Student thesis series INES 490

Detecting bark beetle damage with Sentinel-2 multi-temporal data in Sweden

Shuo Yang

2019

Department of
Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Shuo Yang (2019).
Detecting bark beetle damage with Sentinel-2 multi-temporal data in Sweden.

Master degree thesis, 30 credits in *Master's Programme of Physical Geography and Ecosystem Science, 2019*
Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January* 2019 until *June* 2019

Detecting bark beetle damage with Sentinel-2 multi-temporal data in Sweden

---

Shuo Yang

Master thesis, 30 credits, in *Master's Programme of Physical Geography and Ecosystem Science*

Supervisors:

Per-Ola Olsson
Dept of Physical Geography and Ecosystem Science

Hongxiao Jin
Dept of Physical Geography and Ecosystem Science

Exam committee:

Fredrik Lagergren
Dept of Physical Geography and Ecosystem Science

Erica Jaakkola
Dept of Physical Geography and Ecosystem Science

Detecting bark beetle damage with Sentinel-2 multi-temporal data in Sweden

Abstract

The European spruce bark beetle is considered as one of the most destructive forest insects to Norway spruce trees in Europe. Climate change may increase the frequency and intensity of bark beetle outbreaks. It is therefore of vital importance to detect the bark beetle outbreaks and take it under control to prevent further damages. Remote sensing techniques may provide a cost-efficient solution to the detection of bark beetle outbreaks. In the past years, the detection of bark beetle outbreaks in Northern America has achieved success with the aid of the long time series of LANDSAT satellite images. Sentinel-2 provides satellite images of high spatial and temporal resolution which may be suitable for bark beetle detection in Europe.

The extreme drought and heat in the summer in 2018 favored the outbreaks of bark beetles in central and southern Sweden. In this project, detection of two stages (gray-attack and green-attack stage) of bark beetle outbreaks in southern and central Sweden was carried out separately with Sentinel-2 level 2A satellite multi-temporal images. In bark beetle gray-attack stage detection, the two most commonly used methods: maximum likelihood and random forest classification, were performed and compared on different combinations of Sentinel-2 10m resolution raw bands sensed in March-April and VIs derived from them. Maximum likelihood classification method with EVI and GNDVI gave the highest accuracy: total accuracy of 89% and Kappa of 0.74 (substantial agreement). Random forest classification method with all variables achieved the second best result: total accuracy of 85% and Kappa of 0.62 (substantial agreement). The two best methods were thereafter applied to two test areas in southern (test area 1) and central Sweden (test area 2). Random forest classification method with all variables obtained higher accuracy: total accuracy of 76% and Kappa of 0.53 (moderate agreement) in test area 1 and total accuracy of 71% and Kappa of 0.39 (fair agreement) in test area 2.

Based on detection result from the first part, random forest classification method was employed for bark beetle green-attack stage detection. A series of VIs derived from Sentinel-2 20m resolution bands sensed in the summer in 2018 were calculated and the importance of the VIs and raw bands were ranked with random forest algorithm. The first 13 or 14 most important variables were used for classification. Results show that water content related raw bands and VIs, red-edge VIs and the NIR band are the most sensitive variables to bark beetle green-attack. Bark beetle green-attack stage detection obtained high accuracy in study area 1: total accuracy of 88% and Kappa of 0.67 (substantial agreement) on July 26th and total accuracy of 84% and Kappa of 0.58 (moderate agreement) on October 12th. Relatively low accuracy were achieved in test area 1: total accuracy of 53% and Kappa of 0.03 (no or rarely any agreement). Moderate accuracy were achieved in test area 2: total accuracy of 64% and Kappa of 0.27 (fair agreement) on July 8th, and total accuracy of 71% and Kappa of 0.42 (moderate agreement) on July 31st.

Keywords: Physical Geography and Ecosystem Analysis, Remote Sensing, Bark Beetle, Insect Detection, Sentinel-2, Forest Health, Maximum Likelihood, Random Forest, Spectral Signature, Vegetation Indices.

Supervisors: **Per-Ola Olsson, Hongxiao Jin**
Master degree project 30 credits in the Master's Programme of Physical Geography and Ecosystem Science, 2019
Department of Physical Geography and Ecosystem Science, Lund University. Student thesis series INES 490

NGEM01

Table of content

List of abbreviations ..... **II**
1. Introduction **1**
1.1 Aim 2
1.2 Limitation 2
2. Background **3**
2.1 Bark beetles 3
2.1.1 Biology, population dynamics and seasonal flight pattern of European spruce bark beetle 3
2.1.2 Impact of outbreaks of European spruce bark beetle 5
2.1.3 Climate change and bark beetle outbreaks 6
2.2 Bark beetle outbreak detection 7
2.3 Sentinel-2 MSI data 11
3. Methodology **12**
3.1 Study area 12
3.2 Data 12
3.2.1 Sentinel 2 and landcover data 12
3.2.2 field data 14
3.3 Bark beetle detection methods 15
3.3.1 Bark beetle gray-attack stage detection 15
3.3.2 Bark beetle green-attack stage detection 19
3.3.3 Application in other study areas 21
3.3.4 Accuracy analysis 23
4. Results **23**
4.1 Identification of clear-cut areas 23
4.2 Gray-attack stage detection 24
4.3 Green-attack stage detection 29
5. Discussion **38**
5.1 Gray-attack stage detection 38
5.2 Green-attack stage detection 39
5.3 Threshold 41
5.4 Sources of error and suggestions for improvement 41
6. Conclusions **42**
References **43**
Appendices **49**

List of abbreviations

DGVM: Dynamic Global Vegetation Model

GPS: Global Positioning System

NIR: Near InfraRed

SWIR: Shortwave InfraRed

VI: Vegetation Index

UTM: Urchin Tracking Module

WGS: World Geodetic System

NMD: Nationella Markt äckeData

NDVI: Normalized Differential Vegetation Index

ML: Maximum Likelihood

RF: Random Forest

1. Introduction

The European spruce bark beetle (*Ips typographus*) is considered as one of the most disastrous insects to Norway spruce (*Picea abies*) in Europe (Öhrn 2012; Christiansen and Bakke 1988). Outbreaks of bark beetles have severe impact on both ecosystem and social economy. Forests killed by bark beetles will become sources of carbon release (Lausch et al. 2013), and reduction in timber quality, unusable damaged timber and control managements lead to vast economic losses. Climate change has caused global warming (IPCC 2018). Increased temperature and drought will provoke the increase of the stress of the trees and thereby favor the outbreak of insects (Schlyter et al. 2001). It is impossible to control the happening of the drought, however, it is possible to detect the damages on the spruce trees when the bark beetle outbreaks occur and limit the impact so as to prevent further losses. Therefore, detecting the outbreaks of bark beetles in time is vitally important.

The development of remote sensing techniques provides a cost-efficient way to solve the problem. In the past decades, remote sensing techniques have been proven useful to detect changes in vegetation health and been widely used to detect the infestation of insects (Senf et al. 2017). The method and difficulty of bark beetle outbreak detection is closely related to the stage of the outbreak. The outbreak of bark beetles is usually divided into three stages, green-attack, red-attack and grey-attack stage. In the green-attack stage, bark beetles live and breed inside the infested trees and have not yet moved to other healthy trees. In order to effectively control the outbreaks, the infested trees need to be removed already in the green stage. Therefore, being able to detect bark beetle infestation early in the green-attack stage is critical. Previous studies have confirmed that in the red and grey-attack stages, the changes in biochemical-physical properties in the forest can be successfully detected using low-to-medium resolution remote sensing data (Skakun et al. 2003; Franklin et al. 2003; Wulder et al. 2006; Lausch et al. 2013). However, the detection of green-attack stage remains problematic. Several studies have focused on the green-attack stage detection of the bark beetle outbreaks but achieved limited success. Most of the studies were carried out in central Europe using very high spatial resolution remote sensing data, such as with commercial satellites WorlfView-2 (Immitzer and Atzberger 2014; Filchev 2012), and

RapidEye (Ortiz et al. 2013), as well as HyMAP airborne hyperspectral data(Lausch et al. 2013), or with active remote sensing systems like TerraSAR-X(Ortiz et al. 2013). Abdullah et al. (2018) explored and compared the freely available satellite data of Sentinel-2 and Landsat-8, highlighting the potential of using Sentinel-2 data to detect green-attack stage of bark beetle outbreaks. Sentinel-2 mission, launched in 2015, provides open-access data with high temporal resolution and spatial resolution, it is of interest to assess its potential of detecting bark beetle damages in Sweden.

1.1 Aim

According to the Swedish Meteorological and Hydrological Institute (SMHI), in 2018, southern Sweden experienced the hottest summer in 100 years, accompanied by severe drought. The extreme hot and dry weather increased the stress of the spruce trees and favored outbreaks of bark beetles in the spruce forests. The aim of this study is to explore the potential of Sentinel-2 multi-temporal data for detecting the bark beetle damage in Norway spruce forests in southern and central Sweden. The aim includes two objectives:

1) Developing a method to detect and map the grey-attack stage bark beetle outbreaks in the study area.

2) Exploring the possibility of the bark beetle green-attack stage detection with multi-temporal data.

1.2 Limitation

The swarming of bark beetles usually starts when air temperature is around 16.5 °C (Öhrn 2012). In southern and central Sweden, the mass flying usually happens around mid-April to mid-August (Öhrn 2014). The suitable sampling time for bark beetle green-attack stage outbreaks is usually in late June to July (Abdullah et al. 2018). Carrying out bark beetle green-attack stage sampling is not realistic for the duration of the thesis, and therefore, the bark beetle green-attacked spruce samples are absent. The available field data, however, is not perfect for classification in terms of sample size and sampling time. It is unknown what GPS device was used to locate the bark beetle damaged samples and thus it is hard to control error from input. In addition,

carrying out field sampling on our own is barely possible as it is very time consuming and expensive in such large area. It is better to explore the performance and make full use of the available field data. In terms of satellite images, cloud has always been a problem. In Swedish weather, it is often cloudy throughout the year and hence there is limited amount of useful satellite data.

2. Background

2.1 Bark beetles

There are more than 6000 species of bark beetles described but only a minority of them is able to attack and breed in living trees, and have the ability to kill healthy trees. The tree-killing bark beetles are considered as significant forest pests. They are capable of impacting ecosystem structure and are of significance to the economics. Tree-killing bark beetles are important forest pest, especially in North America and Europe. In North America, the Mountain pine beetle (*Dendroctonus ponderosae*) is the most important pest to the forestry whereas the European spruce bark beetle is the most aggressive and destructive in Europe (Öhrn 2012; Senf et al. 2015).

2.1.1 Biology, population dynamics and seasonal flight pattern of European spruce bark beetle

The adult European spruce bark beetles are small in size, around 4.2 to 5.5 mm long. As insects are ectothermic organisms, the metabolism and development of bark beetles are dependent on air temperature (Öhrn 2012). In the adult stage, bark beetles usually go through hibernation under the bark, or in the soil. The surviving bark beetles during the hibernation experienced a temperature dependent maturation and reach the complete development. The surviving bark beetles emerge and start mass flying in spring, searching for suitable breeding materials when the required thermal sum is fulfilled (Økland et al. 2015). The swarming of adult bark beetles normally starts when the daytime temperature is around 16.5 °C. The flight activities cease when temperature exceed 30 °C (Lobinger, 1994).

During the flight, the male bark beetles disperse in the forest, searching for suitable

spruce trees for breeding. The colonization of healthy spruce trees requires reduced tree vitality or large densities of attack at the same time. Under non-outbreak conditions, bark beetles breed at low population density. The substrate of colonization in this case is usually newly dead or dying spruce trees, fresh wind-felled, cut or unbarked timber. The male bark beetles excavate through the bark and start the construction of a nuptial chamber in the phloem. The communication between bark beetles includes the use of pheromones. For European spruce bark beetles, the male bark beetles are the major pheromone producer. After the initiation of gallery construction, the male bark beetles release aggregation pheromones to attract both sexes of conspecifics (Öhrn 2012; Eidmann 1992). The attracted male bark beetles may excavate new nuptial chambers. Each male will attract one to four females to the chamber. After mating, the female bark beetles excavate tunnels in the gallery and may lay up to 80 eggs in the tunnel, under low population densities (Öhrn 2012; Økland et al. 2015). The productivity of bark beetles is negatively correlated to the population density (Eidmann 1992). High population density may affect the behavior of bark beetles, cause shorter gallery and lower the productivity (Økland et al. 2015). When finished laying eggs, the parental bark beetles may leave the gallery and reemerge to initiate one or more broods, so called sister brood. The timing of reemergence is dependent on intraspecific competition and temperature. Higher attacking density of bark beetles leads to earlier reemergence than lower attacking density, on average. Bark beetles leave the gallery earlier under higher temperature and have better chance of finding a second breeding material (Anderbrant 1989; Öhrn 2014). In Southern Europe and at lower elevations, bark beetles may complete two or three fully developed generations and numerous sister broods in a year. In Sweden, northern Europe and at high elevations areas in southern Europe, one fully developed generation and several sister broods is the most common (Öhrn 2012; Økland et al. 2015).

Öhrn (2014) described that in Southern Sweden, the flight activity of spruce bark beetles lasts a long period, starting from mid-April and lasting till as late as mid-August. The flight activities were assumed to last a much shorter time from mid-May till the beginning of July in the previous time. He related the long period of flight activities to the warm spring and summer due to climate change. His research also confirmed that in southern Sweden, bark beetles initiate a second generation every

year even though a fully developed second generation is uncommon.

2.1.2 Impact of outbreaks of European spruce bark beetle

The outbreak of European spruce bark beetle is usually triggered by high attacking population resulting from the increased productivity in the non-resistant breeding materials. Living spruces have defense mechanism that is resistant to insect attack and other diseases. There are two basic types of defense systems for coniferous trees: the constitutive defense system and the induced defense system. The constitutive defense involves the physical barrier (the bark) and toxic chemicals stored in the phloem and Xylem to defense the potential attackers. The induced defense system, which is triggered by the wound reactions, including mass production of toxic chemicals, is very powerful to a large variety of potential invaders (Öhrn 2012). The defense system of healthy trees usually drives off the bark beetles or kills them. However, when a large number of bark beetles attack simultaneously, especially when the tree is stressed (such as due to drought), the resistance of the living trees can be overcome (Eidmann 1992). At the same time, bark beetles inoculate many types of fungi to the tree stem. Some of the fungi may help bark beetles overcome the defense of the tree (Persson et al. 2009). When the defense is overcome, bark beetles bore through the outer bark, feed and breed in the phloem of the bark, strangling the nutrient transportation from the leaves to the roots. In addition, the fungi vectored by bark beetles may help restraining the water transport in the xylem (Öhrn 2012). Restriction of nutrient and water leads to the death of the tree.

The direct impact of outbreaks of bark beetles is tree loss. During the 1990s, bark beetles have destroyed more than over 30 million m^3 of spruce forests in ten countries of Europe (Økland et al. 2015). Sweden has experienced numerous outbreaks of bark beetles in the history. According to the Swedish Forestry Agency (Skogsstyrelsen), approximately 12 million m^3 in the spruce forests has been killed by outbreaks of bark beetles by 2017. Massive death of trees disrupts the structure of stands and landscapes and leads to loss of potential growth (Eidmann 1992). Forests are of critical significance to ecosystem, economy and human society. Forests do not only provide wood, but also offer ecosystem services that improve air quality, control surface water run-off and lower soil erosion. The loss of trees will therefore disturb the ecosystem

services that the forest provides. Forest also plays an important role in carbon cycle and energy balance (Morris et al. 2017). Healthy forests are known as carbon sinks. The disturbance in the forest caused by insects can reduce the uptake of carbon (Olsson et al. 2016). Mortality of forests can turn it into a source of carbon release (Lausch et al. 2013). In addition, bark beetles kill living trees and reduce timber quality, leading to immense economic loss, not to mention the cost of control management and construction of new roads for the control management (Eidmann 1992). Large scale mortality of spruce forests can therefore bring about great disasters to ecosystem, economy and human society.

2.1.3 Climate change and bark beetle outbreaks

Forest insects are an important part of forest ecosystem. Forest insects consume dead trees and renew the weakened and susceptible forests, recycling nutrients and accelerating carbon and energy cycle. Researches using climate change models revealed that insect outbreaks are very likely to happen more frequently in the future, resulting in severe consequences for the forest ecosystem (Senf et al. 2015; Logan et al. 2003; Volney & Fleming 2000). Specifically, Jönsson et al. (2012) utilized LPG_GUESS, which is a DGVM model, indicating that bark beetle outbreaks are likely to occur more frequently in Sweden, especially in southern and central Sweden and more generations are likely to complete in one year due to the warming climate.

The influence on bark beetles caused by climate change is generally divided into direct effects and indirect effects. Climate and weather directly influence the biology of bark beetles. Increased temperature may provide suitable conditions for maturation of bark beetles and shift the time of swarming to an earlier date. Increased temperature and sunshine cause increased bark temperature, which may speed up the development of bark beetle offspring. Increased temperature therefore may increase the risk of outbreaks resulting from increased population of adult bark beetles when more generations are completed (Økland et al. 2015). On the other hand, the increased frequency or intensity of extreme weather and climate events may indirectly influence bark beetle activities. Windstorm has been the key factor for the spruce bark beetle outbreaks in Sweden in the past century (Økland et al. 2015; Öhrn 2012). Wind-felled trees provide suitable breeding materials for bark beetles. Continuous heat and

drought conditions may favor the outbreaks of bark beetles as it causes water shortage of spruce trees which negatively affects the defense system (Öhrn 2012).

2.2 Bark beetle outbreak detection

The devastating effects of bark beetle outbreaks have drawn broad attentions of researchers to focus on the detection. Based on the knowledge of pheromone, pheromone baited traps have been widely used and proven useful as one of the traditional detection techniques, whereas walking stand transects by foresters is another method (Liu & Dai 2006; Lausch et al. 2013). These methods have achieved success in helping obtain important information about infestation situations as well as behavior of bark beetles (Schlyter 2001; Saeed et al. 2010; Lausch et al. 2013). Nevertheless, these detection methods are expensive in money, time and manpower, yet the investigations are limited in small spatial scales. Remote sensing techniques, on the other hand, may provide a time and cost-efficient alternative. Remote sensing techniques have been widely used in forest health and vitality and insect infestation detection (Senf et al. 2017; Ismail et al. 2007; Lausch et al. 2013; Lamber et al. 2015; Olsson et al. 2016).

Remote sensing methods detect the changes of reflectance in the electromagnetic spectrum in different wavelengths (differences in the spectral signature). The spectral signature of leaves and canopy is determined by the functional and structural properties of plants (figure 2.1.): leaf pigments at visible bands, cell structure at NIR region, water content at SWIR wavelengths (Abdullah et al. 2018). Stressed vegetation shows changes in the biochemical and biophysical properties, for example, stress of vegetation may cause a reduction of leaf pigments, changes in photosynthetic activities, decrease of plant and leaf water content and destruction of cell structure (Lausch et al. 2013). The changes can lead to an increase of reflectance in the visible wavelengths, a decrease in the NIR region and an increase in the SWIR region. It is therefore possible to distinguish healthy and stressed forest based on the differences in the leaf and canopy spectral signature. The visible wavelength has been widely used as a vegetation stress indicator, as the decline of chlorophyll and other leaf pigments will increase the reflectance in this region. More specifically, the red wavelength region is less sensitive whereas the green and red-edge wavelength region shows high

sensitivity to the loss of chlorophyll content (Abdullah et al. 2018). Additionally, the red-edge region shows high sensitivity to the stress in vegetation caused by disease and insect attack (Filchev 2012; Ortiz et al. 2013; Abdullah et al. 2018). The NIR and SWIR wavelength regions, especially the SWIR bands, are sensitive to the changes in the water content. Furthermore, VIs utilize the differences between two or more spectral bands and have been proven useful in previous researches on vegetation health and insect detection (Abdullah et al. 2018; Sonobe et al. 2018).

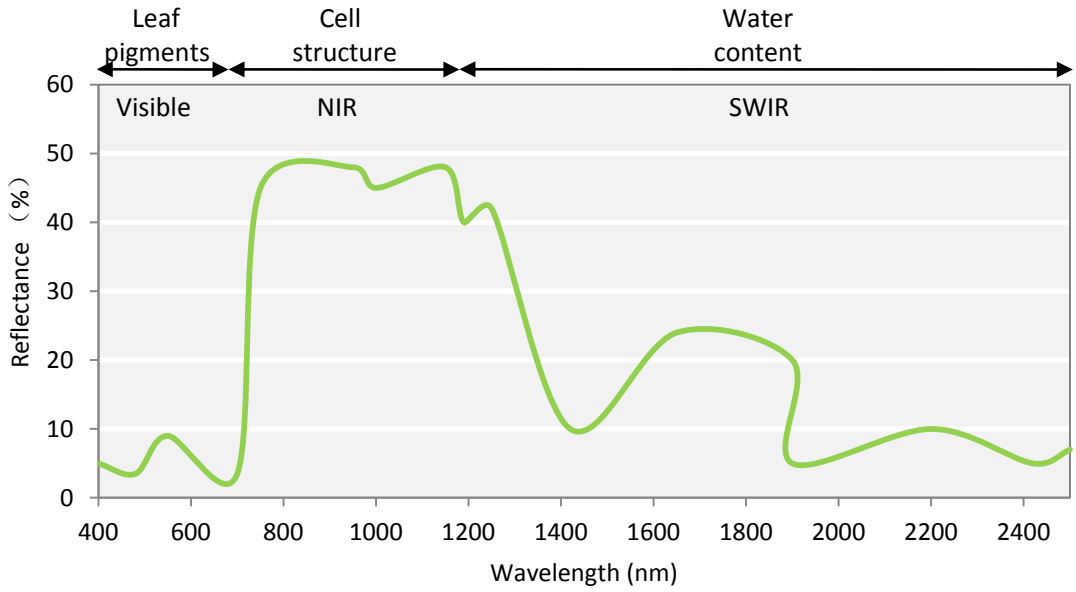


Figure 2.1. An example leaf spectral signature of a healthy plant and the functional and structural properties of the plant that control the reflectance at the corresponding wavelengths, modified from Chuvieco (2016).

The remote sensing detection of bark beetle outbreaks differs in different attack stages. In green-attack stage, bark beetle infested trees remain green and alive, but suffer from the restriction of water transport and decline of chlorophyll content. Detection of this stage is much more difficult than the detection of bark beetle gray-attack stage. Wavelengths related to leaf and canopy chlorophyll content and water content (Green, red-edge and SWIR) and the derived VIs are considered useful in the previous studies (Lausch et al. 2013; Abdullah et al. 2018). In red-attack stage, the leaf pigments severely decrease, which results in as a change of needle color to yellow or brown whereas in gray-attack stage, infested trees have all needles fallen off and eventually reach mortality. In gray-attack stage, bark beetles have left the breeding material. The detection of bark beetle red and gray-attack stage has achieved acceptable results with the utility of raw bands: visible, NIR and SWIR bands from LANDSAT TM,

achieving total accuracy of over 70% (Franklin et al. 2003), VIs derived from water content and greenness relevant bands, such as Normalized Difference Moisture Index (NDMI), Enhanced Wetness Difference Index (EWDI), Tasseled cap wetness component, other tasseled cap transformation indices, NIR and SWIR ratio, Moisture Stress Index (MSI) and Vegetation Chlorophyll Index (VCI), achieving 70% to up to 90% total accuracy (Wulder et al. 2006; Meddens et al. 2013; Havašová et al. 2015).

Based on remote sensing data, a large variety of classification methods to identify bark beetle infestation have been applied and compared in the past a few years (Senf et al. 2017). Those methods can be generally divided in multi-date classification methods (time series methods) and single-date classification methods. Multi-date classification methods are commonly used in detecting mountain pine beetles red and grey attack stages in North America based on LANDSAT long time series data. The most common method is to calculate one VI for several years of remote sensing data using threshold methods to estimate an optimal threshold for classification (Skakun et al. 2003; Meddens et al. 2013). Logistic regression techniques (Wulder et al. 2006) and image differentiation techniques are used to distinguish bark beetle damaged trees (Havašová et al. 2015; Jakus et al. 2003). The most commonly applied VIs are NDMI and EWDI. Single-date classification methods are usually based on spectral reflectance information from more than one band, using either raw bands, a single or combinations of variables (raw bands and VIs). The most commonly applied classification models for single-date classification are maximum likelihood classifier (Meddens et al. 2013; Franklin et al. 2003) and random forest classifier (Ortiz et al. 2013; Immitzer and Atzberger 2014). Single-date classification methods are frequently applied to detect bark beetle infestations in both North America and Europe. Meddens et al. (2013) compared the multi-date classification method and the single-date classification method based on several bands and VIs to detect mountain pine beetle outbreaks in North America. The result indicated that multi-date method achieved higher accuracy at intermediate level of tree mortality while single-date method achieved higher accuracy at high tree mortality levels.

Maximum likelihood is one of the most popular classification models that have been successfully applied in bark beetle outbreak detection (Meddens et al. 2013; Franklin et al. 2003). Maximum likelihood classification is a pixel-based supervised

classification method that assigns each pixel to the class that the pixel has the highest probability of belonging to, based on a Gaussian probability density function. Maximum likelihood considers distribution and covariance and usually achieves good separability between classes. The disadvantage is that maximum likelihood classification demands a large number of training samples.

Random forest, proposed by Breiman (2001), is an algorithm developed for machine learning and is becoming more and more popular in remote sensing image processing. Unlike traditional statistical models, random forest creates a large number of binary decision tree models, learns the relationship between predictor variables (training data) and response variables (e.g. landcover classes) and gives the best prediction result. For each decision tree, two thirds of the training data are randomly selected to create the model whereas the remaining data are used for an accuracy test. At each node of the decision tree, a random subset is selected to determine the split rule. Classification trees create categorical datasets while regression trees create continuous datasets. There are many advantages of using random forest algorithm. It makes use of computer power and machine learning and usually generates high accuracy. With random forest algorithm, it is possible to visualize the variable importance which may be of great interest to researchers. However, the accuracy tends to be low if the training data does not represent the whole area or the random selection over-selected the bad quality training data (Horning 2010).

2.3 Sentinel-2 MSI data

Sentinel-2 is a European wide-swath and multi-spectral imaging mission with high spatial and temporal resolution. The mission contains two twin satellites at the same sun-synchronous orbit and are phased at 180°. Each satellite carries one Multi-Spectral Instrument (MSI) that samples 13 spectral bands with high spatial resolution (see table 2.1.). Sentinel-2 is also designed to revisit at high frequency: under cloud-free conditions, at equator, 5 days with 2 satellites and 10 days with one satellite. Sentinel-2 satellites were first launched in June 2015 (ESA 2019).

Sentinel-2 products are radiometrically and geometrically calibrated and available in two different levels: level-1C and level-2A. Level-1C products contains top-of-atmosphere (TOA) reflectance whereas level-2A products are atmospherically corrected to bottom-of-atmosphere (BOA) reflectance, also known as top-of-canopy (TOC) reflectance. Both Level-1C and Level-2A products are in 100x100 km^2 granules, also called tiles, with reference coordinate system in UTM/WGS84 projection.

Table 2.1. Radiometric and spatial resolution of all 12 bands of Sentinel-2 twin satellites S2A and S2B, carrying the same type of multispectral instruments but with different resolution (ESA, 2019).

		S2A		S2B		
Band number	Description	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	Spatial Resolution (m）
1	Blue	442.7	27	442.2	45	60
2	Blue	492.4	98	492.1	98	10
3	Green	559.8	45	559	46	10
4	Red	664.6	38	664.9	39	10
5	Red-edge	704.1	19	703.8	20	20
6	Red-edge NIR	740.5	18	739.1	18	20
7	Red-edge NIR	782.8	28	779.7	28	20
8	NIR	832.8	145	832.9	133	10
8a	NIR	864.7	33	864	32	20
9	NIR	945.1	26	943.2	27	60
10	SWIR	1373.5	75	1376.9	76	60
11	SWIR	1613.7	143	1610.4	141	20
12	SWIR	2202.4	242	2185.7	238	20

3. Methodology

3.1 Study area

The study area of this project is several forest areas in Sweden, which underwent bark beetle outbreaks in 2018 (figure 3.1.) The main study area (study area 1), which was used for development of the bark beetle detection methods, is a 20 * 20 km square area located between 56 ° 1' 23" N, 13 ° 16' 59" E and 56 ° 12' 25" N, 13 ° 35' 50" E (WGS84) in Scania, southern Sweden. In the bark beetle infested forest areas, the dominant species is Norway spruce, typically in stands with only spruce. Mixed forests with Norway spruce and European beech (*Fagus sylvatica)* are also present in some areas. An accuracy test area 1 (T1) (8 * 4 km) is located around 63 km northeast of study area 1 between 56 ° 15' 59" N, 14 ° 13' 30" E and 56 ° 19' 13" N, 14 ° 17' 23" E. Test area 1 has similar forest vegetation types to study area 1. Accuracy test area 2 (T2) is a 1.6 * 1.6 km square area located between 57 ° 59' 24" N, 15 ° 57' 51" E and 57 ° 59' 15" N, 16 ° 14' 4" E in central Sweden, about 90 km south of Norrköping. Test area 2 is dominated by deciduous and mixed forest with small areas of pure spruce stands.

3.2 Data

3.2.1 Sentinel-2 and landcover data

Sentinel-2 level-2A multi-temporal images with no or few cloud and snow covers were obtained from Copernicus Open Access Hub in UTM zone 33/WGS84 projection. Sentinel-2 based landcover data from Swedish National ground cover data (NMD), with a minimum mapping unit of 10m in SWEREF99TM projection was obtained in order to isolate the spruce forests from other landcovers (table 3.1.).

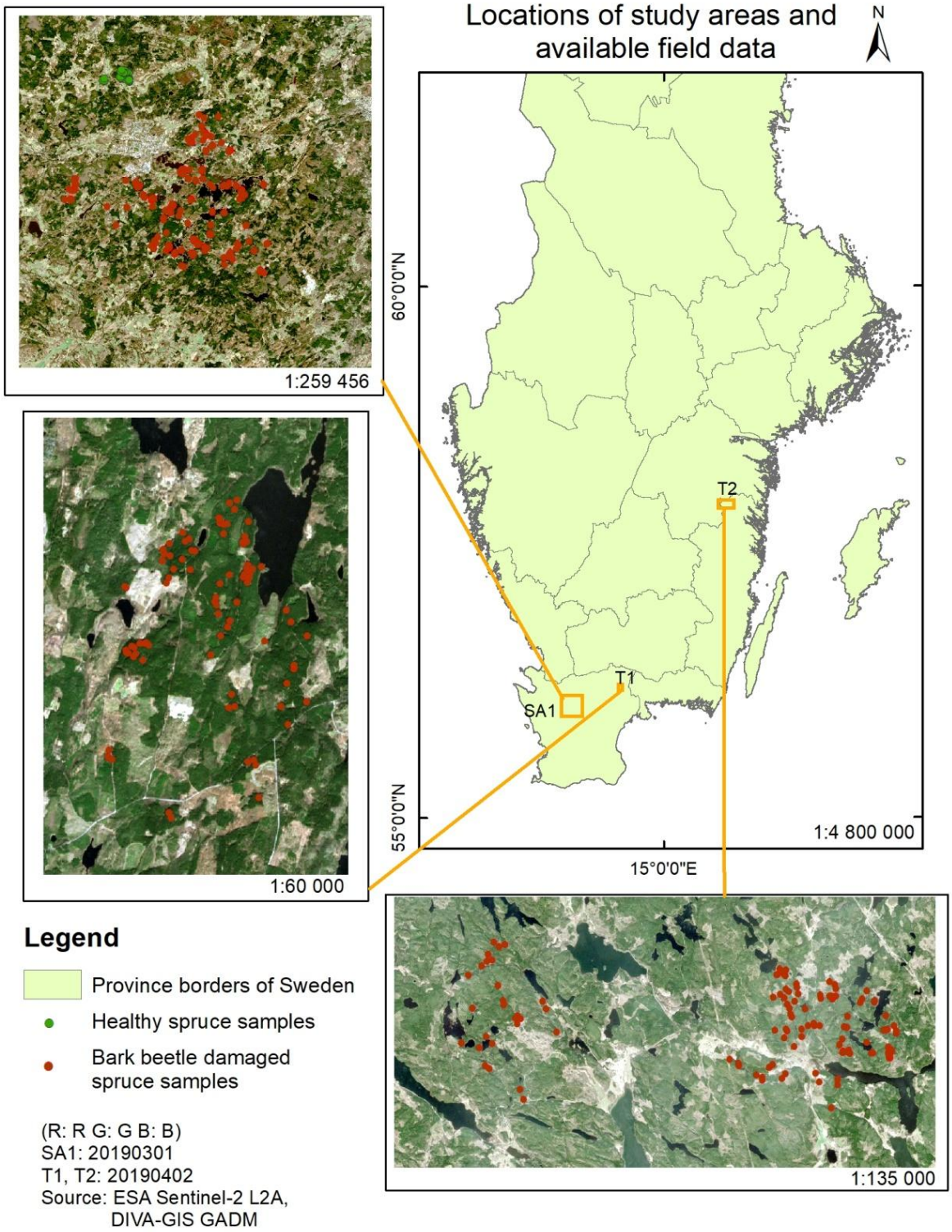


Figure 3.1. Locations of the three study areas and available field data (shown in the zoomed in maps) used in this project, based on satellite images from ESA (2019), administrative boundary data from DIVA-GIS (2019).. The projected coordinate system is SWEREF99TM.

Table 3.1. Information of all data used in this project including the data type, spatial resolution, source of the data, spatial reference information and the time of the sensing date of the satellite images, the newest update time of landcover data and the sampling time of the field data.

Data	Data type	Spatial resolution	Data source	Reference System	Time of acquisition/update
Sentinel-2 level-2A	Raster	10m, 20m, 60m	ESA	UTM/WGS 84	2019, 2018
NMD Landcover data	Raster	10m	NMD	SWEREF 99TM	2018
Bark beetle damages field data	Vector	-	forest managers and field inventory	SWEREF 99TM	2019, 2018

3.2.2 field data

Field data is an important part in image classification both for creating training data and for accuracy test. Due to time limit, field data was not collected myself in this project. Instead, field data was obtained from the forest managers and field inventory, collected by the local foresters and forest inventory students. One way to identify bark beetle damages in the field is to identify the bark beetle holes on the trees. In study area 1, field data was collected in spring 2019 with two strategies: random sampling and sampling of individual trees (figure 3.1.). In the random sampling area, around 20 to 30 spruce trees were examined in random locations in the spruce forests. 60 healthy spruce samples were collected. In the individual sampling area, examination of spruces was carried out on individual tree level. 431 spruce trees were found that were damaged by bark beetles. All available field data is in the form of a vector point layer with attribute information regarding number of damaged trees and collection time. The healthy spruce samples were absent in test area 1 and 2. In test area 1, 86 bark beetle damaged samples were collected whereas 123 damaged samples were collected in test area 2. Locations and number of damaged trees in each location were recorded in the two test areas.

3.3 Bark beetle detection methods

This part is divided into two parts: (1) bark beetle gray-attack stage detection and (2) bark beetle green-attack detection. For bark beetle gray-attack stage detection, only 10m resolution bands were used whereas 20m resolution bands were employed for bark beetle green-attack stage detection. The reason for such choices will be explained in later in this part. Methodology developed for study area 1 is emphasized whereas method applied in the two test areas was briefly described later in *3.3.3 application in other areas* part.

3.3.1 Bark beetle gray-attack stage detection

In the bark beetle gray-attack stage detection, cloud free Sentinel-2 level-2A data (tile 33VVC and 33VUC) sensed on March 1st 2019 was obtained. It was assumed that at this time bark beetle flight activities haven't begun in March-April in 2019, and the spruce trees attacked in 2018 have all needles fallen off or already dead. The workflow of the bark beetle gray-attack stage detection is shown in figure 3.2.

3.3.1.1 Data pre-processing

Data pre-processing is an essential part of image processing as data collected from different sources usually differs in spatial reference system, spatial resolution, spatial and temporal extent, level of generalization, data quality and data type. A series of steps of data pre-processing were required in order to harmonize different levels of field data, satellite images and landcover raster data into a clear and efficient system. In this project, data from different sources is harmonized into Swedish national projected system SWEREF99TM.

Raster data pre-processing

The acquired Sentinel-2 level-2A 10m resolution data has already been rediometrically calibrated from radiance to reflectance value, georectified into UTM zone 33/WGS84 reference system and atmospherically corrected into top-of-canopy reflectance. The first step was the transformation of spatial reference system from UTM zone 33/WGS84 to SWEREF99TM. Thereafter, satellite images as well as

landcover data were cropped into the same extent as the study areas. Binary spruce mask was then created from the NMD landcover map.

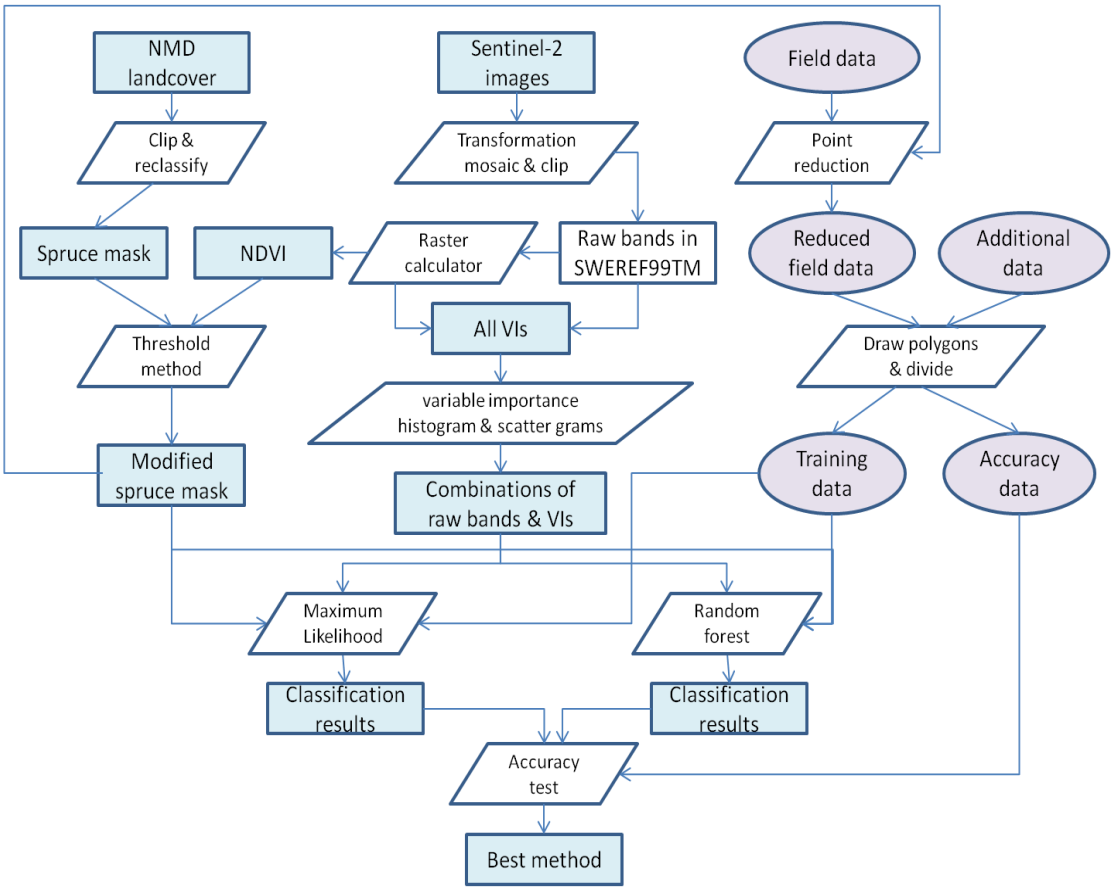


Figure 3.2. Workflow of the bark beetle gray-attack stage detection in this project. The blue squares stand for raster files, the white parallelograms stand for processes and the purple circles stand for vector files.

Spruce mask modification Beside the error produced by the landcover classification, many other factors might cause further error to the landcover map (e.g. clear-cutting activities, change of landcover…). In order to increase the accuracy, modification on the spruce mask derived from the landcover map was carried out. A threshold method was applied based on the difference of NDVI values in the clear-cutting areas and spruce forests. A total of 140 points (70 in the spruce forest areas, 70 in the clear-cutting areas) were randomly taken according to the Sentinel-2 true color composite satellite image. A series of NDVI threshold values were applied for classification and evaluated with error matrix. Receiver operating characteristic (ROC) curves were created for threshold selection.

Field data pre-processing

Field data is a critical part of image processing as it provides the information used for training data and validation which directly influences the final accuracy. It is therefore important to control the error of the field data. In order to reduce error, sample points with obvious positioning errors (points outside spruce forests) or in the clear-cutting areas were removed based on the modified spruce mask. The rest of the bark beetle damaged points on the individual stand level were then aggregated into polygons with a maximum distance of 10 m. Aggregation of points makes it possible to reduce insignificant bark beetle damage samples by only selecting areas with more than 3 infested spruces within one to two pixels. In addition, samples of healthy spruce containing less than 20 trees in the random sampling areas were removed, assuming these samples were taken from mixed forests (spruce and beech). Healthy spruce samples were then randomly taken in the individual sampling areas where no damages were found. A total of 32 bark beetle infested polygons were created manually with the size of 100 to 200 m^2 (1 to 2 pixels of the 10m resolution images). Similarly, polygons covering 1 to 2 pixels were also created centering at the 32 remaining healthy tree samples from the random sampling area and 30 manually taken healthy tree samples. Thereafter, a total of 62 healthy tree samples and 32 bark beetle damaged samples were randomly divided into training data and validation data. 18 bark beetle infested samples and 29 healthy spruce samples were divided into training data and the rest of the data was used for accuracy assessment.

3.3.1.2 Image processing

Outbreaks of barks beetles in Sweden usually occur in small area, sometimes only a few trees are infested within one place. In order to detect the change with remote sensing data, a high spatial resolution is required. Sentinel-2 images provided four 10m resolution bands: band 2 (blue), band 3 (green), band 4 (red) and band 8 (NIR). In the bark beetle gray-attack stage, spruce trees are usually left with no needles or even dead. As mentioned in the background part, visible bands are sensitive to the change of chlorophyll content and NIR is sensitive to change of cell structure. The change of spruce trees should be able to be detected with the four 10m resolution bands and greenness related VIs. With the four 10m resolution bands, four greenness

related VIs were calculated including Normalized Differential Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Green Normalized Differential Vegetation Index (GNDVI), Simple NIR/Red Ratio (SR) are calculated (Equations shown in table 3.2.). Two commonly used single-date classification methods maximum likelihood and random forest were applied to Sentinel-2 images (tile 33VVC and 33VUC) sensed on 1st of March, 2019 and compared. A series of raw bands, single variable or a combination of variables were tested and compared for classification. Maximum likelihood classification enables the visualizations of variable scatter grams. The combinations EVI and GNDVI, GNDVI and NDVI were chosen based on the separability shown from the scatter grams. At the same time, random forest algorithm provides an easy and efficient way to reduce the total amount of variables used for classification through calculating the importance of the variables used for the classification. With the utility of random forest algorithm, it is possible to reduce the variables with low importance. No variables were removed in this part as no particular variables showed low importance. When carrying out random forest classification, it is always a question on how many trees to choose in order to achieve a relatively high accuracy and low computing time. Oshiro et al. (2012) suggested that between 64 to128 is the optimal number of trees to use to achieve a highly accurate result. Accordingly, 100 classification trees were chosen in this project.

Table 3.2. Formulas of the vegetation indices used in bark beetle gray-attack stage detection calculated with Sentinel-2 10m resolution bands. “b” is short for “bands” (for example: b2 = band 2).

Vegetation indices	Full name	Formula	Reference
NDVI	Normalized Difference Vegetation Index	(b8-b4)/(b8+b4)	Rouse Jr 1972
EVI	Enhanced Vegetation Index	2.5*(b8-b4)/(b8+6*b4-7.5*b2+1)	Huete et al.1999
GNDVI	Green Normalized Difference Vegetation Index	(b8-b3)/(b8+b3)	Gitelson et al. 1996
SR	Simple Ratio Vegetation Index	b8/b4	Birth and McVey 1968

3.3.2 Bark beetle green-attack stage detection

The detection of bark beetle green-attack stage is more complex than the gray-attack stage. As only bark beetle gray-attack stage field data is available, it is not known when and how many trees that were under green-attack stage in the past. According to data from the southern most bark beetle trap (Ljungby), we assumed that within one to tow months after the second peak (week 27, 2018), most of the spruce trees were still under green-attack stages (figure 3.3). Sentinel-2 satellite images (tile 33VVC and 33VUC) sensed on 26th of July (week 30) were hence obtained for detection. Satellite images from the same tiles sensed on 12th of October (week 41) was also obtained to test the differences between an early image and a late image.

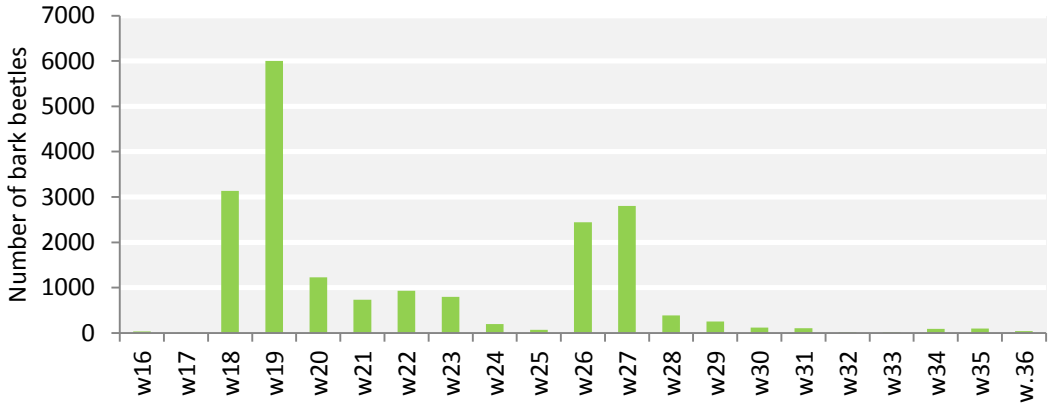


Figure 3.3 Numbers of bark beetles caught by the bark beetle traps in Ljungby station in 2018. "w" is short for "week"(for example: w16 = week 16, the 16th week of the year). Produced based on data from the Swedish Forest Agency (2019).

3.3.2.1 Data pre-processing

Data preprocessing of bark beetle green-attack stage detection followed similar procedures to gray-attack detection with a few steps that are different.

Raster data pre-processing

Sentinel-2 level-2A 20m resolution bands were transformed from UTM zone33/WGS84 projection to SWEREF99TM projection system. Satellite data were thereafter cropped into the same extent as the study areas. Spruce mask created from NMD landcover map from the last part were directly used without modification, due

to that no clear-cutting activity were found in the images used for green-attack detection.

Field data pre-processing

Field data were first overlaid with spruce mask to remove points with positioning errors and then aggregated into polygons with a maximum distance of 20 m. Same as in gray-attack detection, samples of healthy spruce containing less than 20 trees in the random sampling areas were removed. New healthy spruce samples were taken randomly in the individual sampling areas, where no damages were found. Polygons with the size of 400 to 800 m^2 (1 to 2 pixels) were created centered at the remaining healthy and bark beetle infested samples. A total of 26 bark beetle damaged samples and 80 healthy tree samples (34 from the random sampling area and 46 taken manually from the individually sampled area) were created, in which 15 bark beetle damaged samples and 42 healthy spruce samples were randomly divided as training data whereas the rest data were used for accuracy analysis.

3.3.2.2 Image processing

During the bark beetle green-attack stage, the spruce needles remain green in color, making it difficult to detect using the same method as for the gray-attack stage. Green-attack stage detection of bark beetle outbreaks requires involvement of more spectral bands. The Sentinel-2 20m resolution bands provide red-edge and SWIR bands that have been proven important to bark beetle outbreaks green-attack stage detection in previous studies (Abdullah et al. 2018). Based on the previous studies (Lausch et al. 2013; Abdullah et al. 2018) and result from gray-attack stage detection, all 20 m resolution raw bands and a series of VIs related to leaf and canopy chlorophyll content, leaf pigment (anthocyanin and carotinoid), greenness, red-edge bands and water content (all variables) were selected and compared, and ranked using random forest algorithm (table 3.3). The number of variables to achieve a high and robust accuracy and cost lower computing time was estimated using a threshold method. The 13 to 14 most important variables were proven most accurate and efficient for random forest classification.

Table 3.3. Formulas of the vegetation indices used in bark beetle green-attack stage detection calculated with Sentinel-2 20m resolution bands. “b” is short for “bands” (for example: b2 = band 2). CCCI, CIgreen, CIrededge and CVI are chlorophyll content related. ARI, CRI are other leaf pigments related vegetation indices. GNDVI and NDVI are greenness related vegetation indices. NDRE1, NDRE2, RENDVI1 and RENDVI2 are red-edge vegetation indices. DWSI, MSI, NDWI and VMI are water content related vegetation indices.

Type	Vegetation indices	Full name	Formula	Reference
Chlorophyll	CCCI	Canopy Chlorophyll Content Index	((b8a-b5)/(b8a+b5))/((b8a-b4)/(b8a+b4))	El-Shikha et al. 2008
	CIgreen	Chlorophyll Index Green	b8a/b3-1	Gitelson et al. 2003
	CIrededge	Chlorophyll Index RedEdge	b8a/b5-1	Gitelson et al. 2003
	CVI	Chlorophyll Vegetation Index	b8a*b4/b3^2	Vincini et al. 2008
Pigment	ARI	Anthocyanin Reflectance Index	b8a*(1/b3-1/b5)	Gitelson et al. 2006
	CRI	Carotinoid Reflectance Index	1/b2-1/b3	Gitelson et al. 2001
Greenness	GNDVI	Green Normalized Difference Vegetation Index	(b8a-b3)/(b8a+b3)	Gitelson et al. 1996
	NDVI	Normalized Difference Vegetation Index	(b8a-b4)/(b8a+b4)	Rouse Jr 1972
Red-edge	NDRE1	Normalized Difference Red-Edge Index	(b8a-b5)/(b8a+b5)	Barnes et al. 2000
	NDRE2	Normalized Difference Red-Edge Index	(b8a-b6)/(b8a+b6)	
	RENDVI1	Red-edge Normalized Difference Vegetation Index	(b5-b4)/(b5+b4)	Gitelson and Merzlyak 1994
	RENDVI2	Red-edge Normalized Difference Vegetation Index	(b6-b4)/(b6+b4)	
Water content	DWSI	Disease-Water Stress Index	(b8a+b3)/(b11+b4)	Galvao et al. 2005
	MSI	Moisture Stress Index	b11/b8a	Hunt and Rock 1989
	NDWI	Normalized Difference Water Index	(b8a-b11)/(b8a+b11)	Gao 1996
	VMI	Vegetation Moisture Index	((b8a+0.1)-(b12+0.02))/((b8a+0.1)+(b12+0.02))	Ceccato et al. 2002

3.3.3 Application in other study areas

In order to test if the method can be generalized, the same methods were applied on study area 2 and study area 3. In these two areas, only field data containing number of bark beetle damaged trees and locations was available. Thus, we assumed that the

spruce areas with no bark beetle damages discovered were healthy spruces.

Gray-attack stage detection

Sentinel-2 tile 33VVC sensed on April 2nd was acquired for test area 1 whereas tile 33VWE from the same date was acquired for test area 2. The threshold between spruce forest and clear-cutting area from study area 1 was applied in test area 1. However, a new threshold was calculated for test area 2 according to the difference in the vegetation cover. Satellite 10m resolution bands and landcover data went through the same preprocessing procedure as in study area 1. Field data was thereafter reduced and polygons covering 1 to 2 pixels (100 to 200 m^2) were then created centered at the damaged points manually. 38 bark beetle damaged spruces polygons were created for test area 1 whereas 41 polygons were created for test area 2. After that, 40 and 60 healthy spruce polygons were randomly created in test area 1 and test area 2, respectively. Sample polygons were randomly divided into training and validation data. VIs (table 3.2) were calculated from 10m resolution bands for both study areas and methods with the best results from study area 1 are tested in the two areas.

Green-attack stage detection

Sentinel-2 tile 33VVC sensed on 26th of July (week 30) for test area 1 and 33VWE sensed on 8th (week 27) and 31st of July (week 31) for test area 2 were obtained and preprocessed. After the same data preprocessing procedure as described in *3.3.2.1Data pre-processing*, 30 and 39 bark beetle damaged polygons, covering 400 to 1600 m2 area, were created for test area 1 and test area 2, respectively. Healthy spruce polygons covering 400 to 1600 m^2 were created centered at the same areas created in gray-attack stage detection. VIs (table 3.3) were calculated with 20m resolution bands and variable importance were ranked using random forest algorithm. The 13 to 14 most important variables were selected for classification.

3.3.4 Accuracy analysis

The accuracy of the classification results was evaluated using confusion matrix. Confusion matrix is a simple and effective evaluation method that has been commonly used for classification accuracy test. The matrix displays the number of pixels that were assigned to each class on the map in comparison to the reference data (the ground truth) and the total number. The total accuracy is calculated by the number of correctly assigned pixels (the numbers at the diagonal) dividing by the total number of all evaluated pixels. For example, in table 3.4. , the total accuracy is (10 + 22)/ 40 = 0.8.

With confusion matrix, it is possible to calculate Cohen's Kappa. Cohen's Kappa was introduced by Cohen (1960) as a way of measuring the agreement of observing the same phenomenon by different observers (Ben-David 2008). The formula of Cohen's Kappa is:

$$K = (P_0 - P_c)/(1 - P_c)$$

in which P_0 stands for the total agreement and P_c is the agreement happened by chance. For example, in table 3.4. P_0 = 12/40 *16/40 + 28/40 * 24/40 = 0.54. In this case, K = (0.8 - 0.54) / (1 – 0.54) = 0.57.

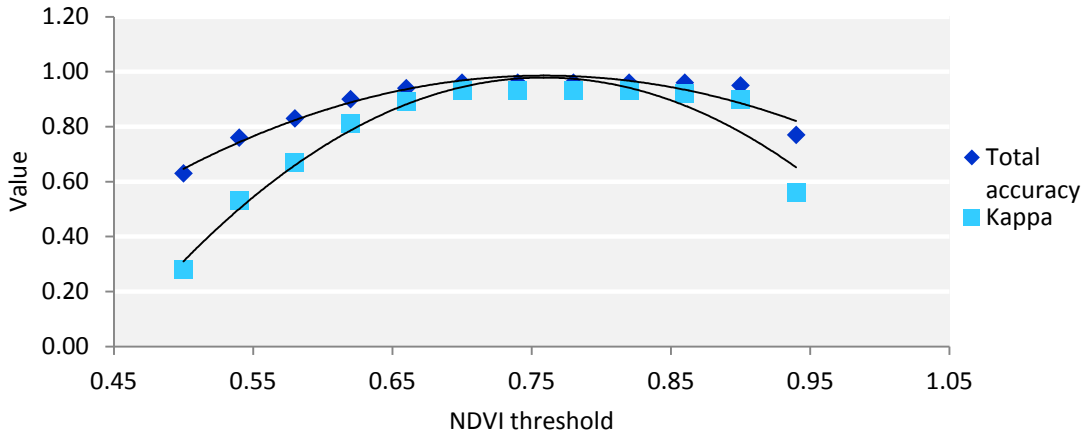
Table 3.4. An example confusion matrix. The class 1 and class 2 in the columns stand for the classes in the ground truth data (the real situation). The class 1 and class 2 in the rows stand for the classes classified in the map (the predicted situation).

		Reference		
	No.	Class 1	Class 2	Total
Map	Class 1	10	6	16
	Class 2	2	22	24
	Total	12	28	40

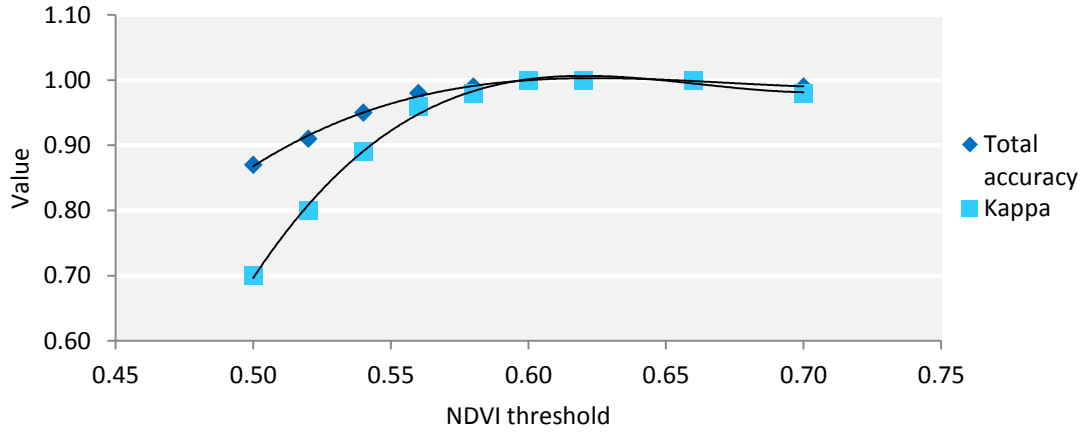
4. Results

4.1 Identification of clear-cut areas

Figure 4.1 shows the result of spruce mask modification on two different dates and study areas. According to the graph a., on March 1st, 2019, the best NDVI threshold is approximately 0.75 in study area 1, which is also applied to test area 1. On April 2nd 2019, the best threshold is about 0.60 in test area 2, according to graph b.



a.



b.

Figure 4.1. The thresholds applied to distinguish clear-cut areas from spruce forests in the masked area and the corresponding total accuracy and Kappa, fitted with 2nd order polynomial for a. and 3rd order polynomial for b. The x-axis represents the NDVI threshold, above which it is assigned to spruce forest and below which it is classified as clear-cutting area. The y-axis exhibits the corresponding total accuracy and Kappa value when applying a certain threshold. a. Result in study area 1 on March 1st 2019. b. Result in test area 2 on April 2nd 2019.

4.2 Gray-attack stage detection

The spectral signatures of pixels with bark beetle infested healthy spruce trees and pixels with only healthy spruce trees are illustrated in figure 4.2. On average, pixels with bark beetle damaged spruce trees has higher reflectance at the visible wavelength region (band 2, 3, 4) and lower reflectance on the NIR wavelength region (band 8) than the healthy spruce pixels.

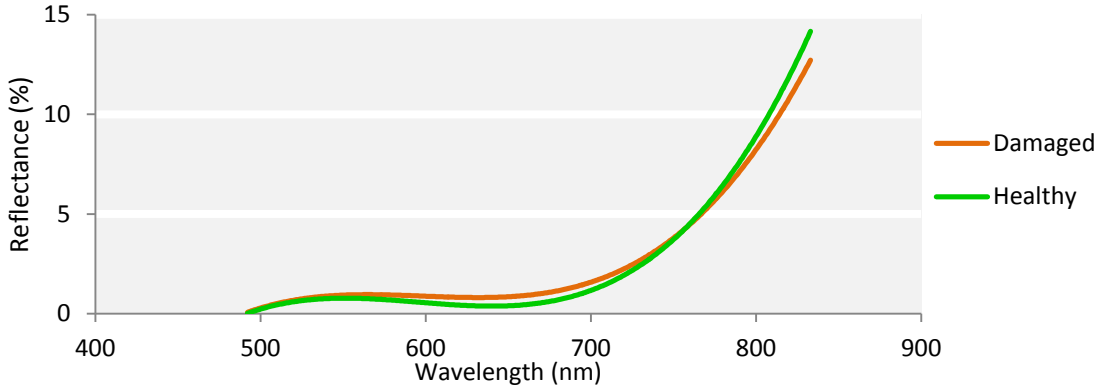


Figure 4.2. Spectral signature for healthy spruce pixels and pixels with damaged spruce trees during bark beetle gray-attack stage, in study area 1. Mean reflectance values of all pixels in the two classes were first plotted with the corresponding wavelengths and then fitted with 3rd order polynomial. Image sensed on March 1st, 2019.

Figure 4.3. depicts the standard deviation, maximum, minimum values, and the mean values of the reflectance of the two classes for different bands. The pixels with bark beetle damaged spruces shows higher standard deviation.

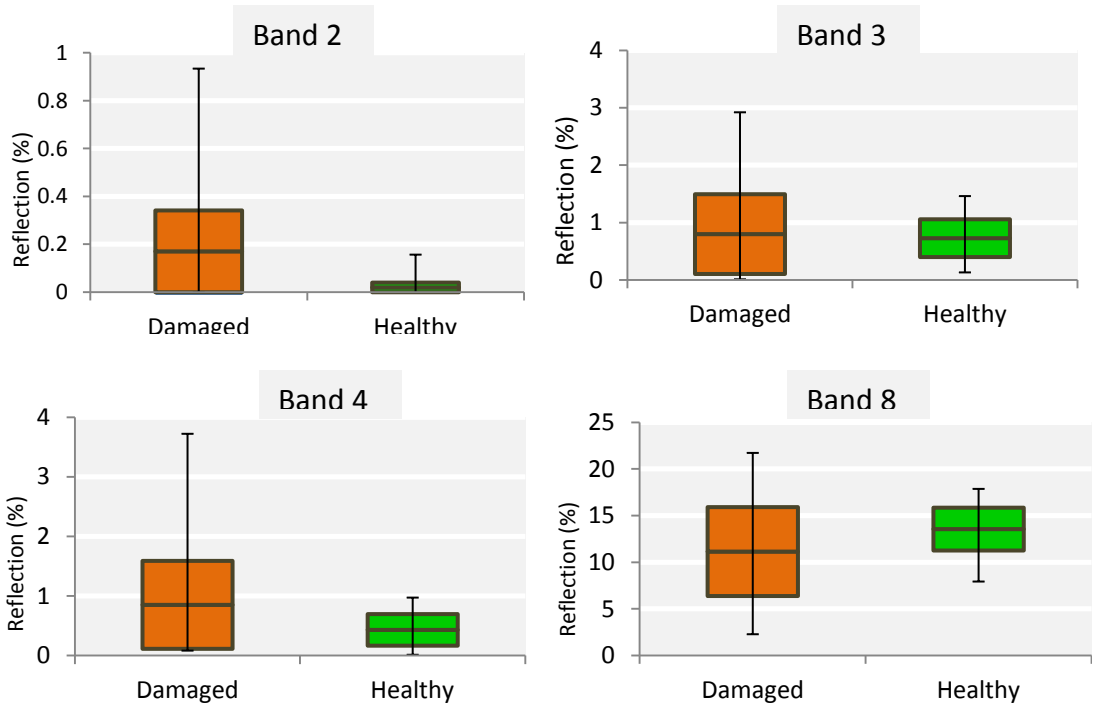


Figure 4.3. Box plots of reflectance of pixels with bark beetle infested spruce trees (orange color) and pixels with only healthy trees (green color) for Sentienl-2 10m resolution bands (band 2, 3, 4, 8) in study area 1, on March 1st. The middle line stands for the mean reflectance value and the box stands for 1* standard deviation above and below the mean value. The error bar stands for the maximum and minimum reflectance values. Band 2 is the blue band, band 3 is the green band, band 4 is the red band and band 8 is the NIR band.

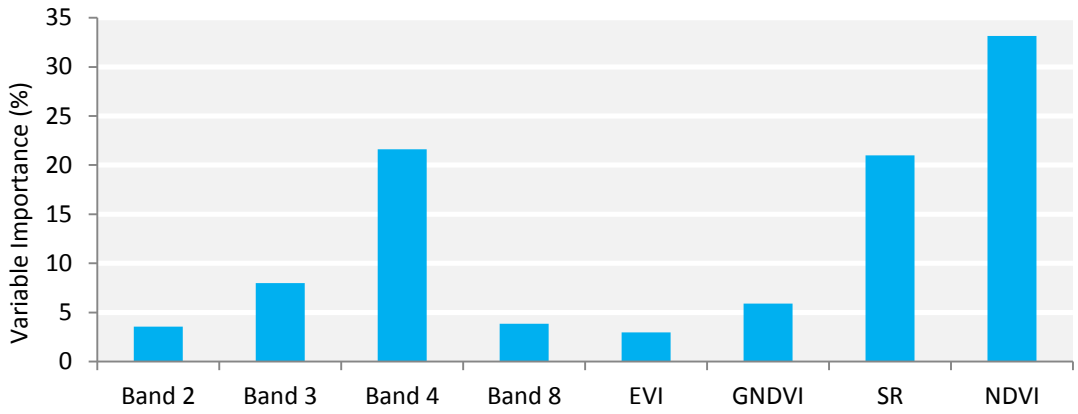


Figure 4.4. Percentage variable importance for distinguishing healthy spruce class and spruce with bark beetle gray-attack damages class ranked by random forest algorithm in study area 1.

The result of random forest classification indicates that NDVI (33.14%), SR (21.01%) and band 4 (NIR) (21.60%) are the most important variables (figure 4.4.). None of the remaining variables, however, shows particularly low importance. Hence, all variables were included in the classification.

Scatter plots (figure 4.5.) with two-two combinations of the four VIs depict the separability between bark beetle infested and healthy spruce trees. As it can be seen from the graph, GNDVI-EVI and NDVI-GNDVI combinations can effectively distinguish the two classes.

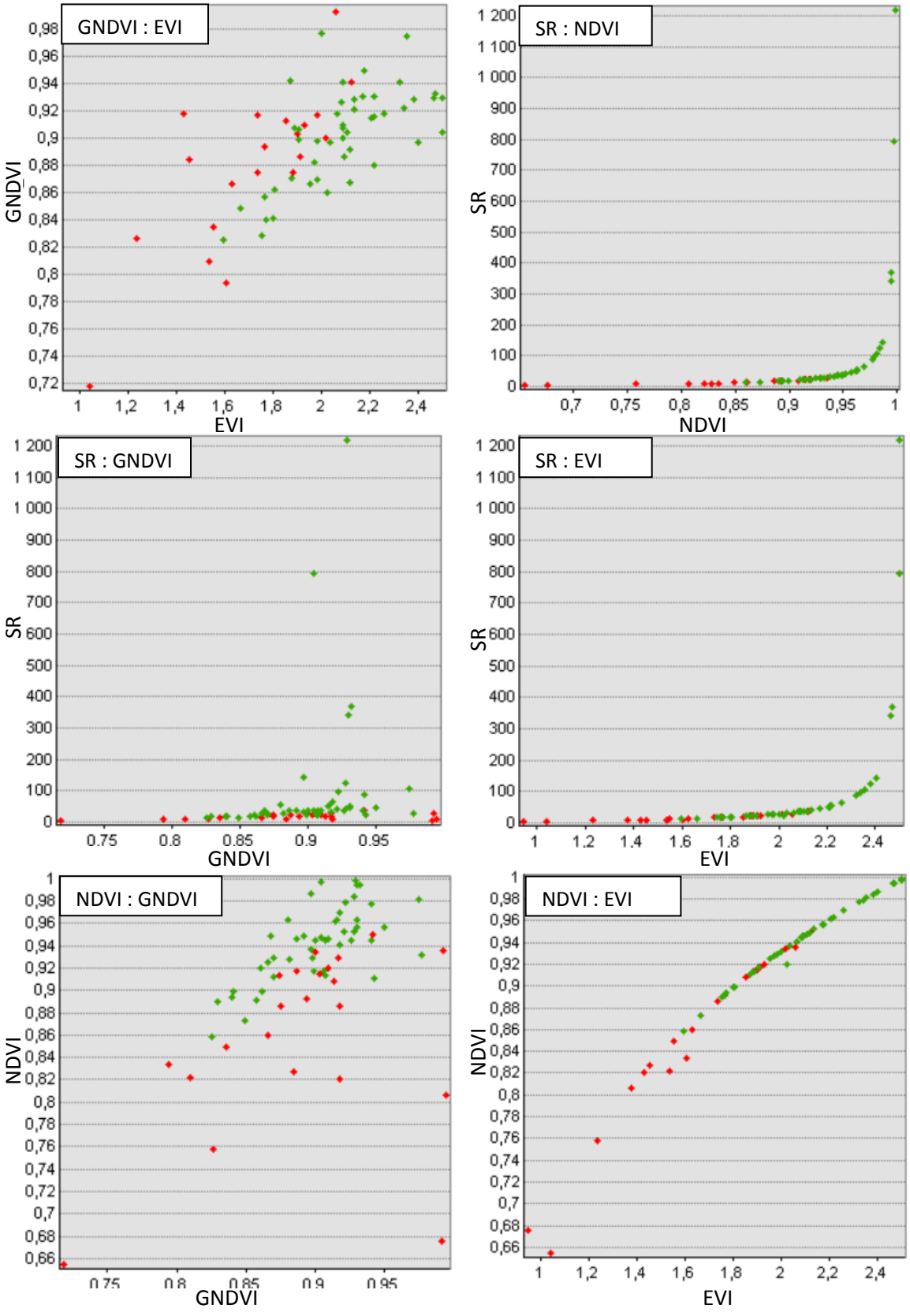


Figure 4.5. Two-two scatter plots of vegetation indices values used in bark beetle gray-attack stage in this project, for pixels with bark beetle damaged spruce (red) and healthy spruces pixels (green).

Table 4.1. displays the accuracy of all methods tested in study area 1. From table 4.1. we can see, the best result is achieved using maximum likelihood classification based on EVI and GNDVI, resulting in 89% total accuracy and substantial agreement (Kappa = 74%). As for random forest classification, the highest accuracy is obtained when all variables were included, achieving 85% total accuracy and substantial agreement (Kappa =62%). Classification with combinations of raw bands or VIs shows higher accuracy and Kappa than with one VI. There is a slight trend that more variables used in the classification will result in higher accuracy.

Table 4.1 Bark beetle gray-attack stage classification total accuracy and Kappa of the corresponding classification method and variables used for classification, in study area 1.The bold values are the highest values within the same classification method.

	Maximum likelihood		Random forest	
	Total accuracy(%)	Kappa	Total accuracy(%)	Kappa
Raw bands	87	0.68	83	0.58
NDVI	83	0.56	74	0.33
GNDVI	83	0.59	70	0.34
EVI	83	0.56	74	0.33
SR	72	0.44	74	0.33
EVI&GNDVI	**89**	**0.74**	83	0.56
GNDVI&NDVI	83	0.56	79	0.45
All variables	83	0.58	**85**	**0.62**

When applying the best methods to test area 1 and test area 2, random forest using all variables achieved higher accuracy and better agreement than maximum likelihood method using EVI and GNDVI (table 4.2.). Detection in test area 1 was more accurate than in test area 2.

Table 4.2. The total accuracy and Kappa resulted from using the best methods selected from table 4.1. (methods with the highest accuracy) in test area 1 and 2. ML: Maximum Likelihood. RF: Random Forest.

	ML- EVI&GNDVI		RF-All variables	
	Accuracy(%)	Kappa	Accuracy(%)	Kappa
T1	61	0.21	76	0.53
T2	57	0.07	71	0.39

4.3 Green-attack stage detection

The result of the threshold method for variable reduction (figure 4.6.) indicates that there is a non linear increase of accuracy with the increasing number of variables. When the number of variables reaches 13, the accumulated variable importance reaches 75%, or the lowest importance is over 2.5%, total accuracy and Kappa will be stabilized as the highest value. Additional information can be found in appendix, table S1.

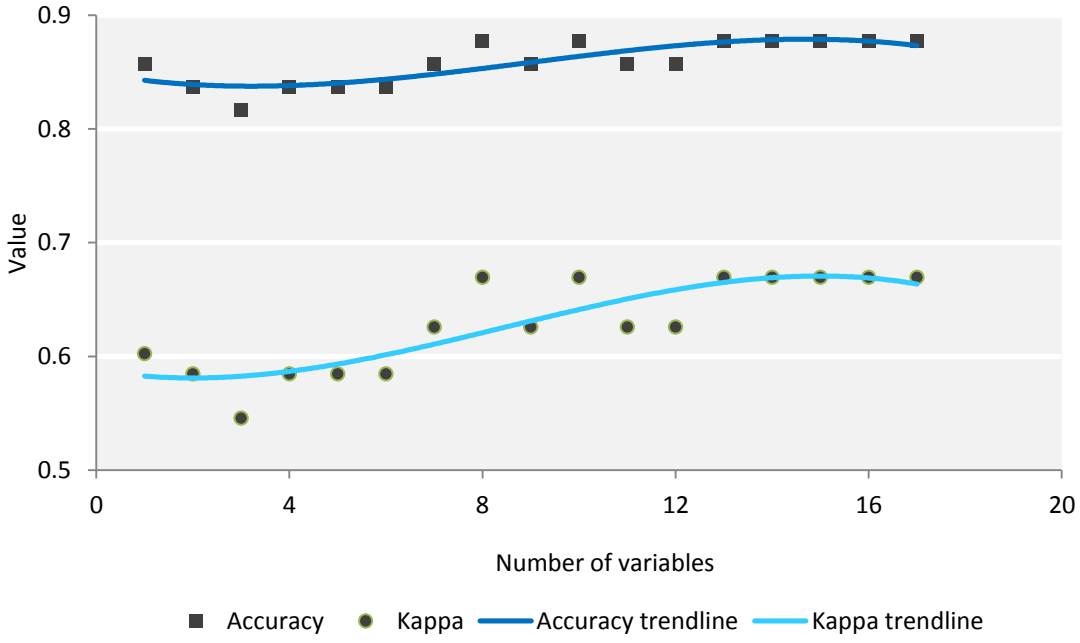
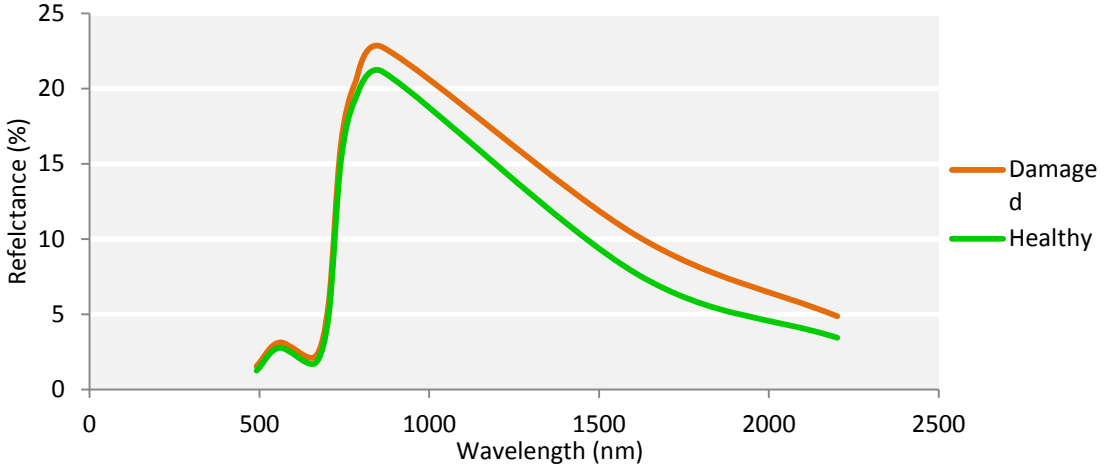
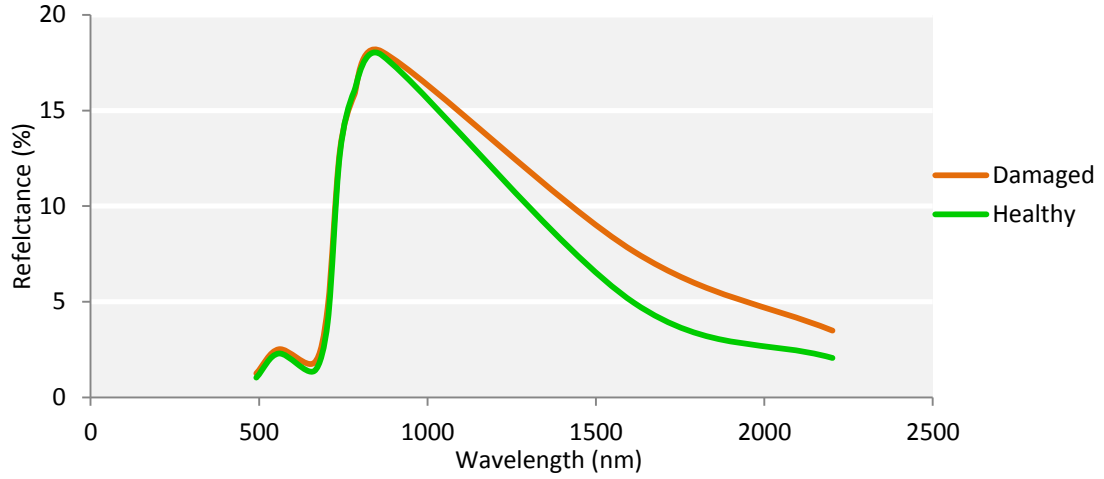


Figure 4.6. Scatter plot of the number of variables used for random forest classification and the corresponding total accuracy and Kappa value, fitted with 3rd polynomial (trendline).

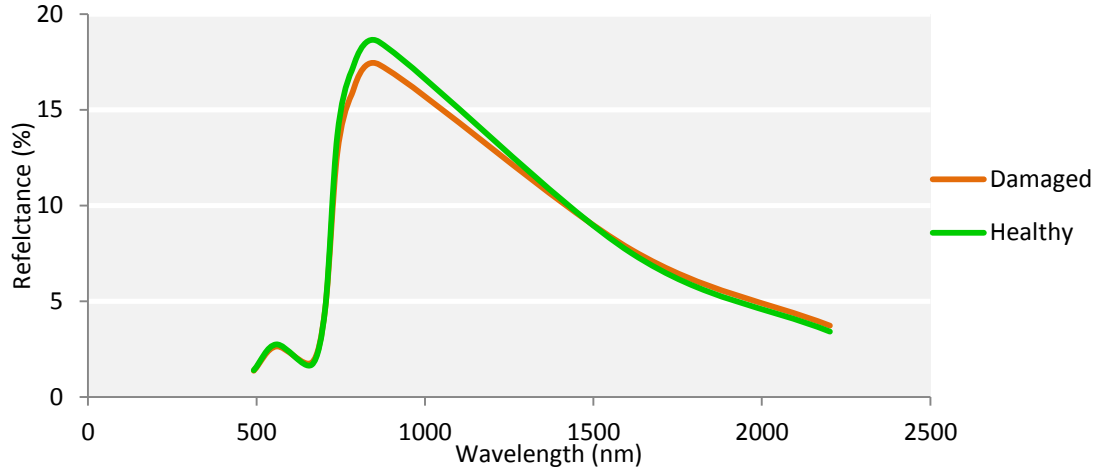
Figure 4.7. illustrate the spectral signatures of pixels with likely bark beetle green-attacked healthy spruce trees and healthy spruce trees in study area 1, test area 1 and test area 2 in last summer after the second swarming peak. In study area 1 on July 26th, on average healthy spruces has lower reflectance than bark beetle damaged spruces (a.). Similar result is also found in test area 2 on July 8th (d.). In both the later image in study area 1 (b.), test area 1 (c.) and test area 2 (e.) on July 31st, the pixels with likely bark beetle green-attacked spruces shows lower reflectance in the NIR wavelength regions and higher reflectance on the SWIR wavelength reflectance regions than pixels with healthy spruce, on average.


25
20
15
10
5
0
Refelctance (%)
0
500
1000
1500
2000
2500
Wavelength (nm)
Damage
d
Healthy

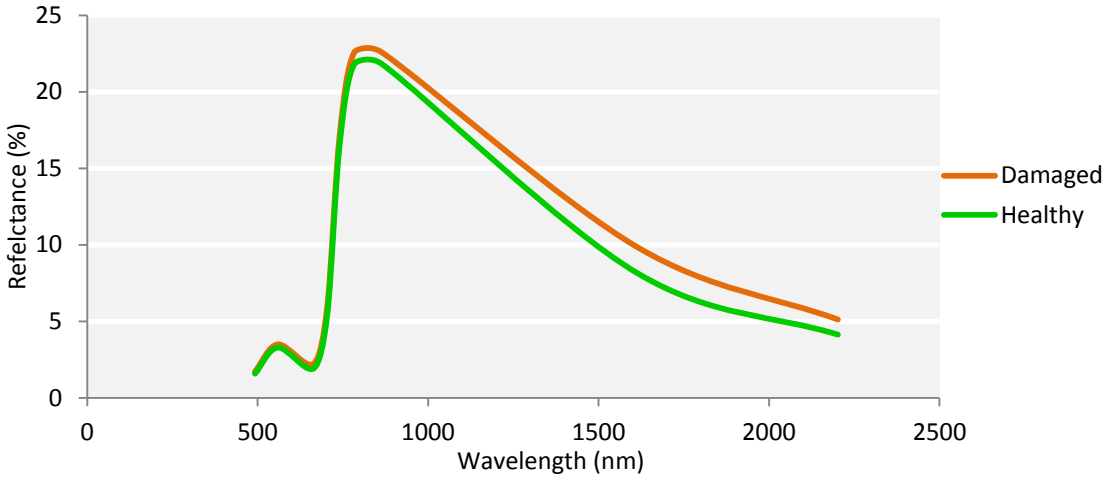
a.


20
15
10
5
0
Refelctance (%)
0
500
1000
1500
2000
2500
Wavelength (nm)
Damaged
Healthy

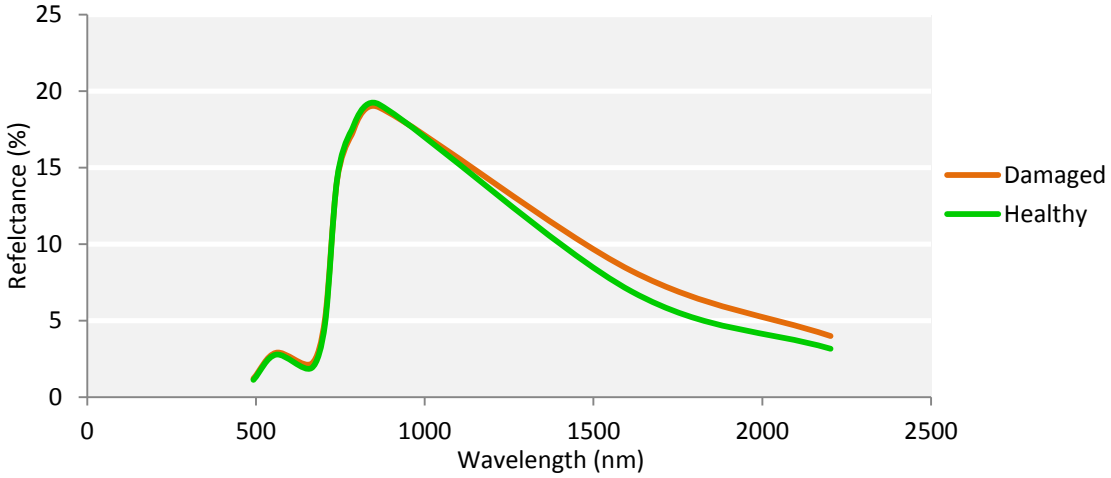
b.


20
15
10
5
0
Refelctance (%)
0
500
1000
1500
2000
2500
Wavelength (nm)
Damaged
Healthy

c.



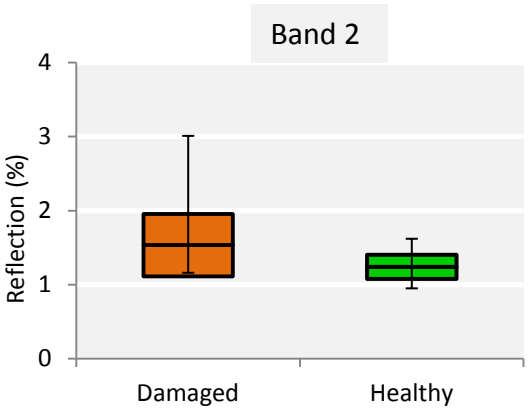
d.

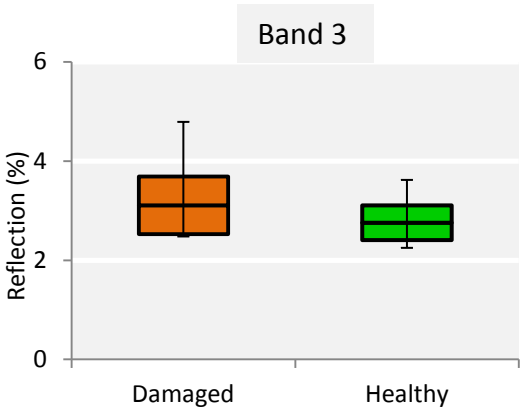


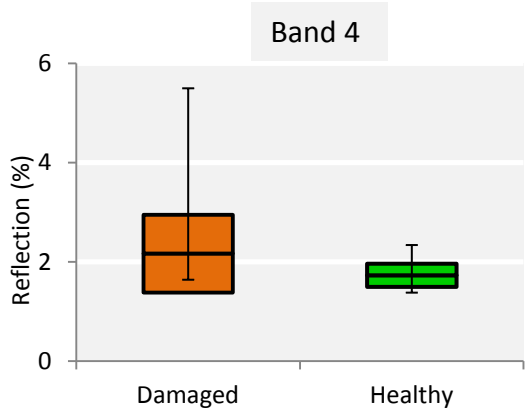
e.

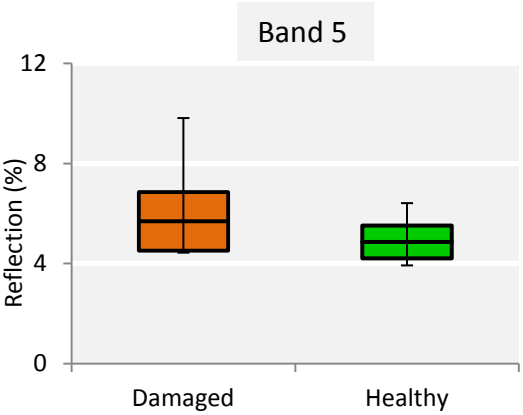
Figure 4.7. Spectral signature for healthy spruce pixels and pixels with damaged spruce trees during bark beetle green-attack stage for different study areas and on different dates. Mean reflectance values of all pixels in the two classes were first plotted with the corresponding wavelengths and then connected with smooth line.
a. Study area 1, July 26th, 2018. b. Study area 1, October 12th, 2018. c. Test area 1, July 26th, 2018. d. Test area 2, July 8th, 2018. e. Test area 2, July 31st, 2018

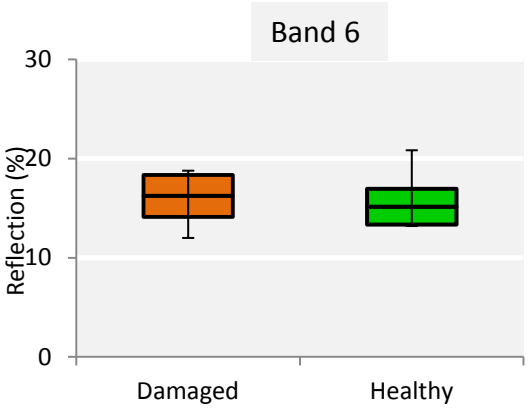
Figure 4.8. and figure 4.9. depicts the mean, minimum and maximum values of reflectance of the two classes for different bands. In general, pixels with bark beetle damaged spruce show larger in-class variation than the healthy spruce pixels for all bands (figure 4.8.). The rest of box plots are displayed in the appendix, figure S1., S2. and S3. One exception was discovered in test area 1 (figure 4.9.), on July 26th, 2018, in which healthy spruce pixels show larger in-class variation than pixels with bark beetle damaged spruce trees in the red-edge and NIR bands (band 6, 7, and 8).

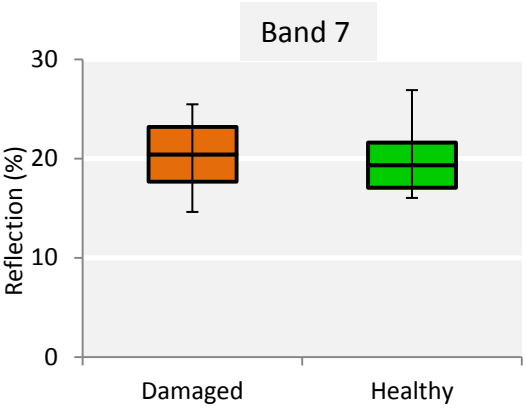

Band 2
Reflection (%)
0
1
2
3
4
Damaged
Healthy

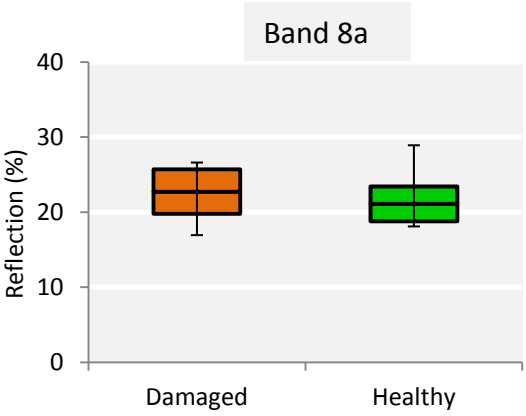

Band 3
Reflection (%)
0
2
4
6
Damaged
Healthy

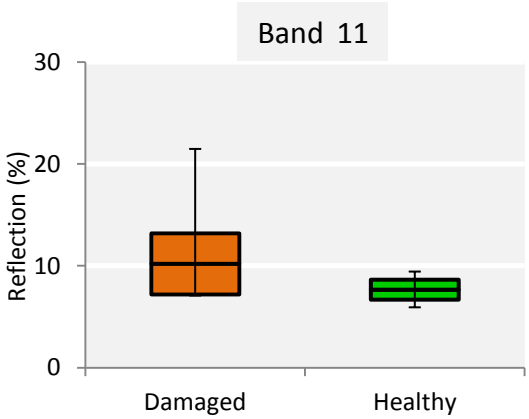

Band 4
Reflection (%)
0
2
4
6
Damaged
Healthy


Band 5
Reflection (%)
0
4
8
12
Damaged
Healthy


Band 6
Reflection (%)
0
10
20
30
Damaged
Healthy


Band 7
Reflection (%)
0
10
20
30
Damaged
Healthy


Band 8a
Reflection (%)
0
10
20
30
40
Damaged
Healthy


Band 11
Reflection (%)
0
10
20
30
Damaged
Healthy

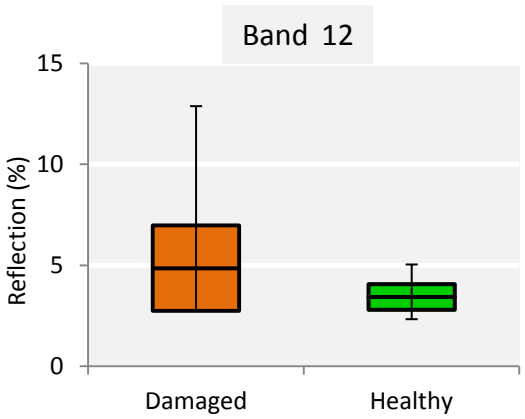
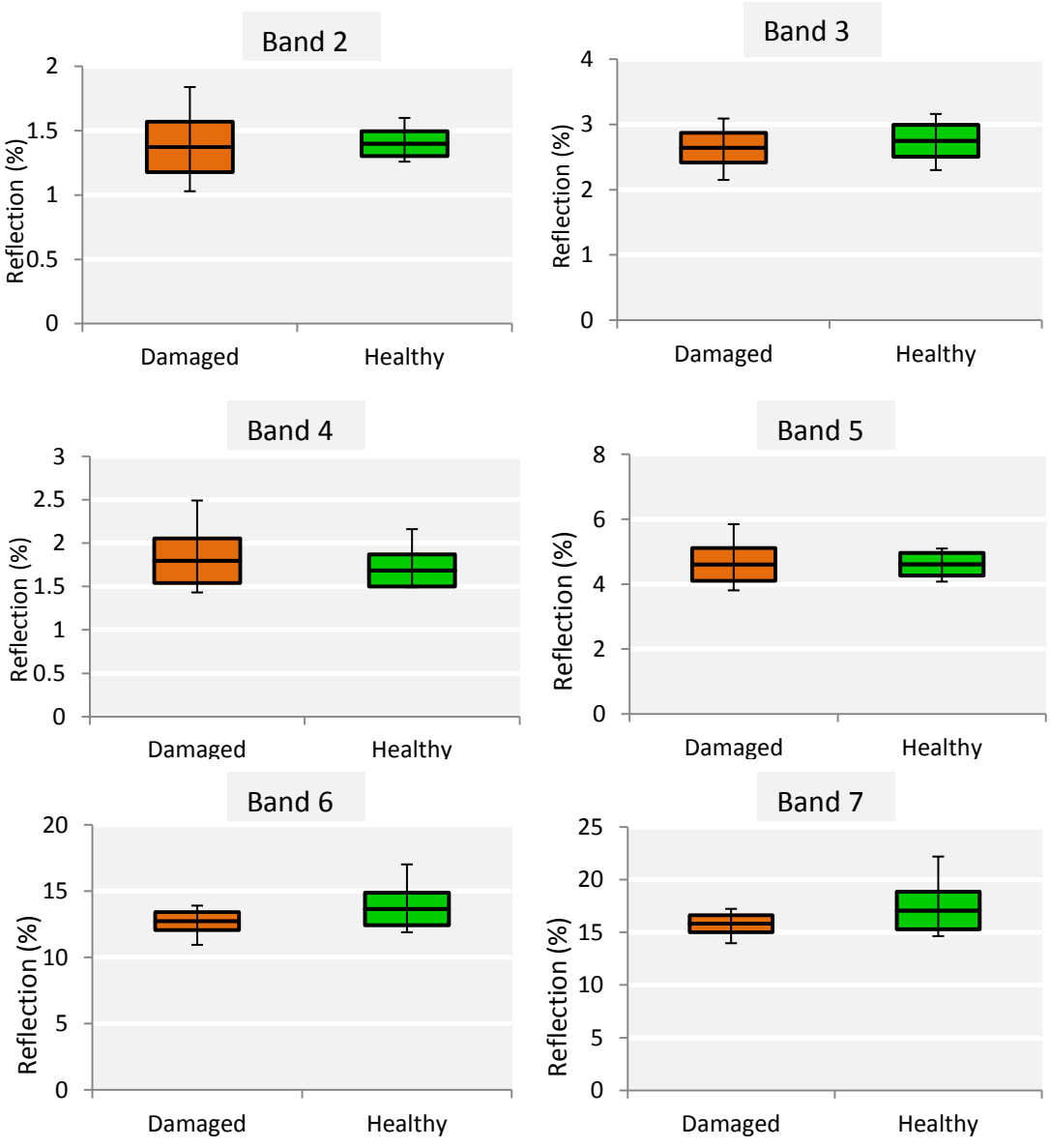


Figure 4.8. Box plots of reflectance of pixels with likely bark beetle green-attacked spruce trees and pixels with only healthy trees for Sentienl-2 20m resolution bands in study area 1, on July 26th, 2018. The middle line stands for the mean reflectance value and the box stands for 1* standard deviation above and below the mean value. The error bar stands for the maximum and minimum reflectance values. Band 2 to 4: visible bands. Band 5-7: red-edge bands. Band 8a: NIR band. Band 11, 12: SWIR bands.



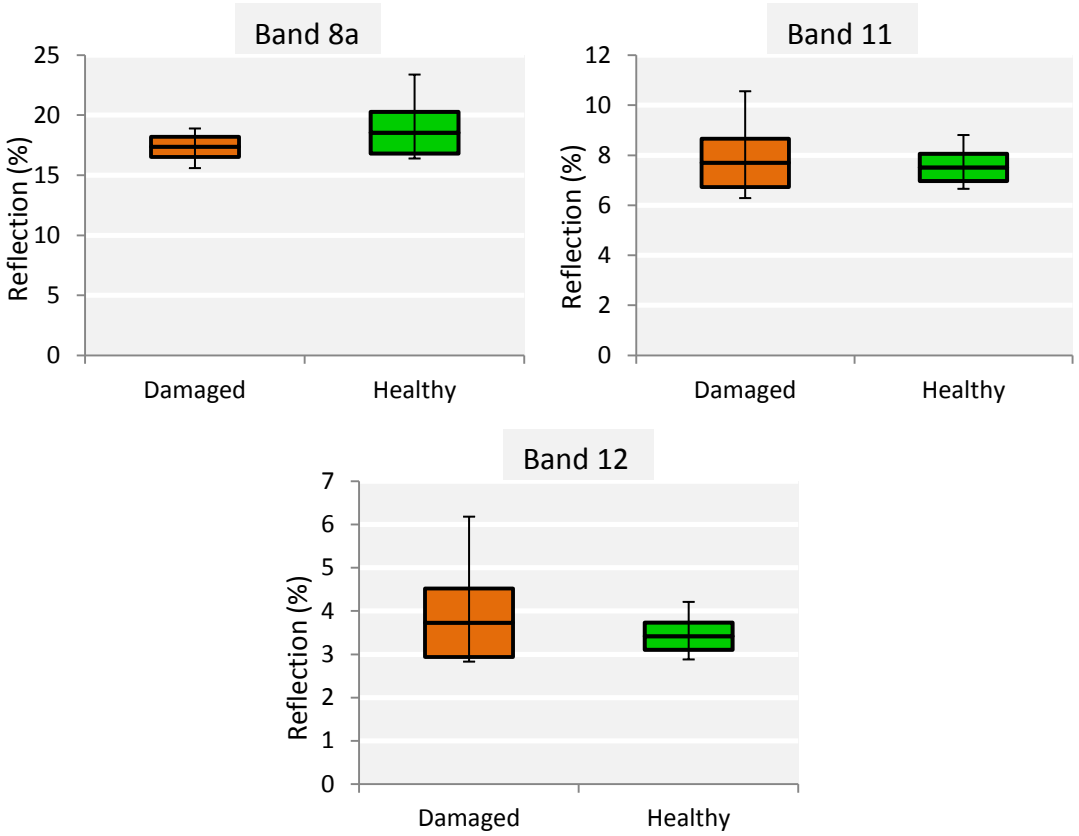
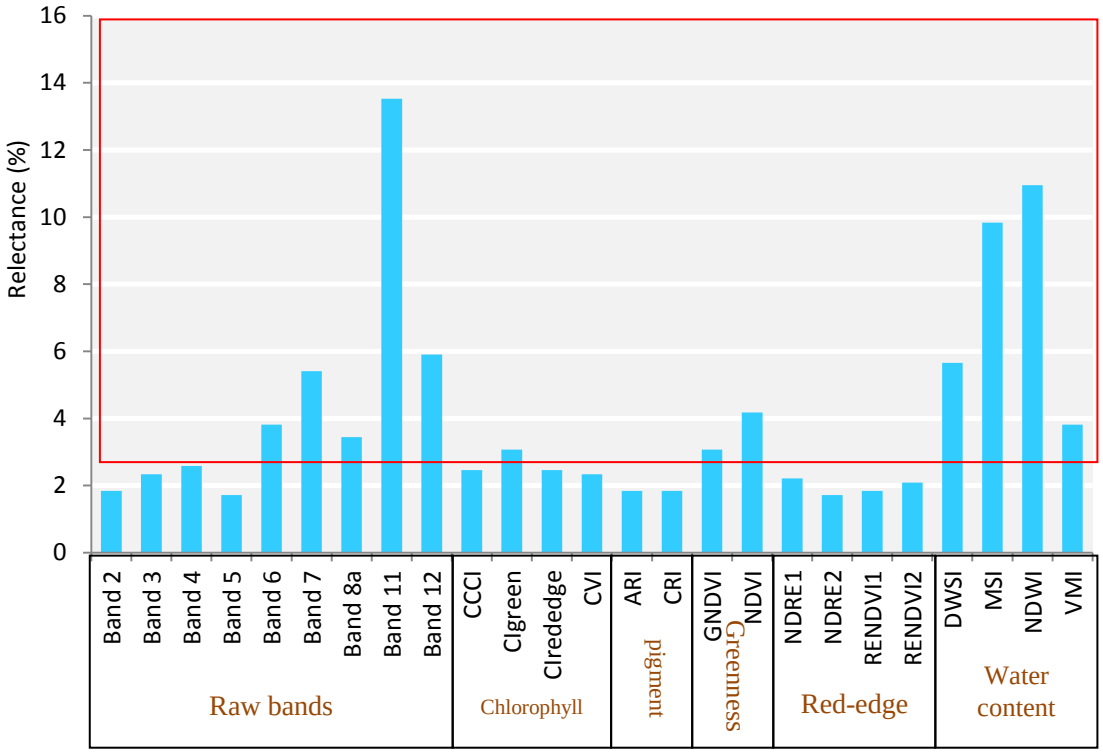
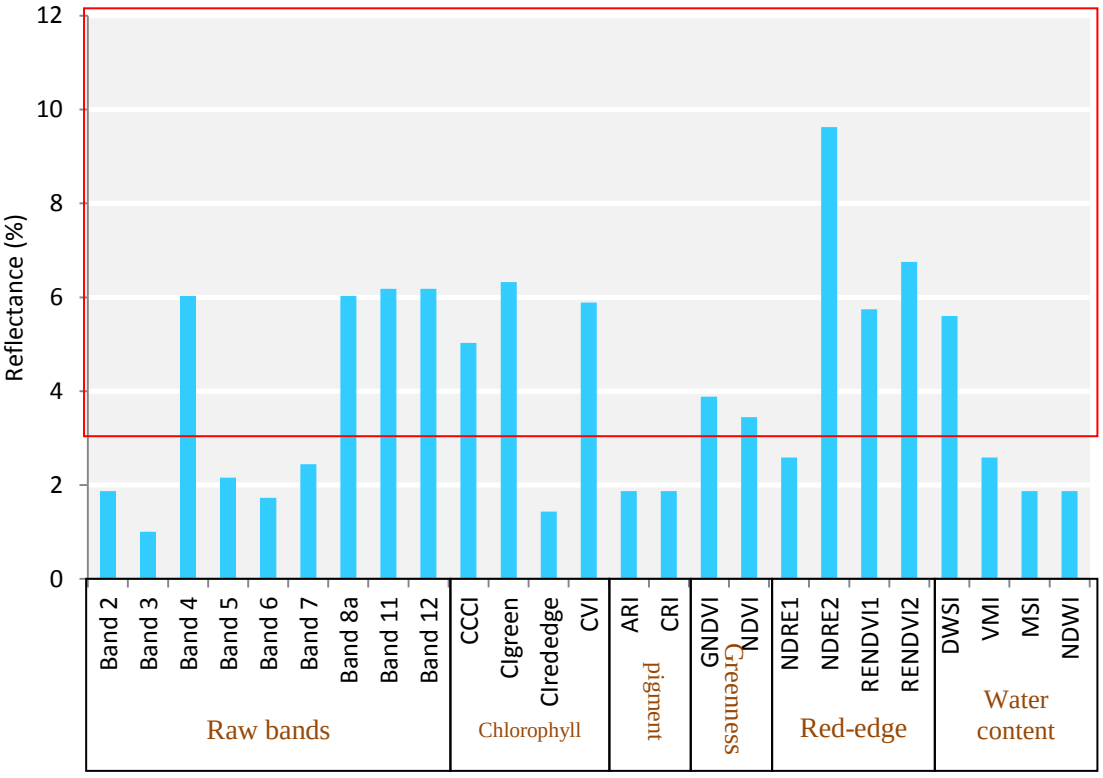


Figure 4.9. Box plots of reflectance of pixels with likely bark beetle green-attacked spruce trees and pixels with only healthy trees for Sentienl-2 20m resolution bands in test area 1, on July 26th, 2018. The middle line stands for the mean reflectance value and the box stands for 1* standard deviation above and below the mean value. The error bar stands for the maximum and minimum reflectance values. Band 2 to 4: visible bands. Band 5-7: red-edge bands. Band 8a: NIR band. Band 11, 12: SWIR bands.

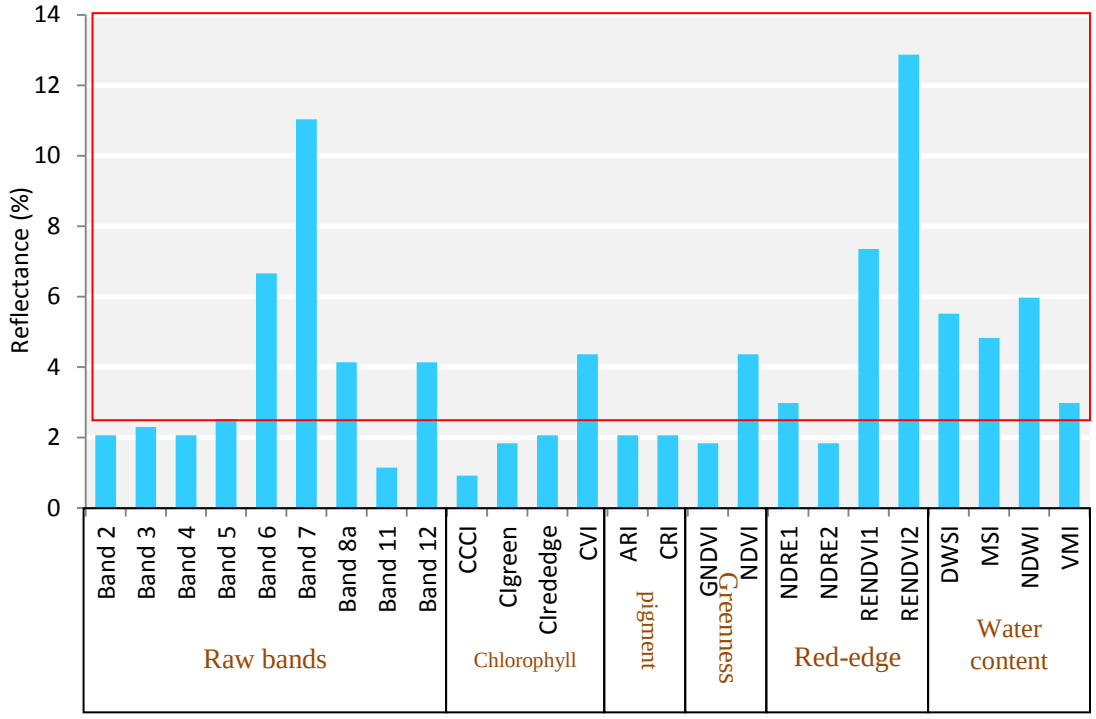
Figure 4.10. depicts the importance of variables calculated with random forest algorithm. The combination of most important variables differs for different areas. On July 26th, 2018, in study area 1 (a.), all water content related VIs and bands, greenness related VIs, red-edge band 6 and 7, NIR band 8 and Clgreen are the 13 most important bands to distinguish the likely green-attacked spruces and healthy spruce. In the later image on October 12th (b.), band 4, band 8a, band 11 and 12, CCCI, Clgreen, CVI, GNDVI, NDVI, NDRE2, RENDVI1 and 2 and NDWI are the most important. In test area 1 (c.), the first 14 most important variables are RENDVI2, Band 7, RENDVI1, band 6, NDWI, DWSI, MS, CVI, NDVI, band 8a, band 12, NDRE1, VMI and band 5. Water content related VIs and bands are less important than red-edge VIs and bands. In test area 2, water related VIs illustrates overwhelming importance on both July 8th and 31st (d. and e.) .


Relectance (%)
16
14
12
10
8
6
4
2
0
Band 2
Band 3
Band 4
Band 5
Band 6
Band 7
Band 8a
Band 11
Band 12
CCCI
CIgreen
CIrededge
CVI
ARI
CRI
GNDVI
NDVI
NDRE1
NDRE2
RENDVI1
RENDVI2
DWSI
MSI
NDWI
VMI
Raw bands
Chlorophyll
pigment
Greenness
Red-edge
Water content

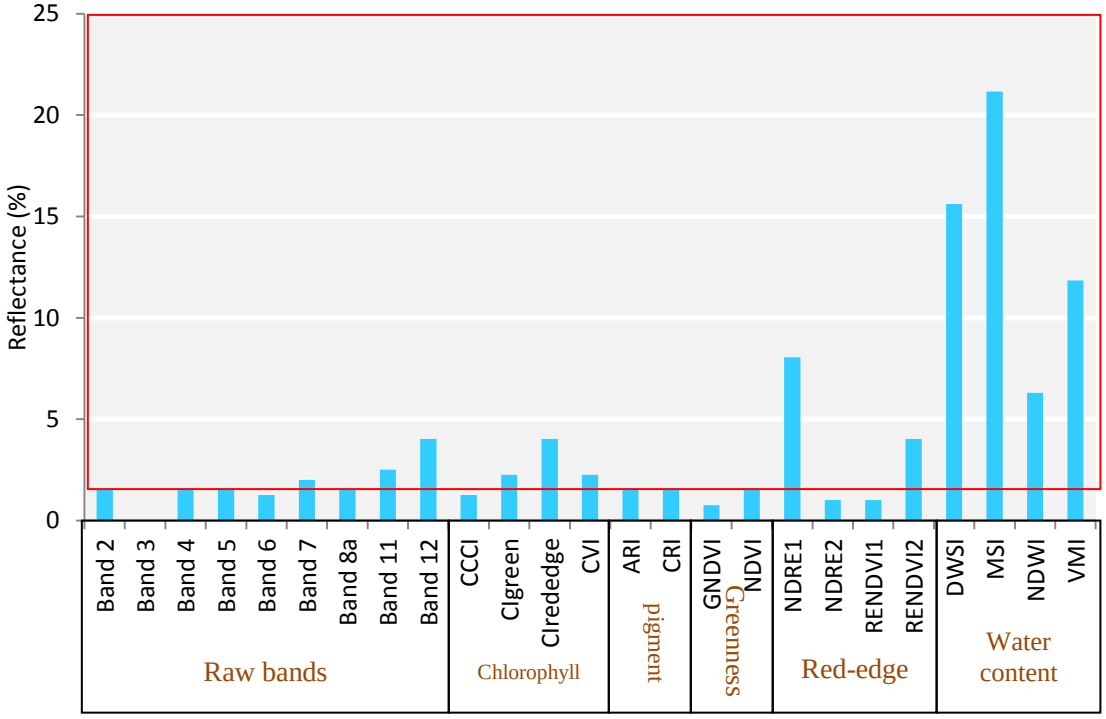
a.


Reflectance (%)
12
10
8
6
4
2
0
Band 2
Band 3
Band 4
Band 5
Band 6
Band 7
Band 8a
Band 11
Band 12
CCCI
CIgreen
CIrededge
CVI
ARI
CRI
GNDVI
NDVI
NDRE1
NDRE2
RENDVI1
RENDVI2
DWSI
VMI
MSI
NDWI
Raw bands
Chlorophyll
pigment
Greenness
Red-edge
Water content

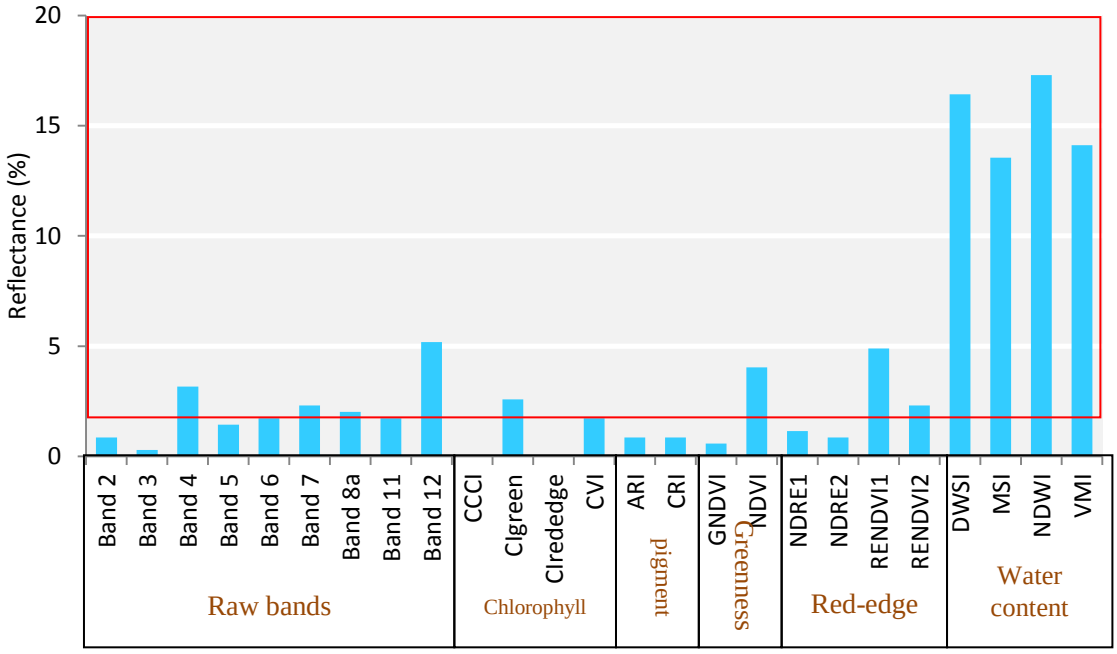
b.


Reflectance (%)
14
12
10
8
6
4
2
0
Band 2
Band 3
Band 4
Band 5
Band 6
Band 7
Band 8a
Band 11
Band 12
CCCI
CIgreen
CIrededge
CVI
ARI
CRI
GNDVI
NDVI
NDRE1
NDRE2
RENDVI1
RENDVI2
DWSI
MSI
NDWI
VMI
Raw bands
Chlorophyll
pigment
Greenness
Red-edge
Water content

c.


Reflectance (%)
25
20
15
10
5
0
Band 2
Band 3
Band 4
Band 5
Band 6
Band 7
Band 8a
Band 11
Band 12
CCCI
CIgreen
CIrededge
CVI
ARI
CRI
GNDVI
NDVI
NDRE1
NDRE2
RENDVI1
RENDVI2
DWSI
MSI
NDWI
VMI
Raw bands
Chlorophyll
pigment
Greenness
Red-edge
Water content

d.



e.

Figure 4.10. Percentage variable importance for distinguishing healthy spruce class and spruce with likely bark beetle green-attack damages class ranked by random forest algorithm. a. Study area 1, July 26th, 2018. b. Study area 1, October 12th, 2018. c. Test area 1, July 26th, 2018. d. Test area 2, July 8th, 2018. e. Test area 2, July 31st, 2018

Table 4.3. displays the accuracy for likely green-attack stage bark beetle detection in study area 1, test area 1 and test area 2. In study area 1 total accuracy is high (88%) and there's substantial agreement (Kappa=0.67) on July 26th and high total accuracy of 84% and moderate agreement is achieved on the later image. However, total accuracy is only 53% in test area 1 and there's barely any agreement shown from Kappa. In test area 2, total accuracy is 71% and moderate agreement is reached on July 31st whereas there is only 64% total accuracy and fair agreement on the earlier image.

Table 4.3. Total accuracy and Kappa values of bark beetle green-attack stage detection in different study areas and on different dates using random forest method. NoV stands for the number of variables used in the classification.

	Total accuracy (%)	Kappa	Sensed date	NoV
SA1	88	0.67	July 26th	13
SA1	84	0.58	October 12th	13
T1	53	0.03	July 26th	14
T2	64	0.27	July 8th	13
T2	71	0.42	July 31st	13

5. Discussion

5.1 Gray-attack stage detection

Sentinel-2 10m resolution raw bands and greenness related VIs shows high capability of distinguishing pixels with bark beetle gray-attacked (dead or with no needles) spruce trees from healthy spruce trees in study area 1. Classification with multi-variables achieved higher accuracy than with a single variable. Maximum likelihood classification achieves higher accuracy in study area 1 than random forest method in general. Maximum likelihood classification seemed to achieve up to 13 % higher total accuracy and up to 25% higher Kappa than random forest classification when applied to each single or combinations of the variables, except to SR and all variables. However, when applying both best methods to test area 1 and test area 2, maximum likelihood method achieved up to 16% lower accuracy and 32% lower Kappa than random forest method. Maximum likelihood assumes normal distribution in multivariate space. As we can see from figure 4.5., when pixel values in the classes resemble normal distribution and are not overlaid with each other, maximum likelihood method can efficiently achieve very high accuracy. However, random forest classification achieves high accuracy and is stable when detecting bark beetle gray-attacked spruces in other study areas. Many previous researches have compared both maximum likelihood and random forest classifications and reached a consistent result that random forest tend to achieve higher accuracy and more robust results than maximum likelihood (Ok et al. 2012; Nitze et al. 2012; Dixon and Candade 2008). On the other hand, classification result of random forest is influenced by the parameters: number of trees and tree depth. Tree depth 30 has been tested to be deep enough to not influence the result whereas the choice of 100 decision trees was based on the research of Oshiro et al. (2012). In their research, numerous different classifications were carried out to calculate the optimal number of classification trees to use. However, a similar threshold method was carried out by Ok et al. (2012) in which they discovered 200 was the optimal amount of trees for the best classification result. This infers that there's a chance for random forest method to achieve even higher

accuracy in this project if the number of the classification trees is increased.

In study area 1, the variation of pixel value in bark beetle damaged spruces class is larger than healthy spruce class. On average, the blue and green wavelength region shows less sensitivity to bark beetle gray-attack than red and NIR wavelength region. Also, variable importance analysis illustrated that red and NIR related bands and VIs are more important than the rest. In previous studies, NDVI has proven to be less effective to identify bark beetle damage especially green-attack stage detection (Havašová et al. 2015; Abdullah et al. 2017). In contrast, results of this project suggest that with the 10m resolution Sentinel-2 bands, when combined raw bands and other greenness related VIs, NDVI is a useful index for bark beetle gray-attack detection.

5.2 Green-attack stage detection

Bark beetle green-attack stage detection has always been problematic. Abdullah et al (2017) has shed light on the potential of Sentinel-2 20m resolution bands in bark beetle green-attack stage detection in Germany, especially with the water related and red-edge VIs, which is confirmed in this project. In general, in study area 1, bark beetle detection achieved very high accuracy and substantial agreement. However, the accuracy is relatively low in test area 1 and the classification is almost random. In test area 2, the classification accuracy and Kappa are more reasonable. Interestingly, the spectral signatures of both classes seem unexpected in study area 1 on July 26th: the average reflectance value is higher for damaged spruces class than healthy spruce class. Similar results were shown in test area 1 on July 8th. According to the theory, reflectance of stressed vegetation will decrease in the visible region, as a result of the weakened absorption caused by the decrease of chlorophyll content. The reflectance increases in NIR region as the destruction of needle cell structure starts to occur through continuous stress (Ortiz et al. 2013). The absorption of SWIR region decreases with the decrease of canopy water content caused by stress, leading to an increase in reflectance (Abdullah et al. 2017). In contrast, spectral signatures in test area 1 and both of the later image in test area 2 and study area 1 are more consistent with the theory. The reason behind the unexpected spectral signature in the early images in both study area 1 and test area 2 may be associated with the drought that

occurred in the summer of 2018. It may also infer that it was too early for bark beetle damages to be detected when bark beetle damages haven't been enough to alter the functional and structural properties of the spruce stands. Except for test area 1, bark beetle damaged spruces has larger spread than healthy spruces in all study areas. This can be explained that bark beetle damaged spruces at that time are in more different situations than healthy spruces: newly attacked, attacked but still green, turning brown etc. Larger variation in bark beetle damaged spruce class is also found in the research of Ortiz et al. (2013). The later image in test area 2 achieves higher total accuracy and better agreement than the earlier image. This is not in consistency with the total accuracy and Kappa in the two images in study area 1. Due to time limit, a second image in test area 1 was not tested. It is unknown if the later detection will lead to higher accuracy, yet it is likely the high accuracy in the early image in study area 1 is a coincidence due to the drought. As mentioned in the study of Ortiz et al. (2013), the spectral reflectance is influenced multiple stressors. Only relying on spectral signature, is it too universal to identify one particular stressor.

Variable importance analysis result corresponds to the spectral signatures of both classes. Water content related VIs (DWSI, MSI, NDWI, VMI) show the highest importance. This result is in line with previous bark beetle detection studies (Lausch el al. 2013; Havašová et al. 2015; Abdullah et al. 2017). Red-edge related bands and VIs especially band 6, 7, RENDVIs are also proven sensitive to bark beetle green-attack (Ortiz et al. 2013). The NIR and SWIR bands also show high sensitivity: bark beetle infestation leads to lower reflectance value in NIR region and higher reflectance in SWIR region (Immitzer et al. 2016; Abdullah et al. 2017). As for the chlorophyll content related VIs, CVI and Clgreen illustrate more sensitivity than the rest. However, the pigment related VIs: ARI and CRI are less effective in bark beetle green-attack stage detection. Similar results were also shown in research of Abdullah et al (2017), in which they found out pigment related VIs are not significant in distinguishing healthy and bark beetle infested spruces. Interestingly, in almost all analysis, NDVI is listed in the first 13 or 14 important variables, which disagrees with previous studies. This might be because the real situation of attacked spruces by that time actually varies from green brown to even dead. Without solid field data, it is hard to choose the best date for bark beetle green-attack detection. Besides, even in bark beetle green-attack stage, the change of water content and chlorophyll content is not

abrupt but rather slow and gradual: bark beetles attack the phloem of spruces, nutrient will be unable to transport to the roots, water content will be gradually reduced in the canopy and chlorophyll content decreases. The restrain of water happens rather early and was therefore proven useful for bark beetle green-attack stage detection. However, there might have taken a certain amount of time after bark beetle attacks to be significant enough for detection with satellite images. In addition, the amount of bark beetle damaged trees that is needed to be detected with sentinel-2 20m resolution bands is also of interest for further research.

5.3 Threshold

The threshold result shows that clear-cutting areas and spruce forests are highly separable with a NDVI threshold. For study area 1, the threshold is 0.75 which is relatively high. This may be related to the growth of grass and bushes on the areas after the clear-cutting activities. According to Nichol and Lee (2005), it is impossible to distinguish grassland from trees with satellite images only based on NDVI values. However, our study seems to have proven it is possible to distinguish clear-cut areas from dense spruce forests. Another threshold method regarding number of variables to use for random forest classification indicates that with the increased number of variables, total accuracy and Kappa has a tendency to go up. Accuracy and Kappa stays stable when the number of variables reaches a threshold, which also might refer to the accumulate variable importance or a lowest variable importance threshold value.

5.4 Sources of error and suggestions for improvement

The error in this project derives from both the input data and the classification method. The error of input data comes from noises in satellite data and error in field data. During the green-attack stage, it is difficult to choose which image to use due to frequent cloud cover. Bark beetle outbreaks in Sweden occurs in relatively small areas. Therefore, the positioning accuracy will be a vital factor to the final accuracy. Without knowing the GPS device and strategy used for field data collection, the positional uncertainty of the field data is hard to control. For test area 1 and test area 2, healthy spruce data is absent, and the assumption that the area without bark beetle damages detected is healthy spruce forest, might also cause error. In future studies, it

is recommended to carry out field survey in a smaller area in bark beetle green-attack stage and collect both bark beetle attacked data and healthy data. If possible, the sampling area should be larger than 20m * 20m and a large amount of pixels should be sampled. High accuracy GPS devices such as differential GPS are highly recommended for sampling. Sampling strategy can refer to research from Abdullah et al. (2017). On the other hand, as mentioned above, the change of spectral signature might be a result of many stress factors which leads to error in the detection result.

6. Conclusions

Sentinel-2 satellite 10m resolution data shows high ability to detect bark beetle attack in gray stage in the study areas in south Sweden, achieving up to 89% total accuracy and substantial agreement. Random forest method shows less limitation and achieves stable and robust result comparing to maximum likelihood method. Classification based on combinations of variables are more accurate than with only one variable.

The bark beetle green-attack stage detection shows reasonable results, highlighting the potential of Sentinel-2 20m resolution bands for bark beetle green-attack stage detection in Sweden. Water content related VIs and bands, red-edge VIs and bands, NIR band are the highly sensitive variables for green-attack stage detection. In addition, detection accuracy of bark beetle green-attack might be related to timing. Due to the limitation of the field data, the potential of Sentinel-2 data may not have been fully uncovered. In practice, remote sensing method might offer a time and cost-efficient alternative to the traditional detection methods. In order to be practically applied in forest management, further research is needed to more effectively and accurately identify bark beetle green-attack damages.

References

Anderbrant, O. 1989. Reemergence and second brood in the bark beetle Ips typographus. Holarct. Ecol. 12: 494-500.

Angst, A., R. Rüegg, and B. Forster. 2012. Declining Bark Beetle Densities (Ips typographus, Coleoptera: Scolytinae) from Infested Norway Spruce Stands and Possible Implications for Management. Psyche: A Journal of Entomology 2012: 1-7.

Barnes, E. M., T. R. Clarke, S. E. Richards, P. D. Colaizzi, J. Haberland, M. Kostrzewski, ... and R. J. Lascano. 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA (Vol. 1619).

Ben-David, A. 2008. Comparison of classification accuracy using Cohen's Weighted Kappa. Expert Systems with Applications 34: 825-832.

Birth, G. S., and G. R. McVey. 1968. Measuring the Color of Growing Turf with a Reflectance Spectrophotometer 1. Agronomy Journal 60: 640-643.

Breiman, L. 2001. Random forests. Machine learning 45: pp.5–32.

Ceccato, P., N. Gobron, S. Flasse, B. Pinty, and S. Tarantola. 2002. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. Remote sensing of environment 82: 188-197.

Christiansen, E. and A. Bakke. 1988. The spruce bark beetle of Eurasia. Dynamics of forest insect populations, Plenum Publishing Corporation.

Chuvieco, E. 2016. Fundamentals of satellite remote sensing: An environmental approach. CRC press.

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

Copernicus Open Access Hub. 2019. Access at https://scihub.copernicus.eu/dhus/#/home

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.

DIVA-GIS. 2019. Access at https://www.diva-gis.org/

Eidmann, H. H. 1992. Impact of bark beetles on forests and forestry in Sweden. Division of Forest Entomology, Swedish University of Agricultural Sciences 114: 193-200.

El-Shikha, D. M., E. M. Barnes, T. R. Clarke, D. J. Hunsaker, J. A. Haberland, P. J. Pinter Jr, ... and T. L Thompson. 2008. Remote sensing of cotton nitrogen status using the Canopy Chlorophyll Content Index (CCCI). Transactions of the ASABE 51: 73-82.

Erbilgin, N., and K. Raffa. 2002. Association of declining red pine stands with reduced populations of bark beetle predators, seasonal increases in root colonizing insects, and incidence of root pathogens. Forest Ecology and Management 164: 221-236.

ESA. 2019. Access at sentinel.esa.int.

Filchev, L. 2012. An assessment of European spruce bark beetle infestation using Worldview-2 satellite data. Proc. of European SCGIS Conf.: "Best practices: Application of GIS technologies for conservation of natural and cultural heritage sites: 9-16.

Franklin, S., M. Wulder, R. Skakun, and A. Carroll. 2003. Mountain Pine Beetle Red-Attack Forest Damage Classification Using Stratified Landsat TM Data in British Columbia, Canada. Photogrammetric Engineering & Remote Sensing 69: 283-288. American Society for Photogrammetry and Remote Sensing.

Galvao, L. S., A. R. Formaggio, and D. A. Tisot. 2005. Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. Remote Sensing of Environment 94:523-534.

Gao, B. C. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote sensing of environment 58:257-266.

Gitelson, A., & M. N. Merzlyak 1994. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. Journal of Photochemistry and Photobiology B: Biology 22: 247-252.

Gitelson, A. A., Y. J. Kaufman, and M. N. Merzlyak. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote sensing of Environment 58: 289-298.

Gitelson, A. A., M. N. Merzlyak, Y. Zur, R. Stark, & U. Gritz 2001. Non-destructive and remote sensing techniques for estimation of vegetation status.

Gitelson, A. A., A. Viña, T. J. Arkebauer, D. C. Rundquist, G. Keydan, & B. Leavitt. 2003. Remote estimation of leaf area index and green leaf biomass in maize

canopies. Geophysical Research Letters 30.

Gitelson, A. A., G. P. Keydan, & M. N. Merzlyak. 2006. Three‐band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophysical research letters 33.

Havašová, M., T. Bucha, J. Ferenčík, and R. Jakuš. 2015. Applicability of a vegetation indices-based method to map bark beetle outbreaks in the High Tatra Mountains. Annals of forest research 58: 295-310.

Horning, N. Random forests: An algorithm for image classification and generation of continuous fields data sets. In Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan, 9–11 December 2010.

Huete, A., Justice, C., and W. Van Leeuwen. 1999. MODIS vegetation index (MOD13). Algorithm theoretical basis document 3 213.

Hunt Jr, E. R., and B. N. Rock. 1989. Detection of changes in leaf water content using near-and middle-infrared reflectances. Remote sensing of environment 30: 43-54.

Immitzer, M. and C. Atzberger. 2014. Early detection of bark beetle infestation in Norway spruce (Picea abies, L.) using worldView-2 data frühzeitige erkennung von borkenkäferbefall an fichten mittels worldView-2 satellitendaten. Photogrammetrie-Fernerkundung-Geoinformation 2014: 351-367.

IPCC, 2018: Summary for Policymakers. In: Global Warming of 1.5 °C. An IPCC Special Report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press.

Ismail, R., O. Mutanga, and U. Bob. 2007. Forest health and vitality: the detection and monitoring of Pinus patula trees infected by Sirex noctilio using digital multispectral imagery. Southern Hemisphere Forestry Journal 69:39-47.

Jakus, R., W. Grodzki, M. Jezik, and M. Jachym. 2003. Definition of spatial patterns of bark beetle Ips typographus (L.) outbreak spreading in Tatra Mountains (Central Europe), using GIS.

Jönsson, A. M., L. M. Schroeder, F. Lagergren, O. Anderbrant, and B. Smith. 2012. Guess the impact of Ips typographus—an ecosystem modelling approach for simulating spruce bark beetle outbreaks. Agricultural and Forest Meteorology 166:

188-200.

Lambert, J., J. P. Denux, J.Verbesselt, G. Balent, and V. Cheret. 2015. Detecting clear-cuts and decreases in forest vitality using MODIS NDVI time series. Remote Sensing 7: 3588-3612.

Lausch, A., M. Heurich, D. Gordalla, H. Dobner, S. Gwillym-Margianto, and C. Salbach. 2013. Forecasting potential bark beetle outbreaks based on spruce forest vitality using hyperspectral remote-sensing techniques at different scales. Forest Ecology and Management 308: 76-89.

Liu, Y. and H. Dai. 2006. Application of bark beetle semiochemicals for quarantine of bark beetles in China. Journal of insect science 6.

Lobinger G. 1994. Air temperature as a limiting factor for flight activity of two species of pine bark beetles, Ips typographus L. and Pityogenes chalcographus L. (Col., Scolytidae). Anzeiger für Schadlingskunde, Pflanzenschutz, Umweltschutz, 67:14-17.

Logan, J. A., J. Régnière, and J. A. Powell. 2003. Assessing the impacts of global warming on forest pest dynamics. Frontiers in Ecology and the Environment 1:130–137.

Meddens, A. J., J. A.Hicke, L. A.Vierling, and A. T. Hudak. 2013. Evaluating methods to detect bark beetle-caused tree mortality using single-date and multi-date Landsat imagery. Remote Sensing of Environment 132:49-58.

Morris, J. L., S. Cottrell, C. J. Fettig, W. D. Hansen, R. L. Sherriff, V. A. Carter, ... and P. E. Higuera. 2017. Managing bark beetle impacts on ecosystems and society: priority questions to motivate future research. Journal of applied ecology 54: 750-760.

Nichol, J., and C. M. Lee. 2005. Urban vegetation monitoring in Hong Kong using high resolution multispectral images. International Journal of Remote Sensing 26: 903-918.

Nitze, I., U. Schulthess and H. Asche. 2012. Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. Proc. of the 4th GEOBIA, 35.

NMD, Nationella marktäckedata. 2019. Access at www.naturvardsverket.se

Ok, A. O., O. Akar, and O. Gungor. 2012. Evaluation of random forest method for agricultural crop classification. European Journal of Remote Sensing 45: 421-432.

Olsson, P., T. Kantola, P. Lyytikäinen-Saarenmaa, A. Jönsson, and L. Eklundh. 2016. Development of a method for monitoring of insect induced forest defoliation –

limitation of MODIS data in Fennoscandian forest landscapes. Silva Fennica 50. Finnish Society of Forest Science

Ortiz, S., J. Breidenbach, and G. Kändler. 2013. Early Detection of Bark Beetle Green Attack Using TerraSAR-X and RapidEye Data. Remote Sensing 5: 1912-1931.

Oshiro, T. M., P. S. Perez, and J. A. Baranauskas. 2012. How many trees in a random forest? In International workshop on machine learning and data mining in pattern recognition. Springer, Berlin, Heidelberg: 154-168.

Persson, Y., R. Vasaitis, B. Långström, P. Öhrn, K. Ihrmark and J. Stenlid. 2009. Fungi vectored by the bark beetle *Ips typographus* following hibernation under the bark of standing trees and in the forest litter. Microbial ecology 58: 651-659.

Rouse Jr, J. W. 1972. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.

Saeed, S., A. Masood, A. Sajjad, and D.M. Zahid. 2010. Monitoring the dispersal potential of bark beetle, hypocryphalus mangiferae stebbing (scolytidae: coleoptera) in mango orchards. Pak. J. Zool. 42: 473–479.

Schlyter, F., Q.-H. Zhang, G.-T. Liu, and L.-Z. Ji. 2001. A successful case of pheromone mass trapping of the bark beetle Ips duplicatus in a forest island, analysed by 20-year time-series data. Integrated Pest Management Reviews 6: 185–196.

Senf, C., D. Pflugmacher, M. Wulder, and P. Hostert. 2015. Characterizing spectral–temporal patterns of defoliator and bark beetle disturbances using Landsat time series. Remote Sensing of Environment 170: 166-177.

Senf, C., R. Seidl, and P. Hostert. 2017. Remote sensing of forest insect disturbances: Current state and future directions. International Journal of Applied Earth Observation and Geoinformation 60: 49-60.

Skakun, R., M. Wulder, and S. Franklin. 2003. Sensitivity of the thematic mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. Remote Sensing of Environment 86: 433-443.

SMHI. 2019. Access at https://www.smhi.se.

Sonobe, R., Y. Yamaya, H. Tani, X. Wang, N. Kobayashi, and K. I. Mochizuki. 2018. Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. Journal of Applied Remote Sensing 12:026019.

Swedish forest agency. 2019. Access at www.skogsstyrelsen.se

Vincini, M., E. R. M. E. S. Frazzi, & P. A. O. L. O. D'Alessio. 2008. A broad-band leaf chlorophyll vegetation index at the canopy scale. Precision Agriculture, 9, 303-

319.

Volney, W. J. A., and R. A. Fleming. 2000. Climate change and impacts of boreal forest insects. Agriculture Ecosystems and Environment 82: 283–294

Weslien, J., E. Annila, A. Bakke, B. Bejer, H. H. Eidmann, K. Narvestad, A. Nikula, and H. P. Ravn. 1989. Estimating risks for spruce bark beetle (Ips typographus (L.) damage using pheromone - baited traps and trees. Scandinavian Journal of Forest Research 4: 87-98.

Wulder, M., J. White, B. Bentz, M. Alvarez, and N. Coops. 2006. Estimating the probability of mountain pine beetle red-attack damage. Remote Sensing of Environment 101: 150-166.

Öhrn, P. 2012. The spruce bark beetle Ips typographus in a changing climate – Effects of weather conditions on the biology of Ips typographus. Department of Ecology, Swedish University of Agricultural Sciencie.

Öhrn, P., B. Långström, B. Lindelöw, and N. Björklund. 2014. Seasonal flight patterns of Ips typographus in southern Sweden and thermal sums required for emergence. Agricultural and forest entomology 16: 147-157.

Økland, B., S. Netherer, and L. Marini. 2015. The Eurasian spruce bark beetle: The role of climate. Climate change and insect pests 202-219.

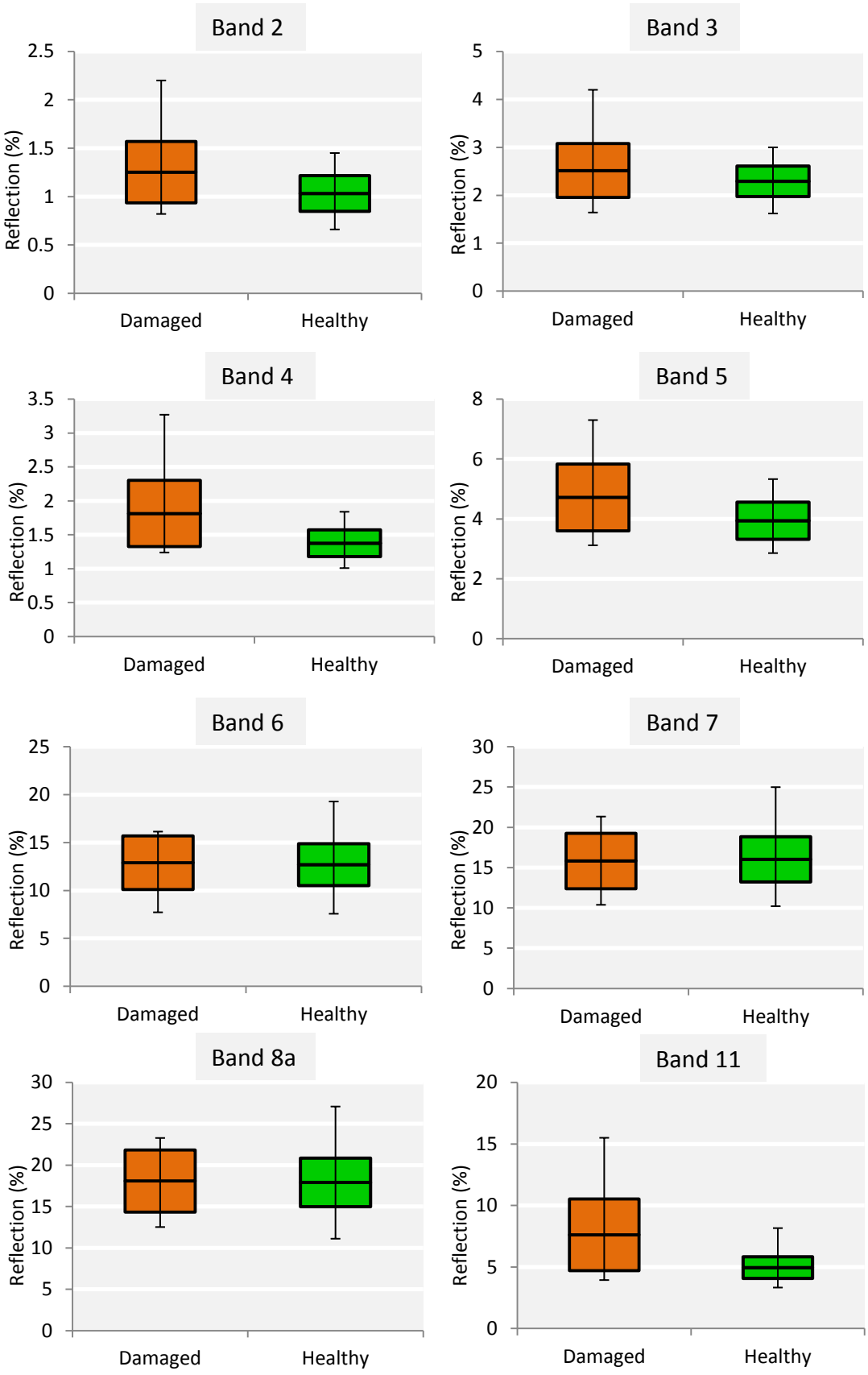
Appendices

Table S1. displays the raw data of result of the threshold method for variable reduction in the section *3.3.2 Bark beetle green-attack stage detection*, including the number of variables used for classification with random forest method and the corresponding achieved total accuracy, Kappa, cumulative importance and variable importance.

Table S1. The number of variables used in random forest classification and the achieved classification total accuracy, Kappa and the corresponding variable importance and cumulative importance resulted from threshold method for variable reduction. The cumulative importance is the add-up value of the corresponding variable importance value and all values above. The row in red color is the row with the selected threshold value.

Number of variables	Total accuracy	Kappa	Cumulative importance	Variable importance
1	0.857143	0.602549	13.53%	13.53%
2	0.836735	0.584746	24.48%	10.95%
3	0.816327	0.545829	34.32%	9.84%
4	0.836735	0.584746	40.22%	5.90%
5	0.836735	0.584746	45.88%	5.66%
6	0.836735	0.584746	51.29%	5.41%
7	0.857143	0.625954	55.47%	4.18%
8	0.877551	0.669663	59.29%	3.81%
9	0.857143	0.625954	63.10%	3.81%
10	0.877551	0.669663	66.54%	3.44%
11	0.857143	0.625954	69.62%	3.08%
12	0.857143	0.6259542	72.69%	3.08%
13	0.877551	0.669663	75.28%	2.58%

Figure S1, S2 and S3 are the box plots for pixels with likely bark beetle green-attacked spruces and healthy spruce pixels in study area 1 on October 12th, in test area 2 on July 8th, and in test area 2 on July 31st, respectively.


Band 2
Reflection (%)
2.5
2
1.5
1
0.5
0
Damaged
Healthy
Band 3
Reflection (%)
5
4
3
2
1
0
Damaged
Healthy
Band 4
Reflection (%)
3.5
3
2.5
2
1.5
1
0.5
0
Damaged
Healthy
Band 5
Reflection (%)
8
6
4
2
0
Damaged
Healthy
Band 6
Reflection (%)
25
20
15
10
5
0
Damaged
Healthy
Band 7
Reflection (%)
30
25
20
15
10
5
0
Damaged
Healthy
Band 8a
Reflection (%)
30
25
20
15
10
5
0
Damaged
Healthy
Band 11
Reflection (%)
20
15
10
5
0
Damaged
Healthy

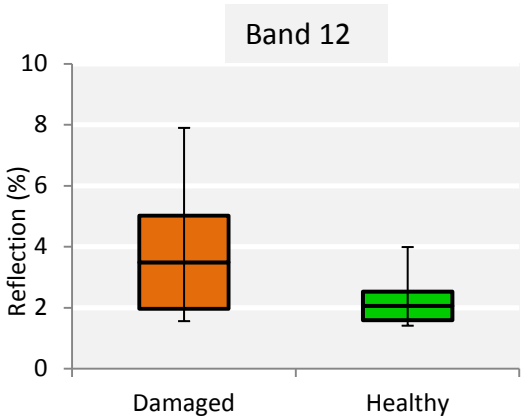
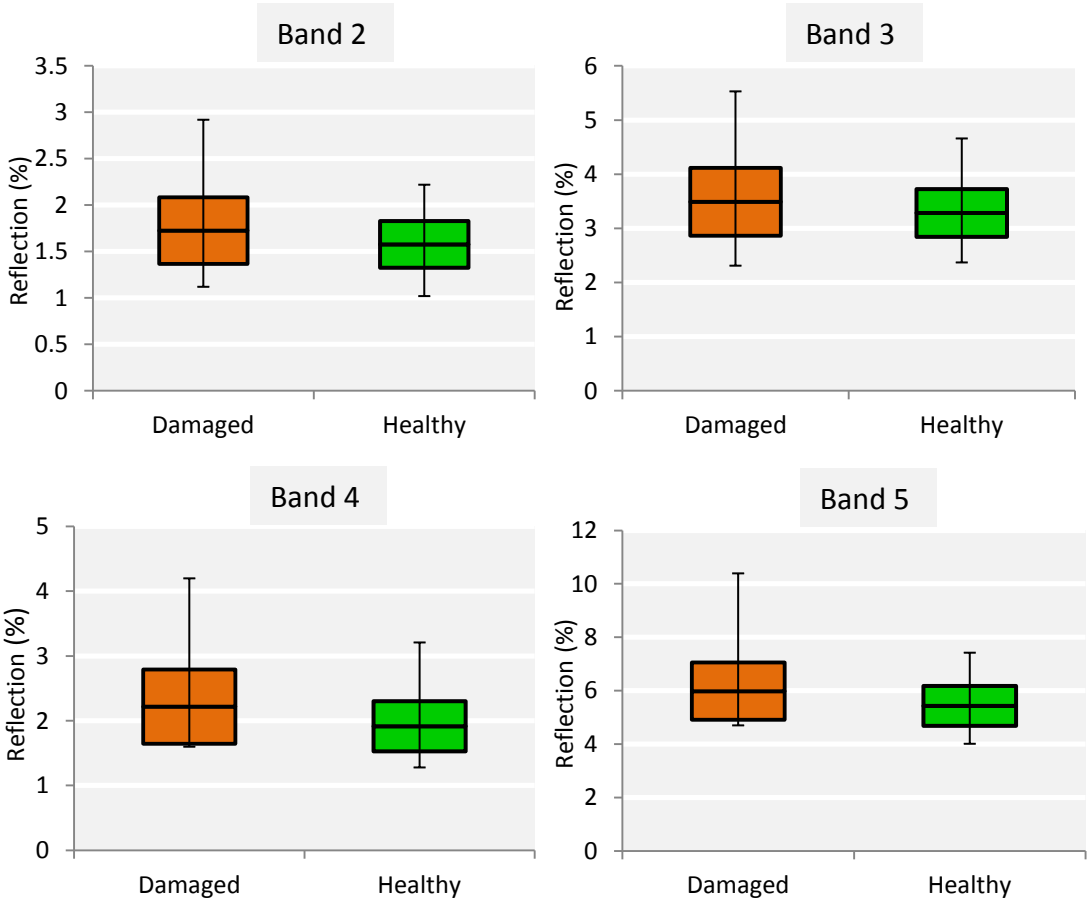


Figure S1. Box plots of reflectance of pixels with likely bark beetle green-attacked spruce trees and pixels with only healthy trees for for Sentienl-2 20m resolution bands in study area 1, on October 12th, 2018. The middle line stands for the mean reflectance value and the box stands for 1* standard deviation above and below the mean value. The error bar stands for the maximum and minimum reflectance values. Band 2 to 4: visible bands. Band 5-7: red-edge bands. Band 8a: NIR band. Band 11, 12: SWIR bands.



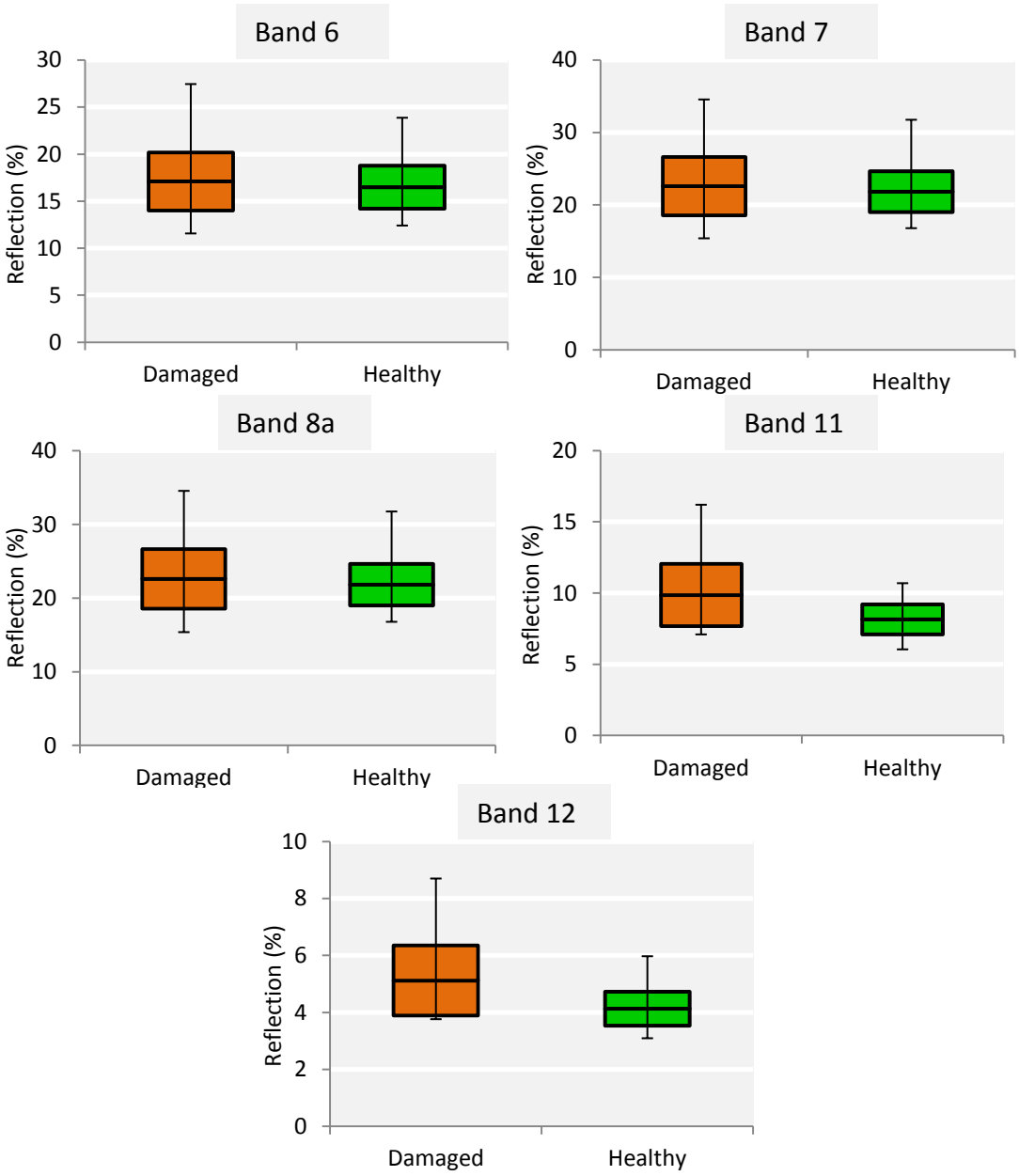
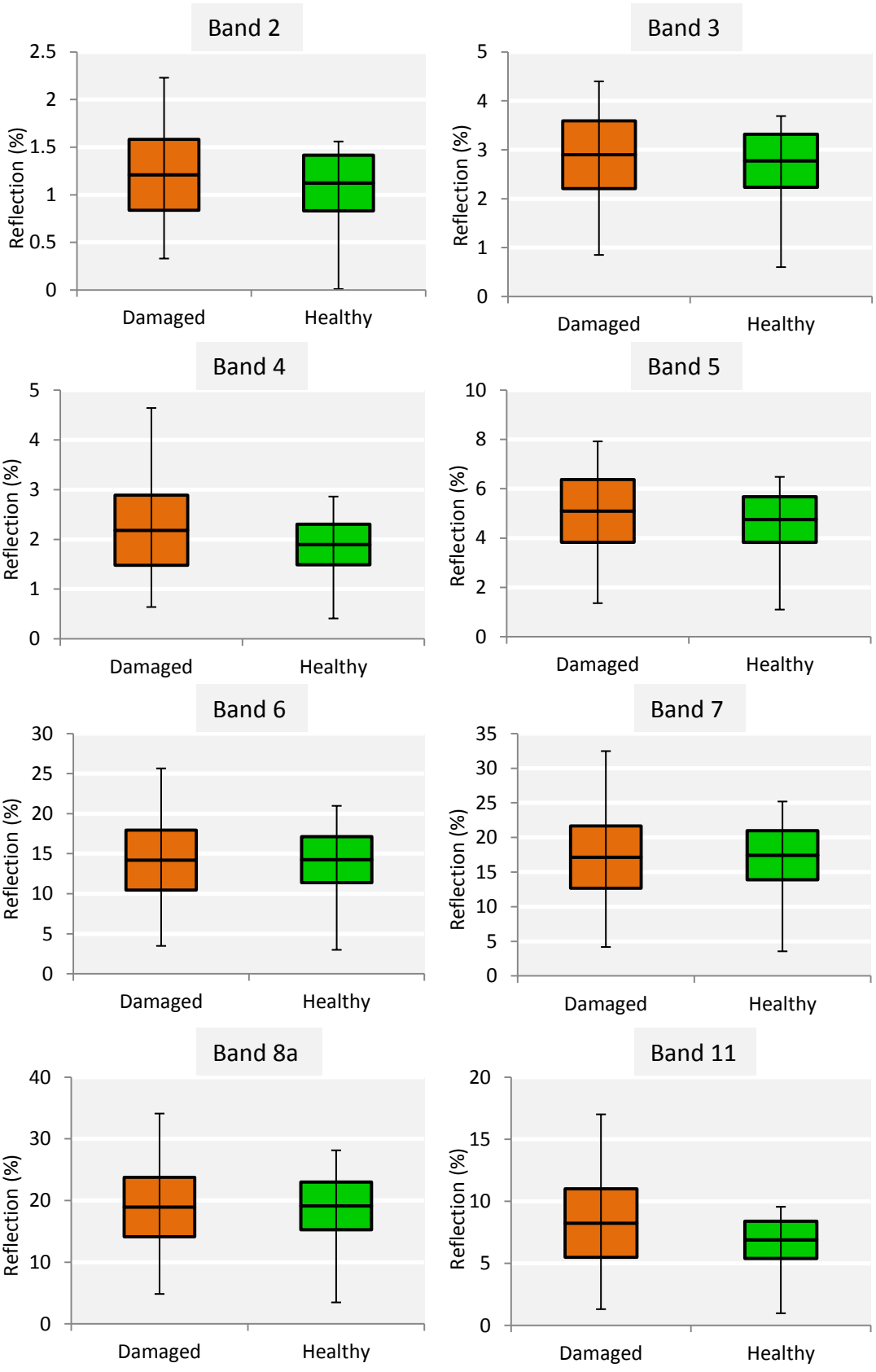


Figure S2. Box plots of reflectance of pixels with likely bark beetle green-attacked spruce trees and pixels with only healthy trees for Sentienl-2 20m resolution bands in test area 2, on July 8th, 2018. The middle line stands for the mean reflectance value and the box stands for 1* standard deviation above and below the mean value. The error bar stands for the maximum and minimum reflectance values. Band 2 to 4: visible bands. Band 5-7: red-edge bands. Band 8a: NIR band. Band 11, 12: SWIR bands.


Band 2
Reflection (%)
2.5
2
1.5
1
0.5
0
Damaged
Healthy
Band 3
Reflection (%)
5
4
3
2
1
0
Damaged
Healthy
Band 4
Reflection (%)
5
4
3
2
1
0
Damaged
Healthy
Band 5
Reflection (%)
10
8
6
4
2
0
Damaged
Healthy
Band 6
Reflection (%)
30
25
20
15
10
5
0
Damaged
Healthy
Band 7
Reflection (%)
35
30
25
20
15
10
5
0
Damaged
Healthy
Band 8a
Reflection (%)
40
30
20
10
0
Damaged
Healthy
Band 11
Reflection (%)
20
15
10
5
0
Damaged
Healthy

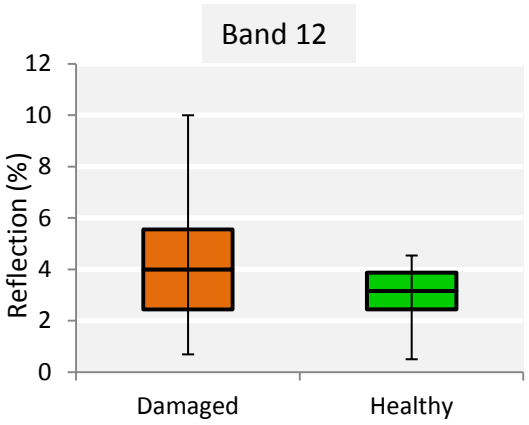


Figure S3. Box plots of reflectance of pixels with likely bark beetle green-attacked spruce trees and pixels with only healthy trees for Sentienl-2 20m resolution bands in test area 2, on July 31st, 2018. The middle line stands for the mean reflectance value and the box stands for 1* standard deviation above and below the mean value. The error bar stands for the maximum and minimum reflectance values. Band 2 to 4: visible bands. Band 5-7: red-edge bands. Band 8a: NIR band. Band 11, 12: SWIR bands.