Engineering High School of Riga Technical University

Forest Damage – Bark Beetle Identification Using Remote Sensing in Latvian Territories

Author: Lukass Roberts Kellijs**,** Secondary school student

Supervisor: Elza Līna Liniņa**,** *M.sc.phys.*, RTU EHS physics teacher

Consultant: Juris Siņica-Siņavskis**,** *Dr.sc.comp.*, IECS researcher

Contents:

Introduction

Global climate change and its impact on the environment are receiving more and more attention every year. Periods of drought and increased temperatures during the vegetation period lead to reduced forest vitality, stressed tree growth and increased susceptibility to damage by insects such as bark beetles [1]. A total of 154 different bark beetle species in Europe are known – each species having adapted to one or a few host tree species [2]. The European spruce bark beetle *(Ips typographus, L.*) is an invasive species considered to be the most critical disturbance agent in European forest ecosystems [1].

Bark beetle infestations in Germany have already destroyed thousands of hectares of spruce stands [2]. More than 50% of Czech forests are seriously threatened by this pest, causing major ecological and economic losses. The total harvested volume of timber per year in the Czech Republic is about 15 million m^3. Around 1 million m^3 of this amount is infested by insects, as well as in the last 5 years a rapid increase in insect-infested wood has been registered [3]. In Latvia, too, the pest's activity and the resulting decline in timber value have been detected, and the infestation of the European spruce bark beetle is being actively monitored [4].

Traditionally, to identify and control these insects, foresters carry out field surveys looking for the characteristic sawdust holes created by the bark beetle or otherwise set pheromone beetle traps [3]. An effective alternative to these labour-intensive methods is remote sensing - the analysis of the physical characteristics of an area by measuring its reflected and emitted radiation at a distance. The biochemical and biophysical properties of the tree needles change during the stages of the insect's attack, thus making it possible to identify the bark beetle using different spectral sensors and the images they produce [1]. An algorithmic model is trained to recognise from various multispectral images whether the tree - represented by a group of pixels of its canopy - is healthy or infested. In essence, it is a binary classification problem of pixels - the smallest image elements that characterise the intensity of the constructed raster graph in a given spectral band at that point. Machine learning algorithms are used to differentiate infested tree image pixels from healthy ones.

There are many potential sources of remote sensing data, such as the freely available Sentinel-2 or Landsat 8 satellite imagery data, whose ability to identify bark beetles has been repeatedly studied in European areas [5]. Unfortunately, this data is limited by its spatial resolution, meaning it is impossible to identify bark beetle infestations in individual trees. However, for various countries there often exist local remote sensing data with much higher spatial resolution. In Latvia such data, consisting of aerial photography orthophotos and LiDAR point cloud data, is provided by the Latvian Geospatial Information Agency. No studies have yet to be conducted on the use of this data to help identify bark beetles.

Aim of the study: To create and evaluate a model, which can help identify forest damage – the bark beetle red-attack stage in Latvian territories.

Research question: How and to what extent can the data of the Latvian Geospatial Information Agency help to identify forest damage - the bark beetle red-attack stage in Latvian territories?

Tasks:

1. To investigate stages of bark beetle attack, the characteristics of bark beetle-infested trees, and bark beetle-infested and damaged areas in Latvia.
2. To obtain, process, and evaluate open access data from the Latvian Geospatial Information Agency for use in bark beetle identification.
3. To research and train a fit-for-purpose algorithmic model using this data.
4. To evaluate the ability of the trained model to identify areas of forest damage – the bark beetle red-attack stage.

1. Literature review

The literature review discusses the characteristics of the European spruce bark beetle and explores its different attack stages, investigates and justifies the use of remote sensing for forest damage – bark beetle identification, as well as analyses the use of a random forest algorithmic model.

1.1 European spruce bark beetle

The migration of the European spruce bark beetle usually starts in mid-March, or mid-April, when air temperatures reach 16.5°C. The optimum flight temperature is 22–26°C. For a successful attack on live trees, the bark beetle needs at least 3-4 consecutive warm days above the threshold temperature. The active flight radius of bark beetles is greater than 500 m. Factors that increase the risk of tree infestation are being within 100 m of the infestation area, tree age greater than 70–100 years, windstorms, dry spells, high temperatures, and snow and ice damage [6].

The attack of the bark beetle (Figure 1.1) can be divided into three stages: green, red, and grey [1]. In the early green-attack stage, the foliage is still alive and visually green. In this stage, a newly hatched larval generation develops in the inner bark of infested trees, attacking the living phloem tissue of the trees, disrupting the transport of nutrients between roots and leaves, resulting in a progressive reduction of chlorophyll in the leaves and ultimately in its decay [2]. The infested tree remains green until mid-June and late July when the foliage turns red. At the red-attack stage, the bark beetles have already abandoned their host trees and have started to attack new trees [1]. Spruce trees suffering from bark beetle infestations, can be identified from August onwards. Only after about 3 years, at the grey-attack stage, the tree is dead and has completely lost its needles [2].

The death of an infested tree is caused not only by *I. typographus L.* feeding but also by blue fungi species associated with the beetle. Among different bark beetle species, *I. typographus L.* was found to transmit more pathogenic fungi than other species [6].

1.2 Remote-sensing for identification of forest damage – bark beetles

Bark beetle infestation leads to changes in leaf water, dry matter and nitrogen content [5]. The spectral characteristics and spectral indices best able to identify the stages of infestation are based on wavebands associated with chlorophyll absorption features in the spectral range of 450 to 890 nm. Only a limited number of useful spectral characteristics and indicators are found in the short-wave infrared region in the spectral range of 1400 to 1800 nm, which reflects the water content of spruce needles [2]. The choice of the specific attack stage to be identified is of great importance for the identification of bark beetles. Most studies deal with identifying the red-attack stage, but recent research indicates the potential to identify bark beetle infestation even in the green-attack stage [5].

The two most common sources of remote sensing data for bark beetle identification studies are Sentinel-2 and Landsat 8 data. Sentinel-2 is an Earth observation satellite system. It offers global coverage of the Earth's surface, high spectral resolution - 13 bands with wavelengths from 422 to 2202 nm - spatial resolution from 10 to 60 m, and temporal resolution of less than 5 days [7]. Similarly, Landsat 8 is a satellite for the analysis of the Earth's surface at different wavelengths, providing 9 bands at wavelengths from 433 to 1390 nm, with a spatial resolution of 15 to 30 m, with repeated flyovers every 16 days [8]. The main drawback of these data sources is their low spatial resolution, which makes it impossible to identify damage to individual trees.

Using these various data sources, the further objective of remote sensing is to develop appropriate techniques for data processing and damage identification. A commonly adopted method is to use spectral indices calculated from multiple spectral bands, such as using the change over time of the NDVI (*Normalised Difference Vegetation Index*) spectral index or selecting thresholds above which an image pixel is classified as infested by the index [5, 9]. Another approach is to use the data to make specific quantitative judgements about the biophysical and biochemical properties of the trees under study, such as canopy chlorophyll content (CCC) products, which are compared with the properties of already known healthy trees [1]. The use of different data mining and machine learning algorithms, such as decision trees or random forests, to identify bark beetles is also widespread [2].

1.3 Random forest machine learning algorithm

The random forest algorithm (Figure 1.2) is widely used in remote sensing to classify areas and pixels according to their properties into different categories, such as tree species [3]. It is one of many supervised classification algorithms, along with other popular methods such as deep learning and neural networks, that can learn features of the target class from established training samples and identify these learned features in an unclassified data sample.

It is derived from decision tree algorithms, which have been widely used in remote sensing since the 1990s [2]. It provides reliable classification using a set of decision trees in which a large number of decision tree classifiers are trained using a random subset of the training samples, with the final predictions resulting from the aggregate voting of these decision trees. This avoids overfitting the model to the training data. The main parameters of the model are the number of decision trees to be generated and the number of variables that are selected when the decision trees are trained [10]. In recent years, the random forest algorithm has been increasingly used in the remote sensing industry due to its accurate results, robustness against overfitting, and short runtime [10].

2. Study area and data used

2.1 Study area

The study area is approximately 100.4 ha of the Daugava Forestry area of SIA "Rīgas meži" 56°47'03"N (longitude) 24°29'56"E (latitude). Within the area, there exist field surveyed territories in which signs of bark beetle damage have been detected in the second half of May 2020 (Figure 2.1) [4]. The dominant tree species in the surveyed areas is Norway spruce *(Picea abies)* and Scots pine (*Pinus sylvestris)*. The dense concentration of spruce stands in the Daugava Forestry area promotes the spread of bark beetles [4].

2.2 Orthophoto map images

The data of the study area is obtained from the Latvian Geospatial Information Agency (LGIA). On average, every 3 years, LGIA produces a full orthophoto map of the entire territory of Latvia. The orthophotos used are obtained from the most recent 7th cycle LGIA aerial photography overflights in 2020, where red, green, blue (RGB) true colour orthophotos (Figure 2.1) and colour-infrared (CIR) orthophotos (Figure 2.2) were acquired with 0.25 m spatial resolution, providing pixel information for a total of 4 spectral bands. The high spatial resolution of the data allows the study to be carried out at the scale of individual trees.

2.3 LiDAR point cloud data

The LiDAR point cloud data is a set of classified aerial laser scanning points, where each point has a known plane coordinate (X,Y) and altitude, provided as open access data by LGIA (Figure 3.2). The points are automatically classified by level: land surface, low vegetation, high vegetation, buildings and so on. The data was acquired by the aerial laser scanning method. The total density of points obtained is at least 4 points/m² [11]. The LiDAR data for the study area are derived from overflights conducted in 2013.

3. Model preparation and training

The methodology used in the study consists of several parts (Figure 3.1). Initially, undesirable pixels in the study area are masked, such as shadows and forest clearings which can degrade the accuracy of the model by introducing noise. Then, healthy and bark beetle-infested canopy samples are manually labelled and categorised from the RGB images. This pixel information is summarised in a single table, together with the predefined category for each pixel. The data in the table is split into training and test data. The training data is used to train the random forest model, while the test data is used to evaluate its accuracy. Finally, a bark beetle infestation map for the study area is generated using the resulting model. The code used and developed in the study is available in a GitHub repository [14].

3.1 Masking unwanted pixels

Masking unwanted pixels involves creating two independent masks, namely a forest clearing mask and a shadow mask. Initially, a forest clearing mask is created using the LiDAR point cloud data (Figure 3.2). Using the set of classified points in LAS format, an intensity image is created from the low and high vegetation classes as a 5x5 pixel grid, where the pixel group value coincides with the relative height.

As the aim of the forest clearing masking is to mask whole forest clearings and not individual trees, the height intensity image is processed with three median 5x5 filters before a threshold is selected, below which pixels are classified as part of the forest clearing. This approach reduces noise from small groups of high-intensity pixels, creating a smooth forest clearing mask.

The creation of a shadow mask consists of selecting a single RGB orthophoto spectral band, from which a threshold value is determined. In this study, the threshold is determined via the blue image band. All pixels below this threshold are classified as shadows and hence masked (Figure 3.3). Combining these two masks results in an overall image mask (Figure 3.4), which is applied to the image to be analysed further in the study.

As can be seen from Figure 3.4, the forest clearing mask has not hidden the clearing areas in several places and has even masked coppices in some places, which can be explained by the underlying LiDAR point cloud data being fairly outdated.

3.2 Selection of pixels of interest

Pixels of interest, or so-called regions of interest, are selected manually (Annex 1). These data are used as the ground-truth data for the study, on which model training and validation are carried out. The regions of interest have two categories - healthy (Figure 3.5) and infested (Figure 3.6) - which are assigned to the regions at the time they are labelled.

The healthy regions are selected with as large a spread as possible and contain the canopy pixels of the healthy trees. The canopies of infested trees are labelled by the same principle, but with an emphasis on the neighbourhoods of the study area, where the bark beetle has been identified via field survey [4].

3.3 Creation of the model training and test datasets

These smaller regions of interest are then used to extract pixel information from a 4-dimensional matrix - a 4-band image containing the red, green, and blue bands of the LGIA RGB orthophoto, plus an additional near-infrared (denoted NIR) band of the CIR image. This matrix is prepared using the open-access software *QGis* [12]. The pixel information is tabulated accordingly (Table 3.1). The pixel intensities of the infested and healthy categories and their relative comparison can be visualised using box plots (Annex 2).

Consequently, 30% of this data is allocated as test data and 70% as training data. The training data is used to train the selected random forest classifier.

3.4 Model development, training, validation, and evaluation

The model - a structure generated by an algorithm that is trained to recognise certain types of relationships in order to make predictions about the problem under study - is created using the free machine learning *python* programming language library *scikit-learn* [13]. The model training parameters were set based on the standard accepted parameters: the number of decision trees to be generated was set as 500, and the number of variables to be selected when growing the trees was set as the square root of the total number of variables [10].

After model training, the test data are used to analyse the ability of the model to reliably identify forest damage – the bark beetle red-attack stage in Latvian territories using data from LGIA. A confusion matrix (Figures 4.1, 4.2) is created, the number of Type 1 and Type 2 errors is analysed and the accuracy of the model is calculated.

Finally, the resulting random forest model is used to classify all pixels in the study area (Figure 2.1). The result is a binary image, where the areas identified as damaged forest - bark beetle-infested are marked with the maximum pixel intensity. The resulting obtained image contains a large number of scattered individual pixels that have been identified as infested by the bark beetle. Knowing the spatial resolution of the images, i.e. 0.25 m, and taking into account that in the case of the red-attack stage of the bark beetle most of the foliage has already changed colour, it can be concluded that these pixels do not provide specific information about an infested tree and can be considered to be noise. To get rid of these individual pixels (as in shadow masking) a median filter is applied, only this time with a size of 3x3, and a threshold is chosen below which the noise pixels are deleted. This image is then combined with the initial RGB orthophoto and transformed into a mapped layer of forest damage – bark beetle red-attack stage, which can be analysed in different GIS systems.

4. Results and analysis

Using the trained model to classify the test data, an accuracy of 99% was obtained, i.e. out of 11874 total pixels, 128 pixels were misidentified. From the confusion matrix (Figure 4.1), it is possible to judge Type 1 and Type 2 errors - respectively pixels predicted to be unhealthy but actually healthy and pixels predicted to be healthy but actually unhealthy. From the given confusion matrix, the number of Type 1 errors is 66 and the number of Type 2 errors is 62, hence the two types of errors are approximately equal.

After the resulting model has been used to classify all pixels in the study area and after noise reduction using a median filter, the resulting total infestation map image more accurately demonstrates the model's ability to identify forest damage - bark beetles. Looking at this resulting image (Annex 3), it can be observed that it is generally successful in identifying bark beetle infestation sites (Figure 4.3).

Attention, however, should be paid to the suspiciously high estimated accuracy of the model obtained from the test data. Although the random forest model is robust to overfitting, other authors have found that the random forest algorithm was not suitable for transfer to other areas [10]. Due to the high estimated accuracy of the model, it would be necessary to validate the model using data from other study areas.

For this reason, following identical procedures, a validation dataset was prepared from another study area where similar bark beetle field surveys had been conducted and the pest identified. The accuracy of the model based on predictions on this novel validation area was once again also 99%, with Type 1 and Type 2 errors of 4 (Figure 4.2). Although, it should be mentioned that the validation data zone was also taken from the Daugava Forestry area of SIA "Rīgas meži" and the differences between the study and validation areas may not have been large enough to categorise them as different.

A major drawback of the current methodology is the incomplete forest clearing masking, which results in many objects in the study area (Annex 3), such as dirt roads or recently felled trees, being identified as infested with bark beetle. These error areas were not included in the model evaluation datasets i.e. the test and validation sets, and thus in the confusion matrix. This is because the main object of the study is the tree canopy - the information of the earth road and forest clearing pixels was not included in the manual data preparation when selecting the pixels of interest. As already mentioned, the incomplete masking of forest clearings can be explained by the outdated nature of the underlying LGIA LiDAR point cloud data - which were acquired in 2013. This presents the need for the renewal of this data. A suitable masking would not only lead to fewer errors in the model's identification of bark beetles but also to faster model performance by reducing the number of pixels to be analysed.

Another methodological problem is the selection of bark beetle-infested trees, i.e. the pixels of interest, when creating datasets. The current approach of manually selecting trees with red foliage from the orthophotos in the surveyed areas is not entirely accurate. This approach could also be one explanation for the high estimated accuracy of the model, as only comparatively easily identifiable trees with red foliage are manually selected, resulting in an easily distinguishable dataset, leading to high accuracy.

The biggest drawback in the use of LGIA data is its very low temporal resolution, i.e. it is only updated every 3 years. This study demonstrates its successful application for tasks such as bark beetle detection, so it is imperative to consider more frequent acquisition of this data for the entire territory of Latvia area or for specific study-related territories.

Overall, the high accuracy of the model and the serviceable generated infestation maps do indicate the suitability of the LGIA data for the identification of forest damage – bark beetle identification at the red-attack stage. The model is successfully able to visually identify bark beetle infested trees, both within and outside the surveyed areas (Figure 4.3).

As the field surveyed areas where bark beetle has been detected are only approximate zones as opposed to individual trees, specific field surveys, where red-attack stage bark beetle infestations are identified at the scale of individual trees, should be carried out to generate suitable training data to thoroughly validate the ability of such a model to identify real red-attack stage bark beetle infestations.

Such individual surveys could also enable new studies, where the ability of LGIA and other data to identify bark beetles at the scale of forest zones or individual trees could be investigated, even at the green-attack stage when the tree is visually green, but the reflectance in bands like the near-infrared has already started to change. As the negative effects of global climate change worsen, bark beetle infestations may become a much more serious problem with the need for more effective bark beetle identification techniques. Local orthophoto imagery like that of LGIA, together with the random forest model, provide one such potential technique.

Conclusions

1. The open-access data of LGIA can be used to successfully identify forest damage and the red-attack stage of bark beetle infestations in Latvia using a random forest machine learning algorithm.
2. To be able to use the LGIA data effectively for bark beetle identification, the temporal resolution of the data needs to be improved.
3. To successfully monitor and study bark beetle damage and the identification of bark beetle attack stages field surveys are needed to identify specific stages of bark beetle attack at the scale of individual trees.
4. Further research should be carried out on the suitability of the LGIA data to identify the green-attack stage of the European spruce bark beetle in Latvia.

List of references

[1] A. M. Ali *et al.* , "Canopy chlorophyll content retrieved from time series remote sensing data as a proxy for detecting bark beetle infestation," *Remote Sens. Appl. Soc. Environ.* , vol. 22, no. August 2020, p. 100524, 2021, doi: 10.1016/j.rsase.2021.100524.

[2] A. Lausch, M. Heurich, D. Gordalla, H. J. Dobner, S. Gwillym-Margianto, and C. Salbach, "Forecasting potential bark beetle outbreaks based on spruce forest vitality using hyperspectral remote-sensing techniques at different scales," *For. Ecol. Manage.* , vol. 308, pp. 76-89, 2013, doi: 10.1016/j.foreco.2013.07.043.

[3] A. Fernandez-Carrillo, Z. Patočka, L. Dobrovolný, A. Franco-Nieto, and B. Revilla-Romero, "Monitoring bark beetle forest damage in central europe. A remote sensing approach validated with field data," *Remote Sens.* , vol. 12, no. 21, pp. 1-19, 2020, doi: 10.3390/rs12213634.

[4] A. Seleznov, "Report on spruce stands in 2020 in areas managed by SIA ' Rīgas meži '," 2020.

[5] H. Abdullah, A. K. Skidmore, R. Darvishzadeh, and M. Heurich, "Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (Ips typographus, L.) compared with Landsat-8," *Remote Sens. Ecol. Conserv.* , vol. 5, no. 1, pp. 87-106, 2019, doi: 10.1002/rse2.93.

[6] B. Wermelinger, "Ecology and management of the spruce bark beetle Ips typographus - A review of recent research," *For. Ecol. Manage.* , vol. 202, no. 1-3, pp. 67-82, 2004, doi: 10.1016/j.foreco.2004.07.018.

[7] M. Drusch *et al.* , "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sens. Environ.* , vol. 120, pp. 25-36, 2012, doi: 10.1016/j.rse.2011.11.026.

[8] I. Yang and T. D. Acharya, "Exploring Landsat 8," *Int. J. IT, Eng. Appl. Sci. Res.* , vol. 4, no. 4, pp. 2319-4413, 2015, [Online]. Available: http://earthobservatory.nasa.gov/IOTD/.

[9] D. F. Gomez, H. M. W. Ritger, C. Pearce, J. Eickwort, and J. Hulcr, "Ability of remote sensing systems to detect bark beetle spots in the southeastern US," *Forests*, vol. 11, no. 11, pp. 1-10, 2020, doi: 10.3390/f11111167.

[10] M. Belgiu and L. Drăgu, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.* , vol. 114, pp. 24-31, 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[11] "Digital Height Model | Latvian Geospatial Information Agency." https://www.lgia.gov.lv/lv/digitalais-augstuma-modelis-0 (accessed Jan. 03, 2022).

[12] "Welcome to the QGIS project!" https://www.qgis.org/en/site/index.html (accessed Jan. 08, 2022).

[13] "sklearn.ensemble.RandomForestClassifier - scikit-learn 1.0.2 documentation." https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (accessed Jan. 08, 2022).

[14] L. Kelly (2022). Forest Damage – Bark Beetle Identification Using Remote Sensing in Latvian Territories [Computer software]. https://github.com/lukass16/Bark-Beetle-Detection-in-Latvia