

* When start/end of season is set to an defined absolute value, the measures of D-MS, D-ME and Max covary and will therefore yield the same linear correlation coefficient (*r*) when performing correlation analysis

seasonal parameters to produce a more stable measure of the seasonality (e.g. Jönsson et al. 2010, Sallaba 2011). To minimise the effort of labour so that the timeframe of the work could be achieved, only one season was chosen for each sample site. The season corresponding to 2006 was primarily chosen; when seasonality could not be modelled for this particular season, the closest earlier or later season was chosen instead. Next follows a list of the additional seasonal parameters with respective equation that was extracted beside those that TIMESAT could write out:

l. VI-values for the start of the season (referred to as S-value).
m. VI-values for the end of the season (referred to as E-value)
n. The difference between Max and S-value; (referred to as D-MS) = Max – S-value[*]
o. The difference between Max and E-value; (referred to as D-ME) = Max – E-value[*]
p. The difference between Max and S-value normalised against the maximum value; (referred to as ND-MS) = (Max – S-value)*Max[†].
q. The difference between Max and E-value normalised against the maximum value; (referred to as ND-ME) = (Max – E-value)*Max[†].
r. Count of days from star of season to mid season (referred to as CD-MS) = Mid-x – S-date
s. Count of days from mid season end of season (referred to as CD-ME) = Mid-x – E-date
t. Mean increase rate for VI-values per day from start of season to mid season, i.e. the seasonal increase rate (referred to as SIR) = D-MS / CD-MS
u. Mean decrease rate for VI-values per day from mid season to end of season, i.e. the seasonal decrease rate (referred to as SDR) = D-ME/ CD-MS

5.4. Statistical analysis and validation

This study investigated how well the phenological parameters mentioned above correlated against any of the extracted variables from the field dataset (Table 1) measured with the linear correlation coefficient (R). This measure is also known as *Pearson's* correlation. To test if the estimated correlation coefficient is significant, i.e. if the real or true correlation coefficient is zero, a two tailed t-test was performed where the significant level (ρ) was indicated if $\rho \leq 0.05$ or $\rho \leq 0.01$ (Rogerson 2010). These analyzes were performed in the software SPSS. However, not all of the sample sites were included in the analysis. 20 % of the data was always reserved for validation so that the validation data are guaranteed to be statistically independent. The selection of validation data was performed via a stratified random selection where the dataset was divided in to two strata on the basis of basal area. The first stratum was defined as samples site were the fraction of the deciduous species was ≥ 0.8 or ≤ 0.2 (i.e. were the fraction of the coniferous species was ≥ 0.8 or ≤ 0.2). The rest of the data set was defined as the second stratum. A random selection was made within each stratum where 20 % of the stratum was selected via a random selection. The reason for making a stratified random

[*] When start/end of seas is set to an defined absolute value, the measures of D-MS, D-ME and Max covariates and will therefore yield the same linear correlation coefficient (r) when performing correlation analysis
[†] When start/end of seas is set to an defined absolute value, the measures of ND-MS and ND-ME covariates and will therefore yield the same linear correlation coefficient (r) when performing correlation analysis

selection rather than simple random selection is that mixed forest are proportionally rare in Sweden (as mentioned in section *4. Study area*), which is also reflected in the sample data. Therefore, to guarantee that both uniform and mixed forest stands were present in the validation dataset, a stratified random selection was preferred.

Regression models were produced for the variables showing the best coefficient of determination (i.e. the best R^2) and the tested correlation analysis was visualised in scatter plots where the independent/explanatory (x) variable is a phenological parameter and the depended/response variable (y) is the fraction of e.g. deciduous species (see Table 1). The predicted values ($\hat{y}$) was evaluated with root-mean-square-error (RMSE) and its relative counterpart ($RMSE_r$) (Muukkonen and Heiskanen 2005). Further, the best performing regression model was tested as if it was used for mapping the variables mentioned in Table 1, and therefore validated as a map with an accuracy assessment via an error matrix. There are three classes represented in the error matrix: coniferous, mixed and deciduous forests. The classes are arbitrarily defined as when the fraction of coniferous or deciduous species exceeds 0.8 (i.e. 80 %); it will be defined as coniferous or deciduous respectively if regarding either basal area or stem count. The overall, user, producer accuracy and Cohen's Kappa coefficient was calculated (Cohen 1960, Congalton and Green 2009, Bakx et al. 2012a). The major difference between the overall accuracy and Kappa is that Kappa takes into account the possibility of which the reference data and the map classification would agree by chance. This means that if the samples were assigned to different classes at random, there would still be some samples that were correctly labelled completely by chance. The Kappa coefficient will have a range from -1 – 1 where the value of 0 will symbolise "as if" labels was assign by chance. A value < 0 is worse than if the samples were assigned classes by chance, > 0 is better than the samples were assigned classes by chance and a value of 1 is produced when the map and the reference data is in complete agreement (Cohen 1960, Congalton and Green 2009, Bakx et al. 2012a). For details, see Appendix 1 or the referred literature.

6. Results

6.1. Fraction of deciduous and coniferous trees

6.1.1. NDVI-seasonality

When using a relative start/end of season: The seasonal maximum NDVI-value seems to be the one of best performing phenological parameter when estimating the correlation between the fraction of deciduous/coniferous species, when considering both seasonality modelling method and data selection made from interpretations at 250 and 500 m (see Table 3). However, the count of day between seasonal start and seasonal maximum (CD-MS) performed best when applying a spatial selection area of 750 m. The correlation seems to be somewhat better when using an absolute value for start/end of season (Table 4), but it should be noted that many sample sites have been excluded from the correlation analysis due to that TIMESAT could not model any seasonality when using an absolute value. The exclusion of samples sites can contribute to noise reduction leading to better correlation.

Table 3: Correlation coefficients (R) between deciduous and coniferous sp. for basal area and stem count when using a relative start/end of season. Table area marked in green is where the correlation is significant at the 0.01 level (2-tailed). Table area marked in blue is where the correlation is significant at the 0.05 level (2-tailed).

Phenological parameter from TIMSAT	Selection criteria	Basal area, Deciduous	Basal area, Coniferous	Stem count, Deciduous	Stem count, Coniferous	N (of total possible N)
Double logistic fitting – with relative start/end of season						
Max	250	0.408	-0.408	0.356	-0.356	332 (332)
CD-MS	250	-0.279	0.279	-0.244	0.244	332 (332)
Lenght	250	-0.25	0.25	-0.216	0.216	332 (332)
Max	500	0.465	-0.465	0.418	-0.418	57 (57)
L-der	500	0.417	-0.417	0.365	-0.365	57 (57)
SIR	500	0.335	-0.335	0.334	-0.334	57 (57)
CD-MS	750	-0.41	0.41	-0.458	0.458	25 (25)
Savitzky-Golay filtering – with relative start/end of season						
Max	250	0.444	-0.444	0.377	-0.377	332 (332)
CD-MS	250	-0.342	0.342	-0.263	0.263	332 (332)
Lenght	250	-0.278	0.278	-0.206	0.206	332 (332)
Max	500	0.6	-0.6	0.534	-0.534	57 (57)
ND-ME	500	0.379	-0.379	0.377	-0.377	57 (57)
D-ME	500	0.313	-0.313	0.318	-0.318	57 (57)
CD-MS	750	-0.58	0.58	-0.66	0.66	25 (25)
SIR	750	0.508	-0.508	0.507	-0.507	25 (25)
Length	750	-0.463	0.463	-0.531	-0.531	25 (25)

When comparing the results for the different NDVI derived seasonal parameters, one finds that the correlation is always somewhat better for basal area compared to stem count, with only one exception. The exception referred to is the Savitzky-Golay filtered data with relative start/end of season (Table 3). It is reasonable to assume that the tree species composition measure that reflects the amount of green biomass or LAI is the one that performs best.

The seasonal maximum NDVI value generally correlated somewhat better when applying the Savitzky-Golay filter, although double logistic fitting correlated better when using absolute start/end for the data selected at 250 m. Again, this can be caused by to reduction of noise related to exclusion of samples sites.

Table 4: Correlation coefficients (R) between deciduous and coniferous sp. for basal area and stem count when using an absolute start/end of season. Table area marked in green is where the correlation is significant at the 0.01 level (2-tailed).

Phenological parameter from TIMSAT	Selection criteria	Basal area, Deciduous	Basal area, Coniferous	Stem count, Deciduous	Stem count, Coniferous	N (of total possible N)
Double logistic fitting – with absolute start/end of season						
ND-MS/ND-ME	250	0.517	-0.517	0.477	-0.477	257 (332)
D-MS/D-ME/Max	250	0.498	-0.498	0.461	-0.461	257 (332)
SDR	250	0.446	-0.446	0.405	-0.405	257 (332)
SDR	500	0.66	-0.66	0.63	-0.63	40 (57)
ND-MS/ND-ME	500	0.615	-0.615	0.568	-0.568	40 (57)
D-MS/D-ME/Max	500	0.596	-0.596	0.554	-0.554	40 (57)
Savitzky-Golay filtering – with absolute start/end of season						
D-MS/D-ME/Max	250	0.4	-0.4	0.343	-0.343	282 (332)
ND-MS/ND-ME	250	0.362	-0.362	0.307	-0.307	282 (332)
ND-MS/ND-ME	500	0.599	-0.599	0.544	-0.544	47 (57)
D-MS/D-ME/Max	500	0.592	-0.592	0.543	-0.543	47 (57)
SDR	750	0.573	-0.573	0.557	-0.557	24 (25)
ND-MS/ND-ME	750	0.564	-0.564	0.534	-0.534	24 (25)
SIR	750	0.562	-0.562	0.539	-0.539	24 (25)
D-MS/D-ME/Max	750	0.561	-0.561	0.538	-0.538	24 (25)

In summary, there is no clear indication of a single seasonal parameter performing overall better compared to another. Even though all result showed in Table 3 and 4 are statistically significant, none of the seasonal parameters showed a particularly strong correlation against the fraction of deciduous/coniferous species. However, when comparing the different seasonal parameters, there always exists a positive correlation between max, L-der, SIR, SDR, D-MS/D-ME, ND-MS/ ND-ME and deciduous tree species composition, and vice versa for coniferous tree species composition. The seasonal parameters showing a negative correlation for deciduous tree species composition are Length of season and CD-MS. That is to say that the more coniferous species a forest stand has, the longer the season will be. This means that

the seasonal variability of NDVI-values is lower the more coniferous species are present in the forest stand.

To get a wider grasp of the results, the mean of the NDVI-values from every sample-site per day for all seven seasons included in the study after being classified into deciduous, mixed or coniferous forest were calculated (for example the mean of sample site 1, 2, 3, 4, etc. for the first day, continuing with the same calculation for the 8_{th} day). The mean of all the seasons were then calculated to generate the graph in Figure 6 (for example the mean value of seasonal mean for 2003, 2004, 2005, etc. the first day of the season, repeating the calculation for the 8_{th} day etc. which results in the graph in Figure 6). In other words, figure 6 visualizes the mean annual NDVI-values for the typical/mean deciduous, mixed or coniferous forest stand included in the study. Anyway, the differences seen in Figure 6 are very small. When visualized as a time series the data intuitively show why the seasonal parameters generated from TIMESAT have a weak correlation to the relative tree species composition. Note that the curves of the time series in Figure 6 are dependent on how the sites were classified before calculating the time series curve.

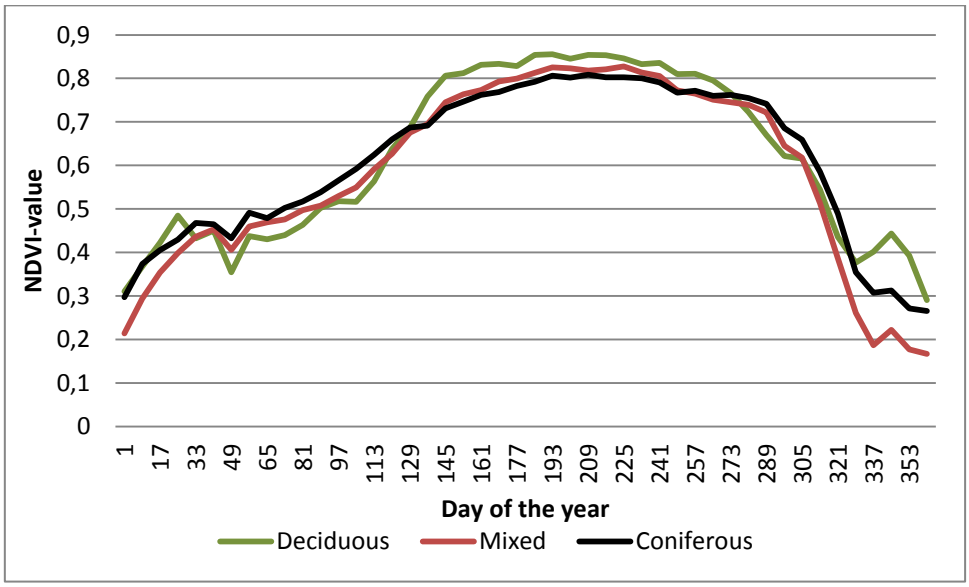

Figure 6: The mean annual NDVI-values for the typical/mean deciduous, mixed or coniferous forest stand included in the study for every time step between 2003 and 2008, N = 415.

6.1.2. EVI –seasonality

When using a relative start/end of season: Seasonality parameters modelled from EVI with double logistic fitting shows generally a strong correlation with the fraction of deciduous/coniferous species. The seasonal parameter of ND-ME and ND-MS for the data selected at 500 m seems to correlate very strongly with the fraction of deciduous/coniferous species, with a correlation coefficient of 0.884 and 0.869 respectively. The seasonal maximum EVI value generally shows a rather strong correlation. The high correlation coefficient received from the EVI extracted seasonality suggests that EVI is a highly dynamic vegetation index when quantifying biophysical parameters in forests.

When using an absolute start/end of season: There are few sample sites excluded from the data when a 250 m criterion was applied when absolute values for start/end of season are used for the seasonality modelling. No samples sites were excluded when applying the 500 m or 750 m criteria. The low number of excluded sample sites gives further evidence that EVI is a highly dynamic vegetation index when modelling seasonality parameters in TIMESAT.

In summary, no specific seasonal parameters derived from time series of EVI-values showed a stronger correlation against the fraction of deciduous/coniferous species. However, the seasonal maximum, ND-ME or MC-MS are almost always present among the top performing parameters. But all results from the correlation analysis presented in Table 5 and 6 are statistically significant when a significant level is set to 0.01(two tailed-test). All presented seasonal parameters are also positively correlated against the fraction of deciduous tree species. There are no negative correlations against the fraction of deciduous tree species, due to the fact that none of the parameters are associated with the length of season. The seasonal parameters are instead always associated with modelled values during the season, and their interrelation. This means that the more the seasonal variability will increase the more deciduous species are present within the forest stand.

The results shows that the seasonal parameters in Table 5 always correlate somewhat better against the fraction of deciduous/coniferous species measured in basal area compared with stem count, without any exceptions.

Table 5: Correlation coefficients (R) between deciduous and coniferous sp. for basal area and stem count when using a relative start/end of season. Table area marked in green is where the correlation is significant at the 0.01 level (2-tailed). Table area marked in blue is where the correlation is significant at the 0.05 level (2-tailed).

Phenological parameter from TIMSAT	Selection criteria	Basal area, Deciduous	Basal area, Coniferous	Stem count, Deciduous	Stem count, Coniferous	N (of total possible N)
Double logistic fitting – with relative start/end of season						
Max	250	0.608	-0.608	0.543	-0.543	332 (332)
ND-ME	250	0.544	-0.544	0.476	-0.476	332 (332)
ND-MS	250	0.51	-0.51	0.455	-0.455	332 (332)
Amp	250	0.454	-0.454	0.373	-0.373	332 (332)
ND-ME	500	0.884	-0.884	0.821	-0.821	57 (57)
ND-MS	500	0.869	-0.869	0.807	-0.807	57 (57)
Max	500	0.834	-0.834	0.762	-0.762	57 (57)
Amp	500	0.798	-0.798	0.76	-0.76	57 (57)
Max	750	0.759	-0.759	0.708	-0.708	25 (25)
ND-ME	750	0.725	-0.725	0.67	-0.67	25 (25)
SIR	750	0.72	-0.72	0.686	-0.686	25 (25)
D-MS	750	0.699	-0.699	0.582	-0.582	25 (25)
Double logistic fitting – with absolute start/end of season						
ND-MS/ND-ME	250	0.637	-0.637	0.554	-0.554	322 (332)
D-MS/D-ME/Max	250	0.611	-0.611	0.538	-0.538	322 (332)
L-der	250	0.447	-0.447	0.37	-0.37	322 (332)
SIR	250	0.417	- 0.417	0.374	-0.374	322 (332)
S-integ	500	0.809	-0.809	0.73	-0.73	57 (57)
Amp	500	0.757	-0.757	0.713	-0.713	57 (57)
ND-MS/ND-ME	500	0.737	-0.737	0.647	-0.647	57 (57)
SIR	500	0.729	0.729	0.65	-0.65	57 (57)
SIR	750	0.741	-0.741	0.696	-0.696	25 (25)
SDR	750	0.713	-0.713	0.66	-0.66	25 (25)
D-MS/D-ME/Max	750	0.593	-0.593	0.518	-0.518	25 (25)
L-der	750	0.579	-0.579	0.553	-0.553	25 (25)

There are noticeable differences between the different tree species communities when looking at the time series for the mean annual EVI-values for the typical/mean deciduous, mixed or coniferous forest stand included in the study in Figure 7 (calculated as Figure 6 described in 7.1.1. NDVI-seasonality). The coniferous species community has generally very low summer values while the mixed species communities have just a little higher mean summer values. The mean deciduous species community has much higher summer values compared with its coniferous and mixed counterparts. Also the period of green up is more much more intense for the deciduous forests compared to the mixed and coniferous forests stands. However, the winter values are much higher for the mixed and the coniferous communities while the deciduous species communities are rather low during winter. The relatively low winter values of the deciduous species communities may be a result of climatic differences due to

geography. Climate may be a factor when looking at the mean values since the majority of the deciduous species communities are located in the southernmost Sweden, while the location of the coniferous species communities are more evenly spread throughout the country. Hence, winter values of the mean coniferous species community may to a large extent be affected by longer periods of snow since the northernmost part of the country generally has longer periods with snow. However, it should be noted that the winter values are of little interest when discussing the results of the correlation analysis since the winter values have been filtered in TIMESAT via the assigned ancillary weights.

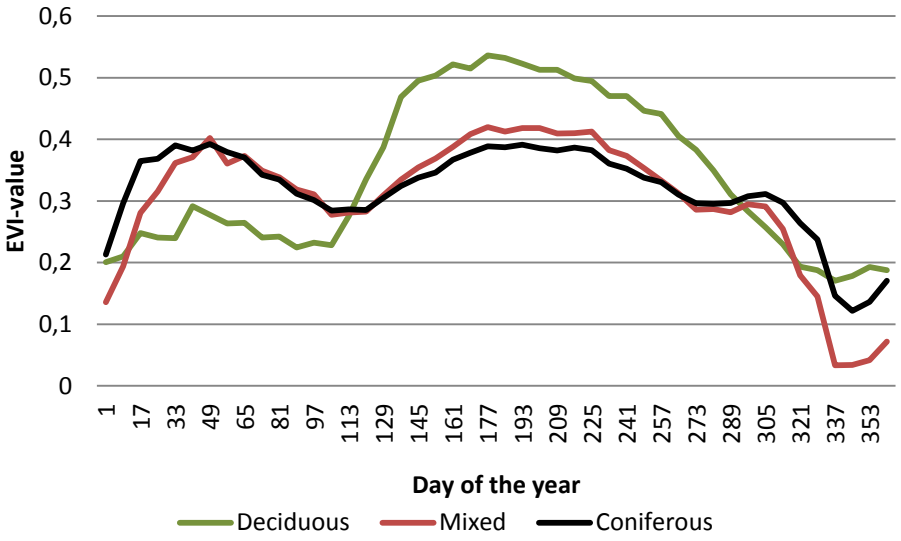

Figure 7: The mean annual EVI-values for the typical/mean deciduous, mixed or coniferous forest stand included in the study for every time step between 2003 and 2008, N = 415.

6.2. Fraction of Norway spruce (*Picea abies)* and pine (*Pinus sp.*)

6.2.1. NDVI-seasonality

When using a relative start/end of season: There are relativity few seasonal parameters that correlate with the relative composition of pine or spruce, especially when considering the few statistically significant relationships. Max has the strongest correlation for the fraction of pine, and the double logistic fitting performed somewhat better. Double logistic fitting was also the best performing for the fraction of spruce but with L-der as the seasonal parameter that correlates the strongest (Table 6).

When using an absolute start/end of season: The seasonal parameters of D-MS/D-ME/Max and ND-MS/ND-ME show also a strong correlation to the fraction of pine. This relatively strong correlation is found when applying a spatial selection area of 500 m. For the spruce species community, CD-MS and L-der show statistically significant correlations when applying a spatial selection area of 750 m that are neither strong nor weak (Table 7).

Table 6: Correlation coefficients (R) between pine and spruce for basal area and stem count when using a relative start/end of season. Table marked with green indicates correlations significant at the 0.01 level (2-tailed). Table area marked in blue indicates correlations significant at the 0.05 level (2-tailed).

Phenological parameter from TIMSAT	Selection criteria	Basal area, Pine	Basal area, Spruce	Stem count, Pine	Stem count, Spruce	N (of total possible N)
Double logistic fitting – with relative start/end of season						
Max	250	-0.44	0.183	-0.407	0.128	332 (332)
CD-ME	250	0.248	-0.138	0.248	-0.13	332 (332)
Max	500	-0.541	No.sig	-0.552	No.sig	57 (57)
L-der	500	No.sig	-0.327	No.sig	-0.301	57 (57)
Max	750	-0.577	No.sig	-0.599	No.sig	25 (25)
L-der	750	No.sig	-0.479	No.sig	-0.458	25 (25)
Savitzky-Golay filtering – with relative start/end of season						
Max	250	-0.425	0.139	-0.393	No.sig	332 (332)
Length	250	0.247	No.sig	0.212	No.sig	332 (332)
MAX	500	-0.486	No.sig	-0.474	No.sig	57 (57)
MAX	750	-0.455	No.sig	-0.453	No.sig	25 (25)
L-der	750	No.sig	-0.43	No.sig	-0.42	25 (25)

Table 7: Correlation coefficients (R) between pine and spruce for basal area and stem count when using an absolute start/end of season. Table marked with green indicates correlations significant at the 0.01 level (2-tailed). Table area marked in blue indicates correlations significant at the 0.05 level (2-tailed).

Phenological parameter from TIMSAT	Selection criteria	Basal area, Pine	Basal area, Spruce	Stem count, Pine	Stem count, Spruce	N (of total possible N)
Double logistic fitting – with absolute start/end of season						
ND-MS/ND-ME	250	-0.494	0.147	-0.454	No.sig	257 (332)
D-MS/D-ME/Max	250	-0.491	0.159	-0.455	No.sig	257 (332)
SDR	250	-0.461	0.166	-0.418	No.sig	257 (332)
D-MS/D-ME/Max	500	-0.702	No.sig	-0.702	No.sig	40 (57)
ND-MS/ND-ME	500	-0.701	No.sig	-0.698	No.sig	40 (57)
SDR	500	-0.552	No.sig	-0.572	No.sig	40 (57)
D-MS/D-ME/Max	750	-0.61	No.sig	-0.624	No.sig	20 (25)
ND-MS/ND-ME	750	-0.608	No.sig	-0.617	No.sig	20 (25)
L-der	750	-0.459	No.sig	No.sig	No.sig	20 (25)
Savitzky-Golay filtering – with absolute start/end of season						
D-MS/D-ME/Max	250	-0.389	0.127	-0.361	No.sig	282 (332)
ND-MS/ND-ME	250	-0.35	No.sig	-0.32	No.sig	282 (332)
D-MS/D-ME/Max	500	-0.487	No.sig	-0.512	No.sig	47 (57)
ND-MS/ND-ME	500	-0.46	No.sig	-0.478	No.sig	47 (57)
D-MS/D-ME/Max	750	-0.457	No.sig	-0.454	No.sig	24 (25)
CD-MS	750	No.sig	0.528	No.sig	0.556	24 (25)
L-der	750	No.sig	-0.51	No.sig	-0.52	24 (25)

6.2.2. EVI-seasonality

When examining the result of the correlation analysis for pine, one sees that the seasonal maximum, ND-MS and ND-ME correlate strongest. The results for the analysis for spruce show no particular seasonal parameter (Table 8).

Table 8: Correlation coefficients (R) between pine and spruce for basal area and stem count when using a relative and absolute start/end of season. Table marked with green indicates correlations significant at the 0.01 level (2-tailed). Table area marked in blue indicates correlations significant at the 0.05 level (2-tailed).

Phenological parameter from TIMSAT	Selection criteria	Basal area, Pine	Basal area, Spruce	Stem count, Pine	Stem count, Spruce	N (of total possible N)
Double logistic fitting – with relative start/end of season						
Max	250	-0.343	No.sig	-0.297	-0.139	332 (332)
ND-MS	250	-0.261	-0.117	-0.222	-0.162	332 (332)
ND-ME	250	-0.235	-0.121	-0.225	-0.141	332 (332)
ND-MS	500	-0.469	-0.309	-0.452	-0.364	57 (57)
MAX	500	-0.464	-0.28	-0.43	-0.34	57 (57)
ND-ME	500	-0.431	-0.366	-0.382	-0.452	57 (57)
Max	750	-0.593	No.sig	-0.558	No.sig	25 (25)
CD-ME	750	0.462	No.sig	0.448	No.sig	25 (25)
D-ME	750	No.sig	-0.479	No.sig	-0.435	25 (25)
L-der	750	No.sig	-0.466	No.sig	-0.499	25 (25)
Double logistic fitting – with absolute start/end of season						
D-MS/D-ME/Max	250	-0.309	-0.117	-0.252	-0.186	322 (332)
ND-MS/ND-ME	250	-0.304	-0.142	-0.243	-0.208	322 (332)
SIR	250	-0.239	No.sig	-0.175	-0.128	322 (332)
SDR	250	-0.221	No.sig	-0.151	No.sig	322 (332)
S-integ	500	-0.471	No.sig	-0.447	-0.289	57 (57)
D-MS/D-ME/Max	500	-0.371	-0.284	-0.321	-0.338	57 (57)
ND-MS/ND-ME	500	-0.366	-0.297	-0.308	-0.349	57 (57)
Amp	500	-0.358	-0.326	-0.327	-0.397	57 (57)
SDR	750	-0.555	No.sig	-0.502	No.sig	25 (25)
D-MS/D-ME/Max	750	-0.49	No.sig	-0.448	No.sig	25 (25)
SIR	750	-0.448	No.sig	-0.414	No.sig	25 (25)
ND-MS/ND-ME	750	-0.435	No.sig	No.sig	No.sig	25 (25)
L-der	750	No.sig	-0.488	No.sig	-0.52	25 (25)
Amp	750	No.sig	-0.415	No.sig	No.sig	25 (25)

It should be noted that the correlation coefficient, R, becomes stronger the stricter the spatial selection criteria becomes. That is to say that the data from the spatial selection criteria of 750 m correlated stronger compared to the data from the spatial selection criteria of 500 m. This is a deviation from the general pattern of the results of the other correlation analyzes (Table 3-7) where spatial selection criteria of 500 m generally show a stronger correlation. The pattern showed in Table 8 was what was otherwise expected beforehand. The correlation results for spruce also deviated from the general pattern when comparing the difference in correlation

between basal area and stem count. Generally, the correlation with the seasonal parameters and the fraction of any tree species communities measured in basal area correlated stronger as compared to stem count (Table 3-7). The results of the correlation analyzes of the seasonality parameter extracted from EVI-values show on the other hand that the fraction of spruce correlated stronger against stem count than basal area.

6.3. Regression analysis and model validation

Scatter plots for the regression analyzes from the strongest correlation are presented in Figure 8-13. The scatter plots on the left side have the fraction of deciduous trees measured in basal area on the y-axel, while the scatter plots on the right has the fraction of deciduous trees measured in stem counts. The scatter plot of Figure 10 has the strongest coefficient of determination with a R^2 of 0.78.

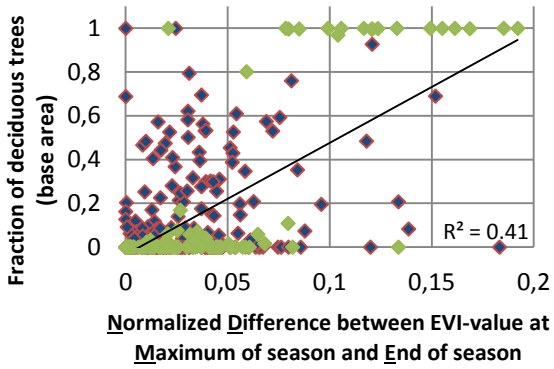

Figure 8: Scatter plot between ND-ME and fraction of deciduous trees (basal area), (green point = monitoring-plots, blue points = NFI plot) N = 322.

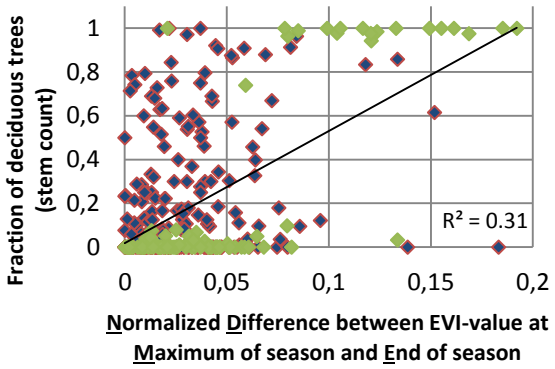

Figure 9: Scatter plot between ND-ME and fraction of deciduous trees (stem count), (green point = monitoring-plots, blue points = NFI plot) N = 322.

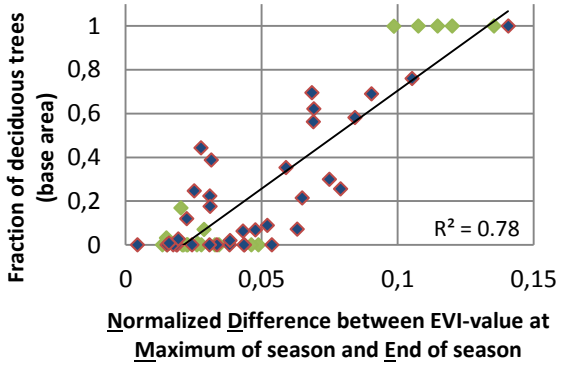

Figure 10: Scatter plot between ND-ME and fraction of deciduous trees (basal area), (green point = monitoring-plots, blue points = NFI plot) N = 57.

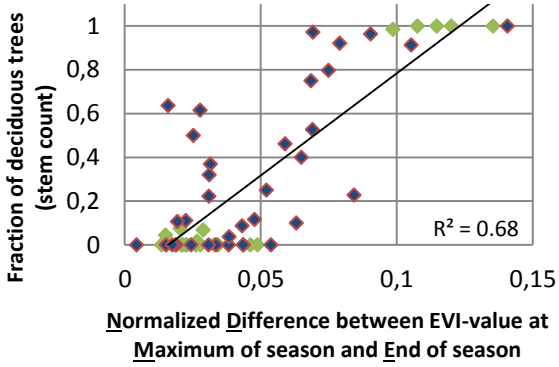

Figure 11: Scatter plot between ND-ME and fraction of deciduous trees (stem count), (green point = monitoring-plots, blue points = NFI plot) N = 57.

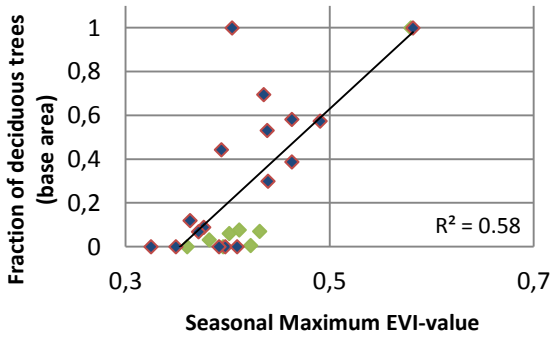

Figure 12: Scatter plot between Max (EVI-value) and fraction of deciduous trees (basal area), (green point = monitoring-plots, blue points = NFI plot) N = 25.

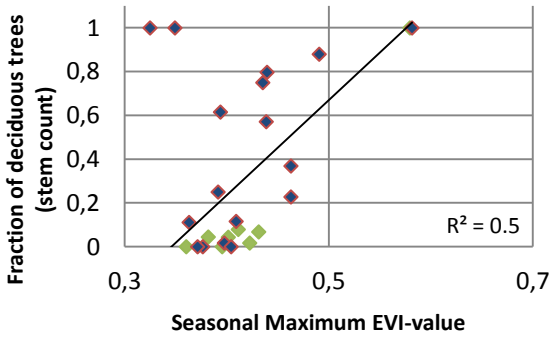

Figure 13: Scatter plot between Max (EVI-value) and fraction of deciduous trees (stem count), (green point = monitoring-plots, blue points = NFI plot) N = 27.

The results of the validation of the regression models when predicting the fraction of deciduous/coniferous measured in basal area are presented in Table 12. The predicted fraction deviated the least from the actual fraction for the dataset selected from the 500 m spatial criteria when calculated as percentage ($RMSE_r$).

Table 9: Validation of the predicted values of the regression models for the fraction of the Deciduous/Coniferous measured in basal area. The regression model starting with a positive value refers to the model predicting the fraction of deciduous and vice versa.

Validation of the regression results					Basal area, Deciduous/Coniferous				
x = Phenological parameter from TIMSAT	Vegetation index	Start/ Stop	Selection criteria	ŷ	RMSE	$RMSE_r$	Training N	Validation N	R^2
ND-MS/ ND-ME	EVI	Absolute	250	5,1021x - 0,0344 -5,1021x – 1,00344	0.306	202.4	332	83	0.41
ND-ME	EVI	Relative	500	8,9499x - 0,1902 -8,9499x + 1,1902	0.303	91.8	57	15	0.78
Max	EVI	Relative	750	4,3032x - 1,5215 -4,3032x + 2,5215	0.246	121.6	25	6	0.58

The results of the validation of the regression models when predicting the fraction of deciduous/coniferous measured in stem count are shown in Table 13. The predicted fraction deviated the least from the actual fraction for the dataset selected from 500 m spatial criteria both in relative and absolute terms.

Table 10: Validation of the predicted values of the regression models for the fraction of the Deciduous/Coniferous measured in Stem count. The regression model starting with a positive value refers to the model predicting the fraction of deciduous and vice versa.

Validation of the regression results					Stem count, Deciduous/Coniferous				
x = Phenological parameter from TIMSAT	Vegetation index	Start/ Stop	Selection criteria	ŷ	RMSE	$RMSE_r$	Training N	Validation N	R^2
ND-MS/ ND-ME	EVI	Absolute	250	5,1271x + 0,0175 -5,1271x + 0,9825	0.352	173.0	322	83	0.31
ND-ME	EVI	Relative	500	9,3432x - 0,1502 -9,3432x + 1,1502	0.337	93.7	57	15	0.68
Max	EVI	Relative	750	4,3489x - 1,5045 -4,3489x + 2,5045	0.34	126.8	25	6	0.5

6.4. Error matrices if method was used for mapping

The error matrices generated from the data selected at 250 m spatial criterion is shown in Tables 14 and 15. The overall accuracy is 73 % for basal area and 69 % for stem count. The overall kappa scored 0.27 and 0.26 for basal area and stem count respectively.

Table 11: Error matrix generated from the predicted and observed classified data when data was selected with 250 m spatial criteria, measured in basal area.

Basal area	Decid-uous	Mixed	Conifer-ous	Sum of rows	Error of commission	User accu racy
Decid-uous	0	0	1	1	100%	0%
Mixed	3	6	9	18	67%	33%
Conifer-ous	3	6	55	64	14%	86%
Sum of columns	6	12	65	83		
Error of omission	100%	50%	15%			
Producer Accuracy	0%	50%	85%			
Overall accuracy	73%					
Kappa	0.27					

Table 12: Error matrix generated from the predicted and observed classified data when data was selected with 250 m spatial criteria, measured in stem count.

Stem count	Decid-uous	Mixed	Conifer-ous	Sum of rows	Error of commission	User accu racy
Decid-uous	0	0	2	2	100%	0%
Mixed	6	6	9	21	71%	29%
Conifer-ous	4	5	51	60	15%	85%
Sum of columns	10	11	62	83		
Error of omission	100%	45%	18%			
Producer Accuracy	0%	55%	81			
Overall accuracy	69%					
Kappa	0.26					

Note the differences in the sum of the columns in Table 14 and 15. Even though the reference data are the same for the two tables above, the differences in the sum of the columns are the result of how a specific reference point is defined. The definition is based on either basal area or stem count, but have the same threshold for when a reference point is referred to as deciduous, mixed or coniferous forest.

The error matrices generated from the data selected at 500 m spatial criteria can be seen in Table 16 and 17. Tables 18 and 19 show the error matrices from the data selected at the 750 m spatial criterion. Observe that the small sample size for the data in Tables 16-18 for the validation data may be the reason why the overall accuracy varies greatly. The overall accuracy varies form 33-67 % and kappa varies from 0-0.4.

Table 13: Error matrix generated from the predicted and observed classified data when data was selected with 500 m spatial criteria, measured in basal area

Basal area	Decid-uous	Mixed	Conifer-ous	Sum of rows	Error of commission	User accu racy
Decid-uous	0	0	0	0	0%	0%
Mixed	3	2	3	8	75%	25%
Conifer-ous	0	2	5	7	29%	71%
Sum of columns	3	4	8	15		
Error of omission	100%	50%	38%			
Producer Accuracy	0%	50%	63%			
Overall accuracy	47%					
Kappa	0.12					

Table 14: Error matrix generated from the predicted and observed classified data when data was selected with 500 m spatial criteria, measured in basal area

Stem count	Decid-uous	Mixed	Conifer-ous	Sum of rows	Error of commission	User accu racy
Decid-uous	0	0	0	0	0%	100 %
Mixed	4	1	4	9	89%	11%
Conifer-ous	0	2	4	6	32%	67%
Sum of columns	4	3	8	15		
Error of omission	100%	67%	50%			
Producer Accuracy	0%	33%	50%			
Overall accuracy	33%					
Kappa	0					

Table 15: Error matrix generated from the predicted and observed classified data when data was selected with 750 m spatial criteria, measured in basal area

Basal area	Decid-uous	Mixed	Conifer-ous	Sum of rows	Error of commission	User accu racy
Decid-uous	0	0	0	0	0%	100 %
Mixed	0	2	2	4	50%	50%
Conifer-ous	0	0	2	2	0%	100 %
Sum of columns	0	2	4	6		
Error of omission	0%	0%	50%			
Producer Accuracy	100%	100%	50%			
Overall accuracy	67%					
Kappa	0.4					

Table 16: Error matrix generated from the predicted and observed classified data when data was selected with 750 m spatial criteria, measured in basal area

Stem count	Decid-uous	Mixed	Conifer-ous	Sum of rows	Error of commission	User accu racy
Decid-uous	0	0	0	0	0%	0%
Mixed	1	1	3	5	80%	20%
Conifer-ous	0	0	1	1	0%	100 %
Sum of columns	1	1	4	6		
Error of omission	100%	0%	75%			
Producer Accuracy	0%	100%	25%			
Overall accuracy	33%					
Kappa	0.1					

Looking at the results from Tables 14-19 class by class, the class of coniferous forest shows a fairly good producer and user accuracy. The coniferous forest measured in basal area shows a producer accuracy ranging from 50-85 %, and the user accuracy varies from 71-100 %. However, when coniferous forest was classified as measured by stem count the producer and user accuracy measured was much lower. The producer accuracy ranges from 25-81 % and the user accuracy ranges from 67-100 %. The class of deciduous forest receives very low producer and user accuracy no matter if the classes were defined by basal area or stem count. The only situation when the observed data and the modelled data agree are when neither of the two can contains any deciduous forest. Mixed forest on the other hand has a very intermediate performance with a producer accuracy of about 50/50 but with generally lower user accuracy. Worth noticing is that when comparing the overall accuracy and kappa between the two methods for definition, the basal area always performs better compared with stem count.

7. Discussion

The results strongly suggest that the seasonal variability in the vegetation indices are higher the more deciduous trees are present, and vice versa with coniferous trees. This is underlined by the fact that a general pattern can be detected no matter whether NDVI or EVI were used, and whether absolute or relative start/stop of season, double logistic filtering or Savitzky-Golay fitting were used, even if the strength of the correlation differed from case to case. The general pattern that was noticed is that when the seasonal parameters of max, amp, S-integ, L-der, D-MS, D-ME, ND-MS, ND-ME, SIR and SDR had a statistically significant correlation between the fractions of deciduous/coniferous, the correlation was always positive for the fraction of deciduous trees and negative for the fraction of coniferous tress. This means that the difference between the vegetation index value at the peak of season and at the start/end of season is always higher the higher the fraction of deciduous trees is within the forest stand. In addition, the correlation was always negative when the seasonal parameters of Length, CD-MS and CD-ME were tested against the fraction of deciduous, and positive when tested against the fraction of coniferous vegetation. This means that the modelled season is always longer the higher the fraction of coniferous trees. These results are in line with what Barr et al. (2009) reported. However, the same author reported that the seasonal differences in NDVI-values between different tree species communities (aspen, spruce and pine) during the growing season was greater compared to this study. The general seasonal pattern for NDVI that Barr et al. (2009) reported resembles more the pattern showed for EVI in Figure 8 which visualises the seasonal signature of EVI for the different tree species communities (deciduous, mixed and coniferous).

The results of this study shows that the seasonal parameters generated from EVI generally correlated stronger against the fraction of deciduous/coniferous forest compared with the seasonal parameters generated from NDVI. The considerably better performance of EVI suggests that EVI is a far more dynamic vegetation index compared to NDVI when used for measuring biophysical parameters in forest stands. Why EVI is the far more dynamic vegetation index may be due to the differences in spectral response when the amount of green biomass becomes high. NDVI tends to saturate when the amount of green biomass and/or LAI becomes high, while EVI has an improved sensitivity in high biomass land covers, such as forested areas. In addition, NDVI is also sensitive to background noise from e.g. understory vegetation due to it being stronger controlled by reflectance in the red part of the EM spectrum. EVI is, on the other hand, more controlled by the NIR band, which explains why EVI may be more dynamic to vegetation (Jensen 2007). Hence, that when measuring the spectral properties of forests, it is of great important to choose a vegetation index that is responsive when the volume of biomass becomes high.

The cases when an absolute threshold was used to define seasonal start/stop and correlated better compared with when a relative start/end was applied was also the cases when many sample sites had been excluded due to the fact that no seasonality could be modelled in TIMESAT. This means that sample sites with much spectral noise may have been excluded, thus generating a high correlation. A support for such an interpretation can be found in the fact that the times when few or none of the sample sites were excluded, even though using an

absolute value for defining start/end of season, the model with relative star/end of season often generated stronger results.

When using a spatial selection criterion of 500 m instead of 250 m, the correlation generally increased which is probably due to the dataset selected with a 500 m criterion is less influenced by noise. Following the same logic when comparing the result of the correlation analysis of the 750 and 500 or 250 m areas, the dataset selected with a 750 m criterion would correlate even better. However, this was not the case. More than anything, this may highlight the uncertainty or insufficiency of the visual interpretation via Google Earth™ that was applied in this study. The different mapping scales of the field and satellite data is clearly a major source of error. It is a source of error in the sense that it is difficult to foresee if the measured field data (represented by a circle of 7 m radius or a square of 30 m in this study) is valid for the whole area covered by the satellite pixel.

Another source of error is the georeferencing of the satellite image. When georeferencing a satellite image a displacement of 0.5-1 pixel can be expected any direction (McCoy 2005). This means that the field data could have been related to the "wrong" pixel, i.e. the wrong spectral signature in both space and time. In addition, the samples sites in this study had been registered with a standard handheld GPS-receiver with a spatial accuracy of at best 5 m, but more realistically at < 15 m. These two errors, derived from the lacking accuracy of georeferencing and GPS-receivesr, can work additively and in this case lead to a positional error of 265 m in any direction. The risk of relating the sample site to the wrong pixel is in this case quite large if sample sites are located at the border between two or more pixels. Normally, the minimum requirement for the area or the field data is set to be 3 times bigger than the pixel size (McCoy 2005, Congalton and Green 2009, Olofsson et al. 2014). In this study, the field data is 36 or 8 times smaller than the pixel size.

Despite these sources of errors, the overall result indicates that a strong correlation exists between modelled seasonality parameters and fraction of deciduous/coniferous tree species. This gives strong reasons to assume that with better data the method proposed in this study could yield better results with higher accuracy when predicting the fraction of deciduous/coniferous tree species in forest stands.

7.1. A future outlook

MODIS data was chosen despite its coarser spatial resolution to gain a higher temporal resolution, which was necessary to reconstruct the phenology. This is a common trade off that scientists are forced to make. However, within the next few years the European satellite system of Sentinel-2 will provide data with high spectral, spatial and temporal resolution over Europe (the first Sentinel-2 satellite is already in orbit and producing useful test data while this text is written). This provides possibilities to map phenology related to specific land covers or vegetation species communities rather than the "ecosystem", as MODIS offers (Eklundh et al. 2012).

If future research will be performed with the same objectives as this study but with Sentinel-2 used as satellite data, a good idea is to construct a field dataset which is specifically designed to be adapted for the spatial properties of Sentinel-2 to avoid some of the sources of error in

this study. The performance of the analysis of the specially designed field data could then be compared with an analysis in which e.g. the Swedish NFI is used instead, to investigate differences. Alternatively, the Swedish NFI could be used as reference data to validate the performance of the specially designed field data. It should be noted that such a field dataset could easily be coordinated with other research programs with other objectives, e.g. mapping LAI with Sentinel-2, phenology studies for tree budburst, etc.

Other factors to consider are what parameters to include in a future field study based on the experiences gained in this study. The seasonal parameters tended to correlate stronger when the fraction of tree species was measured as basal area rather than as stem count. This is logical when considering that the basal area probably correlates better against the total volume of green biomass compared with the stem count. However, it is quite common that the relative fraction of canopy cover is used as a measure when defining tree species communities (Bravo-Oviedo et al. 2014) and land cover classes (McCoy 2005). From the perspective of remote sensing, fractional distribution of canopy cover between tree species may affect the spectral signature to a large extent. Hence, it is recommended to include canopy cover beside basal area and stem count in a study when designing a future field survey.

The tested vegetation indices (NDVI and EVI) in this study are two of the most frequently used vegetation indices to map vegetation properties. Both NDVI and EVI utilize the knowledge of the inverse relationship between the red and the NIR part of the EM spectrum associated with healthy vegetation. But as mentioned and pointed out as one of the problems within this study, the sensitivity when used for high green biomass land covers may be weak. For future studies it may be good to also consider testing other spectral vegetation indices and reflectance models for inverting the satellite measure spectral signature, or other techniques such as a physical based vegetation index (Jensen 2007, Jin and Eklundh 2014). Using physically based models to invert the spectral information to retrieve biophysical parameters such as LAI is a well known and tested method and often yields similar results as simpler statistical method like regression models (e.g. Eklundh, Harrie and Kuusk 2001, Rautiainen 2005, Schlerf and Atzberger 2006). The physically based vegetation index of PPI (Plant Phenology Index, Jin and Eklundh 2014) is, however, a relatively new approach to monitor or map LAI. The PPI has a nearly linear relationship to LAI which means that it does not suffer from the problems of saturation when used in high biomass land covers. PPI is calculated from the red and NIR band but utilize radiative transfer theory rather than empirical approximation of the spectral properties of vegetation (Jin and Eklundh 2014). Jin and Eklundh (2014) show that the PPI has great potential for mapping phenology, and the index may therefore also be suited for phenology based land cover mapping in forested areas.

The results from the regression analysis shows that the general assumption for the analysis is met, and that it in principle works for predicting e.g. fraction of deciduous/coniferous. The accuracy for the prediction is, however, low. The strong correlations showed for some of the parameters indicate however that some of the low accuracy may be the result of the many sources of errors derived from the used data. On the other hand, the choice of statistical analysis can be discussed. Using empirical relationships between forest attributes and satellite derived spectral measurements via regression models, linear or non-linear, is a well tested approach. When using a linear regression model for predicting percent or fractions it is

possible to predict a value outside of the valid range of the model (i.e. <0 or >1). However, even if it is possible obtain values >1 with an ordinary least squares fit as used in this study, it does not make sense to predict values >1 since the forest stand cannot consist of more than 100 % coniferous trees. Also note that the relationship between the variables x and y are not linear, since $\hat{y}$ can be predicted to be <0 or >1, while this is impossible in reality. Hence, it may be less optimal to use a linear regression model as examined in this study. It should be noted that in this study, predicted values <0 and >1 were rare. Even so, for future studies it is preferable to also consider other statistical approaches when modelling the fraction of species composition. The easiest option is using a non-linear regression, where the fitted line takes on a curve with a sigmoid shape (Rogerson 2010). Another approach would be to use neural network modelling. Non-linear regression and neural networks was used by i.e. Muukkonen and Heiskanen (2005) when modelling actual forest properties. Another alternative advocated in the literature when approaching the dilemma of predicting classes or categories are to treat it as binary data, meaning that a forest stand could be either coniferous or not. Then a logistic regression model could be considered (Rogerson 2010). If the outcome have more than two available responses (e.g. coniferous, mixed or deciduous), then a multinomial logistic regression could be considered for mapping relative tree species composition.

8. Conclusion

- The overall result indicates that a strong correlation exists between modelled phenological parameters estimated from vegetation indices extracted from multi-temporal satellite imagery and fraction of deciduous/coniferous tree species.

- The results gives strong reason to assume that if using data of good quality, statistical models can be constructed and used to map e.g. the fraction of deciduous and coniferous trees with phenological parameters extracted from multi-temporal satellite imagery.

- The Enhanced Vegetation Index (EVI) proved to be a more dynamic vegetation index compared to the Normalized Difference Vegetation Index (NDVI) which is probably due the NDVI is subjected to saturation effects at higher amounts of green biomass compared to the EVI.

- Modelled phenological parameters associated with differences between seasonal maximum and seasonal start or end always had a positive correlation against the fraction of deciduous trees species, and a negative correlation against the fraction of coniferous tree species. This means that the seasonal differences in NDVI or EVI within the same year is higher the higher the fraction of deciduous vegetation is.

- Modelled phenological parameters associated with length of season always had a negative correlation against the fraction of deciduous trees species and a positive correlation against the fraction of coniferous tree species. This means that the estimated length of a growing season is longer the higher the fraction of coniferous vegetation is.

9. References

Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). 2015. MODIS subsetted land products, Collection 5. Available on-line. In *ORNL DAAC, Oak Ridge, Tennessee, U.S.A. Accessed July 24, 2014. Subset obtained for M*D13Q1product at time period: 2003-01-01 to 2008-12-361, and subset size: 0.25 x 0.25 km*.

Akselsson, C., G. Phil Karlsson, P. E. Karlsson & J. Ahlstrand (2015) Miljöövervakning i obsytorna 1984-2013 – Beskrivning, resultat, utvärdering och framtid. *Skogsstyrelsens böcker och broschyrer,* 1864**,** 140.

Alkema, D., W. Bijker, A. Sharifi, Z. Vekerdy & W. Verhoef. 2012. Chapter 11: Data integration. In *The core of GIScience: a systems-based approach,* eds. V. Tolpekin & A. Stein, 373-429. Enschede: The International Institute for Geo-Information Science and Earth Observation (ITC).

Atzberger, C., A. Klisch, M. Mattiuzzi & F. Vuolo (2013) Phenological Metrics Derived over the European Continent from NDVI3g Data and MODIS Time Series. *Remote Sensing,* 6**,** 257-284.

Bakker, W., W. Bakx, K. Grabmaier, L. Janssen, J. Horn, G. Huurneman, F. van der Meer, C. Pohl, K. Tempfli, V. Tolpekin & T. Woldai. 2012. Chapter 4: Sensors. In *The core of GIScience: a systems-based approach,* eds. V. Tolpekin & A. Stein, 125-166. Enschede: The International Institute for Geo-Information Science and Earth Observation (ITC).

Bakx, W., L. Janssen, E. Schetselaar, K. Tempfli & V. Tolpekin. 2012a. Chapter 6: Digital image classification. In *The core of GIScience: a systems-based approach,* eds. V. Tolpekin & A. Stein, 205-225. Enschede: The International Institute for Geo-Information Science and Earth Observation (ITC).

Bakx, W., K. Tempfli, V. Tolpekin & T. Woldai. 2012b. Chapter 2: Physics. In *The core of GIScience: a systems-based approach,* eds. V. Tolpekin & A. Stein, 71-92. Enschede: The International Institute for Geo-Information Science and Earth Observation (ITC).

Barr, A., A. T. Black & H. McCaughey. 2009. Climatic and Phenological Controls of the Carbon and Energy Balances of Three Contrasting Boreal Forest Ecosystems in Western Canada. In *Phenology of Ecosystem Processes - Applications in Global Change Research,* ed. A. Noormets, pp. 3-34. Dordrecht Heidelberg London New York: Springer.

Bergil, C., S. Bydén, S. Edman, B. Eknert, C. Jerer, B. Lind, A. Nordström, M. Olsson, B. Sannel & A. Yrgård. 2004. *Mark - Människa – Miljö*. Bohuslän'5.

Bravo-Oviedo, A., H. Pretzsch, C. Ammer, E. Andenmatten, A. Barbati, S. Barreiro, P. Brang, F. Bravo, L. Coll, P. Corona, J. d. Ouden, M. J. Ducey, D. I. Forrester, M. Giergiczny, J. B. Jacobsen, J. Lesinski, M. Löf, B. Mason, B. Matovic, M. Metslaid, F. Morneau, J. Motiejunaite, C. O'Reilly, M. Pach, Q. Ponette & M. d. Rio (2014) European mixed forests: definition and research perspectives. *Forest Systems,* 23**,** 518-533.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20**,** 37-46.

Congalton, R. G. & K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data, Principles and Practices.* Boca Raton. London. New York.: CRC Press.

EEA. 2006. Corine Land Cover 2006. http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster: EEA.

Eklundh, L., J. Ardö, P. Jönsson & M. Sjöström. 2012. High resolution mapping of vegetation dynamics from Sentinel-2. In *European Space Agency, Sentinel-2 Preparatory Symposium, Rome, April 23-27*. Rome.

Eklundh, L., L. Harrie & A. Kuusk (2001) Investigating relationships between Landsat ETM+ sensor data and leaf area index in a boreal conifer forest. *Remote Sensing of Environment,* 78**,** 239-251.

Eklundh, L. & P. Jönsson. 2015. TIMESAT 3.2 with parallel processing Software Manual. ed. S. Lund and Malmö University.

Forest; Encyclopaedia Britannica. 2015. http://academic.eb.com.ezproxy.ub.gu.se/EB-checked/topic/213461/forest. In *Encyclopaedia Britannica*. Encyclopædia Britannica, Inc.

Fridman, J., S. Holm, M. Nilsson, P. Nilsson, A. Hedström Ringvall & G. Ståhl (2014) Adapting National Forest Inventories to changing requirements – the case of the Swedish National Forest Inventory at the turn of the 20th century. *Silva Fennica* 48**,** 29

Fridman, J. & M. Nilsson. 2013. National forest inventory plot data. www.slu.se/PageFiles-/254572/Info_Data_for_calibration.pdf: Anna Stromback, Web Editor at Swedish Forest Agency.

Gustavsson, L. 1996. Växt- och djurgeografika indelningar. In *Sveriges Nationalatlas – Växter och djur,* eds. I. Ahlén & L. Gustavsson. Stockholm: SNA förlag.

Jensen, J. R. 2007. *Remote Sensing of the Environment: An Earth Resource Perspective*. Prentice Hall.

Jin, H. & L. Eklundh (2014) A physically based vegetation index for improved monitoring of plant phenology. *Remote Sensing of Environment,* 152**,** 512-525.

Jönsson, A. M., L. Eklundh, M. Hellström, L. Bärring & P. Jönsson (2010) Annual changes in MODIS vegetation indices of Swedish coniferous forests in relation to snow dynamics and tree phenology. *Remote Sensing of Environment,* 114**,** 2719-2730.

Jönsson, P. & L. Eklundh (2002) Seasonality extraction by function fitting to time-series of satellite sensor data. *Geoscience and Remote Sensing, IEEE Transactions on,* 40**,** 1824-1832.

Jönsson, P. & L. Eklundh (2004) TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences,* 30**,** 833-845.

Lieth, H. 1974. *Phenology and seasonality modeling. Ecological Studies - Analysis and Synthesis Series, Vol 8*. Berlin: Springer-Verlag.

Linderholm, H. W. (2006) Growing season changes in the last century. *Agricultural and Forest Meteorology,* 137**,** 1-14.

McCoy, R. M. 2005. *Field Methods in Remote Sensing*. New York: The Guilford Press.

Muukkonen, P. & J. Heiskanen (2005) Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sensing of Environment,* 99**,** 434-447.

NASA. 2015. Measuring Vegetation. http://earthobservatory.nasa.gov/Features/Measuring-Vegetation/measuring_vegetation_2.php: NASA - The Earth Observatory.

Olofsson, P., G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock & M. A. Wulder (2014) Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment,* 148**,** 42-57.

Olsson, C. 2014. *Tree phenology modelling in the boreal and temperate climate zones : Timing of spring and autumn events*. Lund: Department of physical Geography and Ecosystem Science, Lund University.

Olsson, P. E. 2015. email correspondence 2015-03-29/2015-03-30.

ORNL. 2014. MODIS subsetted land products, Collection 5. Available on-line. In *ORNL DAAC, Oak Ridge, Tennessee, U.S.A. Accessed July 24, 2014. Subset obtained for MOD13Q1 product at time period: 2003-01-01 to 2008-12-31, and subset size: 0.25 x 0.25 km.* . http://daac.ornl.gov/MODIS/modis.html: Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC).

Rautiainen, M. (2005) Retrieval of leaf area index for a coniferous forest by inverting a forest reflectance model. *Remote Sensing of Environment,* 99**,** 295-303.

Rogerson, P. 2010. *Statistical methods for geography*. London: Sage.

Sallaba, F. 2011. The potential of support vector machine classification of land use and land cover using seasonality from MODIS satellite data. In *Department of Earth and Ecosystem Sciences - Physical Geography and Ecosystems Analysis*, 80. Lund University.

Schlerf, M. & C. Atzberger (2006) Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment,* 100**,** 281-294.

Solano, R., K. Didan, A. Jacobson & A. Huete. 2010. MODIS Vegetation Index User's Guide (MOD13 Series). Vegetation Index and Phenology Lab, The University of Arizona.

Tolpekin, V. & A. Stein. 2012. Introduction. In *The core of GIScience: a systems-based approach,* eds. V. Tolpekin & A. Stein, 21-32. Enschede: The International Institute for Geo-Information Science and Earth Observation (ITC).

Tottrup, C., M. S. Rasmussen, L. Eklundh & P. Jönsson (2007) Mapping fractional forest cover across the highlands of mainland Southeast Asia using MODIS data and regression tree modelling. *International Journal of Remote Sensing,* 28**,** 23-46.

Wang, Q. & J. D. Tenhunen (2004) Vegetation mapping with multitemporal NDVI in North Eastern China Transect (NECT). *International Journal of Applied Earth Observation and Geoinformation,* 6**,** 17-31.

Appendixes

Appendix 1. Statistical methods used

The linear correlation coefficient (*R*) used in this study is given by Equation 3:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \qquad \textit{equation 3}$$

where s_x and s_y are the sample standard deviation of variables *x* and *y*. This measure is also known as *Pearson's* correlation (Rogerson 2010).

The predicted value ($\hat{y}$) was evaluated with root-mean-square-error (RMSE) and its relative counterpart ($RMSE_r$):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \textit{equation 4}$$

$$RMSE_r = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}} \times 100 \qquad \textit{equation 5}$$

where y_i is the observed value, $\hat{y}_i$ is the modelled value and $\bar{y}$ is the mean of the observed values (Muukkonen and Heiskanen 2005).

		j = Columns (Reference)			Row total
		1	2	k	n_{1+}
i = Rows (Classification)	1	n_{11}	n_{12}	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	n_{2k}	n_{2+}
	k	n_{k1}	n_{k2}	n_{kk}	k_{k-}
	Coulmn total, n_{+1}	n_{+1}	n_{+2}	n_{+k}	n

Table 17: Principal illustration of an error matrix (Congalton and Green 2009).

The error matrix used in this study can mathematically be represented as Table 17 from which the following accuracy measures could be derived, if letting equation 6 be the number of samples classified as *i* and equation 7 be number of samples classified as *j*

$$n_{i+} = \sum_{j=1}^{k} n_{ij} \qquad \textit{Equation 6}$$

$$n_{+j} = \sum_{i=1}^{k} n_{ij} \qquad \textit{Equation 7}$$

Then the Overall accuracy can be calculated by equation 8, user accuracy with equation 9 and producer accuracy with equation 10

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{k} n_{ii}}{n} \qquad \textit{Equation 8}$$

$$\text{User accuracy} = \frac{n_{ii}}{n_{1+}} \qquad \textit{Equation 9}$$

$$\text{Producer accuracy} = \frac{n_{jj}}{n_{+j}} \qquad \textit{Equation 10}$$

The overall accuracy takes the sum of the correctly classified samples divided by the total number of sample, providing a measure of the probability for all samples to be correctly classified. User and producer accuracy are accuracy measures provided class by class where the user accuracy is the number of times the samples of the reference data and the map classification aggress divided by the total number of samples that the map identified samples as the class at hand. Considering Figure 17, user accuracy of Class 1 is calculated by dividing n_{11} by n_{1+}. Producer accuracy is the number of times the samples of the reference data and the map classification aggress divided by the total number of samples that the reference data assigned the samples as the class at hand. If taking Class 1 again as an example, producer accuracy is calculated by dividing n_{11} by n_{+1} (Congalton and Green 2009). Two other class specific measures are the errors of commission and omission. Errors of commission are the incorrectly classified samples, also known as Type 1 error. In Figure 17, the error of commission for Class 1 is calculated as followed; $(n_{21} + n_{k1})/n_{+1}$. Error of omission, or Type 2 error, refers to the omitted errors of the map classification. In Table 17, the error of omission for Class 1 is calculated by summing $(n_{12} + n_{1k})/n1_{+}$ (Bakx et al. 2012a). The Cohen's Kappa coefficient which is by equation 11:

$$\widehat{\mathrm{K}} = \frac{n\sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k}(n_{i+} \times n_{+i})}{n^2 - \sum_{i=1}^{k}(n_{i+} \times n_{+i})} \qquad \textit{Equation 11}$$

Series from Lund University

Department of Physical Geography and Ecosystem Science

Master Thesis in Geographical Information Science

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland. (2008).
2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research. (2008).
3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing. (2009).
4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia. (2009).
5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).
6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).
7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010)
8. *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010)
9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010)
10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011)
11. *Svajunas Plunge:* Advanced decision support methods for solving diffuse water pollution problems (2011)
12. *Jonathan Higgins:* Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).
13. *Mårten Karlberg:* Mobile Map Client API: Design and Implementation for Android (2011).
14. *Jeanette McBride:* Mapping Chicago area urban tree canopy using color infrared imagery (2011)
15. *Andrew Farina:* Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011)
16. *David Kanyari*: Nairobi City Journey Planner An online and a Mobile Application (2011)
17. *Laura V. Drews:* Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012)
18. *Qaisar Nadeem:* Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012)

19. *Ahmed Mohamed El Saeid Mustafa:* Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012)
20. *Daniel Patrick Taylor*: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013)
21. *Angeleta Oveta Wilson:* A Participatory GIS approach to *unearthing* Manchester's Cultural Heritage '*gold mine'* (2013)
22. *Ola Svensson:* Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylian Hinterlands and the Soulima Valley (2013)
23. *Monika Ogden:* Land use impact on water quality in two river systems in South Africa (2013)
24. *Stefan Rova:* A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013)
25. *Yann Buhot:* Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013)
26. *Christina Fotiou:* Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014)
27. *Inese Linuza:* Accuracy Assessment in Glacier Change Analysis (2014)
28. *Agnieszka Griffin:* Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014)
29. *Brynja Guðmundsdóttir:* Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014)
30. *Oleksandr Nekrasov:* Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014)
31. *Sarah Tressel:* Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure (2014)
32. *Caroline Gevaert:* Combining Hyperspectral UAV and Multispectral Formosat-2 Imagery for Precision Agriculture Applications (2014).
33. *Salem Jamal-Uddeen:* Using GeoTools to implement the multi-criteria evaluation analysis - weighted linear combination model (2014)
34. *Samanah Seyedi-Shandiz:* Schematic representation of geographical railway network at the Swedish Transport Administration (2014)
35. *Kazi Masel Ullah:* Urban Land-use planning using Geographical Information System and analytical hierarchy process: case study Dhaka City (2014)
36. *Alexia Chang-Wailing Spitteler:* Development of a web application based on MCDA and GIS for the decision support of river and floodplain rehabilitation projects (2014)
37. *Alessandro De Martino:* Geographic accessibility analysis and evaluation of potential changes to the public transportation system in the City of Milan (2014)
38. *Alireza Mollasalehi:* GIS Based Modelling for Fuel Reduction Using Controlled Burn in Australia. Case Study: Logan City, QLD (2015)
39. *Negin A. Sanati:* Chronic Kidney Disease Mortality in Costa Rica; Geographical Distribution, Spatial Analysis and Non-traditional Risk Factors (2015)

40. *Karen McIntyre:* Benthic mapping of the Bluefields Bay fish sanctuary, Jamaica (2015)
41. *Kees van Duijvendijk:* Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment (2015)
42. *Sebastian Andersson Hylander:* Evaluation of cultural ecosystem services using GIS (2015)
43. *Deborah Bowyer:* Measuring Urban Growth, Urban Form and Accessibility as Indicators of Urban Sprawl in Hamilton, New Zealand (2015)
44. *Stefan Arvidsson:* Relationship between tree species composition and phenology extracted from satellite data in Swedish forests (2015)