APPLICATION OF RANDOM FOREST CLASSIFICATION TO DETECT THE PINE WILT DISEASE FROM HIGH RESOLUTION SPECTRAL IMAGES

Marian-Daniel Iordache\textsuperscript{1}, Vasco Mantas\textsuperscript{2}, Elsa Baltazar\textsuperscript{2}, Nicolas Lewyckyj\textsuperscript{1}, Niels Souverijns\textsuperscript{1}

\textsuperscript{1}Flemish Institute for Technological Research, Center for Remote Sensing and Earth Observation Processes (VITO-TAP), 2400 Mol, Belgium
\textsuperscript{2}Department of Earth Sciences, University of Coimbra, Portugal

ABSTRACT

Pine Wilt Disease is one of the forest pests with high destructive potential, due to its random spreading and the fast evolution of the symptoms. The correct identification of infected trees is critical for the containment of the pest in affected areas. This paper exploits the capabilities of Random Forest classification algorithms designed to spot the infected trees based on remote sensing images. We use as input both multi- and hyperspectral imagery with high spatial resolution, acquired via remotely piloted airborne systems in infected Portuguese forests. For both imagery types, the classification schemes achieve accuracies higher than 0.91. We conclude that Random Forest classification is a feasible method to detect the Pine Wilt Disease in spectral images acquired over wild forests, even at early stages of the infestation.

Index Terms — Pine Wilt Disease, classification, Random Forest, spectral imagery, RPAS

1. INTRODUCTION

Pine Wilt Disease (PWD) is a forest pest affecting large forested areas on several continents, which translates into a potential global threat on the integrity of coniferous forests. Its presence was confirmed in Asia (Korea, Japan, China, Taiwan), Central and North America from where it is native (Mexico, Canada, U.S.A.), and Europe (Portugal, Spain). The agent responsible for PWD is the nematode \textit{Bursaphelencus xylophilus}. The nematode has a high reproduction rate inside the infected trees and it feeds on the cells of resin canals, thus disturbing the flow of water and minerals through the conductive vessels \cite{1}. This, in turn, translates into a water stress leading to the degradation of the tree health (discoloration, defoliation) and finally to the death of the tree in a relatively short period of time: most of the trees die within one year from the moment of infestation.

One of the particularities of the disease is the erratic spreading pattern. The nematode is transported from an infected tree to a healthy tree by an insect vector. In Portugal, this vector is represented by the \textit{Monochamus galloprovincialis} insect (also known as the pine sawyer beetle). The spreading of the PWD is also facilitated by human (commercial) activities, e.g. transportation of infected wood, wood storage, and worldwide exchange of goods. Every time a new generation of insects emerges, the young beetles leave the infected tree and carry the nematode to other (healthy) trees. The complexity of the variables governing insect vector flight patterns and tree choice, complicate the prediction of disease propagation, resulting in variable distances between infected trees. The neighboring trees of an infected tree are not necessarily infected as the beetle can fly from tens to hundreds of meters (sometimes, kilometers) before touching another tree \cite{2}. This characteristic of the disease implies that spotting the infected trees entails more the analysis of individual trees than that of forest patches.

In practice, the presence of the nematode in pine trees is confirmed via laboratory analysis, using wood samples collected from each suspected tree \cite{3}. Obviously, this type of method cannot be applied to obtain a complete mapping of infected trees in a forest, due to costs, time and resources constraints, and, as explained above, the spreading pattern of the disease. Remote sensing comes, thus, naturally into play as a possible solution to the mapping of infected trees on relatively large areas.

2. PWD DETECTION VIA REMOTE SENSING

Consistent research efforts were devoted to PWD since its discovery. The available literature covers a plethora of topics related to the subject, such as: complete characterization of the nematode and of the insect vector, analysis of nematode-vector-tree interactions, analysis of spatial spreading of the disease, plant defense strategies, wilting mechanisms, management of the disease, among others (see \cite{1,4} and the references therein). However, only
scarce studies reporting disease monitoring via remote sensing are available.
In [5], a fixed, ground-based hyperspectral camera was used to detect PWD symptoms based on spectral characteristics of the observed pixels, and the reflectance at the 668 nm wavelength is found to be the most informative. Significant changes of reflectance in the Red and SWIR spectral regions are reported in [6]. In [7], the authors use multispectral data to assess the capability of four spectral indices in detecting stress in coniferous trees, after simulating a disease outbreak (not necessarily the PWD) by inoculating herbicide to a number of trees in a study area. They find that the red-edge is a strong indicator of the presence of stress in pine trees. In [8], the authors inoculated selected trees with the *Bursaphelencus xylophilus* nematode itself, extended the set of discriminative indices (11 used in total) and performed spectral measurements using a field spectrometer, in a temporal fashion. The most discriminative spectral indices are then highlighted. Another very interesting experiment is presented in [9], where the authors use spectral cameras over a larger pine forest area in Spain. Although the PWD is not confirmed in the test site, extensive ground observations are performed in order to characterize dominant (large) trees w.r.t. their canopy condition (i.e., defoliation, discoloration, canopy die-off). From the acquired airborne hyperspectral data, 49 spectral indices were computed and statistical analysis were performed to test the discriminative power of those indices. The authors do not provide a listing of the best indices as all of them show discriminative power between healthy and infected trees, but useful wavelengths are identified and general recommendations are made w.r.t. minimal data quality, flight costs, image processing requirements for potential operational campaigns.

Our work builds on top of the valuable previous works and tests the capabilities of Random Forest (RF) classification algorithms in the most realistic scenario: the data are acquired over a Portuguese forest where the presence of the PWD is confirmed, the training points are trees identified in the field via long term surveys, a relatively low number of informative indices is selected from a large pool of tested ones, and the result is a detailed map of infected trees that can be directly used by forest managers to keep under control the effects of the disease.

3. SPECTRAL IMAGERY AND TRAINING DATA

The data were acquired during five data acquisition campaigns (May and October 2018; June, September and October 2019) in five sites located in central Portugal, as illustrated in Fig. 1: three near the town of Sertã (Castelo Branco district) and two near the town of Condeixa-a-Nova (Coimbra district). The PWD is confirmed to affect these areas [10]. Remotely piloted aircraft systems (RPAS) were deployed in the field, carrying one multispectral (MS) camera: MicaSense Red-Edge with five spectral bands centered at 475 nm (Blue), 560 nm (Green), 668 nm (Red), 717 nm (Red-Edge) and 840 nm (Near-Infrared), and one hyperspectral (HS) camera: MicroHyperSpec A-Series VNIR with more than 300 spectral bands in the spectral range [380-1000] nm. All acquired data were pre-processed by VITO experts shortly after the flight campaigns. The spatial resolutions of the final datasets are 5 cm and 10 cm for the MS imagery and the HS imagery, respectively.

Training points were then identified via regular field monitoring campaigns. More than 120 trees were monitored, and each one of them was assigned to one of the three following classes: infected – trees in advanced state of degradation, with clear symptoms of the disease, confirmed to contain the nematode by laboratory analysis; suspicious – trees that show degradation symptoms at various stages, but not tested for nematode content in the laboratory; and healthy – trees that do not show any symptom of disease.

4. RANDOM FOREST CLASSIFICATION ALGORITHMS – TRAINING AND PERFORMANCES

The training points were identified in both multi- and hyperspectral imagery wherever possible and the corresponding spectra were extracted. One set of training points was thus compiled for each data type (multi- and hyperspectral data, respectively).

Large sets of suitable indices, partly found in the PWD literature and partly identified in spectral index databases, were computed: 24 for the MS data and 71 for the HS data.

---

1 see: https://support.micasense.com/hc/en-us
2 see: https://www.headwallphotonics.com/hyperspectral-sensors
3 see, e.g., https://www.indexdatabase.de/
A pre-selection of indices was performed in order to keep a reduced set of informative ones. For the MS, an iterative scheme was employed to remove highly correlated indices from the list. For the HS data, the selection is borrowed from the PROBA-V processing chain and it selects indices by ranking their discriminative power in a classification scenario. As a result, two different sets of 13 indices apiece were selected separately for the MS and the HS data.

A balanced selection of points was performed for the separate classes, thus avoiding the use of unbalanced training sets which are known to induce biased classification results. Random Forest (RF) classification schemes were developed separately for the MS and the HS data, respectively. Both schemes were developed in Python 3 programming language, using the scikit-learn machine learning library, with default parameters, except for the number of estimators (set to 500) and activated balanced class weights. From the total number of training points (2055 points per class in MS data; 2344 points per class in HS data), 70% were used for training and 30% for testing purposes.

Table 1 shows the user accuracies (UA), producer accuracies (PA) and overall accuracies (OA) for both multispectral and hyperspectral algorithms. Note that accuracies higher than 0.91 are achieved for both schemes. It can be observed that the accuracy of the HS RF classifier is slightly lower than that of the MS RF classifier, mostly due to a more pronounced confusion between the infected and suspicious classes. However, this confusion is largely tolerated as, in practice, once the tree is classified as suspicious, the forest managers have very limited margins on the actions to be taken. More specifically, due to economic reasons, the tree should be removed from the area in any circumstances once it shows declining symptoms, indistinctly if the symptoms are produced by the nematode or by other causes.

| Data type | UA  | PA  | OA  |
|-----------|-----|-----|-----|
| MS        | 0.95| 0.95| 0.95|
| HS        | 0.91| 0.92| 0.91|

### 5. OUTPUT CLASSIFICATION MAPS

When applied to individual images, the output of the RF classification algorithms is represented by classification maps, in which each pixel is assigned to one of the considered classes. Fig. 2 shows an example of a classified multispectral region in the area of Sertã (October 2018), jointly with an RGB image acquired over the same area. Note that, in the classification map, various post-processing strategies were applied in order to remove unnecessary image pixels (belonging to the background, not to tree crowns). These strategies include surface height thresholding and spatial aggregation based on windowed majority voting. However, as their specific details are out of the scope of this paper and also due to space restrictions, specific details of post-processing are neglected here.

In Fig. 1, the colors assigned to classified pixels have the following meaning: blue = infected pixel; red = suspicious pixel; yellow = healthy pixel. In the example scene, three affected trees and one dead tree are present. All of them are correctly identified by the classification algorithm. The inspection of RGB images acquired in June over the same area revealed that all symptomatic trees were dead by that time, which further confirms the correct functioning of the classification algorithm.

![Example of classification output obtained from MS imagery in October 2018, in the region of Sertã: a) RGB image; b) classification map overlaid onto the RGB representation extracted from MS imagery.](image)

### 6. EARLY DETECTION OF INFECTED TREES

A complete map of infected trees similar to the example in Fig. 2 at the end of the flight season allows the forest managers to deploy intervention teams in the field, in order to remove the affected trees before the next flight period of the insect vector. However, for an optimal planning, the early detection of the trees (i.e., as early as possible) can be of great importance. It is worthy to note that symptoms start developing in a top-down fashion: the superior (crown) branches show degradation before the lower branches. This impedes a ground observer to spot the infected trees even when the symptoms are already present on crown parts.

One of the major contributions of the work presented in this paper refers to this specific issue. In Portugal, the flight season of *Monochamus galloprovincialis* beetle generally starts in early spring and continues for several months. Taking into account that a certain delay exists between the moment of the infestation and the appearance of first symptoms, it results that the symptoms in June are still very scarce. In October, the trees infected along the flight season of the insect generally show clear signs of decline, except for the ones infected at late dates. The proposed classification schemes were able to successfully detect the symptoms in infected but apparently healthy trees, in the data of October 2019, as illustrated in Fig. 3, in both MS and HS data. Despite a larger uncertainty of these detections in comparison to more advanced stages of the disease, it is now confirmed that early detection is indeed possible, based on remote sensing imagery acquired from RPAS.
Fig. 3. Example of early detection: a) RGB image acquired in October 2018 - the tree does not show visible signs of decline; b) RGB image of the tree extracted from MS imagery; c) classification map (red = suspicious pixel; yellow = healthy pixel); d) RGB image acquired in June 2019 (the tree is dead)

7. POSSIBLE IMPROVEMENTS OF THE CLASSIFICATION MAPS

Various aspects can further be improved in the classification maps, mainly related to the background removal. For a more user-friendly appearance, better clearance of sparsely occurring pixels, e.g. isolated pixels classified as infected in healthy homogeneous areas, is needed. Changes in the standard pre-processing of data could be beneficial for a better removal of background pixels. For example, better alignment of the digital terrain and surface models used for data projection is desired, as height thresholding methods are dependent on the accuracy of this alignment. Testing with various methods of spatial post-processing is also needed in order to identify tree locations based on clustered points. More advanced improvements refer to the generation of probabilistic maps of the disease occurrence.

8. CONCLUSIONS

In this paper, RF classification algorithms were applied to spectral imagery acquired from RPAS, in order to derive classification maps of pine forest areas infected by the PWD. For both multispectral and hyperspectral data, the classification accuracies in test data were larger than 0.91. Thus, the feasibility of correctly mapping the diseased trees via the proposed method is proved. Moreover, it is shown that very early detection of diseased trees, before visible signs appear to a human observer, is also possible. The current limitations of the produced maps were identified and possible ways to overcome them were identified.

9. REFERENCES

[1] E. Sousa, F. Vale, and I. Abrantes, eds., “Pine Wilt Disease in Europe: Biological Interactions and Integrated Management”, First Edition, FNAF, FIG-Industrias Graficas SA, ISBN: 978-989-99365-2-2, 2015

[2] S.-R. Kim, E.-S. Kim, Y. Nam, W. I. Choi, and C.-M. Kim, “Distribution characteristics analysis of pine wilt disease using time series hyperspectral aerial imagery”, Korean Journal of Remote Sensing, vol. 31, iss. 5, pp. 385-394, 2015

[3] Y. Nakabayashi, T. Aikawa, M. Matsushita, and K. Hoshizaki, “Sampling Design for Efficient Detection of Pine Wood Nematode, Bursaphelenchus xylophilus, in Diseased Trees using a DNA Detection Kit: Variation Across Branch, Trunk and Tree”, Nematology, vol. 20, issue 7, pp. 641-652, 2018

[4] B.G. Zhao, K. Futai, J.R. Sutherland, and Y. Takeuchi, eds., Pine Wilt Disease, Springer, Japan, 2008

[5] J. Bin Lee, E. Sook Kim, and H. Seung Lee, “An analysis of spectral pattern for detecting pine wilt disease using ground-based hyperspectral camera”, Korean Journal of Remote Sensing, vol. 30, iss. 5, pp. 665-675, 2014

[6] S.-R. Lee, W.-K. Lee, K. Nam, Y. Song, H. Yu, M.-I. Kim, J.-Y. Lee, and S.-H. Lee, “Investigation into reflectance characteristics of trees infected by pine wilt disease”, Journal of Korean Society of Forest Science, vol. 102, issue 4, pp. 499-505, 2013

[7] J. P. Dash, M. S. Watt, G. D. Pearse, M. Heaphy, and H. S. Dungey, “Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak”, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 131, pp. 1-14, 2017

[8] S.-R. Kim, W.-K. Lee, C.-H. Lee, M. Kim, M. C. Kafatos, S.-H. Lee, and S.-S. Lee, “Hyperspectral Analysis of Pine Wilt Disease to Determine an Optimal Detection Index”, Forests, col. 9, issue 3, 115, 2018

[9] P. Beck, Pablo J. Zarco-Tejada, P. Strobl, and Jesus San-Miguel – Ayanz, “The feasibility of detecting trees affected by the pine wood nematode using remote sensing, EUR- Scientific and Technical Research Reports”, Publications Office of the European Union, ISBN: 978-92-79-48946-4, ISSN: 1831-9424

[10] Mota M.M., Futai K., and Vieira P., “Pine Wilt Disease And The Pinewood Nematode, Bursaphelenchus Xylophilus”, in: Ciancio A., Mukerji K. (eds) “Integrated Management of Fruit Crops Nematodes. Integrated Management of Plant Pests and Diseases”, vol. 4, Springer, Dordrecht, 2009