Seminar series nr 220

The potential of support vector machine classification of land use and land cover using seasonality from MODIS satellite data

Florian Sallaba

2011
Department of Earth and Ecosystem Sciences
Physical Geography and Ecosystems Analysis
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Florian Sallaba (2011). The potential of support vector machine classification of land use and land cover using seasonality from MODIS satellite data

Master degree thesis, seminar series 220, 30 credits in Geomatics
Department of Earth and Ecosystem Sciences, Physical Geography and Ecosystems Analysis, Lund University

The potential of support vector machine classification of land use and land cover using seasonality from MODIS satellite data

---

Florian Sallaba

Master's Degree Thesis in the subject: Geomatics in Climate and Environment

Supervisor:
Lars Eklundh

Department of Earth and Ecosystem Sciences
Division of Physical Geography and Ecosystem Analysis
Lund University

Abstract

With respect to climate change it is necessary to study land use and land cover (LULC) and their changes. LULC are related directly and indirectly to climatic changes such as rising temperatures that trigger earlier onset of vegetation growing seasons (IPCC 2007). Land surface phenology refers to the seasonal patterns of variation in vegetated land surfaces over large areas using satellite data (Reed et al. 2009). General variations observed from satellite may also be referred to as seasonality (Jönsson and Eklundh 2002, 2004).

In this study, seasonality was modeled from normalized difference vegetation index time-series derived from Moderate Resolution Imaging Spectro-Radiometer (MODIS) satellite data. Seasonality data contain valuable information about vegetation dynamics of LULC, such as the maximum of a season as well as the season start and end. The specific seasonality data signatures of LULC and may improve LULC classifications compared to multi-spectral satellite data approaches.

Support vector machine classification (SVC) is a machine learning technique that does not require normal distributed input data. A normal distribution of seasonality data cannot be assumed. SVC is superior in comparison to traditional classification methods using multi-spectral satellite data (Tso and Mather 2009). Thus, it is feasible to test the potential of SVC separation of LULC using seasonality data. The most common linear and non-linear SVC methods recommended for satellite data were applied in this study.

The chosen study area is located in southern Sweden, and its LULC classes are well documented by the latest CORINE land cover 2006 data. Thus, it is a good test area for validation of the performance of seasonality parameters for LULC classification using SVC.

In this study, a SVC framework was developed and implemented that: (1) selects the most appropriate input seasonality data, (2) incorporates a direct acyclic graph for multi-classification and (3) validates the SVC outcomes with an accuracy assessment.

The results of the four class SVC show moderate performances with overall accuracies between 61 - 64% and Kappa values ranging from 0.42 – 0.45. The accuracy differences between linear and non-linear SVC are marginal. However, there are potentials to improve the developed methodology, and thus the performance of SVC on seasonality data. In addition, the seasonality data should be tested with traditional parametric classifiers (i.e. maximum likelihood) in order to achieve valuable comparisons.

Key words: Geography, Geomatics, Remote Sensing, MODIS, Support Vector Machine, Phenology, TIMESAT

Sammanfattning

Då klimatförändringar studeras är det viktigt att ha markanvändningarna och deras förändringar i åtanke. Markanvändningarna är direkt och indirekt relaterat till klimatförändringar såsom exempelvis stigande temperaturer, vilket påskyndar starten av växtsäsongen. Fjärranalysdata ger en bättre bild av storskaliga förändringar i vegetationen än vad som är möjligt att observera från jordytan. Vegetationen och dess fenologi kan observeras med satellitdata, och kallas även för årstidsvariationer.

Denna studie använder tidsserier av vegetationsindex från satellitdata för att modellera årstidsvariationer. Dessa matematiska modeller av årstidsvariationerna används sedan för att extrahera olika årstidsrelaterade parametrar som ger värdefull information om vegetationsdynamiken. Dessa årstidsrelaterade parametrar kan förbättra klassificeringen av markanvändningarna.

Studien tillämpar en ny teknik, som kallas ”support vector machine” klassificering, för att klassificera de årstidsrelaterade parametrarna. Studien fokuserade på de vanligaste linjära och olinjära ”support vector machine” tekniker som rekommenderas för satellitdata.

Det studerade området ligger i södra Sverige och dess markanvändningsklasser är sedan tidigare väldokumenterade. Det innebär att området är mycket lämpligt för att testa årstidsrelaterade parametrar genom att tillämpa ”support vector machine” klassificering.

Studien utgick från att: (1) välja de mest lämpliga årstidsrelaterade parametrarna, (2) använda en multi-klassificering, och (3) utvärdera de klassificeringsutfall med noggrannhetsbedömningar baserat från senast dokumenterad CORINE land cover 2006 data.

Resultaten påvisar en måttlig prestanda hos ”support vector machine” klassificering, där den övergripande noggrannheten landar mellan 61 till 64 %, och Kappavärdet varierar mellan 0,49 till 0,52. Skillnaderna mellan de linjära och olinjära ”support vector machine” teknikerna är marginella. Dock bör årstidsrelaterade parametrar klassificeras med traditionella klassificeringsmetoder (t.ex. maximum likelihood-metoden), och jämföras med ”support vector machine” klassificeringsresultaten.

Nyckelord: Geografi, Geomatik, Fjärranalys, MODIS, Support Vector Machine, Fenologi, TIMESAT

Index of Contents

Index of Figures

Index of Tables

Abbreviations

AG	- Agriculture
AVHRR	- Advanced Very High Resolution Radiometer
ANN	- Artificial neural network
CLC	- CORINE land cover
CORINE	- Co-ordination of Information on the Environment
DAG	- Direct acyclic graph
EEA	- European Environment Agency
EOS	- Earth Observing System
EROS	- Earth Resources Observation and Science
EVI	- Enhanced vegetation index
FL	- Forestland
GL	- Grassland
GPP	- Gross primary production
LAI	- Leaf area index
LP DAAC	- Land Processes Distributed Active Archive Center
LSF	- Least square fitting
LSP	- Land surface phenology
LULC	- Land use and land cover
MLC	- Maximum likelihood classification
MODIS	- Moderate Resolution Imaging Spectro-Radiometer
NDVI	- Normalized difference vegetation index
NOAA	- National Oceanic and Atmospheric Administration
PAR	- Photosynthetically active radiation
RBF	- Radial basis function
SAVI	- Soil adjusted vegetation index
SVC	- Support vector machine classification
SVMs	- Support vector machines
SVM-C	- Support vector machine multi-classification
UN	- Unknown
UR	- Urban
USGS	- U.S. Geological Survey
VI	- Vegetation indices
WDRVI	- Wide dynamic range vegetation index
WDVI	- Weighted difference vegetation index

Preface

This master degree thesis is written for awarding degree in the master program of Geomatics in Climate and Environment at the Division of Physical Geography and Ecosystem Analysis - Department of Earth and Ecosystem Sciences at Lund University. It deals with a land use and land cover (LULC) classification study of an area in southern Sweden using support vector machine classification on seasonality data obtained from an Earth observation satellite.

Acknowledgments

I am grateful for the supervision of Lars Eklundh, Division of Physical Geography and Ecosystem Analysis - Department of Earth and Ecosystem Sciences at Lund University, Sweden. I would like to thank Margareta Hellström for the distribution of the MODIS satellite data and for the discussions in the beginning of the master thesis project.

I am very thankful to Harry Lankreijer for the administrative help in order to study at Lund University.

1. Introduction

With respect to climate change phenomena it is necessary to study land use and land cover (LULC) and their changes. LULC are directly and indirectly related to climatic changes such as rising temperatures that trigger earlier onset of vegetation growing season (IPCC 2007). Growing atmospheric CO_2 concentrations lead to an increased biomass production due to CO_2 fertilization (Cramer et al. 2001). The effects of global warming have influences on ecosystems and their vegetation types that may lead to changes in LULC classes. Vegetation adapt well to the seasonality of their environment such as earlier spring start and extended autumn, therefore the seasonal development of the vegetation may change and provide evidences that ecosystems are being influenced by climate change phenomena (Barr et al. 2009).

In northern Europe, vegetation might develop different green biomass (foliage) accumulation patterns, and thus the length of a growing season increases. Earlier starts of the growing seasons (i.e. spring and summer) have been observed (Menzel et al. 2006, Jönsson et al. 2010). Therefore, it is useful to study the response of the vegetation to seasonal climatic cycles in irradiance, temperature and precipitation. This response is called phenology and represents an essential land surface parameter in atmospheric and climate models (Jönsson and Eklundh 2002, 2004).

Over the last decades, satellite data have become a major data source to monitor changes of our environment. Remotely sensed data can provide a better picture of large-scale changes that cannot be observed from Earth's surface.

Land surface phenology (LSP) is the study of vegetation phenology and refers to the seasonal pattern of variation in vegetated land surfaces over larger areas using satellite data (Reed et al. 2009).

Advantages of using satellite data for phenology applications are the ability to capture the continuous development of phenology patterns across the landscape and to create time-series for long term observations (Reed et al. 2009). The satellite derived annual seasonal patterns are referred as seasonality (Jönsson and Eklundh 2002, 2004). Seasonality can be modeled from vegetation indices (VI) computed from satellite data with high temporal resolution such as Moderate Resolution Imaging Spectro-Radiometer (MODIS). VIs such as normalized difference vegetation index (NDVI) time-series, based on MODIS data, have been successfully incorporated to monitor seasonality and phenological events (Jönsson et al. 2010). There is a strong relationship between biomass and NDVI since the latter is a measure

of vegetation greenness and captures the seasonal development of foliage (Jönsson and Eklundh 2002, 2004).

Modeled seasonality can be used to extract seasonality characteristics such as start, end, length, integral and amplitude of a season (Jönsson and Eklundh 2002, 2004). Seasonality characteristics might be useful for improving the separation between LULC classes. In most of the LULC classification studies using MODIS data multi-spectral, multi-spectral transformed data and VI time-series have been applied to separate LULC classes (Gonçalves et al. 2005, Gu et al. 2008, Gu et al. 2010).

So far, the application of seasonality characteristics to classify LULC types has not been tested extensively. Hüttich et al. (2009) successfully applied MODIS and high resolution satellite time-series in combination with in-situ botanical survey data to model seasonality data and to classify the data into vegetation types in dry Savanna ecosystems.

Following the discussion, it is assumed that seasonality data contain valuable information and improve LULC classification approaches. For example, LULC classes such as Forestland, Grassland or cropped Agricultural land have similar spectral signatures (see Figure 1), and a classification of the multi-spectral data may lead to misclassifications between these LULC classes. However, the seasonality parameters of the mentioned LULC classes are theoretically different (see Figure 2), since the LULC classes have different vegetation dynamics or seasonal developments. For example the seasonal development of Forestland shows an earlier onset of foliage growth, higher amplitudes and larger integrals of seasonal development compared to Grassland. Agricultural land has a shorter season length due to harvesting in relation to the other Grassland and Forestland. This seasonality information is valuable and might increase the performance of LULC classification approaches.

However, the statistical distribution of modeled seasonality characteristics is not known (Eklundh 2010, personal communication), thus a normal distribution of the data cannot be assumed, and a feasible classification technique needs to be chosen. In such a case, machine learning algorithms seem to be a reasonable classification technique, since they do not assume normally distributed data (Richards and Jia 2006, Tso and Mather 2009).

Support vector machine classification (SVC) is a recent technique for classifying data that has promising capabilities for improving classification performance of satellite data. It is a non-parametric (Richards and Jia 2006) separation method developed in the field of machine learning theory.

Several studies of multi-spectral data have shown that SVC is superior to traditional classification methods (Pal and Mather 2003, Pal and Mather 2005, Tso and Mather 2009). Thus, it is feasible to test the performance of SVC using seasonality parameters to detect LULC.

The chosen study area is located in southern Sweden, since its LULC classes are well documented by CORINE land cover 2006 – version 13 (CLC 2006). Thus, it is an excellent test area to validate the performance of seasonality parameters for LULC classification using SVC. Gonçalves et al. (2005) employed successfully CORINE land cover 2000 as validation data for SVC applied on MODIS satellite imagery.

The outcomes of this study may be useful for further applications of SVC in large scale LULC mapping and LULC change monitoring using seasonality parameters, for studies of climate change effects.

2. Objectives

The overall objective of this study is to test the potential of SVC to classify seasonality data that was modeled from MODIS satellite time-series. In order to achieve this overall objective it is necessary to develop and implement a methodology that considers the following specific objectives:

- The development of a method that determines the most suitable seasonality characteristics to distinguish between the LULC classes of the chosen study area in southern Sweden. It is assumed that the more seasonality characteristics are used the more complex the classification procedure becomes. Finding a trade-off between computational expenses, complexity and the accuracy of the classification outcomes is an important part of this study.

- The application of three commonly recommended SVC techniques (linear, polynomial function kernel and Gaussian radial basis kernel function) to classify the seasonality parameters in LULC classes.

- The Validation of the classifications results with CORINE land cover 2006 data to determine the best performing of the three commonly used SVC techniques.

3. Theory

3.1 Remote Sensing

3.1.1 Introduction to Remote Sensing

Earth observation is the science of collecting information about the surface of the Earth using satellite platforms. Remote sensing in terms of Earth observation involves two processes that are data acquisition and data analysis (Lillesand et al. 2004).

Data acquisition describes measuring and recording of reflected or emitted energy. The observation takes place on a sensor mounted on an aircraft or spacecraft platform. The sensor records objects of interest, such as clouds or landscapes, beneath the platform. There are two systems of sensing remotely: passive and active way (Lillesand et al. 2004).

The passive way is based on a sensor that records reflected solar radiation from the Earth surface, terrestrial radiation originated from the Earth itself as a radiator in terms of its own temperature. Passive remote sensing can take place in several wavelength bands of the electromagnetic spectrum such as the visible light, infrared and thermal electromagnetic radiation. The active way involves laser and radar devices that can be used to send artificial energy to the Earth and the scattered energy from the surface can be measured (Richards and Jia 2006). Remotely sensed data within a specific electromagnetic bandwidth are called bands or channels and are depending on the design of the corresponding sensors. Kramer (2002) gives a detailed description of all commercial used sensors and their design.

The recorded data are organized in images and give information about reflected or emitted radiance in the chosen spectral wavelength bands (e.g. red light) according to the sensor design. Each surface or object has a unique spectral reflectance pattern, called signature, as shown in Figure 1 and enables to detect the surface or object. The data analysis process contains in brief the processing, interpretation, analysis and mapping of the remotely sensed data (Lillesand et al. 2004).

The selection of a satellite sensor depends on the aim of a study, since sensors have different characteristics such as geometric resolution, temporal resolution and the placement of the sensor wavelength bands in the electromagnetic spectrum (Kramer 2002). This study involves the analysis of data from a passive remote sensing sensor called Moderate Resolution Imaging Spectro-Radiometer (MODIS). MODIS generates data of high temporal resolution, which enables observations of seasonal vegetation dynamics.

3.1.2 Moderate Resolution Imaging Spectro-Radiometer

MODIS is a sensor flown on two satellites of the Earth Observing System (EOS) named EOS-AM Earth and EOS-AM Water programs and is suitable for this study since MODIS records data on a high temporal resolution.

EOS-AM Terra was launched on December 18, 1999 and the mission objectives are the characterization of the terrestrial ecosystems, land use and land cover, soil moisture, terrestrial energy, oceans primary productivity, characteristics of the cloud systems, aerosol and radiative balance. EOS-AM Aqua was launched on May 4, 2002 and its objectives are studies of the water cycle and the radiative energy flux. Both satellites are in a near-polar, sun-synchronous orbit at 705 km altitude and an inclination of 98.2°. EOS-AM Terra has a 10:30 A.M. equatorial crossing time in descending path motion and EOS-AM has a 1:30 P.M equatorial crossing time in ascending path motion (Kramer 2002, Lillesand et al. 2004, Gomarasca 2009).

MODIS has a very high temporal resolution and is designed to measure parameters related to biological and physical processes at global scale every 1 or 2 days (Gomarasca 2009).

MODIS has a total field of view of +- 55° providing a swath width of 2330 km. MODIS measures with 36 spectral bands between 0.405 µm and 14.385 µm at three spatial resolutions (250m, 500m and 1000m) depending on wavelength. In Table 1 the characterizations of the bands used in this study are shown. For further information about MODIS spectral bands please refer to Lillesand et al. (2004).

Table 1 Used MODIS Spectral Bands based on Lillesand et al. (2004)

Band no.	Bandwidth [nm]	Spectral light	Spatial resolution [m]
1	620-670	Red	250
2	841-876	Near Infrared	250

MODIS data have improved geometric rectification and radiometric calibration compared to the data of previous satellite sensors, such as the Advanced Very High Resolution Radiometer (AVHRR) sensor on the National Oceanic and Atmospheric Administration (NOAA) satellites. All 36 MODIS bands have a minimum band-to-band registration of 0.1 pixel. The 20 reflected solar bands are absolutely calibrated radiometrically with a minimum accuracy of 5% (Lillesand et al. 2004).

MODIS data are distributed in the standard tile format and in sinusoidal equal area projection. For further information about MODIS spectral and ancillary data products, tile information and sinusoidal projection please refer to the MODIS product website (LP DAAC 2010).

For this study was used the 8-day 250m ground resolution surface reflectance product (MOD09Q1) and additional quality data from the MODIS sensor on the EOS-AM Terra mission. MOD09Q1 provides data from MODIS bands 1 and 2 at 250m-ground resolution (see Table 1) on an 8-day basis.

The quality data are based on the quality flags of band 1, band 2, the cloud state flag from the 8-day 500m ground resolution surface reflectance (MOD09A1) and quality information for the reflectance measurements (MODLAND QA bits**)**. The bands of MOD09Q1 and MODIS quality data offer the possibility to observe vegetation dynamics and model seasonality with the help of vegetation indices. For more information about MODIS and MOD09Q1 please refer to MODIS product table (LP-DAAC 2011).

3.1.3 Vegetation Indices

Vegetation indices (VIs) are related to cover and status of vegetation and provide information on biomass productivity and its health condition. A direct correlation with leaf chlorophyll content and leaf area index (LAI) exists, which varies in relation to vegetation cycle and phenology. The normalized difference vegetation index (NDVI) bases on measured spectral reflectance. It relates spectral absorption of chlorophyll in the red light spectrum with reflection characteristics of vegetation in near infrared light spectrum (Gomarasca 2009). In Figure 1 spectral signatures of selected surfaces are illustrated, such as deciduous leaves (bright green) and grass (dark green).

In **Equation 1** is shown how NDVI is calculated - based on Gomarasca (2009):

$$NDVI = \frac{\rho NIR - \rho R}{\rho NIR + \rho R} \qquad \textbf{Equation 1}$$

where: ρNIR = reflectance in the near infrared band*;* ρR = reflectance in the red band.

This ratio leads to numbers in an interval between -1 and +1. Vegetated areas will lead to high values of NDVI due to relatively high reflectance of vegetation in the near infrared electromagnetic spectrum and low in the visible spectrum.

Negative NDVI values can be caused by snow, water and clouds that have higher reflectance in the visible spectrum and less reflectance in the near infrared electromagnetic spectrum (compare Figure 1).

NDVI values near zero are based on the similar reflectance of bare soil and rock in the red visible spectrum and near infrared electromagnetic spectrum (Lillesand et al. 2004).

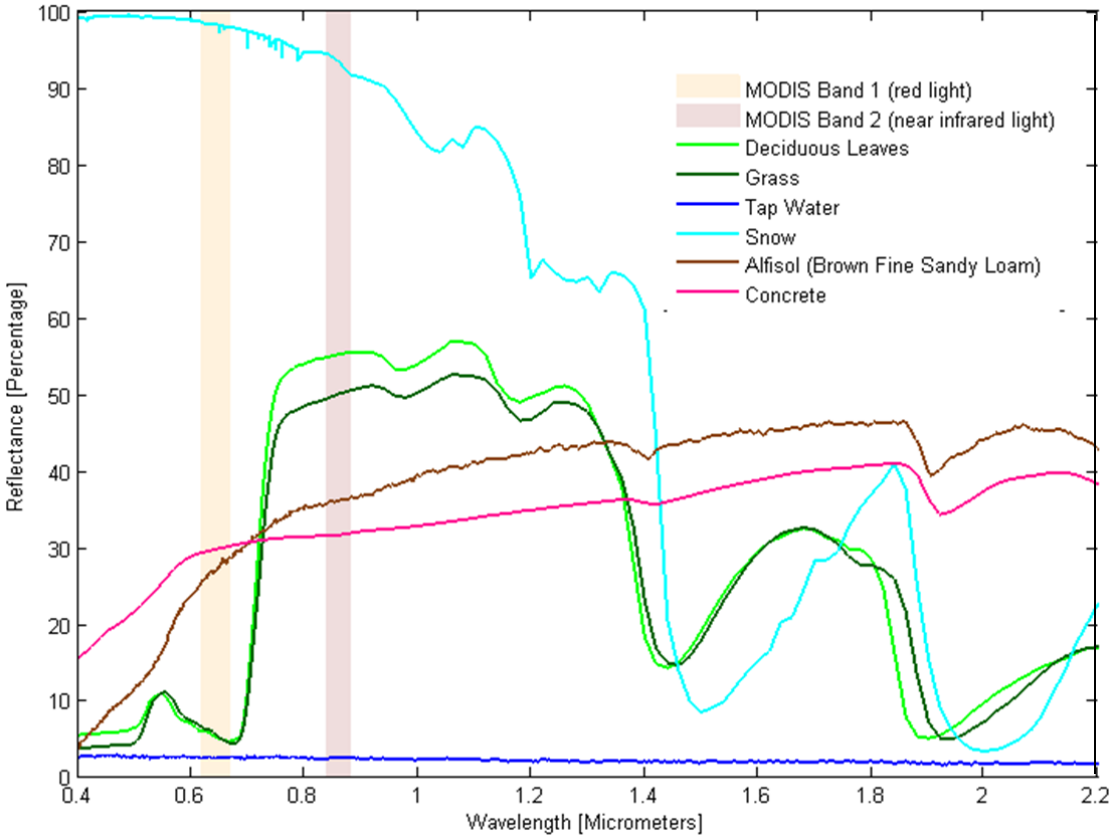

Figure 1 Spectral signatures of selected surfaces and bandwidths of the MODIS channels 1 and 2. Spectral data were obtained from Baldrige et al. (2009).

Advantages of NDVI are a reduced sensitivity, compared to single bands, to background signals such as soil if vegetation does not cover completely the surface; to geometry of view because of sensor angles acquisition; atmospheric effects and to the sun angle. However, a certain sensitivity of NDVI towards the mentioned phenomena exists (Gomarasca 2009).

In this study, MODIS NDVI time-series are used on an 8-day basis and 250m-ground resolution. The NDVI time-series describe the seasonal vegetation development in respect of green biomass and are appropriate to model the seasonality of the chosen study area.

3.2 Seasonality

3.2.1 Introduction

Vegetation index time-series can be used to obtain information on seasonal vegetation development, called seasonality. This information helps to analyze the functional and structural characteristics of the global and regional land cover. Long vegetation indices time-series, such as NDVI, can present information on shifts in the spatial distribution of bio-climatic zones or land-use changes (Jönsson and Eklundh 2002, 2004).

The exploration of NDVI time-series is helpful to study seasonal developments of vegetation over a given period of time. Thus, it is possible to describe how vegetation will produce and lose foliage (green biomass) over the given time period. In Figure 2 seasonal developments of LULC are shown notionally and the quantification of the green biomass measured with a NDVI time-series (period of three years). For this study, the most interesting vegetation types are Conifers-Boreal forest (red line) and Grass-Crops (dark-green line) as shown in Figure 2. There are differences in the seasonal development between Conifers-Boreal forest and Grass-Crops. Conifers-Boreal forest shows an earlier onset of green biomass growth, higher amplitudes and larger integrals of seasonal development compared to Grassland. These differences are valuable to improve the classification of LULC classes.

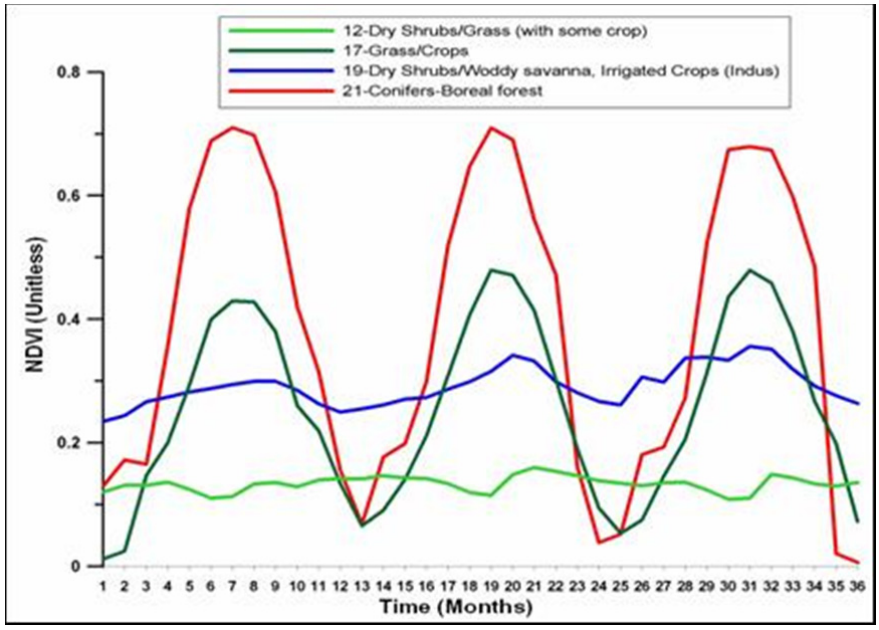

Figure 2 Seasonal development of selected LULC based on NDVI time-series from Thenkabail et al. (2009)

A common seasonal development of a deciduous forest, in southern Sweden, can be described as follows: in the beginning of the season, in springtime, a strong increase of foliage (i.e. leaf- growing and budburst) takes place, this growth peaks in summertime; and

then it starts to decrease slightly until the end of summer is reached. In autumn, a strong decrease of green-biomass occurs due to falling leaves until the end of season is reached. During wintertime no growth or loss of foliage is supposed to occur. The patterns of different vegetation types are diverse, such as crops having a strong and rapid decrease of green-biomass due to harvesting actions.

In Figure 3 the pattern of seasonal developments of vegetation based on MODIS NDVI time-series (period of 5 years) is visualized. There are large differences between Figure 2 and Figure 3. The raw NDVI time-series in Figure 3 contains noise which leads to dips within the seasonal development. The dips during spring and summer are most likely based on cloud coverage. During winter-time the peaks are probably generated by snow coverage and noise due to a weak signal. Large solar zenith angles (i.e. low solar illumination) and long atmospheric path lengths at the high latitudes lead to weak signals and thus erroneous data during the dark season.

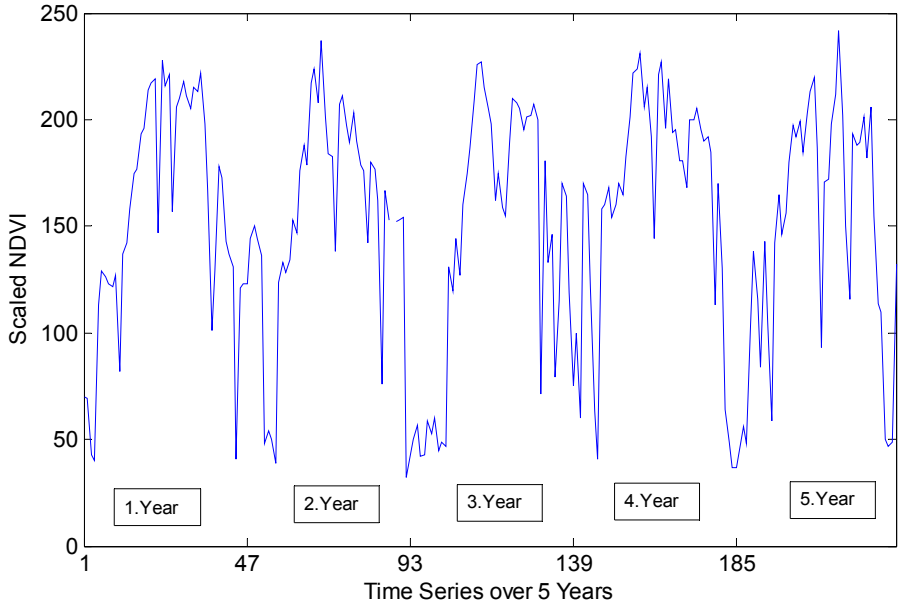

Figure 3 NDVI time-series of a pixel position covered with vegetation over a period of 5 years.

NDVI data are organized in binary images, in other words two-dimensional arrays containing pixels representing NDVI values of defined areas i.e. 250 x 250m. NDVI time-series can be thought as a stack of several NDVI images recorded at different points in time (i.e. based on 8-day cycle). The amount of NDVI images in a time-series is represented by N and a pixel position is denoted (j,k). Each pixel in an image will be overlaid by a pixel of a new image at the same pixel position (j,k) covering the same ground area. The number of images is dependent on the time-series length. In order to achieve a time-series of the same ground

area, the stack is pinpointed at the corresponding pixel position (j,k) and thus a time-series of that pixel position can be extracted (Jönsson and Eklundh 2002, 2004).

NDVI time-series, as shown in Figure 3, are describing annual seasonal developments of the study area, however noise is present. The noise may lead to erroneous conclusions about seasonal development, and thus the aim of this study to classify a pixel to a LULC class according to its seasonal patterns. Hence, it is necessary to process these raw NDVI time-series with modeling approaches and incorporation of ancillary data in order to deal with noise.

3.2.2 Modeling Seasonality

The seasonality modeling and the extraction of seasonality parameters was done in TIMESAT (Jönsson and Eklundh 2002, 2004) ,Version 3.0. TIMESAT has been designed for fitting mathematical functions to remotely sensed time-series to outline seasonal vegetation developments (Jönsson et al. 2010).

This section is limited to an explanation of the most important modeling parameters in TIMESAT and the chosen Saviztky-Golay modeling method. The entire methodology of TIMESAT is well described in the literature (Jönsson and Eklundh 2002, 2004).

TIMESAT incorporates least square fitting (LSF) to the upper envelope of vegetation indices time-series. A time-series of vegetation index values is inserted to a model function containing basic functions and parameters. The best parameter values can be achieved by solving a system of normal equations. The system of normal equations incorporates weights for each observation point in the time-series. The weight values can be information about cloud coverage and determine the uncertainty of NDVI values. In other words, weights are qualitative information about pixel quality, i.e. cloud coverage, from ancillary data in order to improve a fitting.

In case of cloud coverage data, a value of 0 means cloudy conditions, 0.5 means mixed conditions and 1 indicates clear sky conditions. Accordingly, small weight values have less influence on the fit and large weights high influence on the fit respectively. A weight value of 0 will not influence the fitting at all. In case of missing ancillary data, the weights may be set to 0 (Jönsson and Eklundh 2002, 2004), since no information about cloud coverage are available.

The NDVI time-series in Figure 3 contains several outliers; their influence on the fitting procedure can be weakened or they can be taken away by applying small or zero weight

values to the corresponding time-series points. In Figure 4 it is visualized how weights could be set in a NDVI time-series, and the bigger the red circles the higher their weight values. Figure 4 shows that outliers in the summer seasons have low weight values or zero weights.

The remaining positive or negative outliers in a NDVI time-series can be removed with certain methods. This is important since remaining outliers can seriously falsify final function fits (Jönsson and Eklundh 2002, 2004). In TIMESAT a median method is available that uses a moving window in order to compare neighboring values of the tested point. A time-series point will be removed if the neighboring values are considerably different. For further information please refer to the authors (Jönsson and Eklundh 2002, 2004).

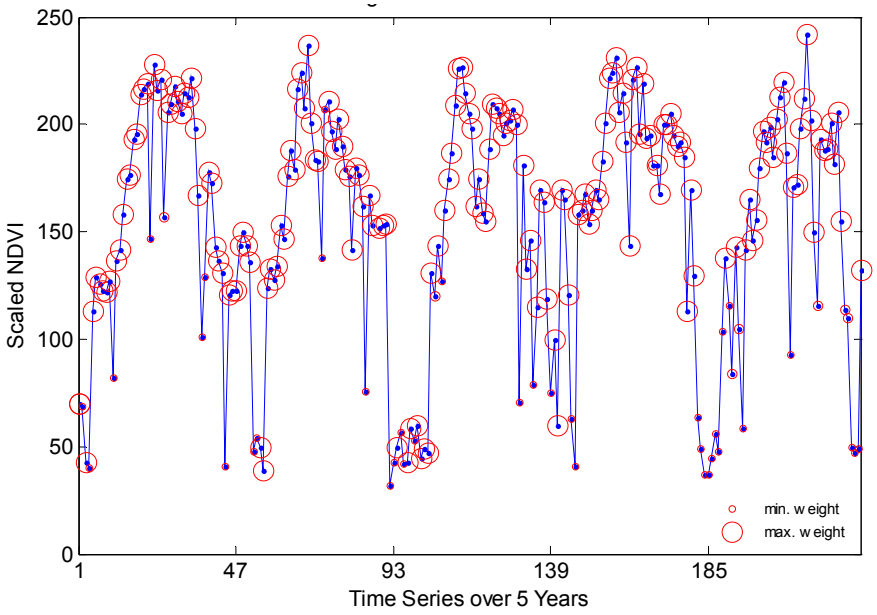

Figure 4 Weighting of a NDVI time-series. The red circles point out the applied weights depending on their circumferences. The bigger the circles the higher their weighting.

Most noise in NDVI, based on remotely sensed data, is negatively biased (Jönsson and Eklundh 2002, 2004). In order to deal with this in the fitting procedure an adaption to the upper envelope is employed in TIMESAT. The parameters of the model function can be determined in two or more steps. During the first step, the system of normal equations with weights is solved in order to determine the parameters of the model function. Data values beneath the achieved model function are considered as less important. In the second step, the system of normal equations is solved with the weights of the low data values decreased by a factor (Jönsson and Eklundh 2002, 2004). Thus, a model function that is adapted to the upper envelope of the data can be obtained.

As mentioned above in this study is used the adaptive Savitzky-Golay technique**.** The Savitzky-Golay uses local polynomial functions to filter data and suppress disturbances, such as noise, by replacing each data value with a linear combination of nearby values in a moving window (Jönsson and Eklundh 2002, 2004). An adaptive Savitzky-Golay filter approximates a value in a moving window computed from a least squares fit to a polynomial. In other words, a quadratic polynomial is used to fit each data value $y_i, i = 1,...,N$ to all $2n+1$ points in the moving window, where *n* denotes the window width or size. Thus, the value y_i will be replaced with the value of the polynomial at the corresponding position (Jönsson and Eklundh 2002, 2004).

Advantages of the adaptive Savitzky-Golay technique are that it preserves the area, width, height and mean position of a seasonal peak. The moving window size sets the degree of smoothing; however, it can also smooth out unintentionally rapid changes that are not considered as noise. Therefore, a global window size *n* is applied to the entire time-series and scans the data for a large increase or decrease in an interval around a data point y_i. In case of a rapid change of values (i.e. decrease or increase) around a certain data point, this data point will be linked to a smaller window size. Then, the filtering will be repeated with new locally adapted window sizes (Jönsson and Eklundh 2002, 2004).

3.2.3 Seasonality Characteristics

The modeled function of seasonal development of the observed ground surface enables the extraction of seasonality characteristics that give information of the annual seasonal development.

The information of the seasonality characteristics are described as numeric (real) values. The seasonality characteristics are described in Table 2 and visualized in Figure 5. Each of the annual seasonality characteristics can be stored as a value in a two dimensional binary image like traditional spectral satellite data. Thus, the seasonality characteristics can be incorporated into a classification, analysis or mapping approaches.

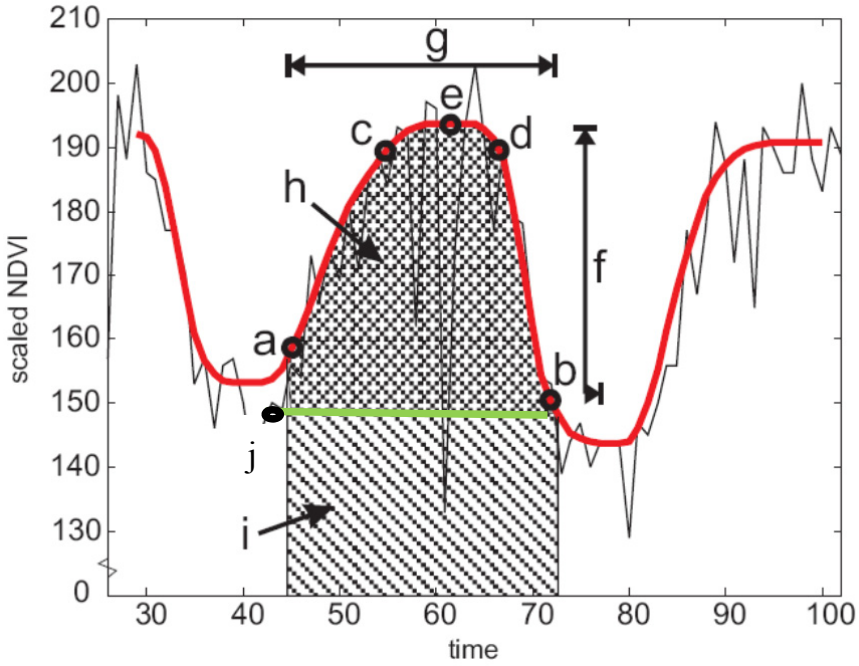

Figure 5 Modeled function of seasonal development based on NDVI time-series from Eklundh and Jönsson (2010). Selection of seasonality characteristic (referenced with small letters from a to j) that can be extracted from the function.

A model function (red line) is visualized in Figure 5. It has been fitted to a NDVI time-series (black line). The seasonality characteristics that can be extracted from the function are referenced with small letters from a to j. All seasonality parameters are explained briefly in Table 2.

Table 2 Seasonality characteristics from Eklundh and Jönsson (2010)

Name	Description of the seasonality characteristic
Start of season	Point in time-series where the rising (left) part of the curve has increased to a user-defined level measured from the left minimum level (point **a** in Figure 5)
End of season	Point in time-series where the sloping (right) part of the curve has decreased to a user-defined level measured from the right minimum level (point **b** in Figure 5)
Season length	Is the length (i.e. time) from the start to the end of a season (object **g** in Figure 5)
Base level	Represents the average of the left and right minimum value (point **j** in Figure 5)
Middle of season	Is estimated by considering the mean value of the times for which, the left edge has increased up to the 80% (point **c** in Figure 5) and the right edge has decreased to the 80% level (point **d** in Figure 5)
Maximum of season	The largest value for the fitted function during the season (point **e** in Figure 5)
Seasonal amplitude	This is the difference between the maximum value and the base value (object **f** in Figure 5)
Small seasonal integral	Integral of the difference between the function describing the season and the base level from the season start to season end (object **h** in Figure 5)
Large seasonal integral	Integral of the function describing the season from the season start to season end (object **i** in Figure 5)
Left derivate	Rate of increase at the start of the season; it is estimated as the ratio of the difference between the left 20% and 80% levels and the corresponding time difference
Right derivate	Rate of decrease at the end of the season; it is estimated as the absolute value (positive) of the ratio of the difference between the left 20% and 80% levels and the corresponding time difference
Derivate ratio	It is calculated as the ratio between the left and right derivate; it gives the skewness of the distribution

3.3 Support Vector Machine Classification

3.3.1 Introduction to Support Vector Machine Classification

The aim of a classification method, in the field of LULC studies, is to separate a set of data points (i.e. seasonality characteristic values) according to their patterns into prior defined LULC classes. Support vector machine classification (SVC) is a non–parametric classification method (Richards and Jia 2006) belonging to the field of machine learning theory. Support vector machines (SVMs) were formulated as a binary classification method in the late 1970s by Vapnik and coworkers and were further developed in the 1990s (Vapnik 1998). The application of SVC consists of two parts. Firstly, the training and validation of SVC classifiers on a subset of the data that is known and well documented (i.e. training dataset). Secondly, the actual classification by applying the previously trained SVM classifiers to the rest of the data that are unknown (i.e. unknown dataset) to the analyst and have not been applied in the training stage.

SVC are gaining increasing attention in research areas such as remote sensing and bioinformatics during the recent years due to their ability to minimize classification errors, using the structural risk minimization concept, and superior generalization characteristics, when solving classification problems. Non-parametric classifiers, such as SVC, do not require prior knowledge of the statistical distribution of a dataset to be classified (Gunn 1998).

According to Tso and Mather (2009) the structural risk minimization is an inductive concept in machine learning theory and decreases the probability of misclassifying a previously unknown data point taken randomly from a fixed but unknown probability distribution.
Generalization in terms of statistical machine learning means the construction of an elastic classifier that is not too fixed on the used training data and can be applied properly to the unknown part of the data. SVC provide proper mathematical foundations to validate how well a classifier generalizes on unknown data (Dixon and Candade 2008) in order to avoid an over-fitting of the classifier due to the training data. For further information about structural risk minimization concept and generalization please refer to the literature (Gunn 1998, Vapnik 1998).

Advantages of SVC are improved classification accuracies (Pal and Mather 2005, Foody and Mathur 2006, Dixon and Candade 2008) in contrast to traditional parametric classifiers such as supervised maximum likelihood classification (MLC) (Tso and Mather 2009) and

unsupervised k-means clustering (MacQueen 1967, Miller and Han 2001). SVC is less sensitive to over-fitting of a classifier during the training stage compared to non-parametric artificial neural network (ANN) classification (Tso and Mather 2009). Additionally, smaller amounts of training data are needed and the data do not have to be normal distributed like it is necessary for parametric classifiers (e.g. MLC). Bagan et al. (2005) state that the multivariate normal model of MLC is not as effective as machine learning techniques in LULC classifications of satellite imagery (e.g. spectral MODIS time-series).

Classification techniques, such as SVC, are based on the pattern of the input data in the feature space. A feature space is an abstract space and has specific characteristics such as data distribution and dimensionality. For example, a two-dimensional feature space contains data points that are described by two attributes (features). The attribute values of a point ($feature_1$, $feature_2$) "act as coordinates" and determine the position of the point in the two-dimensional feature space using both feature space axes. Thus, all samples of a dataset are mapped into the feature space and create a specific pattern (distribution) that can be used to separate the samples into classes. A feature space can be *n*-dimensional depending on the number of features ($feature_1$,…, $feature_n$) that are describing each sample of a dataset. An n-dimensional feature space consists of *n*-feature axes.

In this study, the classification problem is to separate seasonality characteristic values into LULC classes. These seasonality characteristics (features) build up the feature space. Figure 6 shows a two-dimensional feature space containing well spread training data points belonging either to class A or class B. The data points in a feature space are referred as feature vectors[1], since a vector represents *n*-features with its position in the feature space. SVC training aims to construct an optimal separating hyperplane (see Figure 6 hyperplane *b*) between the two classes of feature vectors.

For the construction of the hyperplane the most representative training feature vectors (support vectors) are used according to their distribution in the feature space (Dixon and Candade 2008), shown as solid filled squares and circles in Figure 6. The support vectors are laying at the edge of the class distributions and between the class centroids.

[1] In the following chapters and sections the term feature vector is replaced by point or sample, since it facilitates the differentiation between support vectors and feature vectors (points & samples), and avoids confusion of the reader. Even though the correct terminology of a data sample is feature vector in the feature space.

Support vectors are important to determine the position (i.e. orientation and slope) of an optimal hyperplane. All remaining training samples are not contributing to the computation of the hyperplane location and are omitted (Foody and Mathur 2006).

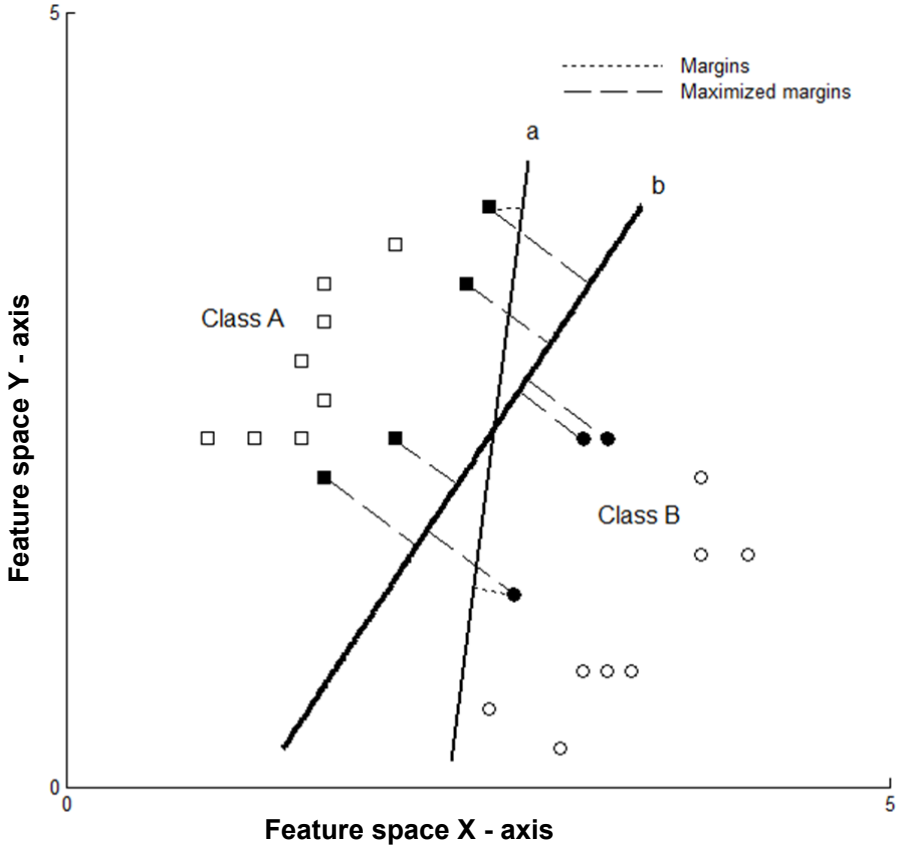

Figure 6 Separating hyperplanes of a classification problem and their margins in a two-dimensional feature space. The solid squares and circles are support vectors.

An optimal hyperplane has to be built so that the distances, called margins, from the hyperplane to closest training feature vectors (support vectors) in both class A and class B are as large as possible (Tso and Mather 2009). In other words, the determination of an optimal separating hyperplane is subjected to the constraint that the margins of separation between class A and class B feature vectors are maximized. Hyperplane *a* (see- Figure 6) is not optimal separating the feature space since its margins (dotted lines) are not maximized, whereas hyperplane *b* has maximized margins (dashed lines) to the support vectors of both classes.

However, to find the optimal separating hyperplane involves a mathematical optimization problem because there are many solutions to separate the feature space with a hyperplane. In case of SVC, a quadratic optimization is applied to find the optimal hyperplane (Tso and Mather 2009). The quadratic optimization is subjected to several constraints depending on the

chosen SVC strategy. Please refer to the literature for further information about quadratic optimization (Bazaraa et al. 1993). The possible solutions of separating hyperplanes are defined by an objective function that is related to the optimization problem. An objective function is a function of costs defined by the hyperplane properties and its constraints (i.e. maximized margins between the classes).

Let us constrain the classification problem in Figure 6 to six possible solutions; its quadratic optimization might lead to an objective function as shown simplified in Figure 7.

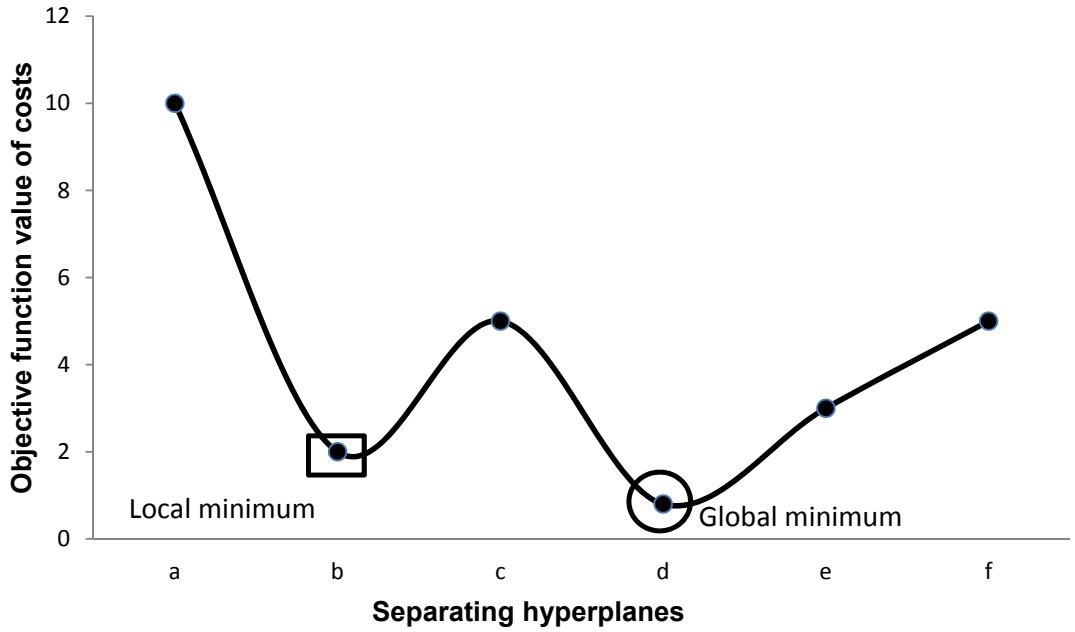

Figure 7 Simplified objective function of the constrained optimization problem in Figure 6 showing six hyperplanes. One hyperplane represents a local and one a global minimum. All separating hyperplanes in a two-dimensional feature space are visualized in Figure 8.

The best solution in the objective function has the lowest value of costs and represents the global minimum of the classification problem. However, there are also local minima that represent the minimum values in a certain neighborhood of the objective function. Some classification methods such as k-means clustering are limited to find local minima (MacQueen 1967, Miller and Han 2001). The global minimum in the objective function with the lowest cost of hyperplane properties is used to construct the optimal hyperplane.

The objective function in Figure 7 contains six possible hyperplanes (*a, b, c, d, e* and *f*) that can be mapped into the feature space of the classification problem as visualized in Figure 8. The global minimum of the objective function is *d* and therefore the best possible separating hyperplane in Figure 8. Hyperplane *d* separates the two classes with the maximal margins and the other hyperplanes do not have as much maximized margins as hyperplane *d* to both sides.

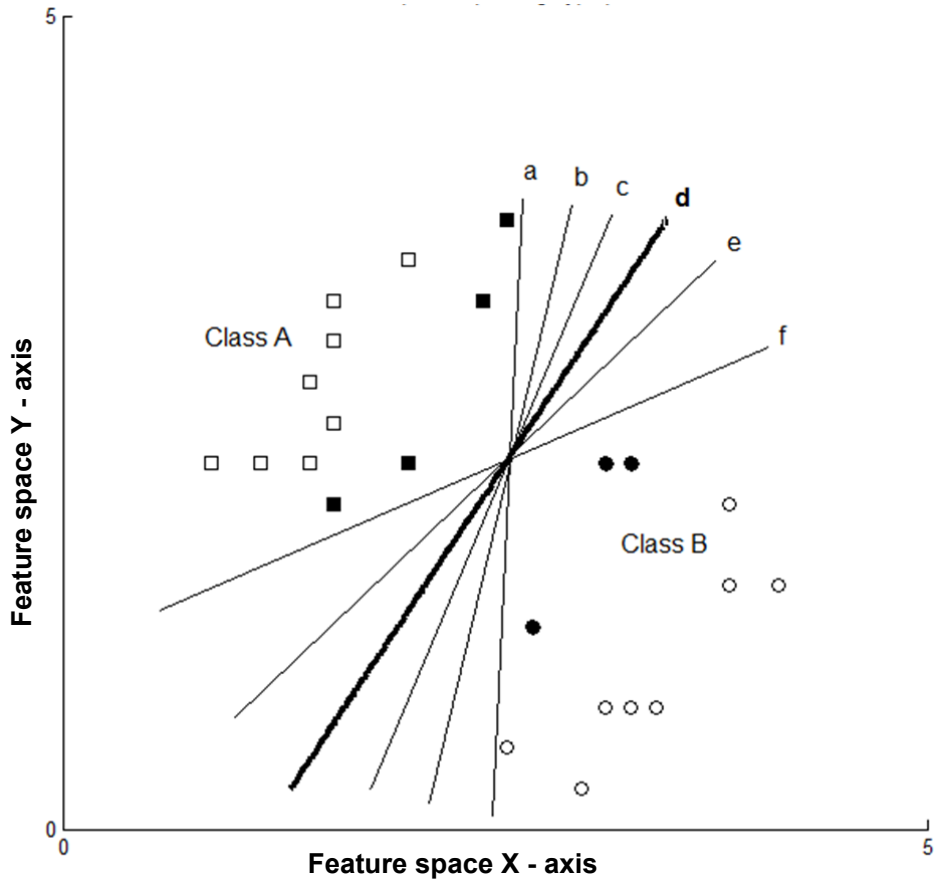

Figure 8 Simplified mapping of the objective function in Figure 7 into the feature space. The dataset is separated in two classes with an optimal separating hyperplane with maximized margins. Hyperplane *d* divides the classes with maximal margins. The solid squares and circles are support vectors.

The trained final classifier of this classification problem is a decision function ($f(x)$) that contains information about the position and orientation of the separating hyperplane *d* in the feature space. Then the final decision function ($f(x)$) is used to test all unknown data points according to their location in the feature space and their relation to the hyperplane, and classifies them into class A or class B.

3.3.2 Linear Support Vector Classification

Linear SVC can handle hard margin (separable) and soft margin (non-separable) cases. In the following will be described the basic mathematical concepts behind linear SVC and visualized with several figures. However, a more thorough description of the derivation and formulation of hard and soft margin linear classification can be found in the literature (Vapnik 1998, Foody and Mathur 2006, Dixon and Candade 2008, Izenman 2008, Tso and Mather 2009).

Hard margin linear classification

Hard margin linear classifiers assume that a dataset of points can be exact linearly separated in two classes as shown in Figure 6 and Figure 8. A hard margin linear classifier is trained with a set of training data points to build an optimal hyperplane. Each point is either a member of class A or class B.

According to Kavzoglu and Colkesen (2009) a training dataset contains k number of data points represented by $\{x_i, y_i\}, i = 1, \dots, k$ where x_i is an element of an n-dimensional feature space $[x_i \in R^n]$ and y_i is the known class label of a point x_i that can either have label -1 (corresponds to class A) or label +1 (corresponds to class B) $[y_i \in \{-1, +1\}]$.

The optimal hyperplane with maximized margins is represented by a hyperplane decision function as shown in Figure 9 and written in **Equation 2** (Kavzoglu and Colkesen 2009):

$$\boldsymbol{w} \times \boldsymbol{x_i} + \boldsymbol{b} = \mathbf{0} \qquad \textbf{Equation 2}$$

where: $\boldsymbol{w}$ is a vector that determines the orientation of the hyperplane, x_i is a point lying on the hyperplane and $\boldsymbol{b}$ is the bias that is the distance of the hyperplane from the feature space origin. Figure 9 shows the basic elements of hyperplane decision function (i.e. $\boldsymbol{w}$, $x_i$and $\boldsymbol{b}$)

For each of the two classes a separating hyperplane can be defined as inequalities, see **Equation 3** and **Equation 4** (Kavzoglu and Colkesen 2009):

$$\boldsymbol{w} \times x_i + \boldsymbol{b} \leq -1 \quad \text{for all y = -1 (class A)} \qquad \textbf{Equation 3}$$

$$\boldsymbol{w} \times x_i + \boldsymbol{b} \geq +1 \quad \text{for all y = +1 (class B)} \qquad \textbf{Equation 4}$$

where: a point (x_i) on the hyperplane is multiplied with the hyperplane orientation vector ($\boldsymbol{w}$) and added to the distance from the feature space origin ($\boldsymbol{b}$). In case of class A (label -1), the result must be either smaller or equal than -1. Considering class B (label +1) the result must be bigger or equal than +1.

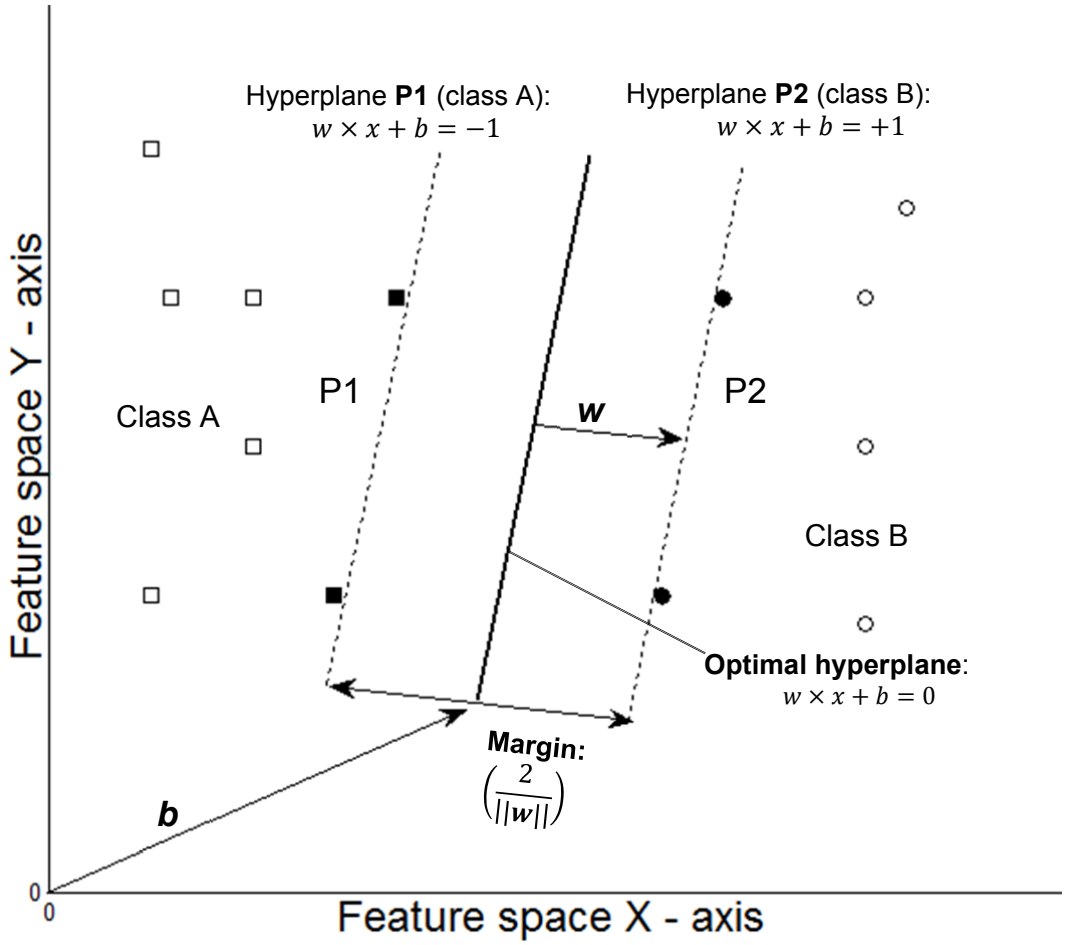

Figure 9 Simplified visualization of a hard margin linear SVC in a two-dimensional feature space. The solid squares and circles are support vectors defining hyperplane P1 and P2, *b* (bias) is the distance of the hyperplane from the feature space origin and *w* is the orientation vector of the optimal hyperplane.

The inequality equations (**Equation 3** and **Equation 4**) can be further combined into a single inequality as in **Equation 5** (Kavzoglu and Colkesen 2009):

$$y_i(\boldsymbol{w} \times x_i + \boldsymbol{b}) - 1 \geq 0 \text{ and } y_i \in \{-1, +1\} \qquad \textbf{Equation 5}$$

where: the class labels y_i (-1 or +1) are incorporated to test points (x_i) for satisfying the inequality. That is, the result must be bigger or equal to 0. If a used point x_i fulfills the inequality than it is a support vector lying on one of the two hyperplanes P1 or P2 (see Figure 9 -solid squares and circles).

The inequalities above take into account that it is a hard margin case and no support vectors of a class are laying in the area of the other class, and thus no training errors are allowed.

The hyperplanes in **Equation 3** and **Equation 4** are parallel to the optimal separating hyperplane as shown in Figure 9 (P1 and P2) and can be defined as functions - see **Equation 6** and **Equation 7** (Kavzoglu and Colkesen 2009):

$$\boldsymbol{w} \times x_i + \boldsymbol{b} = -1 \quad \text{for class A (-1)} \qquad \textbf{Equation 6}$$

$$\boldsymbol{w} \times x_i + \boldsymbol{b} = +1 \quad \text{for class B (+1)} \qquad \textbf{Equation 7}$$

where: $\boldsymbol{w}$ is the orientation vector, x_i is a point on the hyperplane and $\boldsymbol{b}$ is the distance from the feature space origin.

The distance, called margin, between the two hyperplanes P1 and P2 is defined as $\frac{2}{\|\boldsymbol{w}\|}$, where $\|\boldsymbol{w}\|$ is the length of the orientation vector $\boldsymbol{w}$ in the n-dimensional feature vector space. In order to maximize the margin between P1 and P2, $\|\boldsymbol{w}\|$ (length of $\boldsymbol{w}$) in $\frac{2}{\|\boldsymbol{w}\|}$ must be minimized, since a decreasing denominator leads to an increasing result. This maximization is an optimization problem as shown in **Equation 8** (Kavzoglu and Colkesen 2009):

$$\min\left\{\frac{1}{2}\| w \|^2\right\} \qquad \textbf{Equation 8}$$

where: $\|\boldsymbol{w}\|$ is the length of the orientation vector $\boldsymbol{w}$ in the n-dimensional feature vector space.

The optimization problem in **Equation 8** is subjected to the hard margin inequality constraint in **Equation 5** that can be rewritten as in **Equation 9** (Kavzoglu and Colkesen 2009):

$$y_i(\boldsymbol{w} \times x_i + \boldsymbol{b}) \geq 1 \text{ and } y_i \in \{-1, +1\} \qquad \textbf{Equation 9}$$

where: y_i is the class label (-1 or +1), $\boldsymbol{w}$ is the orientation vector, x_i is a point on the hyperplane and $\boldsymbol{b}$ is the distance from the feature space origin. The class labels y_i (-1 or +1) are included to test points (x_i) of satisfying the inequality. Thus, result must be bigger or equal 1 to satisfy the inequality.

The further incorporation of Lagrangian multipliers and quadratic programing optimization (Pal and Mather 2003, Tso and Mather 2009) leads to the exact determination of the support vectors as well as the solutions for the orientation vector ($\boldsymbol{w}$) of the hyperplane and its distance ($\boldsymbol{b}$) to the feature space origin. A detailed determination of $\boldsymbol{w}$, $\boldsymbol{b}$ and final decision rule ($f(x)$) can be found in the literature (Vapnik 1998, Foody and Mathur 2006, Dixon and Candade 2008, Izenman 2008, Tso and Mather 2009).

At the end of the training step, the final decision function ($f(x)$) of the optimal hyperplane can be derived, which is the trained hard margin SVC. The decision function ($f(x)$) enables a linear hard margin separation of the unknown data into class A or class B.

Soft margin linear classification

Remotely sensed spectral data may have distributions that cannot be separated exactly by applying hard margin linear boundaries (Tso and Mather 2009) because the classes are overlapping. In Figure 10 one training sample belonging to class A (a solid square marked with a dashed red circle) is located in the area of class B (circles).

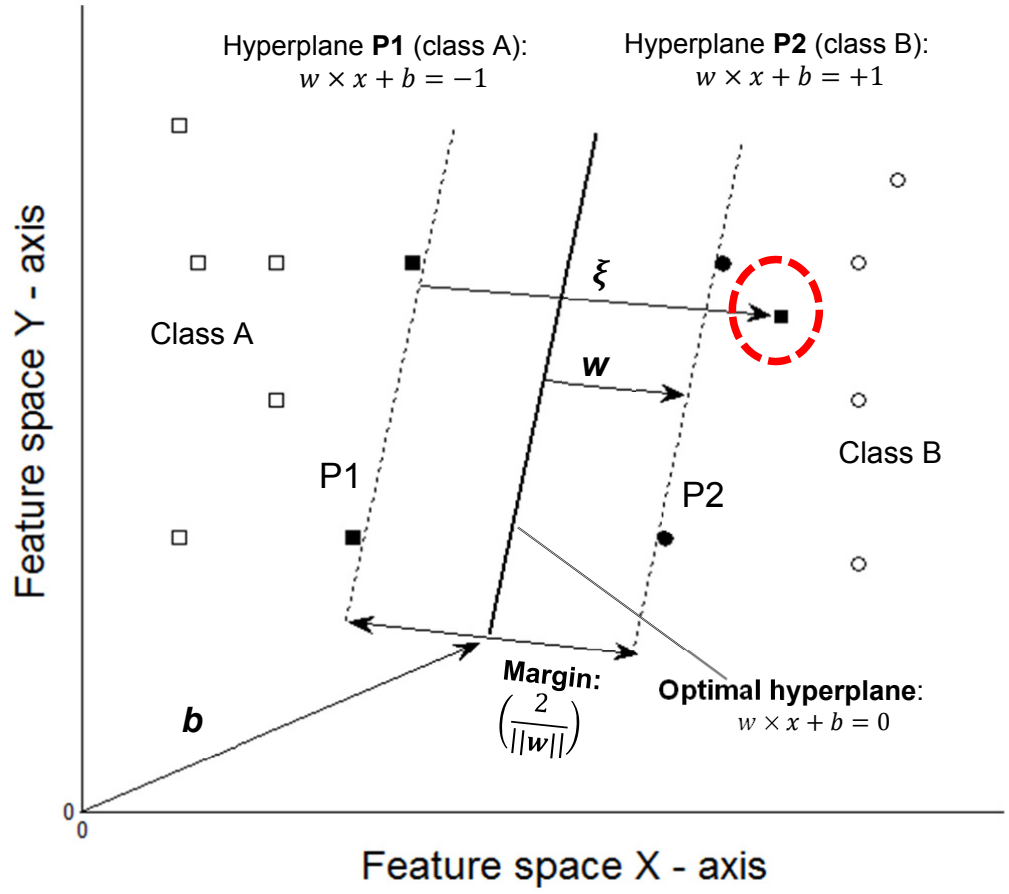

Figure 10 Simplified visualization of a soft margin linear classification problem with the introduction of a slack variable (ξ) in a two-dimensional feature space.

Overlapping classes are also assumed for the modelled phenology characteristic values. Therefore, the hard margin SVC formulations need to be adjusted for data points that are overlapping. The hard margin constraint in **Equation 9** cannot be satisfied and need to be softened by introducing slack variables (ξ_i) as in **Equation 10** (Kavzoglu and Colkesen 2009):

$$y_i(\boldsymbol{w} \times x_i + \boldsymbol{b}) \geq 1 - \xi_i \; ; \quad \xi_i \geq 0 \;\text{ and }\; i = 1, \ldots, N \qquad \textbf{Equation 10}$$

where: the class labels y_i (-1 or +1) are used to test points (x_i) of satisfying the inequality and the slack variable (ξ_i) softens the inequality that has to be met. The result must be bigger or equal to $1 - \xi_i$ to satisfy the inequality and ξ_i must be bigger or equal to 0.

Slack variables are proportional to a measure of cost (Tso and Mather 2009). As shown in Figure 10 the slack variable (ξ) denotes the distance from the outlying training sample (dashed red circle) to its original class hyperplane (P1). In other words, it is a training sample of class A being in the area of class B and its distance to hyperplane P1 is determined by the slack variable (ξ).

In addition, the optimization problem of **Equation 8** needs to be adjusted to the soft margin case as written in **Equation 11** (Foody and Mathur 2006, Kavzoglu and Colkesen 2009):

$$\min\left\{\frac{1}{2}\| w\|^2 + C\sum_{i=1}^{r}\xi_i\right\} \qquad \textbf{Equation 11}$$

where: the first term of **Equation 11** aims to maximize the margins. The second part is a penalty term containing: C which is a penalty parameter constant and ξ_i is the distance of an overlapping training sample to its original class hyperplane. The penalty term introduced to penalize the training samples located on the incorrect side of the hyperplane (compare Figure 10). This optimization problem is constrained by **Equation 10** to consider the slack variables.

The penalty parameter value (C) is used to keep the balance between margin maximization and error minimization. The penalty parameter value (C) is a user-defined constant.

A larger C assigns a higher penalty to errors and may lead to an over-fitting of the classifier to the used training data (Tso and Mather 2009). An over-fitting reduces the generalization capability of a classifier to unknown data points. In other words, the position of the optimal hyperplane would perfectly match the training data but is not flexible to unknown data points. In addition, if the training data are not representative and the hyperplane is fitted perfectly to it, then the classifier might not separate the unknown data correctly.

Similar to the hard margin SVC, Lagrangian multipliers and quadratic programing optimization are further incorporated (Pal and Mather 2003, Tso and Mather 2009) to determine the support vectors as well as the solutions for the orientation vector ($\boldsymbol{w}$) of the hyperplane, its distance ($\boldsymbol{b}$) to the feature space origin, C the penalty parameter and the slack variables (ξ).

The result of the training step is a final soft margin SVC, which is a decision function ($f(x)$) describing the optimal hyperplane. The final decision function allows a linear soft margin separation of the unknown data samples into the two classes.

A further description of $\boldsymbol{w}$, $\boldsymbol{b}$, ξ, C and final decision function ($f(x)$) can be found in the literature (Vapnik 1998, Foody and Mathur 2006, Dixon and Candade 2008, Izenman 2008, Tso and Mather 2009).

3.3.3 Non-linear Support Vector Classification

In cases of considerable class overlapping of the training samples (see Figure 11) linear soft margin SVC classifiers are unable to separate the samples into classes appropriately (Tso and Mather 2009). Therefore, a mapping of the training samples into a higher dimensional feature space is recommended using a vector mapping function (Tso and Mather 2009).

The mathematical concepts and formulations of vector mapping functions in relation to non-linear SVC are described in the literature (Vapnik 1995, Gunn 1998, Richards and Jia 2006, Hsu et al. 2009, Tso and Mather 2009). This section explains the concept of vector mapping function in a visual manner using a simplified illustration.

For example a training dataset cannot be separated linearly in a two-dimensional space, thus the data can be mapped into a higher dimensional feature space (e.g. three-dimensional space). In the higher dimensional feature space the construction of a linear decision boundary can be improved. In such a way the distribution of the training samples might be spread farther apart and may allow a construction of an optimal separating hyperplane.

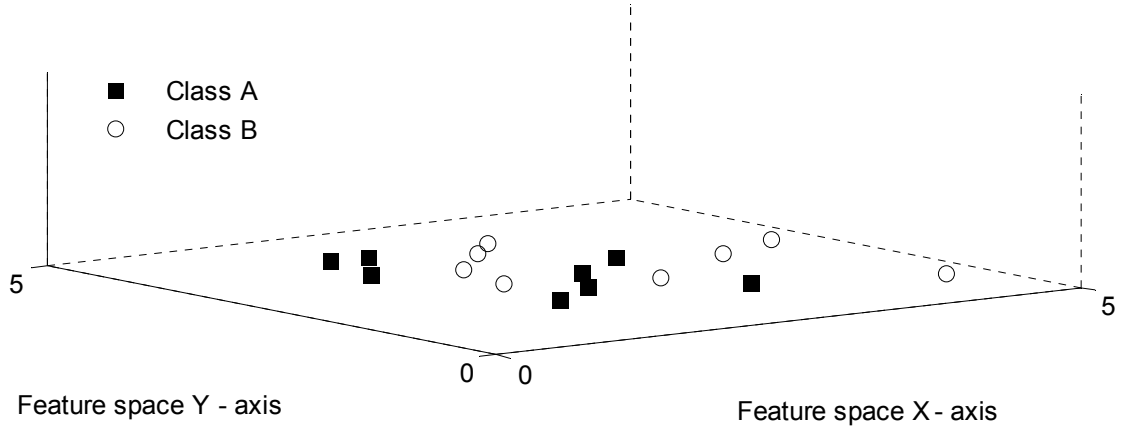

Figure 11 Case of considerable class overlapping representing a non-linear separable classification problem in a two-dimensional feature space. The training samples distribution disables constructions of linear optimal hyperplanes in order to separate the samples into classes.

Figure 11 shows a classification problem in a two-dimensional feature space (X and Y - dimension) where the training samples of class A and class B are overlapping significantly. Considering the overlapping training samples distribution, constructions of appropriate separating hyperplanes are impossible and the outcomes would lead to poor performances of linear SVC classifier. To overcome these concerns, a vector mapping function can be applied in order to project the training data into a higher dimensional space as shown in Figure 12. In detail, the training samples are projected with a non-linear vector mapping function (Φ), and thus a mapped training point x_i is represented as $\Phi(x_i)$ in the higher dimensional feature (Tso and Mather 2009). In the simplified example, the training samples are mapped from a

two-dimensional (X, Y) into a three dimensional space (X, Y, Z) by introducing the Z-dimension. In Figure 12 all training samples of class B moved up in the Z-dimension due to the application of a non-linear vector mapping function (Φ).

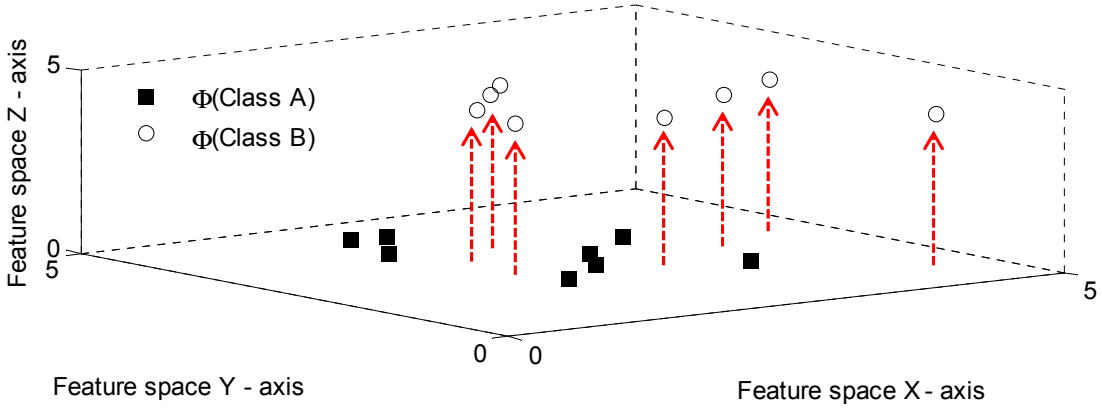

Figure 12 Simplified mapping of the original input training data samples from a two-dimensional into a three-dimensional feature space using a vector mapping function (Φ). The red dashed arrows symbolize the introduction of the Z-dimension. The mapped training samples of the classes are denoted either as Φ (class A) or Φ (class B).

Thus, the training samples are spread farther apart in the three dimensional feature space and the classes do not overlap, considering the Z-dimension. The different distribution in the three dimensional space enables the fitting of an optimal linear hyperplane in order to separate all samples into the classes as shown in Figure 13.

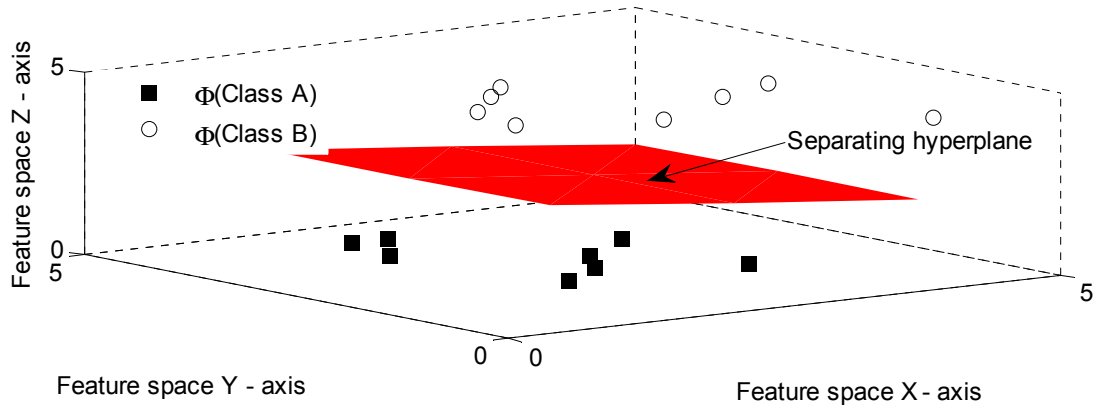

Figure 13 Distribution of the mapped training samples in three-dimensional features space. The three dimensional distribution enables the construction of a linear optimal hyperplane (red surface) in order to separate the training samples into classes.

Figure 14 visualizes an advantage of mapping functions: “back-mapping”. All samples and the fitted linear hyperplane can be mapped back from the higher dimensional space into the original feature space. The linear hyperplane becomes a non-linear decision boundary that can be used to conduct a non-linear separation of the data samples in the original feature space.

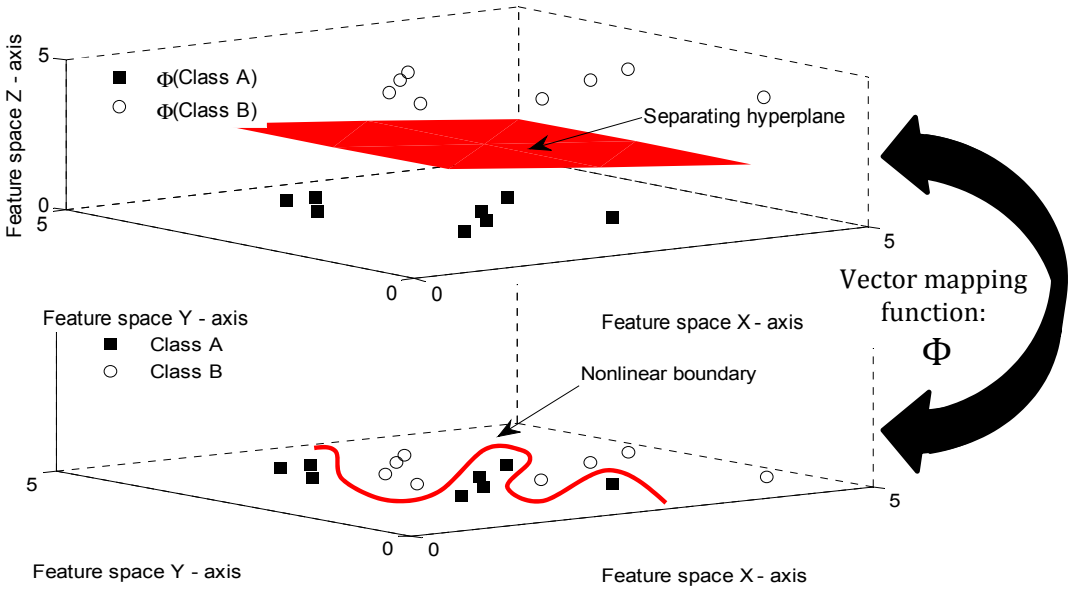

Figure 14 Visualization of the "back-mapping" advantage using vector mapping functions (Φ). Mapping of all samples from the two-dimensional (bottom) into three-dimensional (top) feature space to fit a linear hyperplane (red surface). Then, all samples and the linear hyperplane are mapped back to the original space (bottom) where the linear hyperplane becomes a non-linear decision boundary (red line).

However, the vector mapping function (Φ) leads to high computational expenses that can be reduced by using kernel functions instead (Vapnik 1995). A kernel function allows a more simplified representation of the data (Tso and Mather 2009). There are several kernel functions available; the two most commonly used in remote sensing applications are polynomial and Gaussian radial basis kernels (Richards and Jia 2006).

- The polynomial kernel function uses a polynomial order value (*d*) which is a user-defined constant. An increasing *d* might improve the accuracy of the classification (Tso and Mather 2009). However, an increasing *d* leads to higher computational expenses.
- The Gaussian radial basis function has less numerical difficulties compared to polynomial kernels that might go to infinity (Hsu et al. 2009) and incorporates a user-defined value called rbf sigma (σ).

For further description and derivation of kernel functions please refer to the literature (Vapnik 1995, Gunn 1998, Izenman 2008).

During the training of the SVC in the higher dimensional space, support vectors will be detected, the margins between them will be maximized using quadratic programing optimization and Lagrangian multipliers. This leads to the determination of a decision function ($f(x)$) containing a kernel function.

3.3.4 Parameter Selection

In order to train a support vector machine classifier the analyst needs to define parameters such as the penalty parameter value and if a kernel is incorporated the kernel function related parameters (i.e. polynomial order or rbf sigma value). A suitable parameter selection is very important in order to enhance the generalization capability[2] of a SVC to unknown data points. The parameter selection employs a validation error (Tso and Mather 2009), also known as error rate. It has to be ensured that the validation error is unbiased by using a cross-validation method. The parameter selection is used to determine the parameters with lowest error rates and the respective parameter selection should be used for the final SVC (Tso and Mather 2009).

However, over-fitting issues[3] should be considered (Tso and Mather 2009) in such a way that a parameter selection with the lowest error rate might lead to classifier that do not generalize on the unseen data points. Thus it is recommended to apply a *t*-fold cross-validation, where *t* is defined by the analyst and ten folds are suggested (Tso and Mather 2009). There, the training data is divided in *t*-subsets and a leave-one-out cross-validation strategy is applied. That means, one subset of t-subsets is chosen for validation and the rest (*t-1*) are used to train the classifier during the training stage. This procedure will repeated *t*- times until each subset has been used for validation purposes and then a classification error from all validation sets will be averaged over *t*-trials. The t-fold cross-validation gives a validation error, which is the percentage of validation data points that have not been classified correctly (Tso and Mather 2009).

The range of possible parameters changes can become very large, therefore a grid search method developed by Chang and Lin (2010) is recommended. In case of non-linear SVC, the changing penalty term values (*C*) and the kernel function's related kernel parameters will be tested for their performance. In case of linear SVC the varying penalty term values will be evaluated. The grid search method incorporates the user-defined initial values of the corresponding parameters, which will be t-fold cross-validated, then changed and evaluated again until user-defined stop criteria are reached. The analyst needs to define minimum and maximum values of the corresponding parameters. Over this defined range the values will be

[2] Generalization in terms of statistical machine learning means the construction of an elastic classifier that is not too fixed on the used training data and can be applied properly to the unknown part of the data.

[3] An over-fitting reduces the generalization capability of a classifier to unknown data points. In other words, the position of the optimal hyperplane would perfectly match the training data but is not flexible towards unknown data points.

stepwise incremented and t-fold cross-validated. Thus, it is possible to choose the best performing parameters values of the grid search for the final SVC.

In case of applying a non-linear SVC, a parameter pair[4] (e.g. C and d) needs to be tested in stepwise changes over a certain space. The space is divided into grids; each grid represents different parameter pair values in ascending order. A possible way of a grid search space is shown in Figure 15. The chosen space is limited to minimum and maximum values for both parameters and contains grid points determining parameter value pairs. A grid search could start in the bottom left corner and increase stepwise to the next grid points (e.g. parameter pair values) until the opposite corner is reached. A grid search can become very complex and demand computational resources. Therefore, Tso and Mather (2009) suggest first a coarse grid search with coarser increments of the parameter pair values as shown in Figure 15 (blue line). Then, it continues with a fine search (black line) within the range of the best performing parameter pair values out of the coarse search. The best performing parameter values of the fine grid search are recommended for the training of the final classifier.

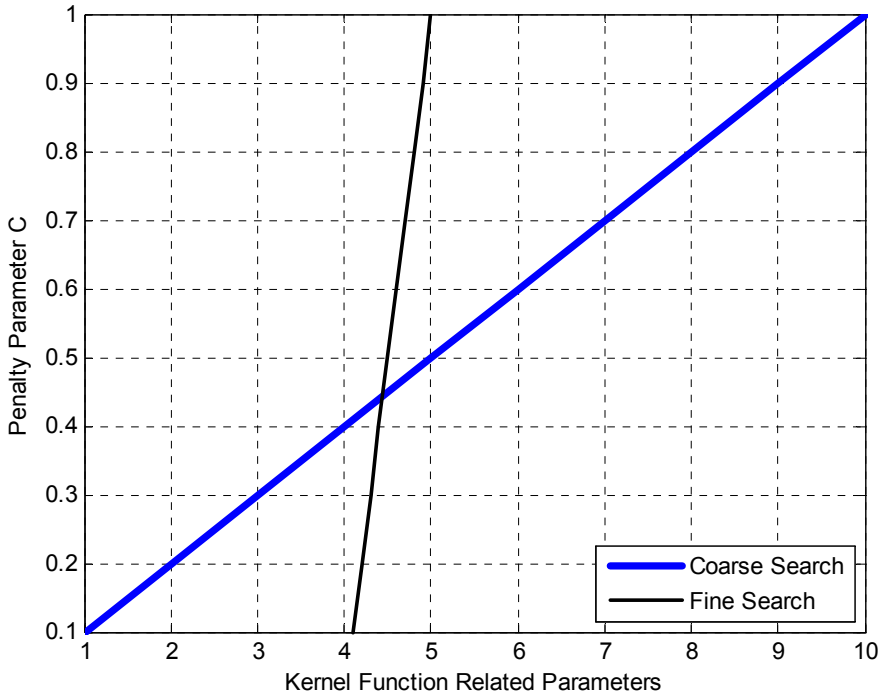

Figure 15 Coarse and fine parameter grid search.

The space shown in Figure 15 can be stretched to larger extents; however the grid search will become more time and resource demanding. More advanced strategies to determine the parameters are mentioned in the literature (Tso and Mather 2009, Chang and Lin 2010).

[4] In this study, the parameter pairs account for the penalty parameter value and a kernel function related parameter. 1) penalty parameter value and polynomial order value (C and d) or 2) penalty parameter value and Gaussian radial basis function (C and σ).

3.3.5 Multi-classification

Originally SVC was developed as a binary classifier; however SVC can be extended to multi-classification purposes using a one-against-one classification strategy. This is a pair wise comparison between all information classes *k* and each classifier will be trained solely on two of the *k* classes. A total number of $k(k-1)/2$ classifiers are needed to ensure all possible one-against-one combinations (Tso and Mather 2009). In order to construct a support vector machine multi-classification (SVM-C) approach Tso and Mather (2009) recommend a directed acyclic graph (DAG) where the one-against-one strategy is incorporated. The $k(k-1)/2$ classifiers are trained beforehand and building up a decision tree.

A DAG SVM-C begins by applying the classifier at the root of DAG (Level 1) to an unseen data point depending on how the data point is classified it will move down to either the left or the right node in the next level. This continues downwards until a final decision is made in the leaf of the graph. Tso and Mather (2009) state that the order of the binary classifiers constructing the DAG is not important and will not affect significantly the classification accuracy, since each possible class combination will be tested.

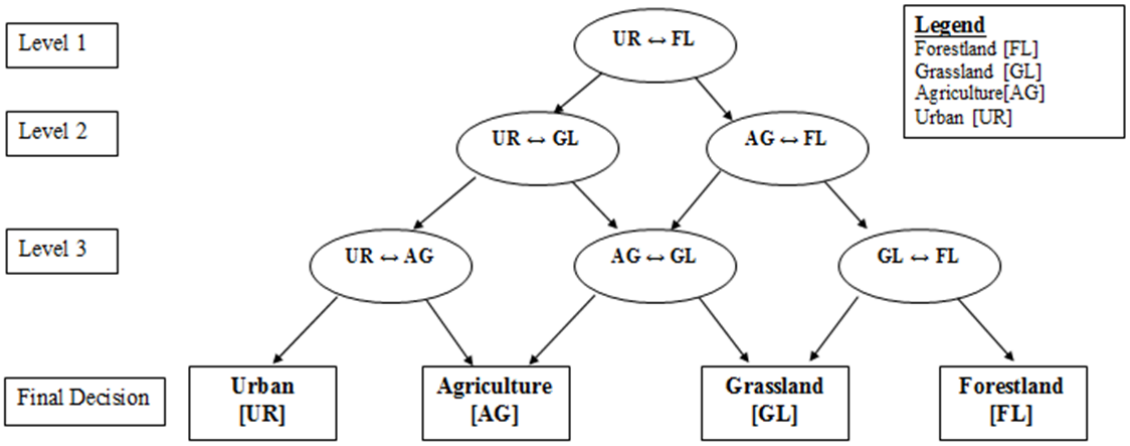

Figure 16 Directed acyclic graph containing six classifiers and four classes to enable multi-classification.

In Figure 16 is shown an example of a SVM-C using DAG. It shows a hypothetical multi-classification problem and contains four classes that are pair wise combined in six classifiers. This approach starts at the root (Level 1) by testing an unseen data sample for being a member of the classes Forestland or Urban. If it belongs to Forestland the procedure will move down to the right (level 2). In level 2 the data sample will be tested for being a member of Forestland or Agriculture. If it is a member of Agriculture it will move down to the left to the next classifier in level 3. In level 3 will be employed the last classifier and it will be tested for a membership in the classes Grassland or Agriculture. Depending on the membership, the data sample will move down to the final decision, which is represented by the leaf nodes of

the DAG. Let us assume the sample belongs to Grassland, thus it would go down to the left and the final decision would be made. The unseen data sample is classified as Grassland. Consequently, it is ensured that each data sample will cascade through the graph and will be tested for being a member of every class due to the pair wise comparison.

3.4 Accuracy Assessment

3.4.1 Reference Data

In LULC studies that are based on classifications of remotely sensed data it is essential to determine how well a chosen classification approach has performed by applying an accuracy assessment. In order to validate a classification approach, reference data of the corresponding study area are needed. There are several techniques described by Congalton and Green (2009) to obtain reference data to validate classification approaches of remotely sensed data.

In this study CORINE land cover 2006 data – version 13 (CLC 2006) in 250m ground resolution are used as raster reference data. CLC 2006 represents the LULC of the year 2006 (EEA 2007) and are distributed in the geodetic reference system called European Terrestrial Reference System 1989 (ETRS1989) and in Lambert Azimuthal Equal Area projection. CORINE stands for Co-ordination of Information on the Environment and is a European Union project, which aims to create a database of LULC for 38 countries in the European Union. The CLC 2006 data are based on satellite imagery interpretations from satellite sensors of high spatial resolution, 25m and finer (EEA 2007).

The LULC classes in CLC 2006 are assigned on three levels. Level 1 contains of five LULC classes, level 2 is more detailed with 15 classes and level 3 contains 44 classes with very detailed information about the land cover (EEA 2007). In the technical guidelines of CLC 2006 (EEA 2007) a minimum thematic accuracy of 85% is stated. The previous CORINE land cover 2000 project was planned to have a minimum thematic accuracy of 85%, and was assessed with a thematic overall accuracy of 87% (EEA 2006). CORINE land cover 200 has been applied successfully as reference and training data in several classification studies using MODIS data (Gonçalves et al. 2005, Heiskanen and Kivinen 2008) However, the accuracies are differing between the LULC classes. A similar thematic accuracy is also assumed for CLC 2006 data.

In this study CLC 2006 is used as reference data to train the SVCs and for validation in the accuracy assessment. It is feasible to apply CLC 2006 data since it has the same ground resolution of 250m as the used MODIS NDVI imagery. For each MODIS pixel in the study

area are reference data available and; thus it is possible to apply an accuracy assessment to the entire study area.

3.4.2 Assessment of Accuracy

In order to perform an accuracy assessment it is necessary to introduce the concept of an error matrix because all following techniques are based on the error matrix. An error matrix, as shown in Figure 17, is a square array containing of n samples, which are distributed into k^2 cells - k represents the number of classes. Every sample is labeled to one of the k classes in the dataset classified by an algorithm and is independently to one of the same k classes in the reference dataset. Let n_{ij} indicate the number of samples predicted into class i $(i = 1,2,...,k)$ in the classified dataset and class j $(j = 1,2,...,k)$ in the reference dataset (Congalton and Green 2009).

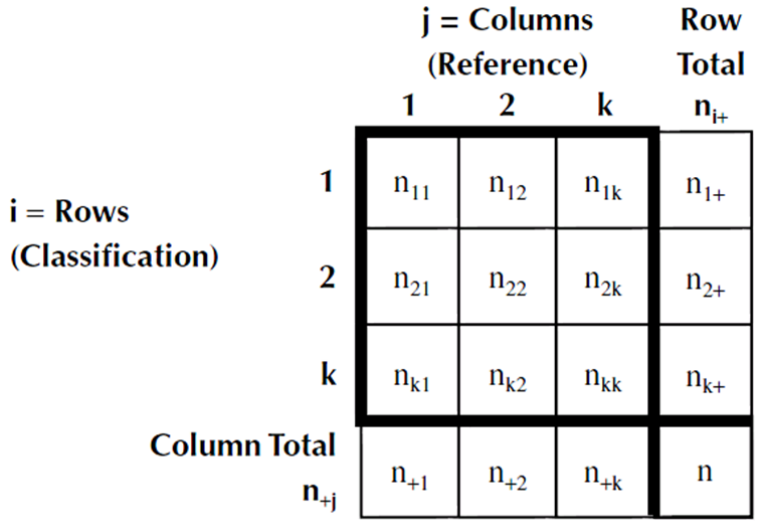

Figure 17 Theoretical example of an error matrix from Congalton and Green (2009)

In Figure 17 a theoretical example of an error matrix is shown. The classified data samples are written row-wise (i) and the known reference data samples column-wise (j). The sums of each column are denoted in n_{+j} and rows in n_{i+} respectively. The sums of n_{+j} and n_{i+} are equal and are represented by n.

Congalton and Green (2009) formulate the number of classified data samples into class i as written in **Equation 12:**

$$n_{i+} = \sum_{j=1}^{k} n_{ij} \qquad \textbf{Equation 12}$$

where: n denotes the total number of samples, n_{i+} is the sum of a row, n_{ij} is a record in the i^{th} row and the j^{th} column, and k is the number of classes.

The known reference data samples of class j are represented in **Equation 13** (Congalton and Green 2009):

$$n_{+j} = \sum_{i=1}^{k} n_{ij} \qquad \textbf{Equation 13}$$

where: n denotes the total number of samples, n_{+j} is the sum of a columns, n_{ij} is a record in the i^{th} row and the j^{th} column, and k is the number of classes. With the help of an error matrix it is possible to calculate measures of quality such as overall accuracy, user's accuracy, producer's accuracy and Kappa coefficient.

Equation 14 shows how to compute the overall accuracy between the classified and reference samples (Congalton and Green 2009):

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{k} n_{ii}}{n} \qquad \textbf{Equation 14}$$

where k is the number of classes, n the total sample number and n_{ii} denotes the number of correctly classified samples of all classes. Increasing i, n_{ii} denotes the major diagonal - compare in Figure 17 (e.g. n_{11}, n_{22} ,…, n_{kk}). Overall accuracy gives a measure of how many samples have been classified correctly in relation to all samples.

The user's accuracy gives a measure of the probability of an unknown sample on the map (reference data) of being correctly classified. User's accuracy is calculated for a specific class and is computed as written in **Equation 15** (Congalton and Green 2009):

$$\text{User's accuracy}_i = \frac{n_{ii}}{n_{i+}} \qquad \textbf{Equation 15}$$

where n_{ii} is the total number of correct classified samples of a class and n_{i+} is the row-wise total sum of samples classified as the same class, thus n_{i+} considers also misclassified samples. The user's accuracy of class 1 in Figure 17 is calculated as follows user's accuracy$_1 = n_{11}/n_{i+}$.

The producer's accuracy illustrates the probability of an unknown sample point in the reference data as being correctly mapped. Also producer's accuracy is applied for a single class and is calculated as formulated in **Equation 16** (Congalton and Green 2009):

$$\text{Producer's accuracy}_i = \frac{n_{jj}}{n_{+j}} \qquad \textbf{Equation 16}$$

where n_{jj} is the total number of samples classified correctly of a class and n_{+j} is the column-wise total sum of samples of the corresponding class indicated by the reference data.
The producer's accuracy of class 1 in Figure 17 is calculated as follows
producer's accuracy$_1$= n_{11}/n_{+j}.

The Kappa analysis was introduced by Cohen (1960); the unweighted Kappa value (Cohen 1960) is a robust measure since it takes the agreement occurring by chance into account and is suggested in the literature (Lillesand et al. 2004, Congalton and Green 2009). Cohen's Kappa gives a measure of how well the classification approach agrees with the reference data. In **Equation 17** is formulated Cohen's Kappa coefficient (Congalton and Green 2009):

$$\hat{K} = \frac{n\sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k}(n_{i+} \times n_{+j})}{n^2 - \sum_{i=1}^{k}(n_{i+} \times n_{+j})} \qquad \textbf{Equation 17}$$

where: $\widehat{K}$ is the Kappa coefficient, k is total number of rows in the error matrix, n_{ii} denotes number of correctly classified samples of all classes, n_{+j} describes all observation in column j, n_{i+} stands for all observations in row i and, n is total number of observations in the error matrix.
For further descriptions of Kappa please refer to the literature (Cohen 1960, Lillesand et al. 2004, Congalton and Green 2009).

4. Study Area

The chosen study area is located in Scania, southern Sweden. The area of interest is situated between the south of Hässleholm and west of Kristianstad. In Figure 18 the study area is visualized. The LULC map is based on CLC 2006 – Version 13 in 250 m ground resolution.

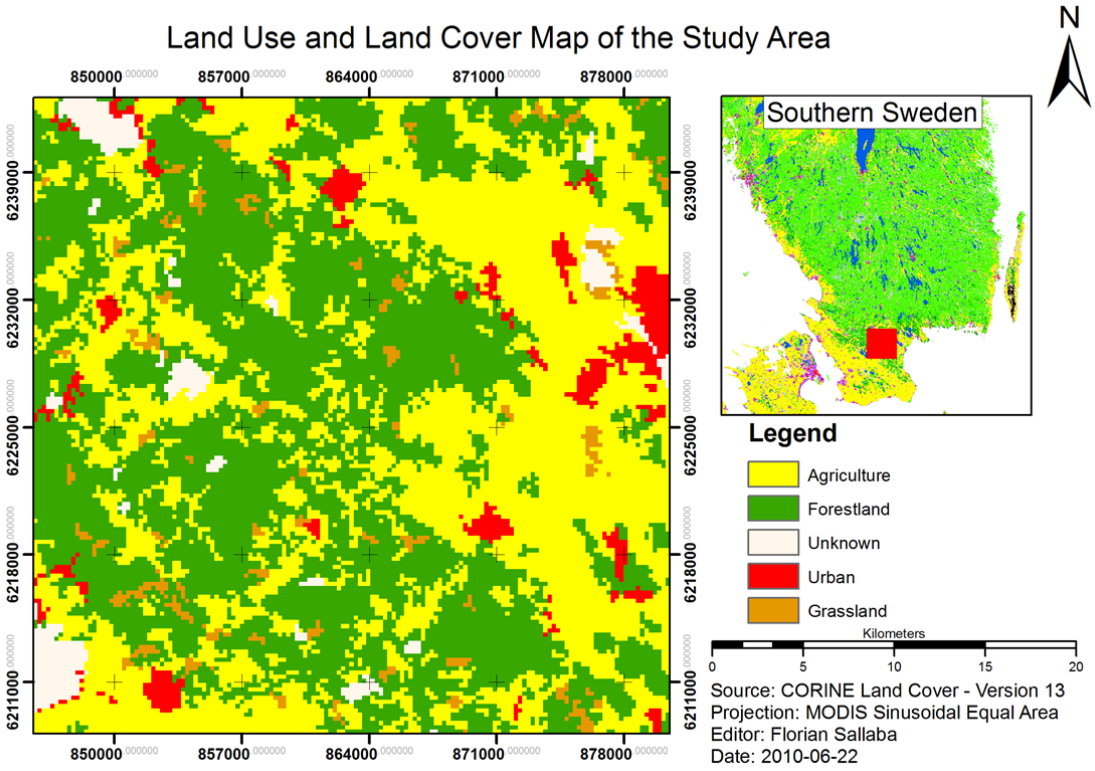

Figure 18 Land use and land cover map of the study area in southern Sweden. The map is based on CLC 2006 Version 13 data in 250 m ground resolution (source: www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster)

The study area is interesting because it is located in a transition zone - compare small map of southern Sweden in Figure 18. That means, to the north of the study area Forestland is dominating and in the south is agricultural land use prevailing. The selection of the study area was based on practical reasons, since the area should be dominated by Forestland and Agriculture. These two LULC classes are assumed to be typical in southern Sweden. In addition, seasonality characteristics (e.g. end and length of season) of Forestland and Agriculture are assumed to be different and effective in the SVC due to the rapid decrease of NDVI of agricultural areas after harvesting. Moreover, LULC classes such as Grassland and Urban cover approximately 6% of the area. Urban areas contain heterogeneous ground cover such as vegetation, concrete or other artificial surfaces. This heterogeneity leads to mixed spectral signatures in a pixel (Lillesand et al. 2004, Foody and Mathur 2006, Richards and Jia 2006, Tso and Mather 2009) and especially in a moderate ground resolution of 250m. The

mixed spectral signatures might initiate NDVI values, seasonality modeling outcomes and finally seasonality characteristics that are difficult to interpret and to classify by the analyst.
Furthermore, Grassland has similar spectral characteristics as Forestland (see Figure 1) and will influence the NDVI calculation and the following seasonality modeling and SVC separation. Even though the seasonal development (seasonality curve) of Grassland is assumed to be on lower amplitudes than in Forestland (see Figure 2).
In addition, the study area is reasonable, since it has small coverage of water areas (inland wetlands and inland waters). In Figure 18 water areas are classified as Unknown and will be mask out for reasons of potential error minimization later in the processing step. A small coverage of Unknown (water) areas leads to a small loss of potential area to study.
Hence, a few of possible error sources are minimized and allow a focus on Forestland and Agriculture with this chosen study area.

The proportions of the LULC information classes of the study area are shown in Table 3. The dominating LULC classes are Forestland and Agriculture that cover approximately 90% of the area. Unknown, Grassland and Urban cover approximately 10% of the area.

Table 3 Proportions of LULC classes in the study area - from CLC 2006 dataset

LULC class	No. of LULC class pixel positions	Percentage of LULC classes [%]
Forestland	10,639	47.28
Agriculture	9,689	43.06
Urban	896	3.98
Unknown	691	3.07
Grassland	585	2.61

Thus, the study area represents almost equal proportions of Forestland (47.28%) and Agriculture (43.06%).

5. Methodology

The developed methodology of this study is shown simplified in Figure 19. Each part will be described more detailed in the corresponding sectionss. In the first part of the study, the used satellite data are described and how it was processed and incorporated into this study. In the second part, the seasonality modeling will be explained, which was applied to generate the seasonality characteristics. The third part contains two subparts: (1) the selection of the best seasonality characteristics as input features for SVC, and (2) the determination of the penalty values and the kernel functions related parameters. Each of the subparts considers twelve replicates due to statistical representativeness reasons. In the fourth part the actual SVM-C in a DAG will be explained, which incorporates twelve replicates for representativeness reasons. The last part of the methodology describes the accuracy assessment of the used SVC strategies (linear and non-linear) and their twelve replicates.

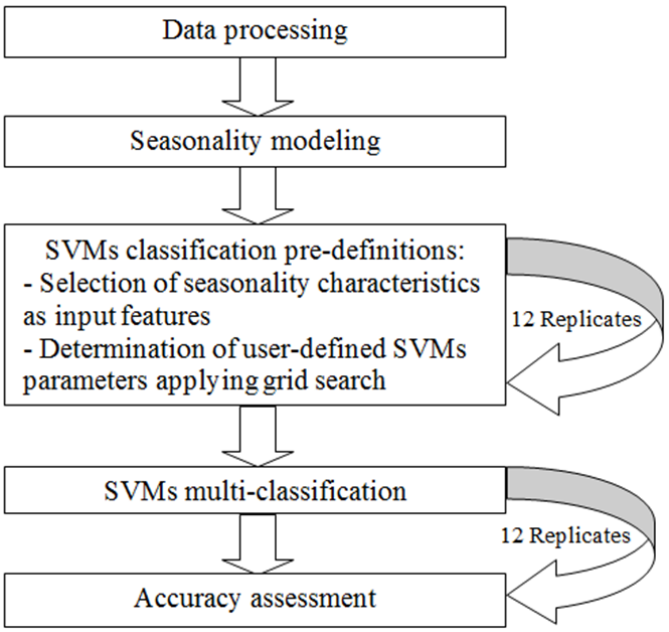

Figure 19 Simplified scheme of the developed methodology for this study.

The entire methodology is implemented in MATLAB (R2010a) using TIMESAT for seasonality modeling and MATLAB's statistical learning functions of the bioinformatics toolbox for SVC. The study's MATLAB code will be provided by the author[5] if asked for.

The most important statistical learning functions used in this study are **svmtrain** and **svmclassify**. The svmtrain function allows training and constructing of binary SVC on a training dataset. The constructed binary SVC is then used by svmclassify to separate the unknown dataset.

[5] Authors email address: Florian.Sallaba(AT)nateko.lu.se

5.1 Data

This study incorporates MODIS NDVI imagery and corresponding overall quality indicator data of the MODIS tile h18v03. The 8-day 250m ground resolution surface reflectance product (MOD09Q1) and quality data were obtained "ready for use" from the research of Jönsson et al. (2010). The raw MODIS Terra spectral (MOD09Q1) and raw quality data were distributed from Land Processes Distributed Active Archive Center (LP DAAC), located at the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (lpdaac.usgs.gov).

The schematic workflow in Figure 20 visualizes the implemented data processing workflow framework and the three incorporated data components. The outcomes of the data processing are used in the subsequent seasonality modeling (described in *Section 5.2*) and the extraction of the seasonality characteristics.

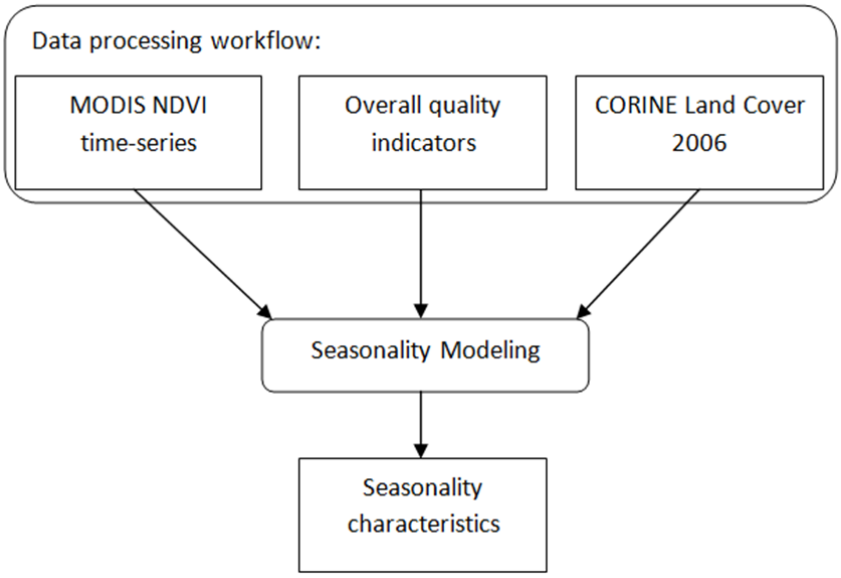

Figure 20 The Schematic workflow represents the implemented data processing and the incorporated three data sources, seasonality modeling, and extraction of seasonality characteristics.

The first data component comprises the NDVI time-series that were computed from MOD09Q1. MOD09Q1 data contain reflectance measurements covering red (band-1) and near-infrared (band-2) spectral ranges (see Table 1) and were used to compute NDVI images. Theoretically NDVI values can range between -1 and +1 and for practical reasons the NDVI values were rescaled according to **Equation 18** (Hellström 2010, personal communication):

$$DN_{NDVI} = NDVI \times 212.5 + 42.5 \qquad \textbf{Equation 18}$$

Where: DN_{NDVI} is the scaled NDVI value, 212.5 and 42.5 are two constants. In addition, negative NDVI values were set to zero and the decimal places of the NDVI values were

truncated by a conversion from floating point format to 8 bit unsigned integer (i.e. uint 8 bit). Hence, the scaled NDVI data contain values between 0 and 255. This was done to reduce the size of the data and enable faster computations (Hellström 2010, personal communication). The applied NDVI time-series are based on 230 NDVI images of the study area and covers a period of 5 years from 01-01-2004 to 31-12-2008.

The second component of the data processing frame (Figure 20) contains the overall quality indicator data, which represents an ancillary data source in order to improve the subsequent seasonality modeling. The overall quality indicator data are based on several MODIS products, such as cloud coverage information (MOD35 cloud) from the 8-day 500m ground resolution surface reflectance (MOD09A1) and MODLAND QA bits. Jönsson et al. (2010) classified the combined quality data in high, reasonable and poor quality. The overall quality indicator data are organized in binary image format and contain pixel that represent quality values (see Table 4). Hence, each pixel position in the NDVI images has a corresponding pixel position in the overall quality indicator binary image having information about the quality. The projection, extent and ground resolution is the same as in the NDVI data. 230 overall quality indicator images were extracted covering the same study area and time period as the corresponding NDVI time-series. For further information about the overall quality indicator data please refer to Jönsson et al. (2010).

The third component of the data processing frame (Figure 20) is the CLC 2006 that is used as ancillary data in seasonality modeling. CLC 2006 data (250m ground resolution) were reprojected from the original Lamberts Azimuthal Equal Area projection (ETRS 1989) to the MODIS sinusoidal projection using nearest neighbor transformation. A nearest neighbor resampling method was feasible, since CLC 2006 and MODIS NDVI datasets have the same ground resolution of 250m. Then, CLC 2006 data were clipped according to the extent of the study area. The original CLC 2006 LULC classes were aggregated to the following LULC information classes: (1) Agriculture, (2) Forestland, (3) Urban, (4) Grassland and (5) Unknown. Inland wetlands and inland waters were linked to class Unknown.

A mask was created from the Unknown class and incorporated in the seasonality modeling to omit seasonality modeling of areas that are potentially covered with water. Seasonality modeling of waters areas may lead to erroneous seasonality characteristics and might also influence the subsequent classification procedure.

CLC 2006 data were distributed from the European Environment Agency (www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster).

5.2 Seasonality Modeling

Seasonality modeling aims to create an overall fitting of the time-series of all pixels in the study area. It was assumed that the LULC classes are not known in this step. Thus, no a priori knowledge of the LULC classes would influence the seasonality parameter modeling, except the water mask.

The seasonality modeling settings were adjusted intuitively in the beginning and then the parameters were fine-tuned iteratively in order to achieve well fitted curves that is more of an art than a science (Eklundh and Jönsson 2010).

Savitzky-Golay fitting was applied in TIMESAT in order to have a local fitting that enables to maintain rapid seasonality changes. Eklundh and Jönsson (2010) recommended a rough estimation of the moving window size, which is the rounded quotient of the number of observations (i.e. 46 observation) per year divided by 4. Thus, the moving window size of the Savitzky-Golay filter was defined to a size of 12.

In the modeling procedure the overall quality indicator data was incorporated, as weights, to weaken the effect of noise such as cloud coverage or poor data quality. The weight settings of the overall quality indicator data are shown in Table 4.

Table 4 Weight settings of quality data. Quality data was obtained from Jönsson et al. (2010).

Quality values	Weight values
0 (poor quality)	0.1
1 (reasonable quality)	0.3
2 (high quality)	1

The influence of poor and reasonable quality data are kept low with corresponding small weights, whereas high quality data have a full effect on the modeling. Figure 4 visualizes the influence of the chosen weight settings on a NDVI time-series; obvious spikes (e.g. cloud coverage) during the growing season have received low weights.

Furthermore, the remaining positive and negative outliers (i.e. spikes) were removed with the median method as described in *Section 3.2.2*.

In addition pixel positions containing bad data due to complete cloud coverage during the growing season were introduced to the water mask and linked to the Unknown LULC class.

In order to deal with negatively biased NDVI time-series, it was used an adaption to the upper envelope, incorporating a fit in two steps as described in *Section 3.2.2*. The adaption strength of the upper envelope was set to a low value (adaption strength value of 2), since a

high value will lead to an emphasis of single high values in the NDVI time-series (Eklundh and Jönsson 2010).

The model was adjusted to one growing season, since just one growing season takes place within a year in the chosen study area. During the winter season NDVI data of the study area may be affected by large solar zenith angles (i.e. low solar illumination and long atmospheric path lengths) and long periods of cloud as well as snow coverage. Under such conditions, spectral sensor measurements are inaccurate and lead to erroneous NDVI data (Jönsson and Eklundh 2002, 2004). The fitted function was forced to a small minimum value (i.e. minimum value of 0.1) in the winter season (Eklundh 2010, personal communication) to treat erroneous NDVI data. In order to determine the start and end of vegetation seasons a method that incorporates the proportion of the seasonal amplitude (Jönsson and Eklundh 2002, 2004) was used. The season start was set when the fitted curve reached a proportion of 0.5 of the seasonal amplitude measured from the left minimum value of the fitted curve. In the same manner was established the season end when the curve reached a proportion of 0.5 of the seasonal amplitude measured from the right minimum.

As mentioned above, a water mask was incorporated to avoid seasonality modeling of water areas because it may lead to erroneous seasonality characteristics that might affect the classification approaches. The water mask covers all areas that are belonging to Unknown class in Figure 18.

In order to summarize, each pixel position in Figure 18 is represented by a NDVI time-series and its overall quality data. The NDVI time-series were used to model seasonality. The modeled functions enable the extraction of seasonality characteristics (compare - Table 2). All modeled seasonality characteristics for three years (2005 - 2007) were extracted and all pixel positions that are assigned to LULC class Unknown achieved seasonality characteristics with zero values. In other words, each pixel position obtained one seasonality characteristic values for each year in the period 2005-2007 (i.e. three seasonality characteristic values). This time frame was chosen because of the available CLC 2006 data. The chosen time period is expected to overlap the inventory and mapping time of the CLC 2006 data. Thus, optimal reference data for the SVC and its accuracy assessment are available.

Finally, an average value of the three seasonality characteristic values for each pixel position was calculated. Consequently, each pixel position is represented by an average seasonality characteristic value. The averaging was applied to all seasonality characteristics in order to incorporate the seasonal development of all three years under the assumption that no LULC changes took place in the study area.

5.3 Support Vector Machine Classification Pre-definitions

5.3.1 Training Data Selection

The training data collection was based on a stratified random selection for each LULC information class. That is, each LULC class achieved 200 randomly chosen pixels according to a priori knowledge from the CLC 2006 information.

Table 5 A training data replicate containing stratified randomly selected pixel positions.

LULC class	No. of pixel positions	No. of replicates
Forestland	200	12x200
Agriculture	200	12x200
Unknown	200	12x200
Grassland	200	12x200
Urban	200	12x200
Sum of training samples	**1,000**	**12,000**

A training dataset achieved 1,000 pixel positions that were stratified randomly chosen as shown in Table 5. One training dataset is statistically not representative and may contain inaccurate modeled seasonality due to uncertainty in the MODIS data and incorrect assigned LULC classes in the CLC 2006 dataset, since 15% of CLC 2006 data are assumed to be incorrect (EEA 2007). To overcome these concerns twelve training dataset replicates were randomly created for each LULC class (see Table 5). In other words, 12,000 pixel positions $(200\times5\times12)$ were chosen. However, the pixel positions that had been chosen in previous replicates were also available to select in a new replicate. Thus, it is possible that the pixel positions were chosen several times, especially in the smaller LULC classes such as Urban and Grassland.

5.3.2 Input Feature Selection

Input features determine the characteristics of the feature space. The selection of input features is very important because SVC is centered on the feature space attributes (e.g. distribution) of the input dataset.

As mentioned before, the nomenclature of SVC describes input features as any kind of data that can be used in a classification (Tso and Mather 2009). In this study, it was important to select the most representative seasonality characteristics for all LULC classes in order to use them as input features for the SVC.

The selection of input features was based on the extracted seasonality characteristics and adapts the technique of the validation error rate as described in *Section 3.3.4*. That means the training and ten-fold cross-validation part of SVC was anticipated to determine the best input feature combinations. It is assumed that the validation error rate gives the best guess, since

there is no established way to determine the best seasonality characteristics for classification approaches (Eklundh 2010, personal communication). The available SVC training function in MATLAB (Version 7.10.0 – R2010a) gives error rates between zero and one. The higher the error rate the less accurate is a trained classifier.

The amount of seasonality characteristics was intuitively reduced to eight characteristics; i.e. season start, season end, season maximum, season length, small integral of season, left derivate, right derivate and derivate ratio. These decisions were done to keep the selection method less complex and the computational expenses as low as possible.

Previous tests of the method have shown that three inputs features are a reasonable amount for the input feature selection. More than three features will lead to a rapidly increasing computation time and memory problems of the used Personal Computer[6].

Figure 22 illustrates a simplified workflow of the input feature selection and is used to describe the following methodology. The eight seasonality characteristics were arranged in three input feature columns, in such a way that each input feature combination contained a unique grouping of seasonality characteristics as presented in Table 6. This led to 56 unique seasonality characteristic combinations.

Table 6 Example of unique input feature combinations and their error rates from linear soft margin SVM-C training and validation.

Combination No.	Input feature 1	Input feature 2	Input feature 3	Final error rates
1	Begin of Season	End of Season	Maximum Season	1.393
2	Begin of Season	End of Season	Length of Season	2.625
3	Begin of Season	End of Season	Small Integral	2.110
…	…	…	…	…
56	Left Derivate	Right Derivate	Derivate Ratio	1.943

Then, 200 training samples were randomly chosen (stratified) from the CLC 2006 reference data for each of the five LULC classes. Each training sample was linked to three seasonality characteristics (input features) values according to its pixel position – as shown in Figure 21.

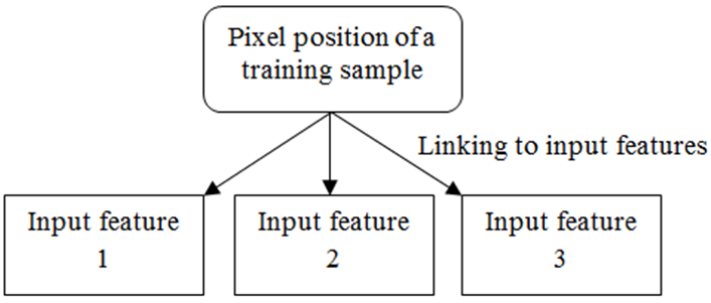

Figure 21 Linking of a training sample pixel position to the corresponding three input features.

[6] The computations were done in MATLAB (R2010a – 32bit version) on a Windows 7 (32 bit) operating system on a Notebook with an Intel Core 2 Duo processor (T66) and 4096 DDR3 1066 MHz RAM.

In case of combination no. 1 in Table 6 each LULC class received 200 Begin of Season, 200 End of Season and 200 Maximum Season values. These seasonality characteristic values were used to build up a three-dimensional feature space for the selection.

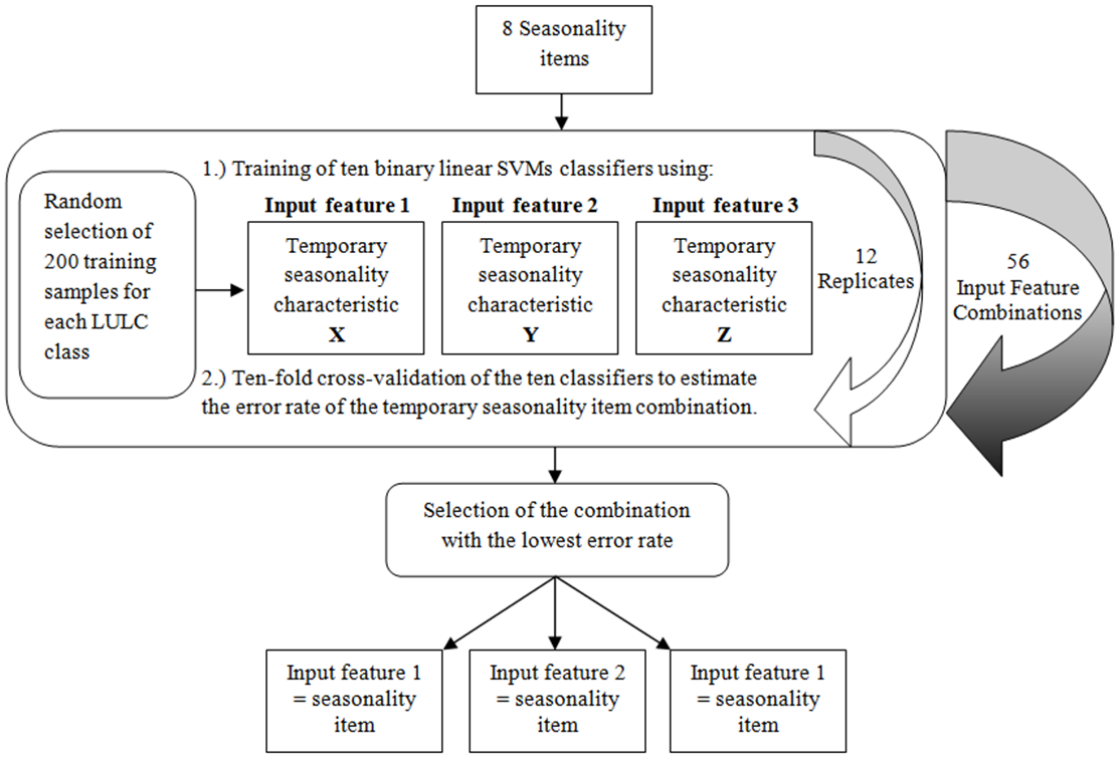

Figure 22 Simplified workflow of the input feature selection.

Each unique combination went through the training and validation part of linear soft margin SVC, since linear SVC needs just the definition of the penalty parameter value (*C*). The selection of the penalty parameter value was based on a constraint[7], a randomly chosen value between 0 and 1, that is $C = 0.2$. Non-linear SVCs incorporate more parameter definitions as described in *Section 3.3.3* and would create a more complex input feature selection, since the kernel related parameter needs to be determined. It was aimed to avoid the introduction of more assumptions and constraints.

This study is dealing with five LULC classes ($k = 5$ classes), therefore was applied the training and validation part of a one-against-one SVM-C (i.e. multi-classification) strategy. Thus, ten binary classifiers cover all possible LULC combinations (i.e. Forestland vs. Grassland etc.) in order to test each unique input feature grouping. The training and validation frame in Figure 22 is described thoroughly in Figure 23.

[7] The decision of that constraint was taken intuitively in order to keep the input feature selection simple. In theory a penalty parameter value can be each possible positive or negative real number.

Figure 23 presents the developed determination of the final error rate of each input feature combination using the training and validation of a SVC classifier.

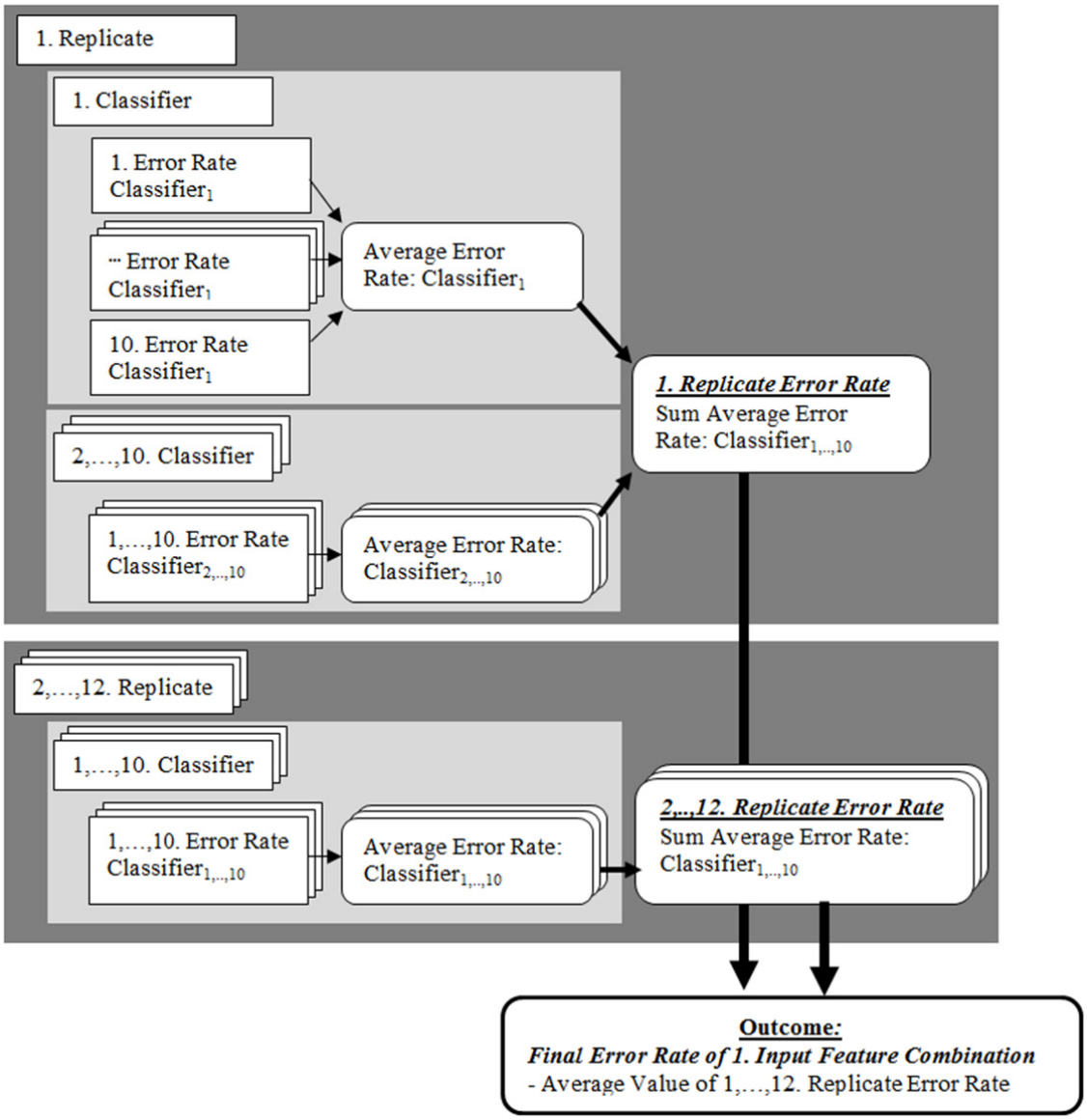

Figure 23 Determination of the final averaged error rate of an input feature combination. The Workflow considers each of the twelve replicates and each of the ten binary classifiers within a training sample replicate. All ten binary classifiers are trained and ten-fold cross-validated that leads to ten error rates for each classifier. The ten error rates of a classifier are averaged. This is repeated for all binary classifier. The resulting ten averaged error rates are summed and assigned as an error rate to a replicate. This is repeated for all 12 replicates until each replicated has a summed error rate. Finally, the summed error rates of all replicates are averaged and used to determine the final error rate of an input feature combination.

In detail, each replicate of the twelve training sample replicates contains ten binary classifiers in order to consider a multi-classification (see *Section 3.3.5*). The randomly chosen training samples were inserted as input features into the feature space of the matching ten binary classifiers. Each of the ten binary classifiers was trained and validated with the ten-fold cross-

validation using a leave-one-out strategy. This strategy divides a training sample in ten equal parts; nine parts were used to train the binary classifier and one part to validate the trained classifier. This continued until each part was used as a validation sample. Hence, ten error rates were estimated and then averaged over the ten trials in order to achieve a representative error rate (see light grey box in Figure 23). All ten binary classifiers were trained and validated in the same manner. Then, the ten averaged error rates of each classifier were summed to express the performance of the training data replicate (compare dark grey box in Figure 23).

As mentioned in *Section 5.3.1* there are concerns about the MODIS data, modeled seasonality and CLC 2006 data quality (assumption of 15% inaccuracy) and a selection of 200 training sample may not be representative. To overcome this, twelve training data replicates were employed to the training and validation of an input feature combination.

The training pixel positions of each replicate were saved in order to apply the same training samples to all input feature combinations. Therefore, reasonable comparisons of each input feature combination performance are guaranteed.

Finally, the twelve error rates of the replicates were averaged to a final error rate for each input feature combination to assign a measure of performance.

This training and validation methodology was applied to all 56 input feature combinations in order to calculate 56 representative measures of performance - final error rates. In the end, the final error rates were compared and the input feature combination with the lowest average error rate was chosen for the subsequent multi-classification approaches – see Figure 22.

5.3.3 Parameter Selection - Grid Search

The chosen three SVC strategies (linear, polynomial kernel and RBF kernel) require pre-defined parameters in order to classify the data accurately. The three SVC strategies that were tested in this study and their user-defined input parameters are listed below:

1. Linear soft margin classification requires a penalty parameter value (*C*).
2. Non-linear classification using polynomial kernel involves a penalty parameter value (*C*) and polynomial order value (d).
3. Non-linear classification using Gaussian radial basis (RBF) kernel function[8], which requires a penalty parameter value (*C*) and a RBF sigma value (σ).

These parameter values may be dependent on the input feature dataset (Tso and Mather 2009) and the values affect the SVC performance. Thus, a grid search, as described in *Section 3.3.4*, was applied to the chosen SVC strategies in order to determine their best user-defined parameters. The grid search anticipates the training and validation part of SVM-C using the validation error rate to measure the quality of the user-defined parameter performance.

The three SVC strategies with changing parameter settings were tested according to the workflow in Figure 24. As Figure 24 visualizes the grid search involves the definition of the individual search space extents for the linear and the two non-linear classification strategies. In case of linear classification a one-dimensional grid space was constructed with increasing penalty values. In case of non-linear classifications two-dimensional grid spaces were created with increasing parameter values, i.e. penalty values in one direction and kernel function related parameter values in the other direction – see Figure 15.

Then, a stratified random selection of 200 training pixel positions for each LULC class was incorporated into the grid searches. The best performing input feature combination (i.e. Maximum Season, Left Derivate and Small Integral) from the previous input feature selection (see *Section 5.3.2*) was used to associate three seasonality characteristic values to each training sample (i.e. 3x 200 seasonality characteristics). The resulting three-dimensional dataset constructed the feature spaces for the three SVM-C strategies.

In case of the non-linear SVM-C strategies, the coarse grid search started at the lower left corner and continued stepwise diagonally, with coarse increasing parameter values, till the

[8] The mathematical form of the polynomial and RBF kernel functions are not provided in the MATLAB help/tutorial for statistical learning functions of the bioinformatics toolbox. However, the author can provide the equations if asked for according to personal communication with a MathWorks Technical Support Engineer (Hempel 2010, personal communication). Authors email address: Florian.Sallaba(AT)nateko.lu.se

upper right corner was reached as shown in Figure 15 (blue line). The corner values defined the chosen extent of the grid spaces. All minimum and maximum values of the parameters and their increment steps are listed in Table 7. The coarse approaches were evaluated for the best performing parameter settings. In the linear case the grid search went stepwise from the lowest to the highest defined value in the one-dimensional space.

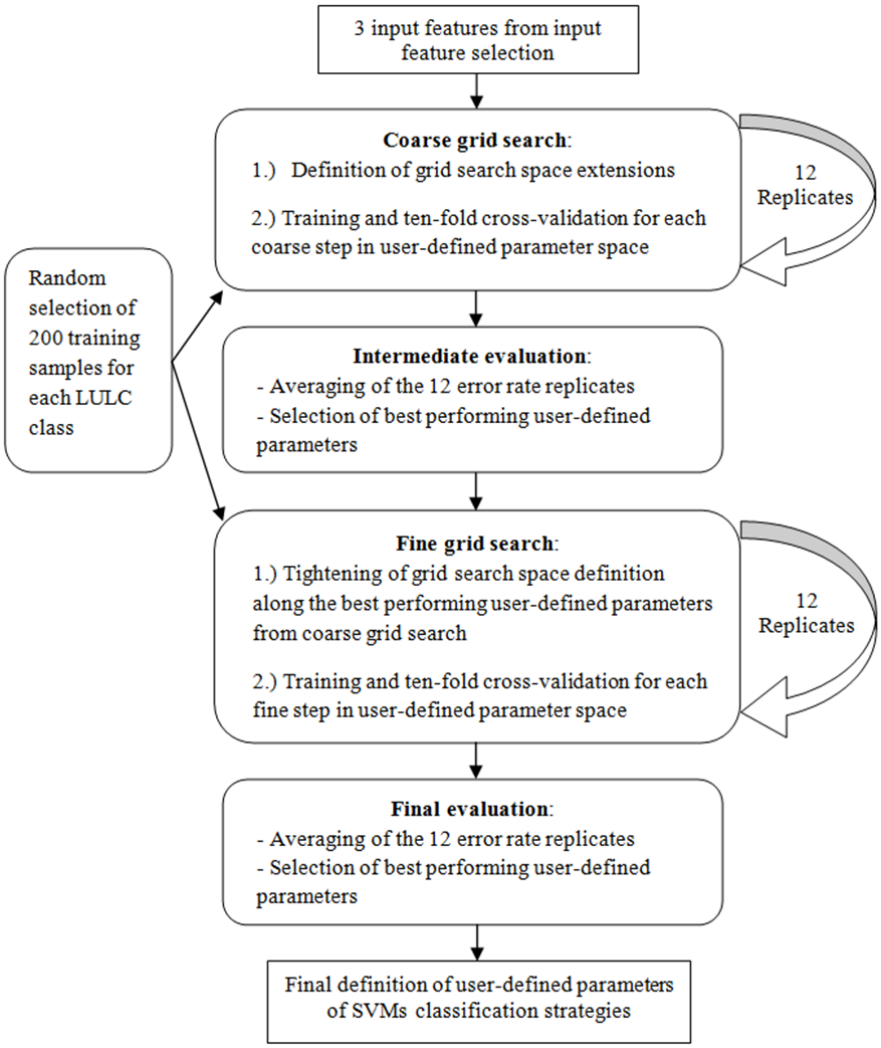

Figure 24 Simplified workflow of the applied grid search.

Each stepwise change incorporated the training and validation part of a one-against-one SVM-C using the corresponding SVC strategy. For each stepwise change ten binary classifiers that cover all possible LULC class combinations were trained and ten-fold cross-validated. In detail, the ten-fold cross-validation led to ten error rates for each binary classifier that were averaged over the ten-fold trials. The resulting averaged error rates of each binary classifier were summed and assigned to each stepwise change i.e. user-defined parameter groupings.

Twelve training data replicates were applied in the same manner to each stepwise change within the coarse grid search as visualized in Figure 24. This was done in order to overcome concerns about the statistical representativeness, and incorrect modeled seasonality due to uncertainty of the MODIS data and the thematic accuracy of the CLC 2006 reference data.

Thus, each stepwise change obtained twelve error rates that were averaged to a final error rate. The training pixel positions of each replicate in the coarse search were saved and applied to the next parameter grouping test in order to guarantee reasonable comparisons of the final error rate of each stepwise change.

All final error rates were linked to the corresponding stepwise changes and ranked in an intermediate evaluation. The user-defined parameters with the lowest error rate were chosen for the fine grid search.

Table 7 Ranges of the penalty parameter (C), polynomial order (d) and RBF sigma (σ) values within the coarse grid search.

User-defined parameters	Minimum Value	Maximum Value	Stepwise Increase
Penalty parameter **(C)**	**0**	**25**	**1**
Polynomial order **(*d*)**	**1**	**6**	**1**
RBF sigma **(σ)**	**0**	**20**	**1**

In the fine grid search, the definitions of the grid spaces were tightened around the best performing parameter value pairs from the coarse grid search – see workflow in Figure 24.

The methodology of the fine grid search is identical to the one in the coarse grid search except that the stepwise changes were taking place on a finer scale. In Figure 15 (black line) is shown the approach of the fine grid search.

In the non-linear SVC cases, it started at lowest value pairs in the grid space and continued diagonally until the highest value pairs were reached. The linear SVC fine grid search took place in a one-dimensional space.

Each stepwise change in the fine grid search was evaluated as in the coarse grid search; averaging the ten-fold error rates for each binary classifier, summing the ten binary classifier error rates and incorporating twelve training data replicates to calculate a representative final error rate for each parameter value pair (compare Figure 24). The new stratified random selections of the training pixel positions in the fine search were saved and applied to the next parameter grouping test in order to guarantee reasonable comparisons of the final error rate of each stepwise change.

In the end, the final error rates of each stepwise change were ranked and compared. The best performing parameter value pairs of the non-linear SVC strategies and the best penalty parameter of the linear SVC were chosen for the final SVM-Cs. The best user-defined parameters are shown in Table 8.

Table 8 User-defined input parameters for SVC

SVC strategy	Penalty parameter (C)	Polynomial order (*d*)	RBF sigma (σ)
Linear soft margin	**10.1**	**---**	**---**
Non-linear polynomial	**0.2**	**2**	**---**
Non-linear RBF	**8.6**	**---**	**7.9**

5.4 Multi-classification

The determinations of user-defined parameter settings and input features enable the final SVM-C separation of the study area's data. In general, a SVM-C consists of two parts, the construction of the classifiers via training and validation; and then the application of the classifiers to the unknown data points that have not been used for training.

The linear and non-linear SVC strategies were applied to a multi-classification (SVM-C), using the one-against-one method. Thus, ten binary one-against-one classifiers were trained in order to satisfy all possible LULC class combinations. A unique DAG containing 4 decision levels was tailor-made and implemented in MATLAB for the SVM-C problem of this study as shown in Figure 25.

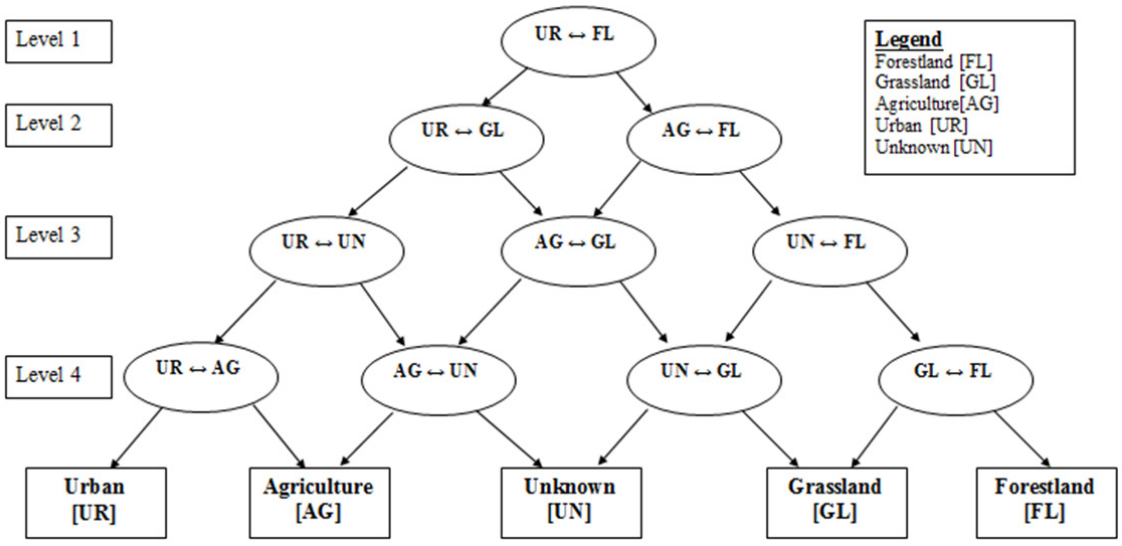

Figure 25 Tailor-made directed acyclic graph containing ten classifiers and five classes

200 training pixel positions for each LULC class were chosen randomly in a stratified manner. The best performing input feature combination were taken to link the three input

feature (i.e. seasonality characteristics) values to each training pixel position. The resulting three-dimensional datasets were used to build the feature spaces of all binary classifiers. Then, the ten binary classifiers of each SVC strategy were trained with the linked input features and the best performing user-defined parameters (e.g. C, d or σ) shown in Table 8 from the grid search (see *Section 5.3.3*).

A leave-one-out strategy (as described above) was applied during the training step. This ensured that all training data in the folds were used for training and validation and thus over-fitting concerns, i.e. reduction of generalization capability on unknown data points, were taken into account.

The ten constructed classifiers of each classification strategy were inserted into the conceptual DAG (see Figure 25) to guarantee that each data point will be tested for being a member in all LULC classes. The SVM-C of the unknown data points, which have not been used for training purposes, started one-by-one at the root (1st level) of the DAG. Here, each unknown data point was tested for being a member of the two LULC classes in the corresponding binary classifier. Depending on the decision the data point went to a classifier in the next level and was tested again until the 4th level was reached in order to determine the final decision. The leaves define the final decision of a data point being a member of a LULC class as shown in Figure 25. In the end, the final decision of LULC was assigned to the corresponding pixel position.

Twelve replicates were employed for each SVM-C strategy (linear, polynomial kernel and RBF kernel) because of statistical representativeness, and concerns about uncertainty of the MODIS data and the modeled seasonality as well as the thematic accuracy of CLC 2006 (15% inaccuracy) reference data. In detail, each replicate contained new stratified random selections of training pixel positions and the unknown dataset. The training pixel positions of each replicate were stored and applied to the other SVM-C strategies in order to guarantee reasonable comparisons of the different SVM-C outcomes in the subsequent accuracy assessment. However, all replicates of the classifiers had the same user-defined parameter settings as shown in Table 8.

5.5 Accuracy Assessment

The twelve replicate outcomes of each SVM-C strategy were assessed for their accuracy using the CLC 2006 reference data (*in Section 3.4.1*). Each classified pixel position was evaluated with the corresponding pixel position in the reference data. Those pixel positions that have been used in training step of each SVC replicate were excluded from the assessment. The accuracy assessments contained four quality measures. Overall accuracy, Cohen's Kappa value, user's and producer's error were calculated as described *in Section 3.4.2*. The estimation of the unweighted Cohen's Kappa in MATLAB incorporated a Cohen's Kappa function programmed by Cardillo (2007). The quality measures of each replicate were averaged over the twelve trials and used for the presentation and interpretation of the results.

6. Results

6.1 Seasonality Modeling

Figure 26 shows modeled vegetation seasonality (brown line) based on the NDVI time-series (blue line) from Figure 3. The time-series covers a period over five years of a chosen pixel in the MODIS NDVI time-series of the study area.

The outliers during the growing season have been smoothed out to a certain degree by help of the weight settings and the local Savitzky-Golay method. The growing seasons have almost similar patterns over the time period. Whereas, the winter seasons are clearly influenced by outliers (e.g. spikes) and this leads to dissimilar patterns compared to the growing season patterns.

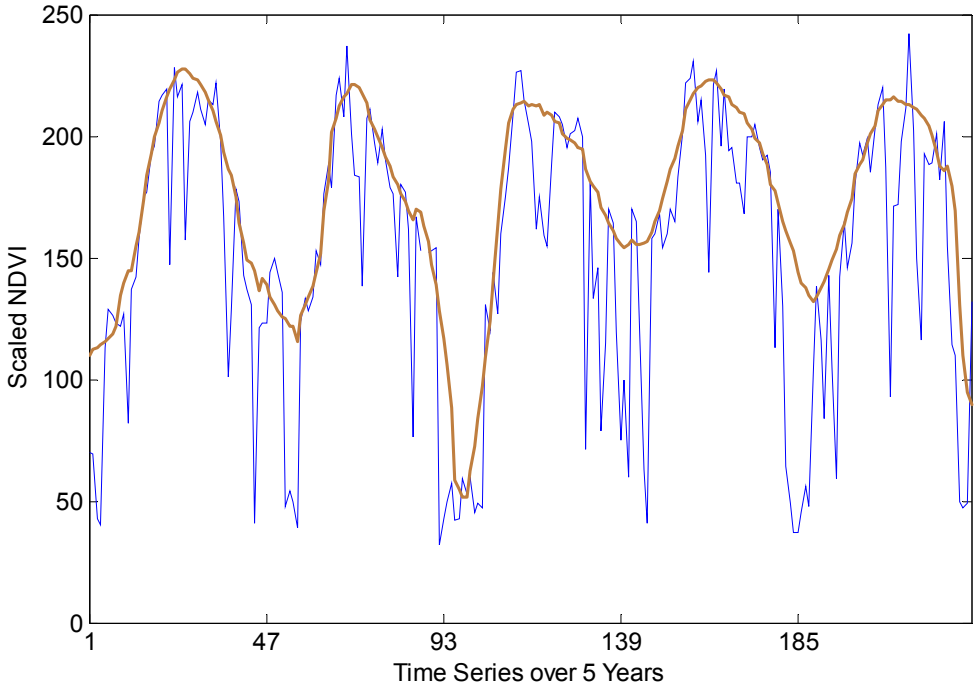

Figure 26 Savitzky–Golay modeling of vegetation seasonality derived from a NDVI time-series.

Different modeling curves based on NDVI time-series were compared and show in most of the cases an acceptable modeling of the growing seasons. However, the winter seasons are generally influenced by noise and may lead to erroneous seasonality parameter values (i.e. start and end of season).

6.2 Input Features

The selection of the three best performing input features in a linear soft margin SVC training (penalty parameter value of 0.2) led to 56 averaged final error rates. All final averaged error rates were ranked according to their performances. The top five input feature combinations are shown in Table 9. The best performing combination consisted of Maximum of Season, Left Derivate and Small Integral with an average error rate of 1.143.

Table 9 Ranking of the top five performing input features

Rank No.	Input Feature 1	Input Feature 2	Input Feature 3	Average Error Rate
1	Maximum Season	Left Derivate	Small Integral	**1.143**
2	Maximum Season	Left Derivate	Derivate Ratio	**1.155**
3	Maximum Season	Right Derivate	Derivate Ratio	**1.168**
4	Maximum Season	Left Derivate	Right Derivate	**1.178**
5	Maximum Season	End of Season	Small Integral	**1.205**

The seasonality characteristics Maximum of Season and Left Derivate appear the most in the top five ranking and seem to be a good choice for input features in a SVM classification. The highest ranked combination was applied to the grid searches of the three SVC strategies.

A test of a unique input feature combination using linear SVM-C led to a computation time of approximately 30 minutes. The computations were conducted in MATLAB (R2010a – 32bit version) on a Windows 7 (32 bit) operating system on a Notebook with an Intel Core 2 Duo processor (T66) and 4096 DDR3 1066 MHz RAM.

6.3 Parameter Selection - Grid Search

The outcomes of the best performing user-defined input parameters of the parameter selection for the three SVC training strategies are shown in Table 8 (*Section 5.3.3*). Table 10 visualizes the top five performing user-defined parameters in order to show how the error rates differ and to enable the recognition of trends in specific value combinations.

Table 10 Top five performing user-defined parameters in the grid search using all three SVC strategies. Penalty value (C), polynomial order *(d*) and RBF sigma (σ) are related to the SVC strategies. The error rates are summed over the ten classifiers of the multi-classification.

Rank No.	Linear		Polynomial		RBF	
	C	Error Rate	C & *d*	Error Rate	C & σ	Error Rate
1	10.1	1.146	0.2 & 2	1.108	8.6 & 7.9	1.045
2	10.4	1.148	1.7 & 2	1.113	6.6 & 7.7	1.048
3	10.0	1.149	19.8 & 2	1.114	8.6 & 8.1	1.049
4	10.7	1.149	0.4 & 2	1.114	9.6 & 7.5	1.051
5	10.9	1.149	0.9 & 2	1.114	8.6 & 7.7	1.051

The differences between the error rates of each SVC strategy are small and are established in the second and third decimal place. The error rates of each SVC training strategy show trends towards specific user-defined parameter combinations. The linear soft margin SVC strategy has the best performing penalty values between 10.1 and 10.9.

The non-linear SVC, using polynomial kernel, has the lowest error rates having a polynomial order of two; however the penalty values are ranging widely and do not show any trend. Non-linear SVC, incorporating RBF kernel, have sigma values between 7.5 and 8.1; whereas the penalty values are ranging between 6.6 and 9.6.

A comparison of all SVC strategies indicates that non-linear multi-classifiers using RBF kernel perform slightly better than the polynomial kernel under the given circumstances.

In addition, it is necessary to present how the best performing user-defined parameter combination has performed on the binary classifier level. In Figure 27 are visualized the average error rates of the best user-defined parameters of all SVC training strategies. Binary classifiers that involve LULC class Unknown have error rates of zero and are not shown in Figure 27.

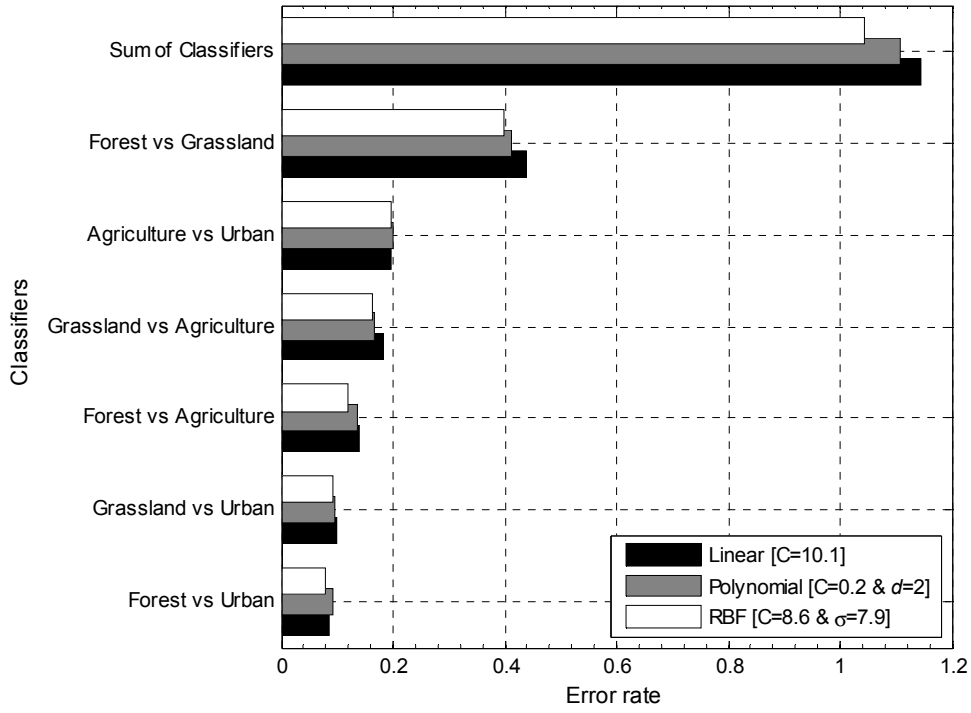

Figure 27 Average error rates of the best performing user-defined parameters for each of the ten classifiers applied on the three SVC strategies.

Each SVC strategy applied on the binary classifiers led to similar trends of error rate values. The highest error rates of approx. 0.4 occur in the binary classifier Forest vs. Grassland.
All other binary classifiers have error rates below 0.2. The error rates are quality measures of the binary classifiers. Thus, the high error rates of the binary classifier "Forest vs. Grassland" during the parameter selection may indicate misclassifications between the classes Forestland and Grassland in the subsequent SVM-C.

A stepwise parameter change within the grid search (including twelve replicates) for the linear classifier took approximately 30 minutes of computation, for the RBF non-linear classifier ~ 45 minutes and for the polynomial classifier ~60 minutes. The computations were conducted in MATLAB (R2010a – 32bit version) on a Windows 7 (32 bit) operating system on a Notebook with an Intel Core 2 Duo processor (T66) and 4096 DDR3 1066 MHz RAM.

6.4 Accuracy Assessment

The SVM-C results of the linear and non-linear strategies were assessed for their accuracy in order to determine the best SVM-C strategy for the input feature combination (Maximum Season, Left Derivate and Small Integral).

Overall accuracy

Table 11 displays two cases of the average overall accuracies of the applied SVM-C strategies. The four class case does not consider the LULC Unknown in the overall accuracy estimation, since its seasonality characteristics have not been modeled in the seasonality modeling due to the mask application. The seasonality characteristics values were uniformly set to zero. The five class case takes LULC class Unknown into account during overall accuracy calculations.

Table 11 Overall accuracies of the applied SVM-C strategies considering the averages, ranges and standard deviations. The four class case excludes the LULC Unknown, since its seasonality characteristics values have not been calculated during the seasonality modeling and were just set to zero.

Overall Accuracies			
Four classes "Unknown class" excluded	RBF [C=**8.6** & σ=**7.9**]	Polynomial [C=**0.2** & d=**2**]	Linear [C=**10.1**]
Average [%]	**63.80**	**63.52**	**60.78**
Range [%]	**4.79**	**4.86**	**5.06**
Standard deviation [%]	**1.68**	**1.89**	**1.53**
Five classes	RBF [C=**8.6** & σ=**7.9**]	Polynomial [C=**0.2** & d=**2**]	Linear [C=**10.1**]
Average [%]	**66.54**	**66.28**	**63.75**
Range [%]	**4.43**	**4.49**	**4.67**
Standard deviation [%]	**1.55**	**1.74**	**1.41**

Four class case

In the four class case, the achieved average overall accuracies of the twelve replicates are ranging between ~61% and ~64% and the differences are small. RBF kernel multi-classification has an averaged overall accuracy of 63.80% and performed slightly better than the polynomial kernel with 63.52%. The linear SVM-C is less accurate with 60.78%; however, the difference to non-linear SVM-Cs is about approx. 3% and is considered to be small.

The overall accuracies ranges of the employed SVM-C strategies are at approximately 5%. The RBF kernel SVM-C has the smallest range of 4.79%. The polynomial kernel SVM-C has an overall accuracy range of 4.86% and the linear SVM-C of 5.06%. Nevertheless, the differences of the overall accuracy ranges between the SVM-C strategies are marginal.

Considering the standard deviations of the overall accuracies, linear SVM-C has the lowest value of 1.53%, followed by RBF kernel SVM-C with 1.68% and polynomial kernel SVM-C has the highest standard deviation with 1.89%.

Five class case

The five class case gives the same pattern of overall accuracies for the corresponding SVM-C strategies as the four class case. The average overall accuracies are inflated by approximately 3%. The RBF kernel has the highest average overall accuracy with 66.54%, the polynomial kernel SVM-C is marginal lower and the linear SVM-C has the lowest average overall accuracy of 63.75%. The overall accuracy ranges are smaller in the five class case. The RBF and polynomial kernel SVM-C have approximately the smallest range of ~4.4%. Linear SVM-C has the largest range of 4.67%, however the lowest standard deviation of 1.41%. RBF kernel SVM-C has the second lowest standard deviation of 1.55% and polynomial kernel SVM-C the highest with 1.74%.

Producer's and user's accuracy

Table 12 displays average producer's and user's accuracies of all LULC classes of the applied SVC-M strategies. In Table 12 the five class case is considered, since the producer's and user's accuracies are theoretically the same for the four class case. LULC Unknown has excellent producer's and user's accuracies that an exclusion of Unknown will lead to almost same percentages of the other LULCs (compare *Section 3.4.2*).

Table 12 Average producer's and user's accuracies of the five LULC classes of the selected SVM-C strategies with following parameter settings Linear [*C*=10.1], polynomial kernel[*C*=0.2 & *d*=2] and RBF kernel [*C*=8.6 & σ=7.9] SVM-C.

Producer's accuracy	Linear [%]	Polynomial [%]	RBF [%]	**User's accuracy**	Linear [%]	Polynomial [%]	RBF [%]
Unknown	**97.62**	**98.53**	**98.90**	Unknown	**91.68**	**96.20**	**95.88**
Forestland	**88.29**	**86.88**	**86.45**	Forestland	**58.85**	**59.00**	**59.97**
Agriculture	**83.70**	**83.50**	**83.20**	Agriculture	**71.93**	**72.87**	**72.80**
Urban	**25.11**	**26.12**	**25.28**	Urban	**73.32**	**72.05**	**73.24**
Grassland	**3.65**	**4.32**	**4.20**	Grassland	**45.54**	**46.39**	**43.37**

The Unknown class has producer's accuracies of ~98- ~99% of being classified correctly considering the reference data and user's accuracies between ~92 - ~96% of being assigned correctly in the classified data. The outcomes of LULC class Unknown are excellent, since it achieved zero values for each seasonality characteristic, and thus an optimal separation

between other values can take place. However a marginal overestimation in the classified data takes place.

The producer's accuracy of LULC class Forestland shows that it has been mapped correctly with accuracies between ~86 - ~88% in the reference data, i.e. CLC 2006 and ~12-~14 % of Forestland pixels have been committed to other classes. On the other hand, the user's accuracies between ~59 - ~60% show how well Forestland has been assigned in the classified data. In other words, 40% of Forestland pixel positions have been classified wrongly. Nevertheless, the results of Forestland are moderate even though it is overestimated in the classified dataset.

Agriculture is mapped correctly in the reference data with producer's accuracies of approximately 84%; however in the classified data it has user's accuracies of approximately 73 %. In other words, approximately 27% of Agriculture pixel positions have been classified inaccurately. In this study Agriculture has been overestimated the least and the results can be considered as good.

LULC class Urban has low producer's accuracies of approximately 25% considering the small amount of available reference data; on the other hand, it has good user's accuracies of ~73% being mapped correctly in the classified data. That means a considerable underestimation of Urban took place in the classified data.

Grassland has poor producer's accuracies of approximately ~4% taking the small amount data to reference with and moderate user's accuracies between ~43 - ~46% in the classified data. Grassland has been underestimated in the classified data.

Kappa analysis

Table 13 displays the average Cohen's Kappa values of the employed SVM-C strategies. The four class case excludes the LULC Unknown in the Kappa analysis, since its seasonality characteristics values were uniformly set to zero. The five class case takes LULC Unknown into account during the Kappa analysis.

Four class case

The best performing strategy is the RBF kernel SVM-C with an average Kappa value of 0.45, followed by the polynomial kernel with an average value of 0.44. The linear classification obtained an average Kappa value of 0.42. The ranges of the Kappa values are equal in all SVM-C strategies. A lower standard deviation of the Kappa value is estimated for the linear SVM-C with a Kappa value of 0.01. However, the Kappa value standard deviations of both non-linear SVM-Cs are equal with 0.02. According to Congalton and Green (2009) Kappa

values greater than 0.40 show moderate agreements. In other words, the estimated Kappa values above 0.40 of this study are moderate but give acceptable results. However, the average Kappa value differences between the applied SVM-C strategies are marginal.

Table 13 Average Cohen's Kappa values of the applied SVM-C strategies considering the averages, ranges and standard deviations. The four class case excludes the LULC Unknown, since its seasonality characteristics values were set to zero.

Cohen's Kappa values			
Four classes "Unknown" excluded	RBF [C=**8.6** & σ=**7.9**]	Polynomial [C=**0.2** & d=**2**]	Linear [C=**10.1**]
Average	**0.45**	**0.44**	**0.42**
Range	**0.05**	**0.05**	**0.05**
Standard deviation	**0.02**	**0.02**	**0.01**
Five classes	RBF [C=**8.6** & σ=**7.9**]	Polynomial [C=**0.2** & d=**2**]	Linear [C=**10.1**]
Average	**0.52**	**0.52**	**0.49**
Range	**0.05**	**0.05**	**0.05**
Standard deviation	**0.02**	**0.02**	**0.01**

Five class case

The RBF kernel and polynomial kernel SVM-C performed similar with average Cohen's Kappa values of 0.52. The linear classification achieved an average Kappa value of 0.49. The ranges of the Kappa values are similar for all employed SVM-C strategies with a value of 0.05. Linear SVM-C has the smallest standard deviation with 0.01 and the non-linear strategies have Kappa value standard deviations of 0.02.

The average Kappa values state the SVM-C methods agree moderately with the CLC 2006 reference data (Congalton and Green 2009). Hence, the performances of SVM-Cs in this study are acceptable considering that the differences of Kappa values are marginal.

7. Discussion

7.1 Data

The high temporal resolution of MODIS data offers possibilities to monitor seasonal development on eight-day resolution and to model seasonality. However, the ground resolution of 250m leads to spectral mixture problems referred as mixed pixel in satellite data classification (Lillesand et al. 2004, Foody and Mathur 2006, Richards and Jia 2006, Tso and Mather 2009). In other words, in a MODIS pixel is aggregated spectral information of several LULC classes and that may be difficult to interpret for the analyst as well as the classification algorithms. Especially agricultural land use with small units or changing LULC types within short distances are leading to mixed pixels. Even though the chosen study site is dominated by 90% of agricultural and forestland, there are small areas with changing LULC classes at a small scale according to the CLC 20069 map. In detail, Grassland islands in Forestland may lead to misinterpretation as well as Forestland areas next or within Agricultural land.

In addition, a MODIS pixel may be influenced to a certain percentage by its neighboring pixels (Tan et al. 2006). Particularly in heterogeneous landscapes, this may lead to mixed pixel and to subsequent misinterpretations.

There are trade-offs that the analyst has to deal with using high temporal resolution at the cost of spatial resolution. For this study, the achieved trade-off is acceptable, since it is assumed that high temporal resolution will add valuable information during the seasonality modeling, and thus the determination of the seasonality characteristics. MODIS satellite data has been stated as a well suited data source for LSP studies (Reed et al. 2009) and Jönnson et al. (2010) applied successfully MODIS data in their phenology study about Swedish coniferous forests. An additional advantage of MODIS data is that they are free of charge.

7.2 Reference Data

CLC 2006 was used as reference data for the training and validation during the input feature selection, user-defined parameter determination as well as the actual multi-classification and its accuracy assessment. Consequently, the study is dependent on the thematic accuracy of CLC 2006, which is assumed to be 85% (EEA 2007).

As mentioned before CLC 2006 250m ground resolution is based on satellite imagery from sensors like Landsat TM or ETM+ having fine spatial resolutions. Hence, the original resolutions were generalized and aggregated to a coarser of 250m ground resolution. During such procedures information get lost and this may led to additional error sources for this study. It is assumed that the LULC classes of CLC 2006 data in the study area are assigned wrongly by a probability of 15%. Thus, the incorporation of CLC 2006 into this study leads to a consistent error source that might falsify the training and validation of each SVM-C strategy and their accuracy assessments.

In addition, CLC 2006 was reprojected to the MODIS sinusoidal projection with a nearest neighbor transformation. In such a way the new calculated raster cells were influenced by its neighboring cells. The reprojection procedure had influence especially on raster cells that were surrounded originally by cells with different LULC classes. Thus, the neighboring cells with different LULC might have led to an assignment of a new LULC class to a raster cell and the original LULC value could have disappeared. However, nearest neighbor transformation alters the less the original input raster cell compared to other resampling methods such as bilinear interpolation or cubic convolution transformation (Lillesand et al. 2004).

One way to deal with these concerns of satellite based reference data is to collect field data for training and validation purposes as described in Congalton and Green (2009). However, it is difficult to collect field data points that are representing a grid of 250mx250m ground resolution in such a heterogeneous study area (compare Figure 18). A set of several field data points within a grid could be selected, and then upscaled with spatial statistical methods. However, the collection of the field data would be very time demanding.

The stratified random selection is a reasonable method to select training data for each LULC class. However the underrepresentation of Urban and Grassland LULC leads to the assumption that some pixel positions had been chosen several times over the twelve replicates. Therefore, the replicates may not be representative and contain wrongly assigned

reference data considering Urban and Grassland. That concern explains the poor producer's and user's accuracies of Grassland. A study area with larger proportions of Urban and Grassland needs to be tested in order to prove this concern.

The decision to overcome the concerns of representativeness and inaccuracy of CLC 2006 reference data with twelve replicates is assumed to be acceptable. It gives representative results and weakens the influence of wrongly assigned LULC classes in the training and reference data.

7.3 NDVI

The feasibility of NDVI time-series to capture phenological events has been discussed in Reed et al. (2009). Disadvantages of NDVI are the sensitivity towards fully and partly soil covered surfaces. The study area contains of approx. 43% agricultural land that gives soil dominated spectral signatures over certain time periods i.e. beginning of season or after harvesting. There are several vegetation indices that can cope with soil covered surfaces or soil background as source of noise. The weighted difference vegetation index (WDVI) or the soil adjusted vegetation index (SAVI) might be better choices for these areas. Reed et al. (2009) mention that SAVI is more linked to structural elements such as biomass.

Furthermore NDVI saturates at a certain level and increasing foliage biomass cannot be represented with NDVI. Wide dynamic range vegetation index (WDRVI) has been reported to reduce the sensitivity towards saturation effects in northern regions (Jönsson et al. 2010). Also the enhanced vegetation index (EVI) has been designed to improve sensitivity in high foliage biomass regions and to decrease background signals and atmospheric influences. Hüttich et al. (2009) applied effectively EVI to model seasonality data in combination with in-situ botanical survey data to map vegetation types in dry Savanna ecosystems.

However in this study, the NDVI time-series are assumed to be appropriate for seasonality modeling approaches and seasonality characteristic extraction. An investigation of the performance of the mentioned vegetation indices in the study area could give comparable results. It would be useful to study how different vegetation indices e.g. WDVI or WDRVI would influence the seasonality modeling, seasonality characteristic extraction and the subsequent SVM-C.

7.4 Seasonality Modeling

The chosen Saviztky-Golay method is considered to be a local fitting and more sensitive to noise (Jönsson et al. 2010) and leads to spikes and peaks (see Figure 26). Methods that keep the NDVI integrity i.e. double logistic fitting may give outcomes that are not that much influenced by noise. On the other hand, such a global approach may lead to information loss of rapid changes such as harvesting of crops. The Saviztky-Golay method is feasible to detect rapid changes that may be useful information for the classification of seasonality parameters. However, it may be interesting to study how other fitting methods will influence the seasonality parameter classification.

An enhancement of the MODIS quality products would help to improve the seasonality modeling in TIMESAT. The overall quality indicator used in this study helped to exclude the influence of cloud coverage and other noise sources during the growing seasons. Nevertheless, the quality indicator performed weak in the winter seasons.
Figure 26 shows how noise influences the modeling of a growing and winter season. The modeled minimum levels are different in each shown winter seasons (Figure 26 NDVI time-series points 47, 93,139 and 185). These different minimum levels, during the winter season, have an influence on the determination of most of seasonality parameters (e.g. season start, season end, season length and base level) and the subsequent SVM-C.
In general, smoothing of peaks and spikes with enhanced quality data can improve the seasonality modeling and the determination of the seasonality parameters.

The averaging of the seasonality characteristic values over the three observed years is an important discussion point in the data processing, which is discussed thoroughly in *Section 7.7*.

7.5 Feature Selections

The input feature selection is based on a ten-fold cross-validation for each possible combination of seasonality parameters. The ten-fold cross-validations of three input seasonality parameters were done over twelve replicates and all resulting averaged error rates were used for a quality measure of the unique combinations.

The range of the error rates differs over the twelve trials; it is caused by the stratified random selection of the training data. This indicates that the training chosen training data contained wrong assigned LULC classes from CLC 2006 considering the probability of 15% (EEA 2007) of choosing a wrong assigned LULC class. Furthermore, it may show the influence of mixed pixel (Lillesand et al. 2004, Foody and Mathur 2006, Richards and Jia 2006, Tso and Mather 2009) that are difficult to handle in the seasonality modeling.

The developed and implemented feature selection of this study is related to the quality of the CLC 2006 reference data. The errors within the reference data might have a strong effect on the LULC classes that are not that well represented in the study data i.e. Grassland and Urban. Since the probability of choosing the same wrong reference point is higher in a smaller dataset than in a larger. This could lead to high error rates of the binary classifiers that involve Grassland and Urban (e.g. "Forestland vs. Grassland" or "Agriculture vs. Urban"). Figure 27 states the highest error rates in the binary classifiers of "Forestland vs. Grassland" or "Agriculture vs. Urban". A study area with larger proportions of Urban and Grassland could be tested in the developed feature selection framework in order to prove this concern. The developed feature selection gives a starting point in picking out the most representative seasonality characteristics as input features. However, the method is complex and leads to high computational burden. Using the validation error rate is an experimental approach that is assumed to give the best guess, since there is no established way to determine the best seasonality characteristics for SVMs classification.

In addition, the penalty parameter value was randomly chosen from a range between 0 and 1, that is $C = 0.2$. A different randomly chosen penalty parameter value, from a larger range, might have led to a different selection of seasonality characteristics.

Moreover, more than three input features might carry more information and may improve the ability to construct a separating hyperplane during SVC training. However, this study was constrained to incorporation of three input features, since the used Personal Computer[9] ran

[9] The computations were done in MATLAB (R2010a – 32bit version) on a Windows 7 (32 bit) operating system on a Notebook with an Intel Core 2 Duo processor (T66) and 4096 DDR3 1066 MHz RAM.

out of memory due to the quadratic optimization of the training samples in higher dimensional spaces.

An alternative to extract the most valuable information is the application of a principal component analysis. The principal component analysis applies transformations that are described with linear combinations of the original data values multiplied by transformation coefficients, also referred as eigenvectors (Lillesand et al. 2004, Richards and Jia 2006, Tso and Mather 2009). All seasonality characteristics could be introduced in a principal component analysis in order to reduce the amount of data. The outcomes would be a new uncorrelated description of the original data in several components. The amount of variation is maximized in first component and decreases successively in following components. The first components could be incorporated in the SVM-C strategies.

Tso and Mather (2009) suggest a feature selection method used for hyper-spectral data in order to determine the best input feature for a SVC. It adopts the concept of weight elimination used in ANN (Tso and Mather 2009).

Instead of applying statistical measures, it might be valuable to conduct a visual inspection of the modeled seasonality data and the analyst decides intuitively which seasonality characteristics to select. For example the seasonal development of Forestland shows an earlier onset of foliage growth, Higher Amplitudes and Larger Integrals of seasonal development compared to Grassland. These seasonality characteristics may be valuable for a SVC. However, this approach will lead to very subjective decisions of the analysis and is more of an art than a science.

7.6 Parameter Selection – Grid Search

The implemented parameter selection using the grid search approach performed well in order to locate the penalty parameter values and the kernel functions related parameters. Tso and Mather (2009) point out that the lowest error rate might not always be the best choice for the final parameter determination, since the trained classifier may not generalize towards unknown data samples. In this study it was tried to overcome these over-fitting concerns by applying the same parameter or parameter pairs to all ten binary classifiers in the SVM-C DAG and their following ten-fold cross-validation. In addition it was tried to avoid over-fitting issues using twelve replicates. In such a way, it is assumed that the calculated error rates are unbiased and not limited to over-fitting. However, the implemented grid search leads to high computational burden and is very time consuming even though the range of the coarse grid search was rather small – compare Table 7.

The results of the grid search for the polynomial kernel function and its penalty parameter value are interesting because the lowest error rates were achieved with a polynomial order value of 2. Tso and Mather (2009) state that increasing polynomial orders lead to a better fitting of the hyperplane and thus a more exact constructed classifier. In this study, a polynomial order of 2 leads to the lowest error rate. The penalty parameter values in this study are ranging between 0 – 11 among the three SVC strategies; compared to other studies (Pal and Mather 2003, Pal and Mather 2005, Dixon and Candade 2008, Kavzoglu and Colkesen 2009, Tso and Mather 2009) larger penalty parameter values have been applied to multi- and hyper-spectral satellite data. A widening of the parameter value ranges during the coarse grid search may improve the selection of the user-defined parameters. The top five performing parameter pairs of each applied SVC strategy have marginal different error rates (compare Table 10). In addition, the penalty parameter values in the polynomial kernel case show no specific trend and slight trend can be supposed for the RBF kernel penalty parameter. This leads to the assumption that are wider grid search considering the penalty parameter could improve the user-defined parameter selection but would be more time and computational demanding.

A different approach of using the grid search could be to apply an independent grid search to each of the ten binary classifiers within the SVM-C DAG. Thus, each binary classifier would be incorporated one-by-one into the grid search in order to find the best penalty parameter and kernel functions related parameter values. For reasons of representativeness twelve replicates need to be applied. Binary classifiers with higher error rates such as Forest vs.

Grassland and Agriculture vs. Urban (compare Figure 27) could be improved with tailor-made penalty parameter and the kernel functions related parameters. This more specific parameter calibration would lead to a more complex grid search combined with higher computational burden and time expenses.

Tso and Mather (2009) mention the gradient descent method as an alternative for the selection of user-defined parameters in SVM classification. The gradient descent method can involve even more complex kernel functions using an error surface function where the slope and the slope direction are used to evaluate the performance of the chosen user-defined parameters. However, the implementation of the gradient descent method is more complex and demands an efficient implementation in a programming language – compared to the grid search.

7.7 Support Vector Machine Classification

The seasonality characteristic of the LULC class Unknown were masked out by setting them to zero and were still introduced into the SVM-C DAG. Consequently the overall accuracies and Kappa values were inflated in the five class case. The LULC class Unknown was introduced for reason of simplifying the programming work in MATLAB. That steered to a very good separation between Unknown and the other LULC classes. This fact is underlined with zero error rates of the binary classifiers that involved the LULC Unknown, during the training and validation of SVM-C. Also the producer's and user's accuracies of Unknown are excellent and underline this discussion point. The exclusion of Unknown shows a small difference of 3 % between the four and five class cases considering the average overall accuracy due to its small proportion of 3% in the study area. This illustrates disadvantages using the overall accuracy (Lillesand et al. 2004, Congalton and Green 2009).

A comparison of the Kappa analysis between the four (exclusion of Unknown) and five class case shows differences of approximately 7% (Kappa value of 0.07), since the Kappa value considers the agreement occurring by chance. The Kappa analysis describes the strong influence of LULC Unknown. Thus it is necessary to describe and discuss the performance of this study in a conservative way using the overall accuracies and Kappa values of the four (Unknown excluded) case.

The results of the three SVC strategies show reasonable average overall accuracies (~61 - 64%) and moderate Cohen Kappa values (0.42 – 0.45). There are negligible differences between the non-linear SVC methods taking the average overall accuracies and Kappa values into account. The differences of the overall accuracies with 3% and the Kappa values with 3% between the linear and the non-linear strategies are marginal and the standard deviations show no significant variations of the values. Considering the spatial resolution of the used MODIS data (250m x 250m) and its vulnerability to the mixed pixel problem (Lillesand et al. 2004, Foody and Mathur 2006, Richards and Jia 2006, Tso and Mather 2009) the results are satisfying.

The results might be improved by incorporating a different input feature selection as discussed in *Section 7.5* and a broader grid search or a different method in the user-defined parameter selection as mentioned in *Section 7.6*.

The

A questionable methodological decision in this study is the assumption that no LULC changes took place in the study area over the three observed years. The assumption initiated

an averaging of the seasonality characteristics values over the three year period. However, different agricultural cultivation strategies and urban sprawl can defeat this assumption. If a LULC change took place in a certain area during the observed period the pixel would have very different seasonality characteristics values. So an average value over the three years would not represent the actual LULC class and could lead to a wrong interpretation of the average value in the corresponding pixel. LULC classes that could be affected by changes may be Forestland, Urban and Grassland; since these classes have large differences in producer's and user's accuracies (compare Table 12). Forestland could be deforested for reasons of bioenergy production or damaged by storm events. Grassland could be used for agricultural land use or vice versa. Growing urban areas need space and thus other LULC classes may become Urban.

An alternative to that chosen assumption is to classify the modeled seasonality characteristics of each year, to compare the classification outcomes and detect changes. This would also incorporate proper reference data for each of the observed years. However, the aim of this study is not a change detection analysis.

The producer's accuracies of all SVC strategies for the Grassland (~4%) and Urban (~25%) LULC classes are poor (compare Table 12). This indicates that these classes are difficult to separate; in case of Urban it can be caused by spectral mixture problems (Lillesand et al. 2004, Foody and Mathur 2006, Richards and Jia 2006, Tso and Mather 2009). Urban areas contain surfaces that are covered with artificial objects (i.e. concrete), soil or vegetation that leads to mixed pixels especially on a MODIS ground resolution of 250m. In case of Grassland it is assumed that NDVI cannot capture the difference in green biomass between Forest and Grassland due to saturation effects of NDVI (Reed et al. 2009, Jönsson et al. 2010). Figure 27 shows that the highest average error rates are located during the training phase in the binary classifier "Forest vs. Grassland".

In addition Urban and Grassland are underrepresented in the study area and may be assigned wrongly in the reference data set (see *Section 7.2*). A more heterogeneous study area with larger proportions of Urban and Grassland might lead to different results. However, this study was constraint to a focus on the separation of seasonality parameters of Forestland and Agriculture in order to minimize error sources of misclassifications.

Even though the results of the seasonality data classifications are moderate it may be useful to combine spectral MODIS (e.g. Band 1 or Band 2 - see Table 1), in combination with several VIs (e.g. EVI, NDVI or WDVI) and seasonality data in order to test synergetic effects

of the data sources and their performance in SVM classification. Hüttich et al. (2009) incorporated successfully MODIS and higher spatial resolution satellite time-series and in-situ botanical survey data sources to classify vegetation types in dry Savanna ecosystems.

Moreover, it is important to check if seasonality data is normal distributed in order to apply traditional supervised (i.e. MLC) and unsupervised (i.e. k-means clustering) classification techniques. Thus, it would be possible to test the performance of MLC on seasonality data.
If the traditional classifiers such as MLC or k-means clustering would lead to the equal classification accuracies it would be more feasible to employ a traditional classifier, since the complex user-defined parameter selection and the training of SVCs lead to high computational and time expenses.

In order to validate the performance of the linear, polynomial kernel and RBF kernel SVM-C it is preferable to select the linear classifier since a one-dimensional grid search within the user-defined parameter selection demands less computational time. The decision on a linear SVM-C would lead to a loss of ~3% considering the Kappa value and a loss of 2% overall accuracy. The trade-off between the losing accuracy and saving computational time is acceptable.

It is important to compare the results of this to other studies that have incorporated SVC. Considering performance comparisons between MLC and SVC, Kavzoglu and Colkesen (2009) classified successfully multi-spectral data (30m ground resolution); and achieved overall accuracies of 87% for MLC and 91% for SVC. In addition, Kappa values of 0.84 for MLC and 0.88% for SVC. Pal and Mather (2003 and 2005) classified multi-spectral data (30m ground resolution) and obtained better performances of SVC in comparison to MLC. A better performance of SVC in comparison to MLC could be assumed for the classification of seasonality characteristics. However, an incorporation of MLC to classify seasonality characteristics needs to be tested in order to prove this assumption.
Considering the use of MODIS generated NDVI data, Goncalves et al (2005) classified a NDVI images, in 500m ground resolution, using SVC with good results i.e. with overall accuracies up to 89%. In addition, Gu et al. (2008) applied SVC to MODIS NDVI data (500m ground resolution) using an extended form of principal component analysis and achieved overall accuracies of 84% and a Kappa value of 0.82. The mentioned studies achieved good results of SVC applied on different satellites sources with different ground resolutions. The results of this study are improvable in comparison to the mentioned studies

although a different kind of data was applied. However, the results of this study does not indicate a weak performance of SVC to separate seasonality characteristics and the potential of seasonality characteristics of being used for separation of LULC, in general. The comparisons to the other studies give the impression that there is potential for improving the developed methodology in this study. The producer's accuracy of Forestland (~88%) and Agriculture (~83) are good and slightly lower in comparison to other studies (Pal and Mather 2003, Pal and Mather 2005, Gu et al. 2008, Kavzoglu and Colkesen 2009). In addition, Agriculture achieved a good user's accuracy with ~72%. The selection of the study area was eventually based on Forestland and Agriculture and SVC has a good potential to separate these two LULC classes. However, the developed methodology of this study needs to be improved in order to increase the performance of SVC to separate Grassland and Urban more correctly. The raised discussion points, in this discussion section, on the development of the study's methodology should be taken into account.

Above all, the most discussable and at the same time improvable parts within the developed methodology are: 1) the averaging of the seasonality characteristics over three years, 2) the experimental based input feature selection and 3) rather small extents of the grid search in the parameter selection. 4) A missing comparison to a traditional classifier (i.e. MLC) and 5) simultaneously to check if seasonality data is normal distributed.

So far, the classification of seasonality data into LULC classes has not been tested extensively. Therefore, this study gives a start with its developed methodology incorporating SVM-C. The methodology needs to be improved and extended in continuing studies, since there is a potential of SVC for separation of modeled seasonality from MODIS NDVI time-series.

8. Conclusion

- In this study was implemented and developed successfully a methodology to classify seasonality data into LULC classes of the chosen study area. The linear, polynomial kernel and RBF kernel SVM-C techniques have been implemented into a multi-classification and successfully applied to separate the seasonality characteristics in LULC classes.

- The determination of the seasonality characteristics as input data for SVM-C is based on an experimental approach and is adapted from the user-defined parameter selection. The decision to use three input features was a trade-off between computational expenses, complexity and the accuracy of the classification outcomes.

- The three SVM-C (linear, polynomial kernel and RBF kernel function) technique outcomes have been validated in an accuracy assessment using CLC 2006. Among the SVM-C techniques performed marginally the RBF kernel function the best compared to the polynomial kernel function. However, it is suggested to apply a linear SVM-C, since it demands less computational time during the training of the SVM-C.

- This study shows that seasonality data can be used for separating LULC classes. There exists a good potential to apply SVM-C for separation of modeled seasonality characteristics from MODIS NDVI time-series. Nevertheless, it needs to be tested how the combination of spectral, VIs and seasonality data performs in a classification. It is assumed that synergetic effects may occur and thus improve a classification.

9. References

Bagan, H., Q. Wang, M. Watanabe, Y. Yang, and J. Ma. 2005. Land cover classification from MODIS EVI timesseries data using SOM neural network. International Journal of Remote Sensing **26**:4999-5012.

Baldridge, A. M., S. J. Hook, C. I. Grove, and G. Rivera. 2009. The ASTER spectral library version 2.0. Remote Sensing of Environment **113**:711-715.

Barr, A., T. A. Black, and H. McCaughey. 2009. Climatic and Phenological Controls of the Carbon and Energy Balances of Three Contrasting Boreal Forest Ecosystems in Western Canada. Pages 3-34 *in* A. Noormets, editor. Phenology of Ecosystem Processes. Springer New York.

Bazaraa, M. S., H. D. Sherali, and C. M. Shetty. 1993. Nonlinear programming : theory and algorithms. 2nd edition. Wiley, New York.

Cardillo, G. 2007. Cohen's kappa: compute the Cohen's kappa ratio on a 2x2 matrix.*in* http://www.mathworks.com/matlabcentral/fileexchange/15365, editor.

Chang, C.-C. and C.-J. Lin. 2010. LIBSVM: a Library for Support Vector Machines. National Taiwan University, Taipei, Taiwan.

Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement **20**:37-46.

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

Cramer, W., A. Bondeau, F. I. Woodward, I. C. Prentice, R. A. Betts, V. Brovkin, P. M. Cox, V. Fisher, J. A. Foley, A. D. Friend, C. Kucharik, M. R. Lomas, N. Ramankutty, S. Sitch, B. Smith, A. White, and C. Young-Molling. 2001. Global response of terrestrial ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models. Global Change Biology **7**:357-373.

Dixon, B. and N. Candade. 2008. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both? International Journal of Remote Sensing **29**:1185 - 1206.

EEA. 2006. The thematic accuracy of CORINE land cover 2000, Assessment using LUCAS (land use/cover area frame statistical survey).

EEA. 2007. CORINE Land Cover 2006. Technical Guide. European Environment Agency - European Topic Centre on Land Cover.

Eklundh, L. and P. Jönsson. 2010. TIMESAT Version 3.0 Software Manual. Lund University, Lund, Sweden.

Foody, G. M. and A. Mathur. 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. Remote Sensing of Environment **103**:179-189.

Gomarasca, M. A. 2009. Basics of Geomatics. Springer, Dordrecht Heidelberg London New York.

Gonçalves, P., H. Carrão, A. Pinheiro, and M. Caetano. 2005. Land cover classification with Support

Vector Machine applied to MODIS imagery.*in* Proceedings of the 25th EARSeL Symposium, Porto, Portugal.

Gu, J., X. Li, and C. Huang. 2008. Land Cover Classification in Heihe River Basin with Time Series - MODIS NDVI Data. Proceedings of the 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery - Volume 02. IEEE Computer Society.

Gu, Y., J. Brown, T. Miura, W. J. van Leeuwen, and B. Reed. 2010. Phenological Classification of the United States: A Geographic Framework for Extending Multi-Sensor Time-Series Data. Remote Sensing **2**:526-544.

Gunn, S. R. 1998. Support Vector Machines for Classification and Regression. University of Southampton, Southampton, United Kingdom.

Heiskanen, J. and S. Kivinen. 2008. Assessment of multispectral, -temporal and -angular MODIS data for tree cover mapping in the tundra-taiga transition zone. Remote Sensing of Environment **112**:2367-2380.

Hsu, C.-W., C.-C. Chang, and C.-J. Lin. 2009. A Prcatical Guide to Support Vector Classification. National Taiwan University, Taipei, Taiwan.

Hüttich, C., U. Gessner, M. Herold, B. Strohbach, M. Schmidt, M. Keil, and S. Dech. 2009. On the Suitability of MODIS Time Series Metrics to Map Vegetation Types in Dry Savanna Ecosystems: A Case Study in the Kalahari of NE Namibia. Remote Sensing **1**:620-643.

IPCC, editor. 2007. Climate change 2007: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press., Cambridge, UK.

Izenman, A. J. 2008. Modern multivariate statistical techniques : regression, classification, and manifold learning. Springer, New York ; London.

Jönsson, A. M., L. Eklundh, M. Hellström, L. Bärring, and 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. and L. Eklundh. 2002. Seasonality extraction and noise removal by function fitting to time-series of satellite sensor data. IEEE Transaction of Geoscience and Remote Sensing **40**:1824 - 1832.

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

Kavzoglu, T. and I. Colkesen. 2009. A kernel functions analysis for support vector machines for land cover classification. International Journal of Applied Earth Observation and Geoinformation **11**:352-359.

Kramer, H. J. 2002. Observation of the Earth and Its Environment: Survey of Missions and Sensors. 4th Edition edition. Springer, Berlin Hong Kong London Milano Paris Tokyo.

Lillesand, T. M., R. W. Kiefer, and J. W. Chipman. 2004. Remote Sensing and Image Interpretation. 5th edition. Wiley, New York.

LP-DAAC. 2011. MOD09Q1 - (Surface Reflectance 8-Day L3 Global 250m). MODIS Products Table,. Land Processes Distributed Active Archive Center (LP DAAC) at the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls USA. Last date of access: 2011-02-15 from URL: https://lpdaac.usgs.gov/lpdaac/products/modis_products_table/surface_reflectance/8_day_l3_global_250m/mod09q1.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. Pages 281-297 *in* Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability. University of California Press., Berkeley,.

Menzel, A., T. H. Sparks, N. Estrella, E. Koch, A. Aasa, R. Ahas, K. Alm-KÜBler, P. Bissolli, O. G. BraslavskÁ, A. Briede, F. M. Chmielewski, Z. Crepinsek, Y. Curnel, Å. Dahl, C. Defila, A. Donnelly, Y. Filella, K. Jatczak, F. MÅGe, A. Mestre, Ø. Nordli, J. PeÑUelas, P. Pirinen, V. RemiŠOvÁ, H. Scheifinger, M. Striz, A. Susnik, A. J. H. Van Vliet, F.-E. Wielgolaski, S. Zach, and A. N. A. Zust. 2006. European phenological response to climate change matches the warming pattern. Global Change Biology **12**:1969-1976.

Miller, H. J. and J. Han. 2001. Geographic data mining and knowledge discovery. Taylor & Francis, London ; New York.

Pal, M. and P. M. Mather. 2003. Support Vector classifiers for Land Cover Classification.*in* http://www.gisdevelopment.net/technology/rs/pdf/23.pdf, editor. GIS Development, Haryana, India.

Pal, M. and P. M. Mather. 2005. Support vector machines for classification in remote sensing. International Journal of Remote Sensing **26**:1007 - 1011.

Reed, B. C., M. D. Schwartz, and X. Xiao. 2009. Remote Sensing Phenology. Pages 231-246 *in* A. Noormets, editor. Phenology of Ecosystem Processes. Springer New York.

Richards, J. A. and X. Jia. 2006. Remote Sensing Digital Image Analysis : An Introduction. 4th edition. Springer, Berlin.

Tan, B., C. E. Woodcock, J. Hu, P. Zhang, M. Ozdogan, D. Huang, W. Yang, Y. Knyazikhin, and R. B. Myneni. 2006. The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. Remote Sensing of Environment **105**:98-114.

Thenkabail, P. S., C. M. Biradar, P. Noojipady, V. Dheeravath, Y. Li, M. Velpuri, M. Gumma, O. R. P. Gangalakunta, H. Turral, X. Cai, J. Vithanage, M. A. Schull, and R. Dutta. 2009. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. International Journal of Remote Sensing **30**:3679 - 3733.

Tso, B. and P. M. Mather. 2009. Remote Sensing in the Optical and Microwave Regions, Classification Methods for Remotely Sensed Data. Second Edition edition. CRC Press.

Vapnik, V. N. 1995. The nature of statistical learning theory. Springer, New York.

Vapnik, V. N. 1998. Statistical learning theory. Wiley, New York.

Lunds Universitets Naturgeografiska institution. Seminarieuppsatser. Uppsatserna finns tillgängliga på Naturgeografiska institutionens bibliotek, Sölvegatan 12, 223 62 LUND. Serien startade 1985. Hela listan och själva uppsatserna är även tillgängliga på http://www.geobib.lu.se/

The reports are available at the Geo-Library, Department of Physical Geography, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden.
Report series started 1985. The whole complete list and electronic versions are available at http://www.geobib.lu.se/

157 Öberg, Hanna (2009): GIS-användning i katastrofdrabbade utvecklingsländer
158 Marion Früchtl &Miriam Hurkuck (2009): Reproduction of methane emissions from terrestrial plants under aerobic conditions
159 Florian Sallaba (2009): Potential of a Post-Classification Change Detection Analysis to Identify Land Use and Land Cover Changes. A Case Study in Northern Greece
160 Sara Odelius (2009): Analys av stadsluftens kvalitet med hjälp av geografiska informationssystem.
161 Carl Bergman (2009): En undersökning av samband mellan förändringar i fenologi och temperatur 1982-2005 med hjälp av GIMMS datasetet och klimatdata från SMHI.
162 Per Ola Olsson (2009): Digitala höjdmodeller och höjdsystem. Insamling av höjddata med fokus på flygburen laserskanning.
163 Johanna Engström (2009): Landskapets påverkan på vinden -sett ur ett vindkraftperspektiv.
164 Andrea Johansson (2009): Olika våtmarkstypers påverkan på CH_4, N_2O och CO_2 utsläpp, och upptag av N_2.
165 Linn Elmlund (2009): The Threat of Climate Change to Coral Reefs
166 Hanna Forssman (2009): Avsmältningen av isen på Arktis - mätmetoder, orsaker och effekter.
167 Julia Olsson (2009): Alpina trädgränsens förändring i Jämtlands- och Dalarnas län över 100 år.
168 Helen Thorstensson (2009): Relating soil properties to biomass consumption and land management in semiarid Sudan – A Minor Field Study in North Kordofan
169 Nina Cerić och Sanna Elgh Dalgren (2009): Kustöversvämningar och GIS - en studie om Skånska kustnära kommuners arbete samt interpolations-metodens betydelse av höjddata vid översvämningssimulering.
170 Mats Carlsson (2009): Aerosolers påverkan på klimatet.
171 Elise Palm (2009): Övervakning av gåsbete av vass – en metodutveckling
172 Sophie Rychlik (2009): Relating interannual variability of atmospheric CH_4 growth rate to large-scale CH_4 emissions from northern wetlands
173 Per-Olof Seiron and Hanna Friman (2009): The Effects of Climate Induced Sea Level Rise on the Coastal Areas in the Hambantota District, Sri Lanka - A geographical study of Hambantota and an identification of vulnerable ecosystems and land use along the coast.
174 Norbert Pirk (2009): Methane Emission Peaks from Permafrost Environments: Using Ultra–Wideband Spectroscopy, Sub-Surface Pressure Sensing and Finite Element Solving as Means of their Exploration
175 Hongxiao Jin (2010): Drivers of Global Wildfires — Statistical analyses

176 Emma Cederlund (2010): Dalby Söderskog – Den historiska utvecklingen
177 Lina Glad (2010): En förändringsstudie av Ivösjöns strandlinje
178 Erika Filppa (2010): Utsläpp till luft från ballastproduktionen år 2008
179 Karolina Jacobsson (2010):Havsisens avsmältning i Arktis och dess effekter
180 Mattias Spångmyr (2010): Global of effects of albedo change due to urbanization
181 Emmelie Johansson & Towe Andersson (2010): Ekologiskt jordbruk - ett sätt att minska övergödningen och bevara den biologiska mångfalden?
182 Åsa Cornander (2010): Stigande havsnivåer och dess effect på känsligt belägna bosättningar
183 Linda Adamsson (2010): Landskapsekologisk undersökning av ädellövskogen i Östra Vätterbranterna
184 Ylva Persson (2010): Markfuktighetens påverkan på granens tillväxt i Guvarp
185 Boel Hedgren (2010): Den arktiska permafrostens degradering och metangasutsläpp
186 Joakim Lindblad & Johan Lindenbaum (2010): GIS-baserad kartläggning av sambandet mellan pesticidförekomster i grundvatten och markegenskaper
187 Oscar Dagerskog (2010): Baösberggrottan – Historiska tillbakablickar och en lokalklimatologisk undersökning
188 Mikael Månsson (2010): Webbaserad GIS-klient för hantering av geologisk information
189 Lina Eklund (2010): Accessibility to health services in the West Bank, occupied Palestinian Territory.
190 Edvin Eriksson (2010): Kvalitet och osäkerhet i geografisk analys - En studie om kvalitetsaspekter med fokus på osäkerhetsanalys av rumslig prognosmodell för trafikolyckor
191 Elsa Tessaire (2010): Impacts of stressful weather events on forest ecosystems in south Sweden.
192 Xuejing Lei (2010): Assessment of Climate Change Impacts on Cork Oak in Western Mediterranean Regions: A Comparative Analysis of Extreme Indices
193 Radoslaw Guzinski (2010): Comparison of vegetation indices to determine their accuracy in predicting spring phenology of Swedish ecosystems
194 Yasar Arfat (2010): Land Use / Land Cover Change Detection and Quantification — A Case study in Eastern Sudan
195 Ling Bai (2010): Comparison and Validation of Five Global Land Cover Products Over African Continent
196 Raunaq Jahan (2010): Vegetation indices, FAPAR and spatial seasonality analysis of crops in southern Sweden
197 Masoumeh Ghadiri (2010): Potential of Hyperion imagery for simulation of MODIS NDVI and AVHRR-consistent NDVI time series in a semi-arid region
198 Maoela A. Malebajoa (2010): Climate change impacts on crop yields and adaptive measures for agricultural sector in the lowlands of Lesotho
199 Herbert Mbufong Njuabe (2011): Subarctic Peatlands in a Changing Climate: Greenhouse gas response to experimentally increased snow cover
200 Naemi Gunlycke & Anja Tuomaala (2011): Detecting forest degradation in Marakwet district, Kenya, using remote sensing and GIS
201 Nzung Seraphine Ebang (2011): How was the carbon balance of Europe affected by the summer 2003 heat wave? A study based on the use of a Dynamic Global Vegetation Model; LPJ-GUESS
202 Per-Ola Olsson (2011): Cartography in Internet-based view services – methods

to improve cartography when geographic data from several sources are combined

203 Kristoffer Mattisson (2011): Modelling noise exposure from roads – a case study in Burlövs municipality

204 Erik Ahlberg (2011): BVOC emissions from a subarctic Mountain birch: Analysis of short-term chamber measurements.

205 Wilbert Timiza (2011): Climate variability and satellite – observed vegetation responses in Tanzania.

206 Louise Svensson (2011): The ethanol industry - impact on land use and biodiversity. A case study of São Paulo State in Brazil.

207 Fredrik Fredén (2011): Impacts of dams on lowland agriculture in the Mekong river catchment.

208 Johanna Hjärpe (2011): Kartläggning av kväve i vatten i LKAB:s verksamhet i Malmberget år 2011 och kvävets betydelse i akvatiska ekosystem ur ett lokalt och ett globalt perspektiv.

209 Oskar Löfgren (2011): Increase of tree abundance between 1960 and 2009 in the treeline of Luongastunturi in the northern Swedish Scandes

210 Izabella Rosengren (2011): Land degradation in the Ovitoto region of Namibia: what are the local causes and consequences and how do we avoid them?

211 Irina Popova (2011): Agroforestry och dess påverkan på den biofysiska miljön i Afrika.

212 Emilie Walsund (2011): Food Security and Food Sufficiency in Ethiopia and Eastern Africa.

213 Martin Bernhardson (2011): Jökulhlaups: Their Associated Landforms and Landscape Impacts.

214 Michel Tholin (2011): Weather induced variations in raptor migration; A study of raptor migration during one autumn season in Kazbegi, Georgia, 2010

215 Amelie Lindgren (2011) The Effect of Natural Disturbances on the Carbon Balance of Boreal Forests.

216 Klara Århem (2011): Environmental consequences of the palm oil industry in Malaysia.

217 Ana Maria Yáñez Serrano (2011) Within-Canopy Sesquiterpene Ozonolysis in Amazonia

218 Edward Kashava Kuliwoye (2011) Flood Hazard Assessment by means of Remote Sensing and Spatial analyses in the Cuvelai Basin Case Study Ohangwena Region –Northern Namibia

219 Julia Olsson (2011) GIS-baserad metod för etablering av centraliserade biogasanläggningar baserad på husdjursgödsel.

220 Florian Sallaba (2011) The potential of support vector machine classification of land use and land cover using seasonality from MODIS satellite data

221 Salem Beyene Ghezahai (2011) Assessing vegetation changes for parts of the Sudan and Chad during 2000-2010 using time series analysis of MODIS-NDVI

222 Bahzad Khaled (2011) Spatial heterogeneity of soil CO2 efflux at ADVEX site Norunda in Sweden

223 Emmy Axelsson (2011) Spatiotemporal variation of carbon stocks and fluxes at a clear-cut area in central Sweden