

# Monitoring vegetation structure on Natura 2000 areas using Remote Sensing

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

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

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

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

<sup>δ</sup> Both authors contributed equally to this paper

## ABSTRACT

Nature areas in the Netherlands are under threat of nitrogen deposition as a result of agricultural activity, traffic and industry. Regular monitoring of biodiversity in these areas is, therefore, important to establish proper mitigation actions. Monitoring of nature areas is traditionally done by field surveys which is time consuming and has a low land coverage. As a result, these areas are on average monitored once every six years. Ideally, changes in biodiversity should be more frequently monitored to regularly follow the effects of adopted measures. In the present study, we combined high-resolution satellite imaging and Lidar data to predict vegetation structure in two Natura 2000 sites in the Province of Zuid-Holland. A random forest model was used to detect different vegetation classes. Training data was obtained by annotating satellite images with the help of field experts. Results showed that for the Coepelduynen area, the classes grass, forest, wet dune valley and sand showed a high predicting accuracy ( $F_1$ -score > 0.9) while shrub showed an intermediate accuracy ( $F_1$ -score = 0.6). For the Voornes Duin area, the classes forest, shrub, water and sand showed a high accuracy ( $F_1$ -score > 0.8) while grass showed an intermediate accuracy ( $F_1$ -score= 0.6). Sandy areas are particularly sensitive to overgrowth by plant species which benefit from nitrogen such as grasses, shrub and trees. For protected areas in the Netherlands, remote sensing may be an important addition to field surveys, allowing to monitor nature sites more regularly and point out areas which need further inspection by specialists.

KEYWORDS: Remote sensing; Vegetation structure; Natura 2000; Random forest; Time-series.

## 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 (of the European Communities 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 (of the European Communities 2009) 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.

decisions having to be made about the trade-off between biodiversity, construction and agriculture activity (Stokstad 2019).

Within the Province of Zuid-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 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 et al. 2019, Reis et al. 2019, Koma et al. 2021, Moudrì et al. 2021, Zhang and Shao 2021, Wenge Ni-Meister 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

The Netherlands has designated 162 Natura 2000 sites, 21 of these sites are located in the Province of Zuid-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). Biodiversity has been, recently, subject of increasing debate with difficult

\*✉ vwdh@pzh.nl



Figure 1: Location of the study sites.

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 multispectral satellite imagery with Lidar technology to monitor vegetation structure in two Natura 2000 sites in the Province of Zuid-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 aim is to develop a tool to help monitoring biodiversity in Nature areas sensitive to nitrogen deposition.

## 2 MATERIALS AND METHODS

### 2.1 Study areas

This study focused on two Natura 2000 sites in the Province of Zuid-Holland, the Coepelduynen and the Voornes Duin (Fig. 1). These are young calcareous coastal dunes characterized by a wide variety of landscape types. The 188 ha Coepelduynen site is a narrow strip of dunes located between the municipalities of Katwijk and Noordwijk. 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 has an area of 1432 ha and is located on the island of Voorne-Putten (Fig. 1). It has a large variety of landscape types from dune grasslands and wet dune valleys (i.e. humid dune slacks) to large areas of forest and shrub, salt-marshes 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. This site 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 per-

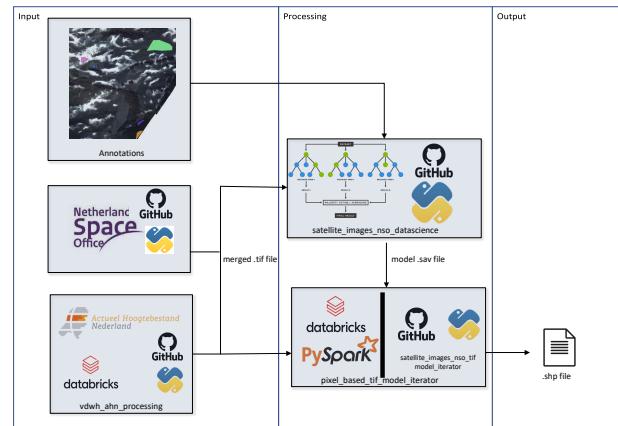


Figure 2: Remote Sensing component overview.

formed 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, the middle panel the used model and the multiprocessing/parallelism implementation, and the right panel the output. In the following sections each component is described in detail (see section 5.1 for documentation in 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 an 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, B et al. 2019, Sajjad 2021, R et al. 2022). 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. Only images 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 for all satellite images, which was acquired for the study areas in 2020. 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 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 calculations were done using PySpark (Fouilloux 2017) due to the long computation time.

### 2.2.4 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 (QGIS Development Team 2021) 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. See section 5.4 for access to annotation data.

## 2.3 Model selection and training

The selected model is a pixel based model, which means that we predicted every pixel in an image. 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. Grid search was performed using the RandomizedSearchCV class from

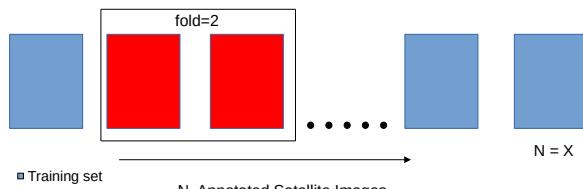


Figure 3: Cross-validation with random annotated satellite image sampling as folds

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 images as training set (see Fig. 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 average F1-score over all folds 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 computation time, multiprocessing and parallelism was used to generate model predictions. For the Coepelduynen, multiprocessing was done locally using the Multiprocessing module within Python. Due to the large area of the Voornes Duin, parallelism was used for this area 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. A list of the selected images is presented in Table 4 (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,  $p>0.95$ ), and between NIR and NDVI ( $p=0.8$ ). 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, 8 satellite images were annotated for the Voornes Duin and 7 for the Coepelduynen (Appendix Table 4). For the Coepelduynen, the following classes were defined: forest, grass, shadow, sand, wet dune valley (also named humid dune slacks), 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. Values are deter-

225 mined based on spectral and Lidar data of annotated pixels in prepared satellite images (see sections 2.2.2 and 2.2.3).

**Table 1: Mean values of the blue spectrum, NDVI and height (m) for the different classes in the Coepelduynen and Voornes Duin sites.**

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.

For the Coepelduynen area, prediction accuracy (F1-score) was the highest for the classes grass, forest, wet dune valley 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 = 0.6).

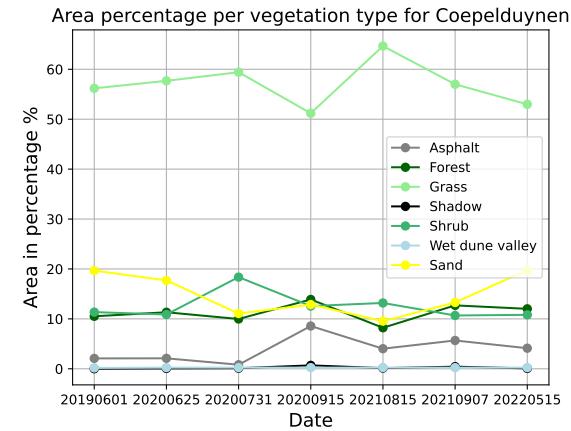
**Table 2: Mean values of performance metrics for the Coepelduynen.**

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

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 = 0.2; Table 3). Grass showed an intermediate accuracy (F1-score = 0.6).

**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



**Figure 4: Time-series of vegetation cover in the Coepelduynen**

### 3.4 Temporal changes in vegetation structure

In order to monitor changes in vegetation structure over time, the area occupied by each class was determined for each image. In the Coepelduynen site, more than 50% of the vegetation was identified as grass (Fig. 4). Forest, sand and shrub cover 10-20% of the area while wet dune valley only a few percent. Overall, the trends for the different vegetation classes remain rather constant over time. However, between June 2009 and July 2020 a decrease in sand area and increase in grass is predicted while the opposite occurs in 2021 and 2022, with a increase in sand and decrease in grass.

To visualize the difference in sand between two years, satellite images from August 2021 and May 2022 were compared (Fig. 5). The increase in predicted area covered by sand (upper images) from 2021 to 2022 is also seen in the original satellite images (lower images).

In the Voornes Duin site, the model classified most vegetation as forest and shrub, covering 30-35% of the area (Fig. 5). Grass covers about 10-20% of the area, while sand, water and wet dune valley less than 10%. Overall, trends do not fluctuate much over time although between March 2020 and April 2021 there seems to be a decrease in forest area and increase in shrub area. Due to the large area of the Voornes Duin, in some months a small part of the area (the northern point) is missing in the satellite image. Therefore, in order to calculate the percentage coverage this area was excluded from the images, to allow a more accurate comparison between dates.



**Figure 5:** Comparison of images from the Coepelduynen from August 2021 and May 2022. The upper panel shows the model annotated images while the lower panel shows the original satellite images.

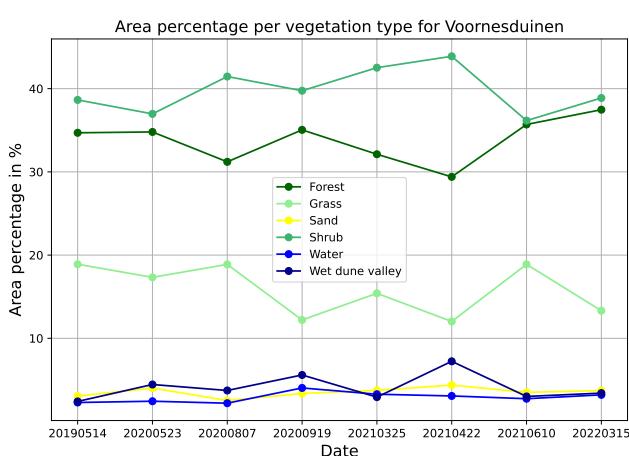
A cross-section of two satellite images from the Voornes Duin from March 2020 and April 2021 shows a example of the original satellite image and the model annotated image. 275

## 4 DISCUSSION

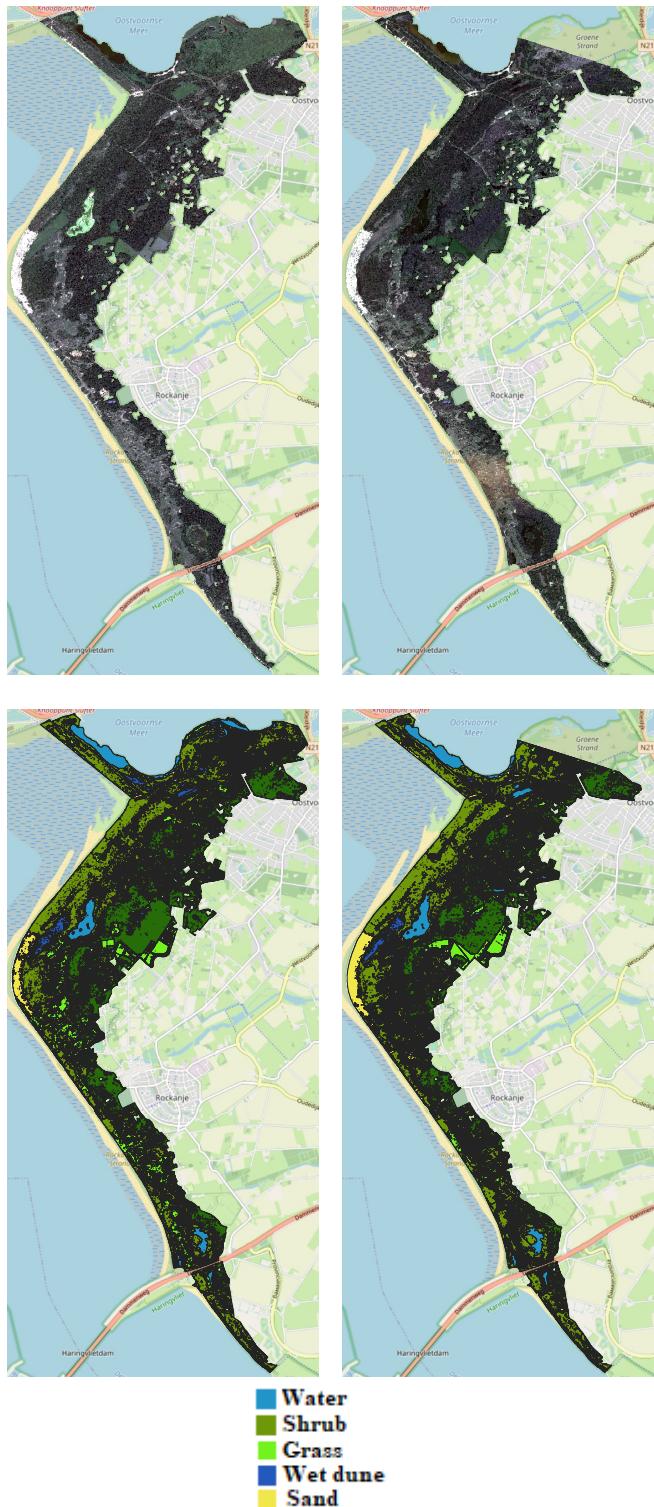
### 4.1 Vegetation structure

Dune areas and heathlands are particularly sensitive to overgrowth by plant species which benefit from nitrogen, such as grasses, shrub and trees (Stokstad 2019), leading to a decrease in habitat specific species and an overall decrease in biodiversity. In the Voornes Duin site, the negative effects of nitrogen deposition have already been observed some decades ago: dune grassland and open dune areas, characterised by a rich biodiversity, have been overgrown by shrub and forest leading to a decrease in habitat quality (Ministerie van Landbouw 2010). Between 2005 and 2008, restoration projects were carried out in some areas to mainly remove shrub and trees. In a short period of time, pioneer plants typical from sandy bottoms increased, followed by an increase in insects and breeding birds (Ministerie van Landbouw 2010). Recent reports based on field observations suggest that the restored areas are being invaded again by nitrogen-loving species and new actions are being considered (Arcadis 2022a, Zuid-Holland 2022). Recent environmental analysis in the Coepelduynen site has also shown that this area is under pressure of increasing nitrogen deposition (Arcadis 2022b, Zuid-Holland 2022). Also here, restoration actions are planned to decrease overgrowth by grasses, shrubs and trees. 280  
285  
290  
295  
300

Our results show that the combination of high-resolution satellite imaging and Lidar data is useful to access changes in vegetation structure, as reported in many other papers (muLLcher2017ontwikkelen, Bolton et al. 2018, Jongman et al. 2019, Mücher et al. 2019, Rapinel et al. 2019, Reis et al. 2019, Zhang and Shao 2021, among others). For the studied areas, random forest showed a good accuracy to predict sand, water, shrub, grass and forest. Overall accuracy is in line with other vegetation structure studies based on the same approach (around 80%; Mücher et al. 2019, Rapinel et al. 2019, Reis et al. 2019, Zhang and Shao 2021, Marcinkowska-Ochtyra et al. 2019). The prediction accuracy for wet dune valley showed a discrepancy between areas with a high F1-score in the Coepelduynen and a low one in the Voornes Duin. Wet dune valleys (humid dune slacks) are wet depressions left between dunes where the groundwater reaches or approaches the surface of the sand. They may contain water pools during winter and spring and vegetation, such as low pioneer vegetation and grasslands, in the summer (Houston 2008). Differences in the time of the year height data were obtained or in the way images were annotated could be responsible for the differences in accuracy between areas. For the Coepelduynen the height for these class was zero, identical to water, while for the Voornes Duin height was similar to the one of grass. It could also be that in the former site annotations were done in the center of the dune valley, which is more moist, and in the latter more on the sides which have more vegetation. 305  
310  
315  
320  
325



**Figure 6:** Time-series of vegetation cover in the Voornesduinen



**Figure 7:** Comparison of images from the Voornes Duin from August 2020 and April 2021. The upper panel shows the satellite images and the lower panel the model annotated images.

For protected areas in the Netherlands, the approach presented in this study may be an important addition to field surveys, allowing to monitor nature sites more regularly and point out at an earlier stage which areas need further inspection by specialists and eventually mitigation actions.

#### 4.2 Annotated satellite images as training data instead of ground truth labels

In the present paper, we chose to use annotations based on satellite images instead of the commonly used ground truth labels to train our model. This was done together with experts involved in the monitoring of the studied areas and, in our case, had several advantages in relation to using ground truth labels:

- **Discrepancy in time:** ground truth labels in the Netherlands are taken once every six years per area and, therefore, there is a time difference between ground annotations and satellite image. This can lead to noise and bias in the data.
- **Training data over time:** since ecologists need to cover a large area in order to make ground truth annotations it is impossible to produce a lot of training data across different time periods. Annotating satellite images is less time consuming which results in more annotations over time and, therefore, more training data for the model. A large amount of training data may be helpful to deal with noise due to different atmospheric conditions.
- **Discrepancy in orthorectification:** orthorectification is the process of converting images into a format suitable to be used in maps. This is done by removing distortions, due to the sensor system, the satellite motion and changes in the geometry of the terrain, from the raw image (Leprinse et al. 2007). Although the downloaded SuperView satellite images are already orthorectified, there is still some discrepancy in coordinates between satellite images. By using ground truth annotations which could be not properly aligned with the satellite image would lead to bias and noise in the training data.

#### 4.3 Leave-two-out cross-validation approach instead of random sampling

In the present paper, we chose to randomly pick different satellite images to serve as a test set instead of doing random sampling. Earlier tests revealed that when the model was trained using random cross-validation sampling, performance was better. This is because random sampling will usually take training (annotated) data from all satellite images and, therefore, data from one satellite image appears both in the test and training sets. Each satellite image has its own RGBI value ranges due to different reasons such as the position of the sun (Sekrecka et al. 2020) or atmospheric influences (Huete and Jackson 1988, Tyagi and Bhosle 2011). If the model has been trained on annotated data from one image, it is logical that it will perform better on the test set of the same image than if it had not seen that image before. probably already figured out it's unique RGBI value range. However, in a production

environment, a trained model is used on new satellite images, which it has not seen before. It is, therefore, more accurate to access its performance by excluding images during cross-validation than by using random sampling.

#### 4.4 Further work

This paper has shown that the use of remote sensing techniques on Natura 2000 areas in the Netherlands may be useful to access changes in vegetation structure and monitor of biodiversity. Random Forest modeling is widely used for classification problems similar to the one in the present paper. Nevertheless, deep learning methods have also been successfully implemented in the agricultural sector to estimate e.g. land cover or crop type (Kamilaris and Prenafeta-Boldú 2018, and references therein). We chose as a first modeling approach random forest but deep learning methods could lead to a better performance. These methods are, however, more time consuming and have higher associated costs. In the present paper, a pixel based model is used due to its easy integration with traditional machine learning algorithms such as random forest. However, other authors point out that an object based model could lead to a better performance (Hussain et al. 2013, Zerrouki and Bouchaffra 2014). We aim in a near future to test different methodologies to achieve better predictions.

### 5 SOURCE CODE

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

#### 5.1 Components overview

For a overview and links to all the components, see [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 [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 see [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 via <https://github.com/Provincie-Zuid-Holland/satellite-images-nso-datascience>.

For training and cross-validation code for the Random Forest model go to [https://github.com/Provincie-Zuid-Holland/satellite-images-nso-datascience/blob/main/annotations\\_models/Coepelduynen/random\\_forest/make\\_train\\_model\\_on\\_annotations\\_coepelduynen.ipynb](https://github.com/Provincie-Zuid-Holland/satellite-images-nso-datascience/blob/main/annotations_models/Coepelduynen/random_forest/make_train_model_on_annotations_coepelduynen.ipynb).

Annotations are found in <https://github.com/Provincie-Zuid-Holland/satellite-images-nso-datascience/tree/main/data/annotations>.

#### 5.5 Pixel Based Model Iterator

This repository contains a stand alone pixel based model iterator for running multiprocessing locally, which may be used with any model as long as a .predict function is available. The PySpark code used for parallelism is also available. See [https://github.com/Provincie-Zuid-Holland/satellite\\_images\\_nso\\_tif\\_model\\_iterator](https://github.com/Provincie-Zuid-Holland/satellite_images_nso_tif_model_iterator).

### ACKNOWLEDGEMENTS

The authors would like to extend their appreciation to Wendy Jesse, Gijs Koene, Huibert van Rossum and Daniël van der Staak for their help with annotating satellite images. The authors would also like to thank the vDWH-team colleagues for their support during this study and the fruitful discussions. This research was funded by the Province of Zuid-Holland.

### REFERENCES

- Tucker, C. J. (1979). "Red and photographic infrared linear combinations for monitoring vegetation". *Remote sensing of Environment* 8(2), pages 127–150.
- 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.
- Of the European Communities, C. (1992). "Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora". *Official Journal of the European Union* L 206, pages 0007–0050. doi: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31992L0043>.
- 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.
- 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.
- Pettorelli, N., J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth (2005). "Using the satellite-derived NDVI to assess ecological responses to environmental change". *Trends in ecology and evolution* 20(9), pages 503–510.
- Leprince, S., S. Barbot, F. Ajoub, 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.
- Houston, J. A. (2008). *Management of Natura 2000 habitats. 2190 Humid dune slacks*. Technical report. European Commission.
- 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.

- 490 Of the European Communities, C. (2009). "Directive 2009/147/EC of the European Parliament and of the Council of 30 November 2009 on the conservation of wild birds". *Official Journal of the European Union* L 20/7. doi: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009L0147>.
- 495 Ministerie van Landbouw Natuur en voedselkwaliteit, N. (2010). *Onderzoeks-monitoring Voornes Duin 2004-2008. Den Haag, juni 2010.*
- 500 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.
- 505 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.
- 510 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.
- 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.
- 515 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.
- Fouilloux, A. (2017). *Introduction to PySpark Version 2017.01.*
- Mücher, S., H. Kramer, R. van der Wijngaart, and R. Huiskes (2017). *Ontwikkelen van een Remote Sensing monitoringssysteem voor vegetatiestructuur: pilotstudie: detectie verruiging Grijze Duinen (H2130) voor het Natura 2000-gebied Meijendel-Berkheide*. Technical report. Wageningen Environmental Research.
- 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 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.
- 520 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.
- B, C., S. F., and P. A. S.-H. F. A. H. R. O. A. M. M. A. M. S. V. P. S. H (2019). "Chapter 17 - Effects of drought on vegetative cover changes: Investigating spatiotemporal patterns." *Extreme Hydrology and Climate Variability*, pages 213–222. doi: <https://doi.org/10.1016/B978-0-12-815998-9.00017-8>.
- 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.
- Jongman, R. H., C. A. Mücher, R. G. Bunce, M. Lang, and K. Sepp (2019). "A Review of Approaches for Automated Habitat Mapping and their Potential Added Value for Biodiversity Monitoring Projects". *Journal of Landscape Ecology* 12(3), pages 53–69. doi: [doi:10.2478/jlecol-2019-0015](https://doi.org/10.2478/jlecol-2019-0015).
- Marcinkowska-Ochtyra, A., K. Gryguc, A. Ochtyra, D. Kopeć, A. Jarocińska, and Ł. Ślawik (2019). "Multitemporal Hyperspectral Data Fusion with Topographic Indices—Improving Classification of Natura 2000 Grassland Habitats". *Remote Sensing* 11(19).
- Mücher, C. A., H. Kramer, M. R. Najafabadi, L. Kooistra, A. T. Kuiters, and P. A. Slim (2019). "Exploiting low-cost and commonly shared aerial photographs and LiDAR data for detailed vegetation structure mapping of the Wadden Sea island of Ameland". *Journal of Earth Sciences and Environmental Studies* 4(1).
- Rapinel, S., C. Mony, L. Lecoq, B. Clément, A. Thomas, and L. Hubert-Moy (2019). "Evaluation of Sentinel-2 time-series for 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, 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.
- 525 Stokstad, E. (2019). "Nitrogen crisis threatens Dutch environment and economy". *Science* 366(6470), pages 1180–1181. doi: [10.1126/science.366.6470.1180](https://doi.org/10.1126/science.366.6470.1180).
- Sekrecka, A., D. Wierzbicki, and M. Kedzierski (2020). "Influence of the sun position and platform orientation on the quality of imagery obtained from unmanned aerial vehicles". *Remote Sensing* 12(6), page 1040.
- Koma, Z., A. Zlinszky, L. Bekő, 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.
- Moudrý, V., L. Moudrá, V. Barták, V. Bejček, K. Gdulová, M. Hendrychová, D. Moravec, P. Musil, D. Rocchini, K. Štastný, 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.
- 530 QGIS Development Team (2021). *QGIS Geographic Information System*. Open Source Geospatial Foundation.
- Sajjad, S. D. S. R. S. C. H. (2021). "Chapter 3 - Analyzing seasonal variation in the vegetation cover using NDVI and rainfall in the dry deciduous forest region of Eastern India." *Forest Resources Resilience and Conflicts*, pages 33–48. doi: <https://doi.org/10.1016/B978-0-12-822931-6.00003-4>.
- Zhang, Y. and Z. Shao (2021). "Assessing of urban vegetation biomass in combination with LiDAR and high-resolution

- remote sensing images". *International Journal of Remote Sensing* 42(3), pages 964–985.
- 605 Arcadis Royal Haskoning DHV, S. (2022a). *Natuurdoelanalyse Natura 2000. 100 Voornes Duin, Provincie Zuid-Holland*. Den Haag, 14 maart 2022.
- (2022b). *Natuurdoelanalyse Natura 2000. 96 Coepelduynen, Provincie Zuid-Holland*. Den Haag, 14 maart 2022.
- 610 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".
- 615 R, K., N. A. J. N. A. S. N, and P. R (2022). "Landsat-based multi-decadal spatio-temporal assessment of the vegetation greening and browning trend in the Eastern Indian Himalayan Region". *Remote Sensing Applications: Society and Environment* 25(9), page 113147. doi: <https://doi.org/10.1016/j.rsase.2022.100695>.
- 620 Taddeo, S. (2022). "Leveraging time series of satellite and aerial images to promote the long-term monitoring of restored plant communities". *Applied Vegetation Science* 25(2), e12664.
- 625 Wenge Ni-Meister Alejandro Rojas, S. L. (2022). "Direct use of large-footprint lidar waveforms to estimate above-ground biomass". *Remote Sensing of Environment* 280, page 113147.
- Zuid-Holland, P. (2022). *Gebiedsplan Stikstof 0.5*. Den Haag, mei 2022.

**APPENDIX**

- 630      SuperView satellite images used in the present paper  
Satellite images Coepelduynen  
20190601\_105844\_SV1-  
04\_50cm\_RD\_11bit\_RGBI\_KatwijkAanZee.tif



640

20200915\_112329\_SV1-  
04\_SV\_RD\_11bit\_RGBI\_50cm\_KatwijkAanZee.tif

- 635      20200625\_112015\_SV1-  
03\_SV\_RD\_11bit\_RGBI\_50cm\_Rijnsburg.tif



645

20210815\_111051\_SV1-  
03\_SV\_RD\_11bit\_RGBI\_50cm\_Oegstgeest.tif

- 20200731\_112003\_SV1-  
03\_SV\_RD\_11bit\_RGBI\_50cm\_Rijnsburg.tif



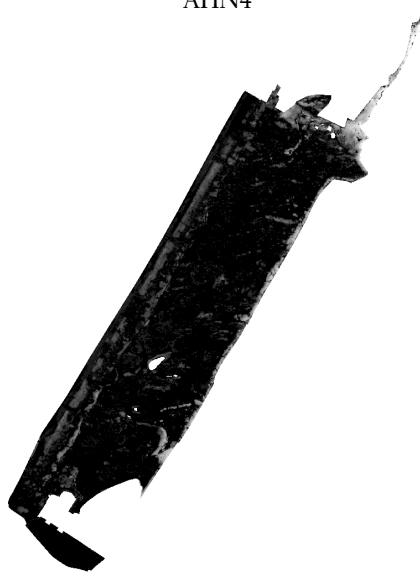
650  
20210907\_112017\_SV1-  
04\_SV\_RD\_11bit\_RGBI\_50cm\_KatwijkAanZee.tif



AHN3



20220515\_113347\_SV1-  
02\_SV\_RD\_11bit\_RGBI\_50cm\_KatwijkAanZee.tif



AHN4

655



Satellite images Voornes duinen  
20190514\_110143\_SV1-  
03\_50cm\_RD\_11bit\_RGBI\_Rockanje.tif



660  
20200523\_110107\_SV1-  
01\_SV\_RD\_11bit\_RGBI\_50cm\_Rockanje.tif



20200919\_112329\_SV1-  
03\_SV\_RD\_11bit\_RGBI\_50cm\_Rockanje.tif



665  
20200807\_112237\_SV1-  
03\_SV\_RD\_11bit\_RGBI\_50cm\_Rockanje.tif



670  
20210325\_114359\_SV1-  
02\_SV\_RD\_11bit\_RGBI\_50cm\_Rockanje.tif



20210422\_112448\_SV1-  
04\_SV\_RD\_11bit\_RGBI\_50cm\_Rockanje.tif



20220315\_112039\_SV1-  
04\_SV\_RD\_11bit\_RGBI\_50cm\_Rockanje.tif

680



20210610\_111718\_SV1-  
04\_SV\_RD\_11bit\_RGBI\_50cm\_Hellevoetsluis.tif

675

