

# Monitoring vegetation structure on Natura 2000 areas using Remote Sensing

M. R. de Winter \*<sup>α</sup>, Joana F. M. F. Cardoso<sup>α</sup>, K. Hoogenboom,<sup>β</sup> O. Kavi<sup>α</sup>, and D. van Bentum<sup>γ</sup>

<sup>α</sup> Department Informatics and Automation, Province of South Holland

<sup>β</sup> Department Geo Information Systems (GIS), Province of South Holland

<sup>γ</sup> Department Water and Nature, Province of South Holland

## ABSTRACT

A machine learning model was made to detect different types of vegetation structures in satellite images combined with LIDAR (3D) data for measuring biodiversity purposes. Here we propose a machine learning model made with hand annotated satellite images instead of ground truth annotations i.e. annotations made in the field. Currently at Province South Holland every 8 years a field study is done to observe biodiversity, with satellite images every 3 months we want to increase this interval of measuring biodiversity. There is not yet consensuses about which variables to measure in remote sensing for biodiversity, with the currently lowest 50cm resolution of satellite images, we cannot measure small fauna and flora variables but can detect changes in vegetation structures. The current model accurately detects Gras, Forest, Thicket, Sand and Water regions, with sand being especially interesting. Even with 50cm resolution we run into big data challenges, with a relative small Natura 2000 area we already have 3 billion pixels. Here we want to share not only our methodology but also our (python) code and annotated training's data which hopefully will make this easier to implement for other organisations as well.

KEYWORDS: Remote sensing; Natura2000; Biodiversity; Object detection; Satellite images; Nitrogen crisis; Big data.

## 1 INTRODUCTION

Natural ecosystems are under pressure from human activity. Climate change, intensive agriculture, pollution, ocean acidification, invasive species and eutrophication are some of the problems land and water ecosystems face. At the European level, the European Union (EU) has been committed to the protection of nature and biodiversity by adopting several laws. The Habitats Directive (Council of the European Communities 92/43/EEC, 1992), adopted in 1992, aims to protect a wide range of rare, threatened or endemic animal and plant species and rare habitat types. Within this directive, a wide network of protected areas, the Natura 2000, has been established, covering 18% of the EU's land area and more than 8% of its marine territory. Threatened species and habitats listed under both the Birds Directive (Council of the European Communities 79/409/EEC, 1979) and the Habitats Directive are included in this network. EU member states are required to report every six years on the conservation status of habitats and species and on the actions taken on Natura 2000 sites.

monitoring cannot be done yearly. Besides, only a small area of the site is covered during such surveys. Ideally, changes in biodiversity in Natura 2000 sites should be more frequently monitored so that the effects of adopted measures are regularly followed.

Satellite imagery and Lidar (light detection and ranging) technology have been widely used to monitor changes in vegetation structure and biomass Brock et al. 2002, Xie et al. 2008, Zhao et al. 2012, Bolton et al. 2018, Mücher, Desk, et al. 2019, Reis et al. 2019, Koma et al. 2021, Moudrj et al. 2021, Zhang and Shao 2021, Ni-Meister et al. 2022, Taddeo 2022 among many others.

The use of satellite images to study vegetation structure is based on the fact that healthy vegetation absorbs more solar radiation in the red and blue light spectrum and reflects more near-infrared and green light. Satellite images capture wavelengths of reflected light with high accuracy and can, therefore, be used to determine vegetation's spatial distribution. Lidar, on the other hand, is an optical remote-sensing method that can be used to record vegetation height across large areas of the earth's surface. It works by emitting light from a laser and measuring the time for the reflected light to return to the sensor, which is then used to calculate elevation. Lidar technology produces highly accurate topographic positions (i.e., x, y, z coordinates) of objects and surfaces.

In the present paper, we combine high-resolution multi spectral satellite imagery with Lidar technology to monitor vegetation structure in two Natura 2000 sites in the Province of South-Holland. Our aim is to: 1) use satellite images and vegetation height data to predict vegetation structure in these sites, 2) follow temporal changes in vegetation structure to access quantitative changes in habitat structure.

The Netherlands has designated 162 Natura 2000 sites, 21 of these sites are located in the Province of South-Holland. The main threat for nature in these areas is nitrogen deposition as a result of agricultural activity, traffic and industry. Long-term exposure to large amounts of nitrogen leads to a decrease in biodiversity due to the decrease in sensitive (rare) species and an increase in generic species Marra et al. 2022. Within the Province of South-Holland, 12 Natura 2000 sites are, currently, sensitive to nitrogen deposition. Regular monitoring of the biodiversity in Natura 2000 sites is important to establish proper mitigation actions. Monitoring biodiversity in these sites is traditionally done by field surveys. These field surveys, done by experts, are important but time-consuming and

\*✉ vwdh@pzh.nl



Figure 1: Location of the study sites.

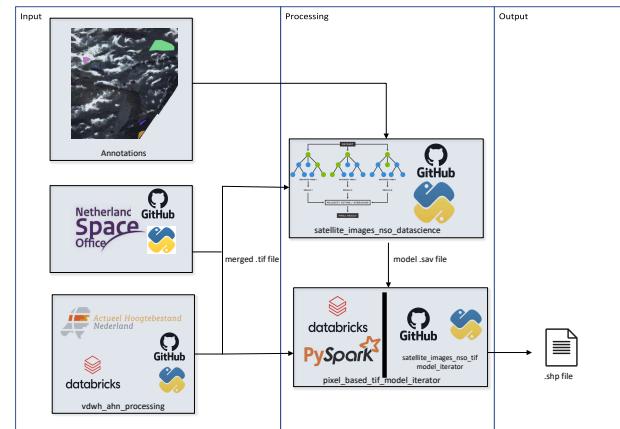


Figure 2: Remote Sensing component overview.

## 2 MATERIALS AND METHODS

### 2.1 Study areas

This study focused on two Natura 2000 sites in the Province of South-Holland, the Coepelduynen and the Voornes Duin. These are young calcareous coastal dunes characterized by a wide variety of landscape types. The Coepelduynen site is a narrow strip of dunes located between the municipalities of Katwijk and Noordwijk (Fig. 1). It is characterized by a fine-scale mosaic of open and closed grassland, forming an important habitat for flora and fauna. The calcareous dune grassland contains many rare plant species and the relatively open dunes are key breeding areas for several bird species. The Voornes Duin site is located on the island of Voorne-Putten, see Figure 1, and has a great variety of landscape types from dune grasslands and wet dune valleys to large areas of forest and shrub, saltmarshes and two large lakes. It is the most valuable area in terms of plant species in the Netherlands, with a high species diversity and a wide variety of biotic communities. It is also an important area for breeding bird colonies.

### 2.2 Data sources and pre-processing

This paper uses two types of remote sensing data, high-resolution multispectral satellite data and Lidar point cloud data, as well as image annotations. All analysis were performed using Python version 3.7 Van Rossum and Drake Jr 1995.

#### 2.2.1 Component Overview

An architectural overview of the different components used for model implementation is seen in Figure 2. The left panel shows the input data used in the model while the middle panel shows the used model and the multiprocessing/parallelism implementation. In the following sections each component is described in detail. (See section 5.1 for documentation on github)

#### 2.2.2 Multispectral data

SuperView satellite images from 2019 up to 2022 were downloaded for each of the study sites from the Netherlands Space

Office (NSO) website (<https://www.spaceoffice.nl/>). Available images were already pre-processed in terms of geometric and radiometric corrections (including orthorectification). Images included four spectral bands at 0,5m spatial resolution in the blue (B1: 450 - 520 nm), green (B2: 520 - 590 nm), red (B3: 630 - 690 nm) and near-infrared (NIR) (B4: 770 - 890 nm) spectra. Downloaded images were cropped to match each of the two study sites using the 'satellite\_images\_nso\_extractor' pipeline (see section 5.2). In this pipeline, the normalized difference vegetation index (NDVI) is calculated and added as extra band to the cropped image. The NDVI is a widely used remote sensing index for monitoring vegetation cover and biomass (Tucker, 1979; Pettorelli et al., 2005; Choubin et al., 2019; Dutta et al., 2021; Kumar et al., 2021). Healthy vegetation reflects more near-infrared light and absorbs more red light while unhealthy or sparse vegetation reflect more red light and less near-infrared light. The NDVI is, therefore, useful to distinguish between live green vegetation and non-vegetation such as water bodies, sand or rocks. In addition, higher NDVI values indicate a higher vegetation density such as dense forests. The NDVI was calculated, per pixel, based on the values of the NIR and red bands as:

$$NDVI = \frac{(NIR - red)}{(NIR + red)} \quad (1)$$

A set of cropped images from different years and seasons was selected for further analysis (Appendix Table 1). Only images covering the complete site and with a cloud-cover less than 10% were selected.

#### 2.2.3 Lidar data

Lidar data were used to calculate vegetation height. Lidar point cloud data were obtained from the Current Dutch Elevation (Actueel Hoogtebestand Nederland, AHN) website which contains detailed and precise altitude data for the Netherlands (<https://www.ahn.nl/>). We used the AHN4 dataset, acquired between 2020 and 2022 (for our study areas data was acquired in 2020) for satellite images between 2020 and 2022 and AHN3 dataset for satellite images taken in 2019. Lidar point density is between 10-14 point per m<sup>2</sup>. Lidar points are classified in two classes: 1) ground and 2) unclassified, which includes objects

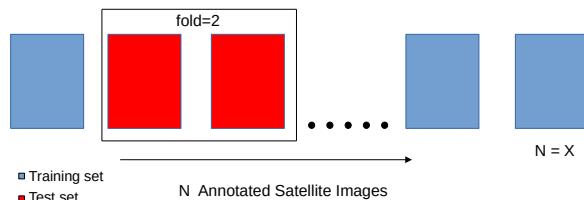


Figure 3: Cross validation with random annotated satellite image sampling as folds

such as tree, buildings and water. Processing of Lidar data was done in the 'vdwh\_ahn\_processing' pipeline (see section 5.3).

In this pipeline, points are aggregated per 0,5m and vegetation height is calculated as the difference between class 2 and 1. Missing pixels were filled based on the median value of the pixels around the missing pixel. Data was then cropped to match the study sites and height data per pixel was added as extra band to the cropped satellite image. All of the calculations were done in pyspark Fouilloux 2017 because of it's large compute time.

#### 2.2.4 Multiprocessing/Parallelism Model Iterator

Our model is a pixel based model, which means that we predict every pixel in a Satellite image with Lidar data. More in the discussion about why we picked pixel based.

Due to the large computational time of iterating over each pixel, multiprocessing and parallelism was used. Multiprocessing can be run locally which was used on Coepelduynen since it is a rather small area but pyspark parallelism was used for Voornesduinen because the area was to large for a acceptable processing time.

The 'pixel\_based\_tif\_model\_iterator' repository contains a stand alone pixel based model iterator on tif file, which can accept any model as long as it has a .predict function for running multiprocessing locally.

The pipeline also contains pyspark code which was used for parallelism for Voornesduinen.

The github python code for this can be found in Section 5.5

#### 2.2.5 Annotations of vegetation structures on Satellite Images

From the selected satellite images, a set of images from different years and seasons was selected and annotated. Image annotations were made together with ecologists well familiarized with the study areas and involved in the monitoring of these areas. For each area, classes were defined based on the landscape in that area. In each area, images were annotated by drawing polygons in QGIS 3.20 software and assigning a vegetation class to each polygon. Spectral data in the polygons was converted to pixels in order to be used in the model.

For a discussion why we use annotated images instead of ground truth labels as training data look in the discussion section.

We share these annotations freely look in Section 5.4.1

### 2.3 Model selection and training

Data per pixel were used in the model. For each pixel, the following data were available: red, green, blue, NIR, NDVI, height (independent variables) and vegetation label (dependent variable). Independent variables were normalized between 0 and 1 to correct for scale differences. Multi-collinearity between variables was checked using a correlation matrix to choose the optimal set of independent variables to use in the model.

Random Forest was used to classify the vegetation type in satellite images. Random forest is a robust machine-learning algorithm which is often used to classify satellite images (Pal

2005, Reis et al. 2018, Fritz et al. 2019, Rapinel et al. 2019). The random forest classifier generates multiple decision trees and combines their output to reach a single result (Breiman 1999). The classifier used for this study was built using the Scikit-Learn library (Pedregosa et al. 2011) of Python. A 70-30 split was used to partition the data into training and validation sets based on Satellite Images. Grid search was performed using the RandomizedSearchCV class from Scikit-Learn to find the optimal hyperparameter settings. The selected model was run using a leave-two-out cross-validation approach by randomly selecting 2 images as test set and the remaining 10 as training set (see Figure 3). The model was run using 4-fold cross-validation. Since some labels occur more often than others, training data was balanced by over-sampling the data set. The F1-score was used as performance metrics.

The model was further used to predict vegetation structure in a time series of satellite images for each site. Due to the large number of pixels and the long processing time, multiprocessing was used to generate model predictions. For the Coepelduynen, multiprocessing was done using the Multiprocessing module within Python. Due to the large area of the Voornes Duin, multiprocessing was done using PySpark (Fouilloux 2017) in Azure Databricks, which resulted in a faster output. See section 5.5 for both source codes.

The area occupied by each class was calculated and temporal changes in occupied area per class were analyzed.

## 3 RESULTS

### 3.1 Satellite and Lidar data

From the downloaded satellite images, a set of images was selected based on image quality and cloud-cover. Only images with less than 10% cloud cover were selected for further analysis. A list of the selected images is presented in Table 1 of the Appendix. Each prepared image has six bands, the original four spectral bands (red, green, blue and NIR), and the added NDVI and height. Multi-collinearity analysis revealed a high correlation between red, green and blue (Pearson correlation coefficient, Coepelduynen:  $x$ , Voornes Duin:  $p>0.95$ ), and between NIR and NDVI (Coepelduynen:  $x$ , Voornes Duin:  $p=0.8$ ). The independent variables blue, NDVI and height were selected for the model since they showed higher correlations with the dependent variable.

### 3.2 Annotations

In total, 9 satellite images were annotated for the Voornes dunes and 12 for the Coepelduynen. (see Appendix table 1). For the Coepelduynen, the following classes were defined: forest, grass, shadow, sand, wet dune valley, shrub and asphalt. For the Voornes Duin, the defined classes were: water, forest, grass, wet dune valley, shrub and sand.

Average values for the variables blue, NDVI and height for each class and area are shown in Table 1 based on the annotated extracted pixels from the satellite images with Lidar data.

Table 1 - Mean values ( $\pm$  SE) of the blue spectrum, NDVI and height for the different classes in the Coepelduynen and Voornes Duin.

Class	Blue	NDVI	Height
<i>Coepelduynen</i>			
Asphalt	435.909	99.233	6.454
Forest	220.604	128.121	143.794
Grass	325.834	125.809	2.027
Sand	756.046	111.723	2.043
Shadow	320.109	82.992	22.621
Shrub	253.889	125.690	61.502
Wet dune	313.373	121.767	0
<i>Voornes Duin</i>			
Forest	208.618	126.418	202.173
Grass	264.988	137.096	1.618
Shrub	223.799	118.844	98.191
Wet dune	266.383	127.965	1.881
Water	192.958	67.546	0.062
Sand	628.650	109.776	2.369

### 3.3 Vegetation structure classification

Grid search results were analysed to choose the set of parameters to use in the Random Forest model. Since the difference in the F1-score between 100 trees (n\_estimators) and 10 trees was less than 5% we chose to use the lowest amount of trees in favour of the prediction speed. For the other parameters we followed the results of the grid search: the minimum number of samples required to split an internal node (min\_samples\_split) was set to 10 and for the remainder parameters the default value was used.

Cross validation was used with folds being based on randomly picked satellite images which served as a test sets. Cross validation is the practice of rotating dividing the data into training and test sets based on the number of folds and then evaluating a statistical model Berrar 2019.

We randomly pick satellite images because of data spillage i.e. pixel data from a specific satellite image appearing in the training data as well as in the test set. This is done because the model seem to have better performance when if it already trained on pixels from a specific satellite image to also predict other pixels from this image see more in discussion and Figure 3.

The performance metrics here are averaged across all folds, look for all the results of all the folds in the appendix.

### 3.3.1 Coepelduynen results

For the Coepelduynen area, prediction accuracy (F1-score) was the highest for the classes grass, forest, Wet dunes and sand (F1-score > 0.8; Table 2). Shadow and asphalt showed a low accuracy (F1-score < 0.5) while shrub showed an intermediate accuracy (F1-score = 0.58).

Table 2 - Mean values of performance metrics for the Coepelduynen.

Class	Precision	Recall	F1-score	Support
Asphalt	0.302	0.597	0.395	8254
Forest	0.919	0.828	0.869	45734
Grass	0.984	0.900	0.939	180642
Shadow	0.061	0.042	0.046	1432
Shrub	0.510	0.692	0.578	15469
Wet dune	1	1	1	340
Sand	0.914	0.955	0.932	36280

### 3.4 Voornesduinen results

For the Voornes Duin area, the classes forest, shrub, water and sand showed a high accuracy (F1-score > 0.8) while the class wet dune valley showed the lowest (F1-score = 0.24; Table 3). Grass showed an intermediate accuracy (F1-score = 0.64).

Table 3 - Mean values of performance metrics for the Voornes Duin.

Class	Precision	Recall	F1-score	Support
Forest	0.932	0.833	0.880	215860
Grass	0.673	0.629	0.644	46884
Shrub	0.793	0.868	0.828	159752
Wet dune	0.188	0.332	0.239	14074
Water	0.949	0.956	0.952	31082
Sand	0.970	0.858	0.901	8433

### 3.5 Temporal changes in vegetation structure

Monitoring vegetation structure depends on monitoring changes over time. Ideally we want to see that changes in vegetation structures and thus (hopefully) biodiversity stay the same across time.

Here we make plots about the percentage of each vegetation structure that occupies a natura 2000 region. Note that we make percentages and not absolute numbers because sadly not all the satellite images always 100% contain the same amount of space. For instance sometimes the satellite misses a north tip of the Coepelduynen in it's fly over thus a more robust way of measuring changes over time would be to do it in percentages.

#### 3.5.1 Temporal changes in Coepelduynen vegetation structure

Figure 4 represents the developments over time for the different vegetation structures for the Coepelduynen

#### 3.5.2 Time Series Voornesduinen Vegetation structures

Figure 5 represents the developments over time for the different vegetation structures.

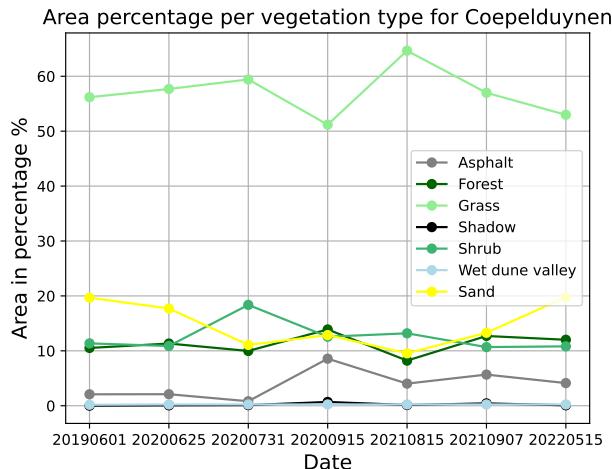


Figure 4: Temporal Developments Coepelduynen

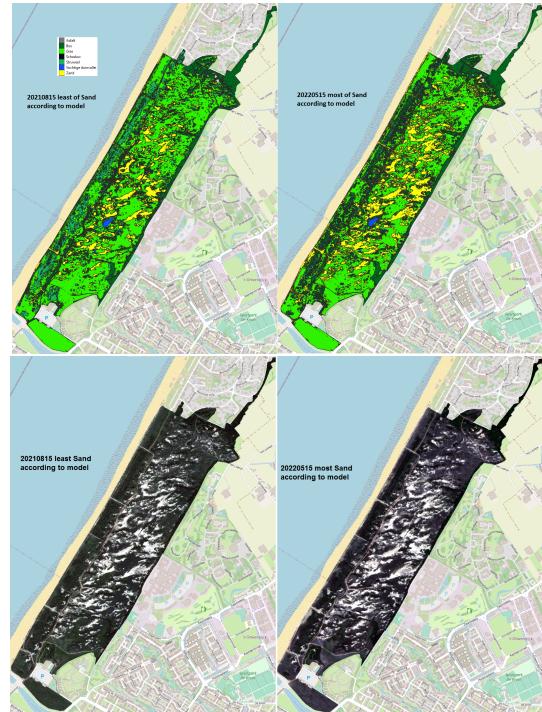


Figure 6: Comparison of two model annotated and original satellite images

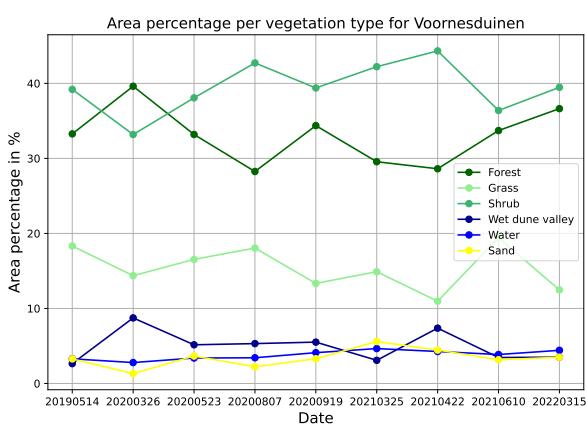


Figure 5: Temporal developments Voornesduinen

### 3.5.3 Examples of model annotations

Figure 6 highlights where the model saw big differences in Sand.

315

Here we see that indeed the most correctly saw big differences in Sand between these 2 satellite images. In the left images we see the lowest amount of Sand for each satellite image while the right had the most.

320

Figure 7 displays examples annotations of the Voornesduinen here we see a slight cut off from the top of the image. As stated earlier it is hard to find a satellite image which fully contains Voornes duinen, it seems to be a too big area for Superview.

325

## 4 DISCUSSION

### 4.1 Annotated satellite images as traning data instead of ground truth labels

330

Ground truth labels are currently the standard in remote sensing machine learning Satyanarayana et al. 2011. We propose several reasons why handmade annotating a satellite images is better a approach than using ground truth annotations i.e. annotations that were made in the field by a ecologist.

335

- **Discrepancy in time:** There is a gap in time between when the annotations were made on the ground and when the satellite image was shot. None of the ground truth annotations that we have were made right when a satellite flown over. This can lead to noise and bias in the data.

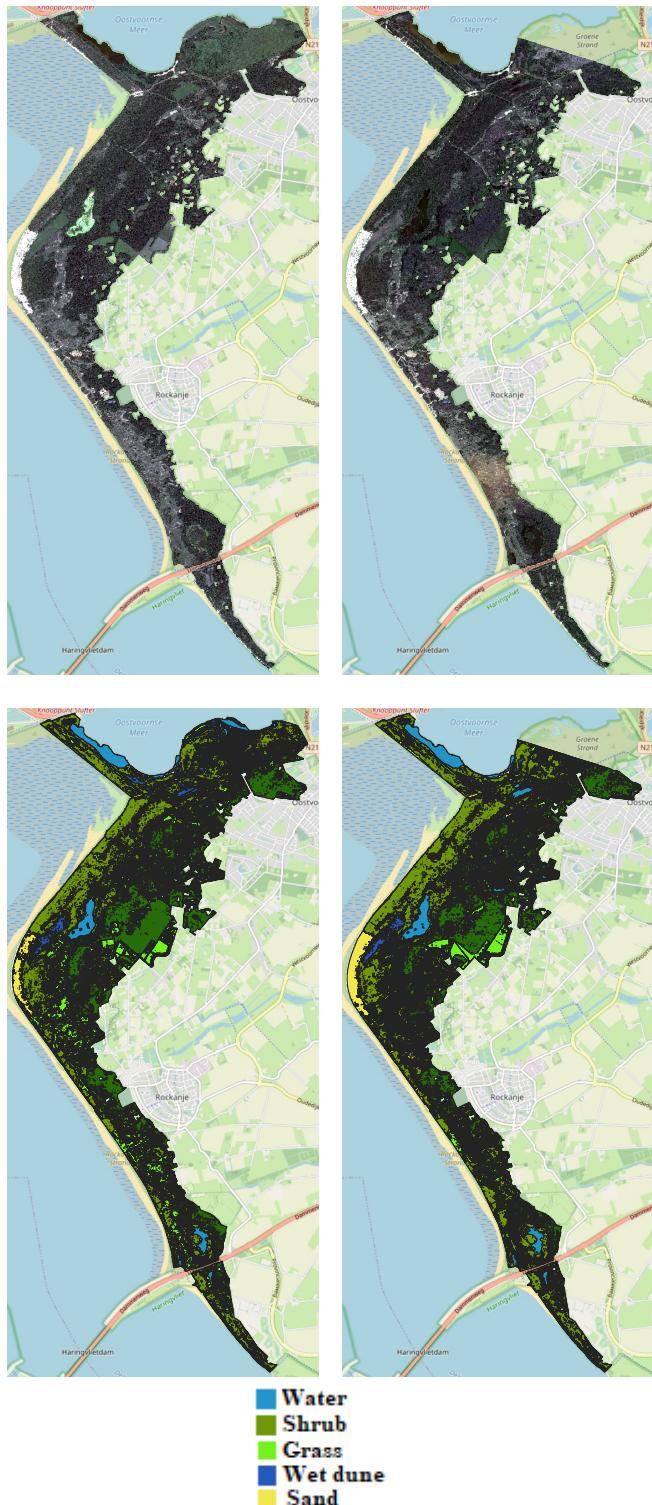


Figure 7: Comparision between Voornesduinen 20200807 and 20210422 model and satellite image

- **Less (temporal) training data :** Since an ecologist needs to cover a lot of ground making these ground truth annotations it is hard to produce a lot of training data across different time periods. Annotating satellite images is less time consuming which in turn will give us more training data for the model. Due to its ease of making data it can help to make more data across time. Which in turn can help the model deal better with unique atmospheric influences which randomly occur with each satellite image.
- **Orthorectification discrepancy with ground truth annotations:** Orthorectification is the process of converting images into a form suitable for maps by removing sensor, satellite/aircraft motion and terrain-related geometric distortions from raw imagery Leprince et al. 2007. This has been done with the Superview satellites but we still see some discrepancy between coordinates between satellites it seems that the orthorectification still left some noise. If we use ground truth annotations for this it would lead to bias and noise in the training data since the ground truth labels are not properly aligned with the satellite images. Thus it would be a better idea to annotated the satellite images themselves since we would then have no aligned problems.

## 4.2 Pixel based vs Object based classification

Doing prediction with remote sensing can be done with pixel based vs object based Zerrouki and Bouchaffra 2014. Currently we chose pixel based because of its ease of integration with traditional machine learning algorithms and our speculation that its performance time would be lower. In further work can hopefully compare pixel based vs object based performance

### 4.2.1 Data spillage in cross validation

For the folds in cross validation we at random pick different satellite images to serve as a test set instead of random sampling.

We do this because random sampling almost always gives us sampled data from all satellite images, thus giving us data spillage i.e. pixel data from one satellite image appearing in the test and training set. In our earlier experiments we saw that when a model has trained on annotated sampled training data from a satellite image it had better performance when predicting other pixels of the same satellite image in the test set.

Our reasoning is that this probably has to do with each unique RGBI range values a specific satellite images has due to different influences such as position of the sun Sekrecka et al. 2020 , atmospheric influences Huete and Jackson 1988 Tyagi and Bhosle 2011 amongst others. If the model has already been trained on sampled training data from this image, it probably already figured out its unique RGBI value range. But in a production environment the model can not first pre-trained on satellite image thus we base our training and test sets on including/excluding images in fold rather than random sampling which gives a more realistic approach.

Which why it was also important for annotated as much satellite images as possible instead of annotating much in one satellite image.

### 4.3 Further work

We did not thoroughly test deep learning parameters due to time constraints and custom code required to make deep learning run paralleled and distributed. Deep learning has been successfully implemented in other sectors in agriculture Kamilaris and Prenafeta-Boldú 2018 and thus could lead to better performance for us as well.

A pixel based model was used but a object based model could lead to better performance Hussain et al. 2013. Due to it's integration with tradition machine learning algorithms and data parallelism, we choose pixel based instead.

## 5 SOURCE CODE

Code and data examples associated with this study are open-source and can freely be obtained from GitHub.

### 5.1 Components overview

For a overview and links of all the components, check this repository: ([https://github.com/Provincie-Zuid-Holland/remote\\_sensing\\_natura2000](https://github.com/Provincie-Zuid-Holland/remote_sensing_natura2000))

### 5.2 NSO Satellite images extractor

For code related to the download and processing of satellite images go to here: ([https://github.com/Provincie-Zuid-Holland/satellite\\_images\\_nso\\_extractor](https://github.com/Provincie-Zuid-Holland/satellite_images_nso_extractor))

### 5.3 Lidar processing

For code related to the processing of Lidar data go to here: ([https://github.com/Provincie-Zuid-Holland/vdwh\\_ahn\\_processing](https://github.com/Provincie-Zuid-Holland/vdwh_ahn_processing))

### 5.4 Model and Annotations

Code for preparing image annotation data and modelling is available in here: (<https://github.com/Provincie-Zuid-Holland/satellite-images-nso-datasience>)

#### 5.4.1 Annotations

With the annotations itself being here: (<https://github.com/Provincie-Zuid-Holland/satellite-images-nso-datasience/tree/main/data/annotations>)

#### 5.4.2 Random forest model

With the random forest model being trained and cross validated here: ([https://github.com/Provincie-Zuid-Holland/satellite-images-nso-datasience/blob/main/annotations\\_models/Coepelduynen/random\\_forest/make\\_train\\_model\\_on\\_annotations\\_coepelduynen.ipynb](https://github.com/Provincie-Zuid-Holland/satellite-images-nso-datasience/blob/main/annotations_models/Coepelduynen/random_forest/make_train_model_on_annotations_coepelduynen.ipynb))

### 5.5 Pixel Based Model Iterator

Code for running a pixel based model in a multiprocessing distributed way can be found here: ([https://github.com/Provincie-Zuid-Holland/satellite\\_images\\_nso\\_tif\\_model\\_iterator](https://github.com/Provincie-Zuid-Holland/satellite_images_nso_tif_model_iterator))

## REFERENCES

- Huete, A. and R. Jackson (1988). "Soil and atmosphere influences on the spectra of partial canopies". *Remote Sensing of Environment* 25(1), pages 89–105. 445
- Van Rossum, G. and F. L. Drake Jr (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.
- Breiman, L. (1999). "Prediction games and arcing algorithms". *Neural computation* 11(7), pages 1493–1517. 450
- Brock, J. C., C. W. Wright, A. H. Sallenger, W. B. Krabill, and R. N. Swift (2002). "Basis and methods of NASA airborne topographic mapper lidar surveys for coastal studies". *Journal of Coastal Research*, pages 1–13.
- Pal, M. (2005). "Random forest classifier for remote sensing classification". *International journal of remote sensing* 26(1), pages 217–222. 455
- Leprince, S., S. Barbot, F. Ayoub, and J.-P. Avouac (2007). "Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements". *IEEE Transactions on Geoscience and Remote Sensing* 45(6), pages 1529–1558. 460
- Xie, Y., Z. Sha, and M. Yu (2008). "Remote sensing imagery in vegetation mapping: a review". *Journal of plant ecology* 1(1), pages 9–23. 465
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. (2011). "Scikit-learn: Machine learning in Python". *the Journal of machine Learning research* 12, pages 2825–2830. 470
- Satyanarayana, B., K. A. Mohamad, I. F. Idris, M.-L. Husain, and F. Dahdouh-Guebas (2011). "Assessment of mangrove vegetation based on remote sensing and ground-truth measurements at Tumpat, Kelantan Delta, East Coast of Peninsular Malaysia". *International Journal of Remote Sensing* 32(6), pages 1635–1650. 475
- Tyagi, P. and U. Bhosle (2011). "Atmospheric correction of remotely sensed images in spatial and transform domain". *International Journal of Image Processing* 5(5), pages 564–579. 480
- Zhao, X., D. Zhou, and J. Fang (2012). "Satellite-based Studies on Large-Scale Vegetation Changes in China F". *Journal of integrative plant biology* 54(10), pages 713–728. 485
- Hussain, M., D. Chen, A. Cheng, H. Wei, and D. Stanley (2013). "Change detection from remotely sensed images: From pixel-based to object-based approaches". *ISPRS Journal of photogrammetry and remote sensing* 80, pages 91–106.
- Zerrouki, N. and D. Bouchaffra (2014). "Pixel-based or object-based: Which approach is more appropriate for remote sensing image classification?" In: *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, pages 864–869. 490
- Fouilloux, A. (2017). *Introduction to PySpark Version 2017.01*.
- Bolton, D. K., N. C. Coops, T. Hermosilla, M. A. Wulder, and J. C. White (2018). "Evidence of vegetation greening at alpine treeline ecotones: Three decades of Landsat spectral

- 500 trends informed by lidar-derived vertical structure". *Environmental Research Letters* 13(8), page 084022.
- Kamilaris, A. and F. X. Prenafeta-Boldú (2018). "Deep learning in agriculture: A survey". *Computers and electronics in agriculture* 147, pages 70–90.
- 505 Reis, J., M. Amorim, N. Melão, and P. Matos (2018). "Digital transformation: a literature review and guidelines for future research". *Trends and Advances in Information Systems and Technologies: Volume 1 6*, pages 411–421.
- Berrar, D. (2019). *Cross-Validation*.
- 510 Fritz, S., L. See, J. C. L. Bayas, F. Waldner, D. Jacques, I. Becker-Reshef, A. Whitcraft, B. Baruth, R. Bonifacio, J. Crutchfield, et al. (2019). "A comparison of global agricultural monitoring systems and current gaps". *Agricultural systems* 168, pages 258–272.
- Mücher, C., S. Desk, et al. (2019). "Exploiting low-cost and 515 commonly shared aerial photographs and LiDAR data for detailed vegetation structure mapping of the Wadden Sea island of Ameland". *Journal of Earth Sciences & Environmental Studies* 4(1).
- Rapinel, S., C. Mong, L. Lecoq, B. Clément, A. Thomas, and L. Hubert-Moy (2019). "Evaluation of Sentinel-2 time-series for 520 mapping floodplain grassland plant communities". *Remote sensing of environment* 223, pages 115–129.
- Reis, B. P., S. V. Martins, E. I. Fernandes Filho, T. S. Sarcinelli, 525 J. M. Gleriani, H. G. Leite, and M. Halassy (2019). "Forest restoration monitoring through digital processing of high resolution images". *Ecological Engineering* 127, pages 178–186.
- Sekrecka, A., D. Wierzbicki, and M. Kedzierski (2020). "Influence of the sun position and platform orientation on the 530 quality of imagery obtained from unmanned aerial vehicles". *Remote Sensing* 12(6), page 1040.
- Koma, Z., A. Zlinszky, L. Bekó, P. Burai, A. C. Seijmonsbergen, and W. D. Kissling (2021). "Quantifying 3D vegetation structure in wetlands using differently measured airborne laser scanning data". *Ecological Indicators* 127, page 107752.
- 535 Moudrý, V., L. Moudrá, V. Barták, V. Bejček, K. Gdulová, M. Hendrychová, D. Moravec, P. Musil, D. Rocchini, K. Šťastný, et al. (2021). "The role of the vegetation structure, primary productivity and senescence derived from airborne LiDAR and hyperspectral data for birds diversity and rarity on a restored site". *Landscape and Urban Planning* 210, page 104064.
- Zhang, Y. and Z. Shao (2021). "Assessing of urban vegetation 540 biomass in combination with LiDAR and high-resolution remote sensing images". *International Journal of Remote Sensing* 42(3), pages 964–985.
- Marra, W., S. Hazelhorst, K. Brandt, R. Wichink Kruit, and J. Schram (2022). "Monitor stikstofdepositie in Natura 2000-gebieden 2022. Uitgangssituatie voor de Wet Stikstofreductie en Natuurverbetering".
- Ni-Meister, W., A. Rojas, and S. Lee (2022). "Direct use 545 of large-footprint lidar waveforms to estimate above-ground biomass". *Remote Sensing of Environment* 280, page 113147.
- Taddeo, S. (2022). "Leveraging time series of satellite and 550 aerial images to promote the long-term monitoring of re-stored plant communities". *Applied Vegetation Science* 25(2), e12664.