Synergistic use of Sentinel-2 and UAV-derived data for Plant Fractional Cover distribution mapping of coastal meadows with Digital Elevation Models.

Ricardo Martínez Prentice[1], Miguel Villoslada Peciña[2,1], Raymond D. Ward[1,3], Thaisa F. Bergamo[1,2], Chris B. Joyce[4], and Kalev Sepp[1]

[1]Institute of Agriculture and Environmental Sciences, Estonian University of Life Sciences, Kreutzwaldi 5, EE-51006 Tartu, Estonia.
[2]Department of Geographical and Historical Studies, University of Eastern Finland, P.O. Box 111, 80101 Joensuu, Finland.
[3]School of Geography, Queen Mary University of London, London E1 4NS, UK.
[4]Centre for Aquatic Environments, School of Applied Sciences, University of Brighton, Cockcroft Building, Moulsecoomb, Brighton BN2 4GJ, UK.

Correspondence: Ricardo Martínez Prentice (ricardo@emu.ee)

Abstract. Coastal wetlands provide a range of ecosystem services, yet are currently under threat from global change impacts. Thus, monitoring and assessment is vital for evaluating their status, extent and distribution. Remote sensing provides an excellent tool for evaluating coastal ecosystems, whether with small scale studies using drones or national/regional/global scale studies using satellite derived data. This study used a fine-scale plant community classification of coastal meadows in Estonia derived from a multispectral camera on board Unoccupied Aerial Vehicles (UAV) to calculate the Plant Fractional Cover (PFC) in Sentinel-2 MultiSpectral Instrument sensor (MSI) grids. A Random Forest (RF) algorithm was trained and tested with vegetation indices (VI) calculated from the spectral bands extracted from the MSI sensor to predict the PFC. Additional RF models were trained and tested after adding a Digital Elevation Model (DEM). After comparing the models, results show that using DEM with VI can increase the prediction accuracy of PFC up to two times (R^2 58-70%). This suggests the use of ancillary data such as DEM to improve the prediction of empirical machine learning models, providing an appropriate approach to upscale local studies to wider areas for management and conservation purposes.

1 Introduction

Vegetation is the main target of study to monitor ecosystem changes caused by drastic environmental shifts, because it is the key structural component of ecosystems (Diekmann, 2003; Van der Maarel, 2005). Variations in the distribution patterns of vegetation over an area depend on climate, environmental factors and human activities (Gardner et al., 2009) and can be assessed by identifying plant communities. These are assemblages of plant species at a place and time (Magurran, 1988; Spiegelberger et al., 2012), and are considered as ecosystem service (ESs) - providing units, as their structure and function directly underpins the supply of ESs (Luck et al., 2003; Burkhard et al., 2012).

Wetlands are one of the most degraded type of ecosystems in the world and natural wetlands have experienced a 50% decline in total area since 1900 (Davidson, 2014), continuing to decline by 35% between 1970 and 2015, mainly due to agricultural expansion and intensification, and drought (Courouble, 2021). Among wetlands, coastal wetlands have increasingly received attention due to their capacity for carbon sequestration (Hopkinson et al., 2012; Ward, 2020), coastal protection (Gedan et al., 2011) and biodiversity maintenance (Sutton-Grier and Sandifer, 2019).

Boreal Baltic coastal meadows, as stated in Annex I of the EU Habitats Directive (1992), are semi-natural wetlands managed for centuries with low-intensity activities (Paal, 1998) such as grazing and mowing. Many such meadows along the Baltic Sea coast currently show a degraded ecological status as a consequence of agriculture intensification or abandonment of traditional management (Henle et al, 2008, Rannap et al, 2004). Since the 1960s, the total area of coastal meadows has decreased by 34000 ha in Estonia, affecting a range of breeding and migratory bird species listed in the Birds Directive (Rannap et al, 2004, Leito et al, 2014). To assess the effectiveness of conservation efforts, previous studies on coastal meadows have focused on the effects of different environmental and management factors on the distribution of plant communities, such as sea level rise, microtopography (Ward et al., 2016), grazing abandonment (Burnside et al., 2007), and mowing (Berg et al., 2012).

Remote Sensing techniques are increasingly used to map the distribution of coastal meadow plant communities (Villoslada et al., 2020; Martínez Prentice et al., 2021) and to estimate biomass and sward structure using Unoccupied Aerial Vehicles (UAVs) (Villoslada Peciña et al., 2021). The very high spatial resolutions supplied by UAV-borne sensors also allow fine-grained ecosystem properties to be unveiled, which otherwise that remain concealed under the coarse spatial resolution of satellites, such as plant fractional cover, soil organic carbon, or aboveground biomass (Heil et al., 2022). In addition, near-real-time monitoring routines and the avoidance of the effect of clouds are among the advantages of UAVs over satellite sensors (Colomina and Molina, 2014; Díaz-Delgado et al., 2019). Conversely, the disadvantages are not only their limited coverage and battery capacity but also the legislation restrictions and their dependency on the weather conditions, as well as the requirement to be in the field (Cracknell, 2017; Emilien et al., 2021).

On the other hand, Earth Observation satellites capture images with large swaths and a high temporal resolution, which allows the consistent study of large extents of ecosystems over multiple years. In the last decade, the idea of combining the high spatial resolution derived from UAVs with the large swath and regular revisit times of satellites has gained momentum. Some studies have successfully addressed the potential upscaling of UAV multispectral images to satellite image resolutions in order to address wetland biophysical variables at multiple scales (Laliberte et al., 2011; Díaz-Delgado et al., 2019) with UAV as a support for ground-truth observations. The accurate geometrical and radiometrical overlapping allows UAV imagery values to be aggregated into satellite pixel grids (Padró et al., 2018).

These remotely sensed data in combination with artificial intelligence is essential to supply comprehensive assessments of these shifts (Knight et al., 2006; Adam et al., 2010; Pettorelli et al., 2014), are playing a key role in wetland mapping, ecosystem monitoring and trend detection, overcoming some of the difficulties of local wetland surveys with traditional in-situ field methodologies in large areal extents, remoteness and inaccessibility (Mahdianpari et al., 2020). As the growing impacts of land-use intensification and climate change become more conspicuous and widespread (Findell et al., 2017), local-scale field survey methods may not adequately reveal plant community shifts in a spatially-explicit manner to study spatio-temporal patterns in plant community distribution, environmental monitoring, and biodiversity conservation.

Modelling plant community coverage with remote sensing data is one of the main goals in ecological assessments and monitoring (Corbane et al., 2015). Combining remotely sensed data with Machine Learning (ML) algorithms shows robust performance due to their ability to deal with non-parametric distribution of the ground-truth data as well as the multicollinearity of variables (Rodriguez-Galiano et al., 2012; E Thessen, 2016; Maurya et al., 2021). ML-based models are used to predict the presence of vegetation using indices as the input for different algorithms (Maurya et al., 2021), and Random Forest (RF) has been shown to be an accurate algorithm to predict Plant Fractional Cover (PFC) over large areas (Zhang et al., 2019; De Simone et al., 2021). Moreover, ancillary data such as Digital Elevation Models (DEM) interpolated from Light Detection and Ranging (LiDAR) point clouds have been successfully used for mapping plant communities in coastal wetlands (Ward et al., 2013) together with multispectral data from remote sensing platforms. This combination provides an enhancement on the detection of new plant distribution patterns (Okolie and Smit, 2022).

The present study compared two PFC models of five plant communities in Estonian coastal meadows from Vegetation Indices (VI) calculated with Sentinel-2 MultiSpectral Instrument (MSI) sensor and ancillary data from a DEM. High-resolution UAV imagery was used as the reference for PFC within the spatial resolution of a Sentinel-2 image. The main objectives were to: (1) quantify the relationship between UAV imagery and MSI imagery values; (2) predict and test ML models to predict individual PFC per plant community with VI derived from Sentinel-2 spectral values; (3) build and test the performance of the models by adding DEM data to the VI.

To improve the article's readability, we have included a list of abbreviations in Table 1.

Table 1. List of abbreviation and acronyms used in the paper

Abbreviation	Definition	Abbreviation	Definition
UAV	Unmaned Aerial Vehicle	N	Number of Estimators
MSI	MultiSpectral Instrument	MF	Maximum Features
PS	Parrot Sequoia	RMSE	Root Mean Squared Error
ESs	Ecosystem Services	MBE	Mean Biased Error
VI	Vegetation Index	LS	Lower Shore
PFC	Plant Fractional Cover	OP	Open Pioneer
DEM	Digital Elevation Model	US	Upper Shore
dGPS	Differential Global Positioning System	TG	Tall Grassland
DI	Initial Dataframe	RS	Reed Swamp
ID	Unique Identifier	KUD	Kudani
DF0	Sampled Data Frame	TAN	Tahu North
DF1	Data Frame 1	TAS	Tahu South
DF2	Data Frame 2	RAL	Rälby
RF	Random Forest	MAT	Matsalu

2 Materials and Methods

2.1 Study areas

Six coastal meadow study sites located in protected areas on the west coast of Estonia were selected for this study. Kudani (KUD), Tahu North (TAN) and Tahu South (TAS) belong to the Silma Nature Reserve; Rälby (RAL) and Rumpo (RMP), to the Vormsi Landscape Protection Area; and Matsalu (MAT), to the Matsalu National Park (Figure 1). These landscapes are characterized by coastal meadows extended over a gradual transition from the sea to terrestrial ecosystems with a low variation of topography, typically 0 to 2 metres above mean sea level (Ward et al., 2016). Sites were chosen based on their near-continuous management history, high conservation value for wading birds, and presence of endangered plant species (Rannap et al., 2004; Berg et al., 2012).

The plant communities under study are very characteristic of Estonian coastal meadows and have been previously grouped following a phytosociological classification by Burnside et al (2007): Lower Shore (LS), Open pioneer (OP), Upper shore (US), Tall grassland (TG), and Reed Swamp (RS). This classification has been used in various studies in these coastal meadows and the plant communities have proven to be differentiable from high-resolution images (ca. 10 centimetres per pixel) (Ward et al., 2013; Villoslada et al., 2020; Martínez Prentice et al., 2021). The distribution of plant communities shows site-specific patterns due to local variations in the inundation levels and flood frequencies (Rivis et al., 2016), sediment accretion, microtopography (Ward et al., 2016), and grazing regimes (Berg et al., 2012). Figure A1 shows the plant community distribution in relation

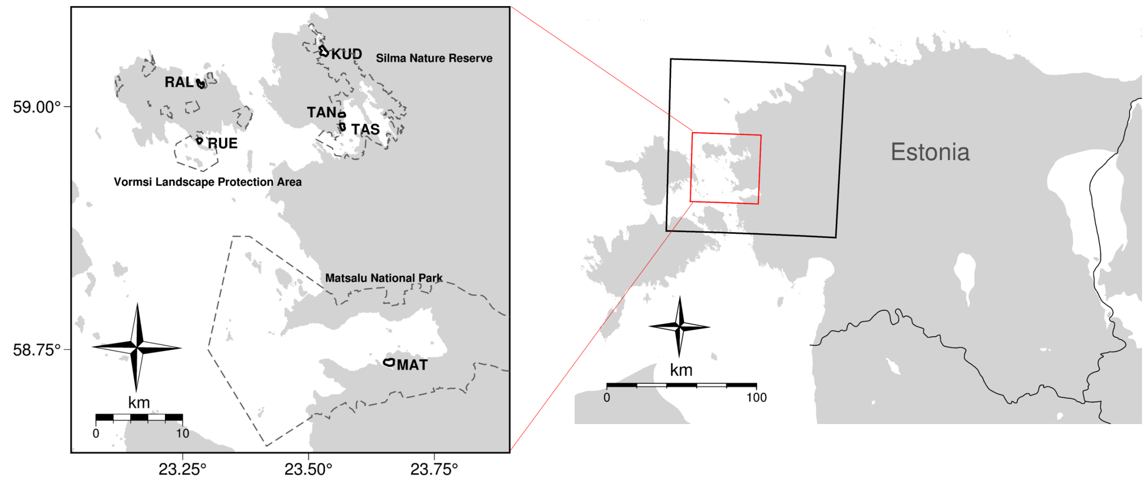

Figure 1. Location of the study sites in the western coast of Estonia. Matsalu (MAT), Tahu South (TAS), Tahu North (TAN), Kudani (KUD), Rälby (RAL) and Rumpo (RMP). The black square shows the Sentinel-2 tile footprint, whereas the extent of all the study areas is in red.

to this microtopography in a boxplot. Floods depend mostly on the meteorological conditions across the North Atlantic and Fennoscandia (Kont et al., 2003) and the maximum level is reached in April after the snow melts. This is followed by the growing season, which is characterized by the maximum plant activity occurring from May until September. During this period, plant communities rely on mean temperatures above 10 °C (Maasing and Paal, 1998).

Table 2 provides a distribution description of each plant community in this study.

Table 2. Summary of the description of plant communities in this study (Berg, 2008; Ward, 2012). Lower Shore (LS), Open Pioneer (OP), Upper Shore (US), Tall Grassland (TG) and Reed Swamp (RS).

Plant Community	Description
Lower Shore (LS)	This community has adapted to significant variations in hydrological conditions, which result in the accumulation of litter and waterlogged soil, leading to considerable salinity concentrations.
Open Pioneer (OP)	This community of halophytic species is located in low-lying areas subject to prolonged inundation during the growing season, with its distribution primarily influenced by salinity. These specific locations exhibit the highest proportion of bare ground cover and highest salinity levels.
Upper Shore (US)	This community is established in higher elevations, characterized by less frequent and shorter floods. US is relatively more species rich and productive than LS.
Tall Grassland (TG)	This community is located on the highest elevations within the coastal meadows. Flooding is less pronounced and frequent, and vegetation is dense and very species-rich.
Reed Swamp (RS)	This community consists of extensive reedbeds along the coastline, which are influenced by more frequent inundations of brackish water.

2.2 UAV classification data

A classification of the plant communities from high-resolution images (Martínez Prentice et al., 2021) was used as training/-validation. A multispectral Parrot Sequoia (PS) camera was carried on board of an eBee fixed-wing drone controlled remotely with the software SenseFly eMotion (Parrot S.A. Paris) over the six study areas at an altitude of 120 m to obtain Ground Sample Distance of 10 cm. The UAV flights were conducted during the dates corresponding to the growing season, carefully chosen to minimize the impact of inundation effects (Table 3). The images were radiometrically corrected with Airinov calibration panels and a sunshine sensor to produce multi band orthoimages that were merged in Pix4D v.4.3.31 software. VI were calculated based on all the spectral bands (green, red, red edge and near infrared spectral bands) and used as input for two different workflows: a pixel-based classification, where the pixels were classified with a Random Forest and K-nearest neighbours' algorithms; and segmentation for an object-based classification with the same algorithms. The highest accuracy was achieved by a Random Forest pixel classification (accuracy and kappa greater than 90% and 0.85, respectively), calculated from a confusion matrix constructed using 140 vegetation survey quadrats as training samples (Figure A2), where all the species with coverage above 5% were recorded within the quadrats. A Sokkia GSR2700 ISX differential Global Positioning System (dGPS) was used to record the location and elevation7 of each plant community.

The plants in OP community were recorded as a result of the low cover of all species and predominance of bare ground. Not all the plant communities were present in each site (Table A1). Further details of this methodology and results can be obtained in Martínez Prentice et al (2021).

2.3 Satellite imagery

Recent studies have shown that images taken by light-weight cameras in the visible and near infrared spectrum on board of UAV have a good correlation with satellite images, especially with MSI images of Sentinel-2 (Zabala, 2017; Zhu et al., 2021). Thus, one Sentinel-2 Level 2A image covering the six study areas (Figure 1) with the closest date to the drone flights was used (Table 2), with an estimated cloud cover of 19%. The tile number was T34VFL and its date, 24th of June of 2019. The level 2A was chosen because the orthorectified Bottom-of-Atmosphere reflectance values are comparable with PS reflectance (Fawcett et al., 2020). This image product is radiometrically corrected by the Payload Data Ground Segment with Sen2Cor algorithm (Main-Knorn et al., 2017) and available online via the Copernicus Scientific Data Hub tool (Copernicus Hub).

Table 3. Drone flight dates. Matsalu (MAT), Tahu South (TAS), Tahu North (TAN), Kudani (KUD), Rälby (RAL) and Rumpo (RMP).

Study Area	Drone flight date
MAT	29 June 2019
TAS	23 July 2019
TAN	30 June 2019
KUD	30 June 2019
RAL	04 July 2019
RMP	02 July 2019

Band 6 from MSI is in the Red-Edge region (Table 4) and contains valuable information of vegetation, avoiding background reflectance that affects wetlands especially (Turpie, 2013). Its spatial resolution is 20 metres per pixel. To use its reflectance values with the highest spatial resolution corresponding to the VNIR bands at 10 meters, an enhancement process based on a super-resolution method was applied, instead of using a panchromatic band to carry out a pan-sharpening since this band does not exist in MSI. The super-resolution algorithm (Brodu, 2017) is available in the SNAP software (ESA, 2014) and combines the geometric and radiometric information of target bands to increase the spatial resolution.

Table 4. Comparison of spectral resolution of bands in both sensors: MultiSpectral Instrument (MSI) on board of Sentinel-2 and Parrot Sequoia (PS) on board of eBee. First number is the central wavelength and the second one is the wavelength width. Units are in nanometers (nm).

Band	MultiSpectral Instrument	Parrot Sequoia
Green	559.8, 35	550, 40
Red	664.6, 30	660, 40
Red Edge	740.5, 14	735, 10
Near Infrared	832.8, 105	790, 40

2.4 Digital Elevation Models

DEMs constitute a powerful co-predictor in species distribution models, due to the prominent role of elevation in the distribution patterns of coastal plant communities (Ward et al., 2013). This holds especially true for coastal meadows, characterized by pronounced salinity and moisture gradients due to small variations of elevation, called microtopography (Ward et al., 2016).

Thus, a high-spatial resolution DEM was included in the models to test whether prediction accuracies improved. A LiDAR-derived DEM was downloaded from Eesti Maa-amet (Maa-amet geoportaal) with a spatial resolution of 1 m. The DEM was interpolated from a LiDAR point cloud of density of 2.1 points per square using a streaming triangulation (Isenburg et al., 2006). The vertical error was calculated with RMSE between in-situ elevation points of the DEM (Figure 2).

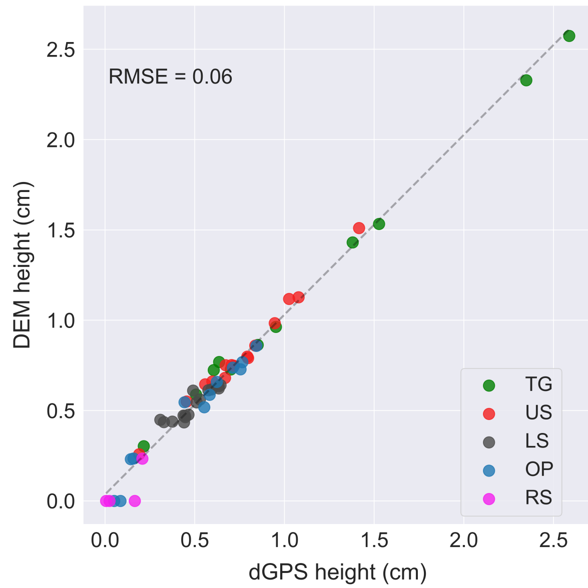

Figure 2. Vertical error between the measured heights with the differential Global Positioning System (dGPS), Sokkia GSR2700 ISX in each sampled plant community and the Digital Elevation Model (DEM). Units are centimeters (cm). Lower Shore (LS), Open Pioneer (OP), Upper Shore (US), Tall Grass (TG), Reed Swamp (RS).

2.5 Image processing and upscaling

A key process required to perform upscaling of remote sensing images is the aggregation of pixel values from a high-resolution image to the geographically coincident pixels of coarser resolution image. Several studies have performed the aggregation process to a common geographical data frame in the form of a quasi-continuous grid, where all the spectral data is stored (Padró et al., 2018; Riihimäki et al., 2019; Mao et al., 2022; Bergamo et al., 2023). In the present study, the grid was constrained to the limits of each study area, avoiding those overlapping with the edges, excluding transitional areas that do not correspond to the extent of plant communities of interest and submerged areas. In total, 9766 MSI pixels cover the study areas (Figure A3).

A band-to-band comparison between the PS bands used for the final classification in Martínez Prentice et al (2021) and MSI reflectance values was undertaken to assess the potential differences in both sensors caused by different temporal, spectral or spatial resolutions (Padró et al., 2018; Fernández-Guisuraga et al., 2018; Jiang et al., 2022; Isgró et al., 2022). To carry out this process, PS and MSI reflectance values were transferred into a polygon grid generated with the exact cell size as the MSI image pixels covering the study areas with an associated Unique Identifier (ID) for each row of the data frame (Figure 3). Level

2A MSI reflectance values were transferred to each cell of the polygon grid and the PS values were aggregated calculating the average mean (Figure 3). This approach generalizes the reflectance within a unit of grid, reducing noise from high-resolution images of PS and resulting in more predictable behavior (Blan and Butler, 1999). This aggregation criteria was also used to integrate the DEM values into the polygon grid (Figure 3).

The comparability and consistency of the spectral data from PS and MSI bands was analyzed by fitting the values in a linear model, calculating the coefficient of determination (R^2) and Root mean squared error (RMSE). The p-value showing the significance of the relation between PS and MSI.

The PFC was the response variable under assessment for each plant community. It was calculated by intersecting the UAV-derived classification maps (Martínez Prentice et al., 2021) within each polygon grid (MSI pixel) and applying equation 1. All the grids sum a total of 1 (100% PFC).

For all the operations, all the pixels completely covered by each grid were extracted.

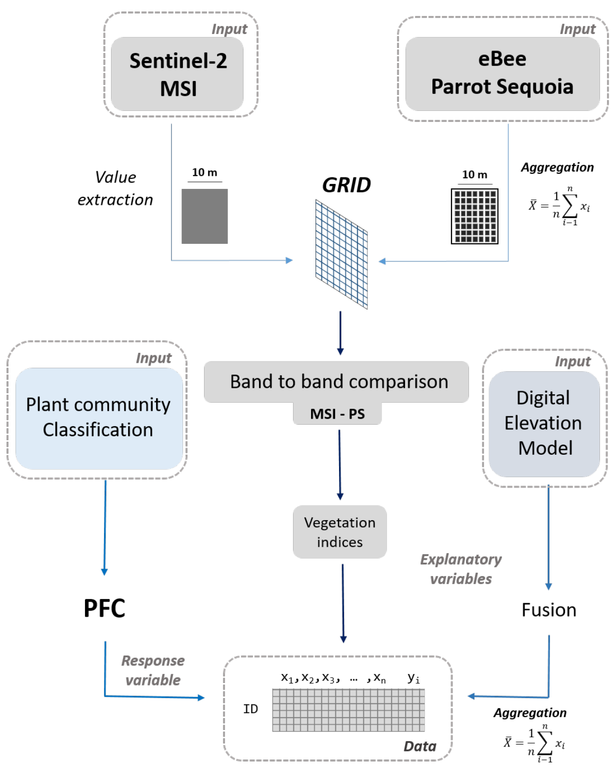

Figure 3. General workflow. The source data is marked as "Input" and the output Data Frame is DI. The final data frame contains the explanatory variables (x_n) and response variable (y_i) of Plant Fractional Cover (PFC), where i = Lower Shore (LS), Open Pioneer (OP), Upper Shore (US), Tall Grass (TG), Reed Swamp (RS).

PFC was calculated within each MSI pixel in the main data frame after an overlay process of the classification and MSI pixel extents. The Equation (1) was applied to each plant community of study.

$$PFC = \frac{Area\,of\,plant\,community\,within\,pixel\,extent}{Area\,of\,MSI\,pixel} \times 100 \tag{1}$$

All processes were carried out using the open source Python packages NumPy (Harris et al., 2020), GeoPandas (Jordahl et al., 2020) and rasterio (Gillies, 2013).

2.6 Vegetation indices

VI are quantitative and dimensionless mathematical combinations of spectral bands, related to vegetation structural properties (Lima-Cueto et al., 2019). VIs have been used to monitor vegetation cover by the enhancement of spectral contrast between photosynthetically active vegetation and other components (Andreatta et al., 2022). In this study, these band combinations may unveil vegetation patterns related to different levels of flooding and phenological activity, even though the flight dates correspond to the growing season and water presence is at its lowest in the study areas (Table 3). Because of variations in the amount of bare ground within each plant community (Table 2), VI are also used for their sensitiveness to this type of ground cover. In total, 14 VI were calculated (Table 3) from all the MSI bands of this study. The red edge MSI band was included in the calculations of VI because its reflectances show the highest photosynthetical activity and thus, better differentiation between plant communities (Schuster et al., 2012; Turpie, 2013). The indices in Table 5 were calculated by combining the features in the data frame (Figure 3) using the Pandas Python package (McKinney, 2010).

Table 5. List of fourteen vegetation indices used as explanatory variables in this study. G: Green band; R: Red band; Rre: Red Edge band; NIR: Near Infrarred band.

Vegetation Index	Calculation	Reference
Normalized Difference Vegetation Index	$NDVI = \frac{NIR-R}{NIR+R}$	Rouse et al. (1973)
Green Normalized Difference Vegetation Index	$GNDVI = \frac{NIR-G}{NIR+G}$	Gitelson et al. (1996)
Chlorophyll Vegetation Index	$CVI = \frac{NIR \times R}{G^2}$	Vincini et al. (2008)
Modified Simple Ratio (red edge)	$MSRred = \frac{(NIR/Rre)-1}{\sqrt{(NIR/Rre)+1}}$	Wu et al. (2008)
Red edge triangular vegetation index (core)	$RTVI_{\text{core}} = 100 \times (NIR - Rre) - 10 \times (NIR - G)$	Chen et al. (2010)
Canopy Chlorophyll Content Index	$CCCI = \frac{(NIR-Rre)/(NIR+Rre)}{(NIR-R)/(NIR+R)}$	Barnes et al. (2000)
Chlorophyll Index (red edge)	$CI_{\text{re}} = \frac{NIR}{Rre} - 1$	Gitelson et al. (2003)
Chlorophyll Index (green)	$CI_{\text{g}} = \frac{NIR}{G} - 1$	Merzlyak et al. (2003)
Red edge normalized difference vegetation index	$NDVI_{\text{re}} = \frac{NIR - Rre}{NIR + Rre}$	Gitelson and Merzlyak (1994)
Datt4	$datt_4 = \frac{R}{(G \times Rre))}$	Datt (1998)
Modified Green Red Vegetation Index	$MGRVI = \frac{(G^2-R^2)}{(G^2+R^2)}$	Bendig et al. (2015)
Modified Soil Adjusted Vegetation Index	$MGRVI = \frac{2 \times NIR+1-\sqrt{(2 \times NIR+1)^2-8 \times (NIR-R)}}{2}$	Qi et al. (1994)
Red Edge Ratio	$SR = \frac{NIR}{Rre}$	Gitelson and Merzlyak (1994)
Green-red vegetation index	$GRVI = \frac{G-R}{G+R}$	Chen et al. (2019)

2.7 Machine Learning models

A ML algorithm was chosen to build each PFC model because this approach has been successfully used in various ecological applications with Remote Sensing data (Olden et al., 2008; E Thessen, 2016). More specifically, the RF algorithm is widely accepted because of its high performance in modelling species occurrence and distribution with remote sensing data without making assumptions of data distribution (Evans et al., 2011; Shiferaw et al., 2019; Valavi et al., 2021). This algorithm was chosen to build ten regression models.

To build the training and test samples, a stratified sampling from the Initial Data Frame (DI, Figure 3) was carried out. The values of PFC were grouped in four bins created for this purpose ('0-25', '25-50', '50-75' and '75-100'). The bin '0-25' contained the majority of PFC values in all the plant communities, causing an imbalanced distribution. This is due to the presence of minimum cover or absence of plant communities in a large share of the grid cells. Imbalanced distributions are common in ecological data (Tang et al., 2023). To account for this, an under-sampling strategy was done by reducing the number of values in each of the bins to the number of values in the minority bin (Table 6).

Table 6. Balanced training dataset per plant community with the number of training rows considered in each bin and the proportion of all the bins in relation to the number of all Sentinel-2 MultiSpectral Instrument (MSI) pixels (9766). Plant communities are: Lower Shore (LS), Open Pioneer (OP), Upper Shore (US), Tall Grass (TG), Reed Swamp (RS).

Plant	0 - 25	25 - 50	50 - 75	75 - 100	Total	Proportion (%)
LS	823	823	823	823	3292	34
OP	178	178	178	178	712	7
US	1169	1169	1169	1169	4676	48
TG	711	711	711	711	2844	29
RS	100	100	100	100	400	4

This procedure balanced the number of rows per bin, avoiding overfitting of the models on the skewed bin of values. Two models were built for each plant community (LS, OP, US, TG and RS) from the Sampled Data Frame (DF0, Figure 4), one trained with the list of 14 VI as explanatory variables (Data Frame 1, DF1, Figure 4) and the other one adding the DEM to the explanatory variables (Data Frame 2, DF2, Figure 4).

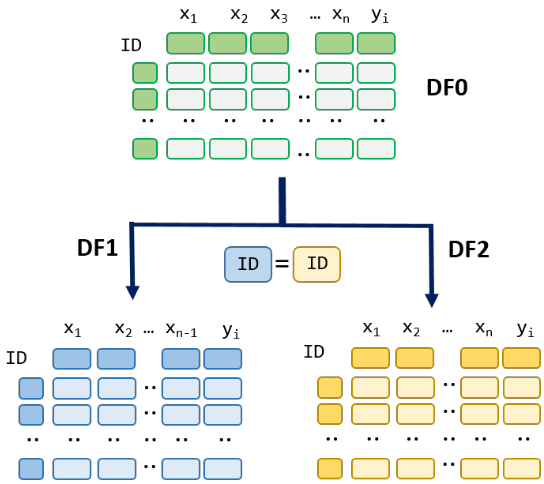

Figure 4. Diagram of the two training datasets used for the Random Forest (RF) models per plant community. DF0 is the sampled dataset after the under-sample strategy. DF1 has the same structure as DF0 except the DEM variable (x_{n-1}) and DF2 has all the explanatory variables (x_n). y_i is the response variable (Plant Fractional Cover, PFC), where i = Lower Shore (LS), Open Pioneer (OP), Upper Shore (US), Tall Grass (TG), Reed Swamp (RS). DF1 and DF2 have the same samples (rows) matching the Unique Identifier (ID) column derived from DF0.

A fraction of 80% was used to train the RF regression models with DF1 and DF2 (Figure 4). A Grid Search Cross-Validation strategy was implemented to search for the best hyperparameters and tune a RF model (Figure 5). This method iterates through a grid of predefined hyperparameters and tests the results with a 10-fold cross validation (Figure 5). The hyperparameters used to carry out the grid search approach were the number of estimators (N) and maximum features (MF) used to find the best split to grow each tree in the forest. The standard parameters for RF (Probst et al., 2019) were not used in this study because preliminary results did not show acceptable R^2 and RMSE scores on the training dataset. The remaining 20% of the samples were used to test the trained model with the best hyperparameters. Using this approach, training and testing dependencies are removed, ensuring the robustness of the final model. In order to compare the RF models of plant communities trained with each dataset (DF1 and DF2), R^2, RMSE and Mean Bias Error (MBE) metrics were reported to quantify deviations between actual and predicted PFC. To account for the contribution of each variable to the models, the variable importance was also extracted. Variable importance ranges from 0 to 1, indicating the contribution of each single variable to each of the tree's total impurity reduction. In the RF algorithm, the importance is calculated as the average of importance over all trees. The models with the best scores were used to predict the PFC of each plant community over the whole polygon grid (Figure 5). RF models were programmed using the package scikit-learn in Python (Pedregosa et al., 2018).

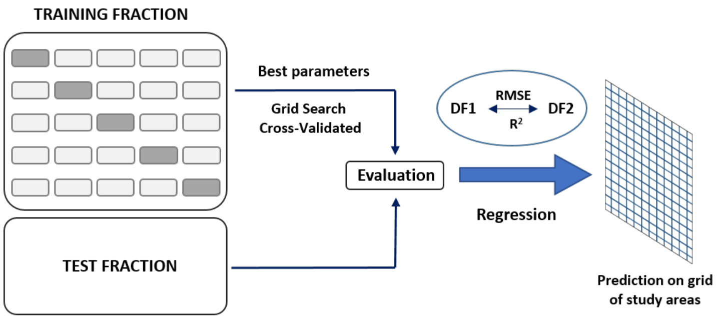

Figure 5. Machine learning algorithm training and testing process. A 10-fold cross validation on the Training fraction (80% of the input dataset) was used to search for the best hyperparameters for the Random Forest (RF) model and the 20% for Test fraction was used to test the trained model. The lowest Root Mean Square Error (RMSE) in the different RF models was used to predict the plant community distribution values on the polygon grid.

3 Results

3.1 Inter-sensor comparison

Figure 6 shows a quantitative comparison of spectral overlapping bands between MS and PS with R^2, RMSE and the significance level. Although the spectral resolution of PS and MSI sensors do not overlap completely (Table 4), the PS values aggregated by average into the MSI pixel show a significant positive correlation as well as low RMSE (Figure 6).

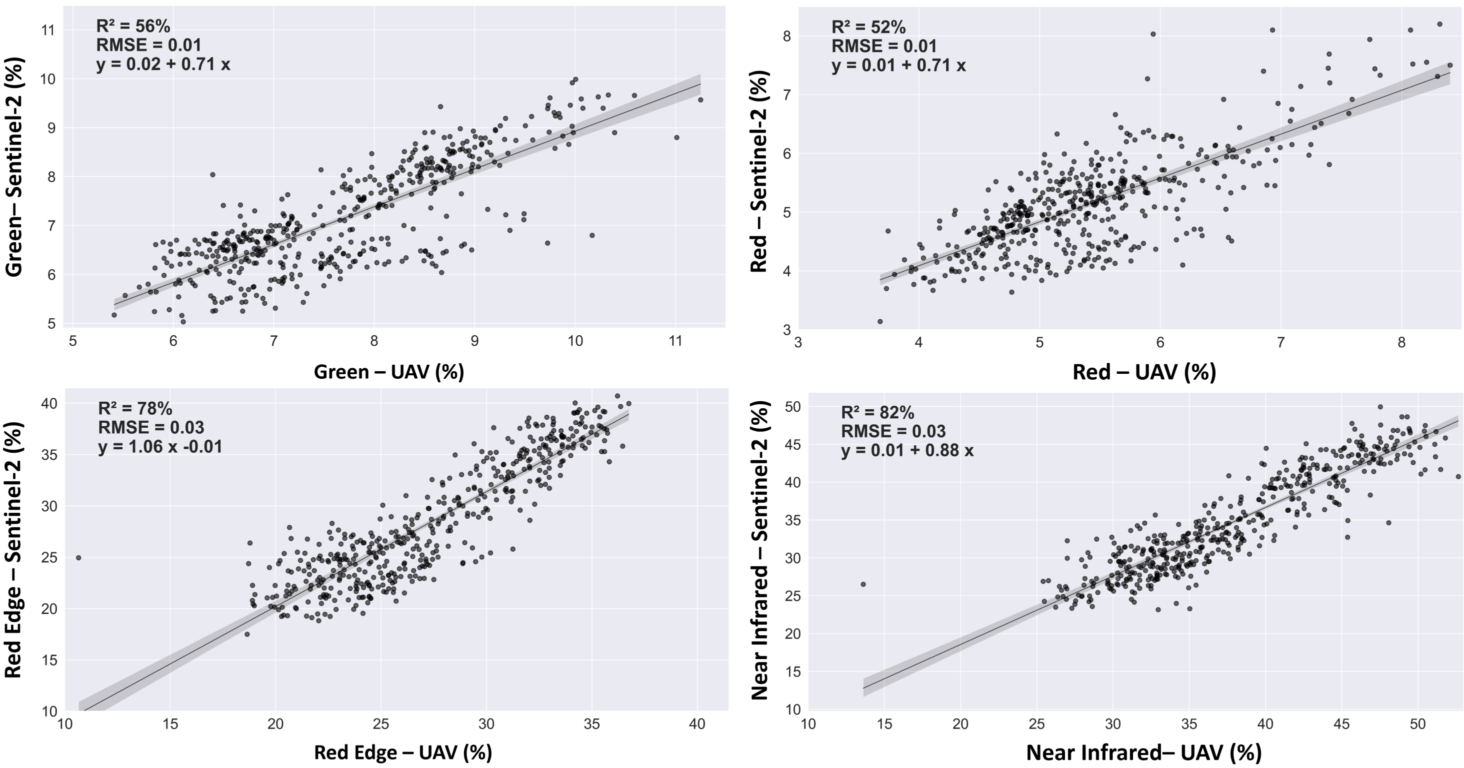

Figure 6. R^2 and RMSE obtained from the linear fitting between bands. X and y axes are in reflectance units (%) as well as RMSE. Correlations in all the cases are significant (p-value < 0.0001).

3.2 Random Forest regressions

There was a clear imbalance for categories of PFC between 0-25 % of cover of each plant community. This was taken into account and each RF model was finally built on different sizes of training datasets because they were set to the minimum size of categories used per distribution (Table 6).

The Grid Search Cross-Validation procedure enabled the selection of the best hyperparameters to build RF regression models with the lowest errors, leading to the identification of a minimum of 325 N. For models built on DF1, N was 500 except in RS (325) and OP (375) and 11 MF considered for the best split. Figure 7 shows the overall results of each RF regressor model with the best hyperparameters after the Grid Search in 10-fold cross validation. The models using only the VI calculated from MSI bands (DF1) show a R^2 score under 57% and RMSE above 22% (units of PFC), resulting in a moderate to low prediction capability. Having 48% of the total samples to train, the RF model of the US community performed the worst with DF1, followed by TG and LS communities, which also had a greater percentage of samples to train the models (29 and 34%, respectively, Figure 7). The models built on DF2 required more N, from 400 to 500 and the same MF. These models showed a higher performance, where R^2 scores increased on average 20 units and RMSE decreased 5% on average (Figure 7). The best improvement of RF models is in US because the model trained and tested with DF2 increased its R^2 by 2.15 times (Figure 7). The highest R^2 was achieved by the RF model of TG, reaching 70% after training and testing with DF2. Its RMSE decreased the most between models DF1 and DF2, from 27 to 19 % PFC. RF models trained and tested with DF1 and DF2 for LS, OP and RS show the lowest differences of R^2 and RMSE despite having 34%, 7% and 4% of the samples for training and testing.

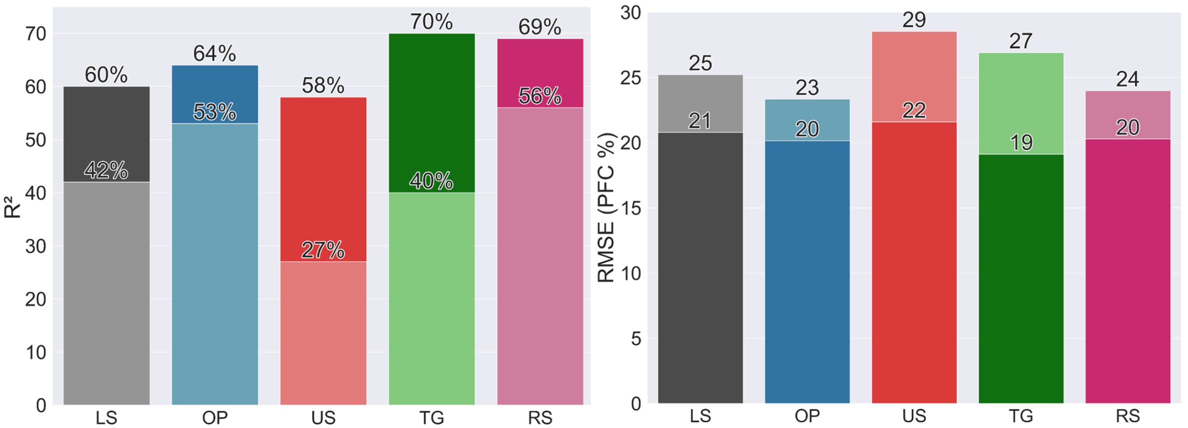

Figure 7. R^2 and Root Mean Squared Error (RMSE) retrieved by each Random Forest (RF) regressor on plant communities. Darker colours correspond to the model scores from Data Frame 2 (DF2) and lighter shades are model scores from Data Drame 1 (DF1). Plant communities are: Lower Shore (LS), Open Pioneer (OP), Upper Shore (US), Tall Grass (TG), Reed Swamp (RS).

Variable importance measured by the RF models to split the nodes did not show a common variable used in the models built with DF1 (Figure 10). Oppositely, the models trained with DF2 show the DEM as a common important explanatory variable used to split the nodes, except for the model to predict PFC of OP (Figure 11). Because a higher variance was explained by the models trained with DF2 their lower RMSE values, they were used to predict the distribution of PFC in the whole dataset (Figure 12).

The prediction errors in the RF models show a scattered distribution between the predicted and real PFC (Figures 8 and 9). In general, models tend to overestimate PFC below 50% of the real value and underestimate above it, according to the differences between the best fit line of the point distribution and the identity line (perfect prediction). This is more evident on the extreme values, around 0 and 100 PFC (Figures 8 and 9). These results improved in models on DF2, which showed the best-fit line closer to the identity line than those on DF1. The MBE metric indicated that the models of LS, US and TG under and overestimate the same way, either trained with DF1 or DF2 (Figures 8 and 9). On the contrary, the models of OP and RS underestimated the predicted values mostly and did not show any improvement.

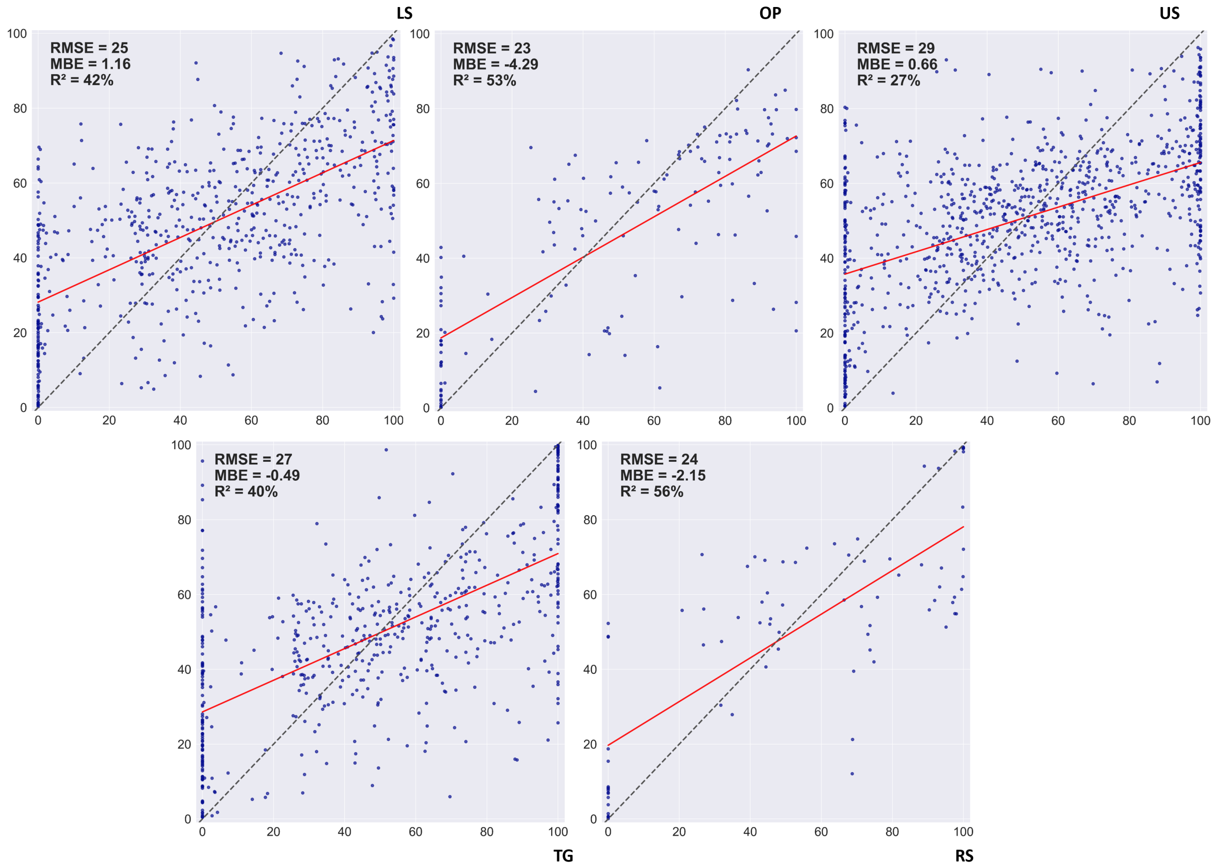

Figure 8. Prediction errors per plant community derived from Random Forest (RF) regressions with Data Frame 1 (DF1). On the x axis, actual values of Plant Fractional Cover (PFC, %) and on the y axis, predicted values of PFC (%). Black dotted lines show the best fit estimated from the correlation between the predicted and measured value of the PFC (%). Red dotted lines represent the over or under estimation of the predictions with its quantification with Mean Biased Error (MBE) and Root Mean Squared Error (RMSE).

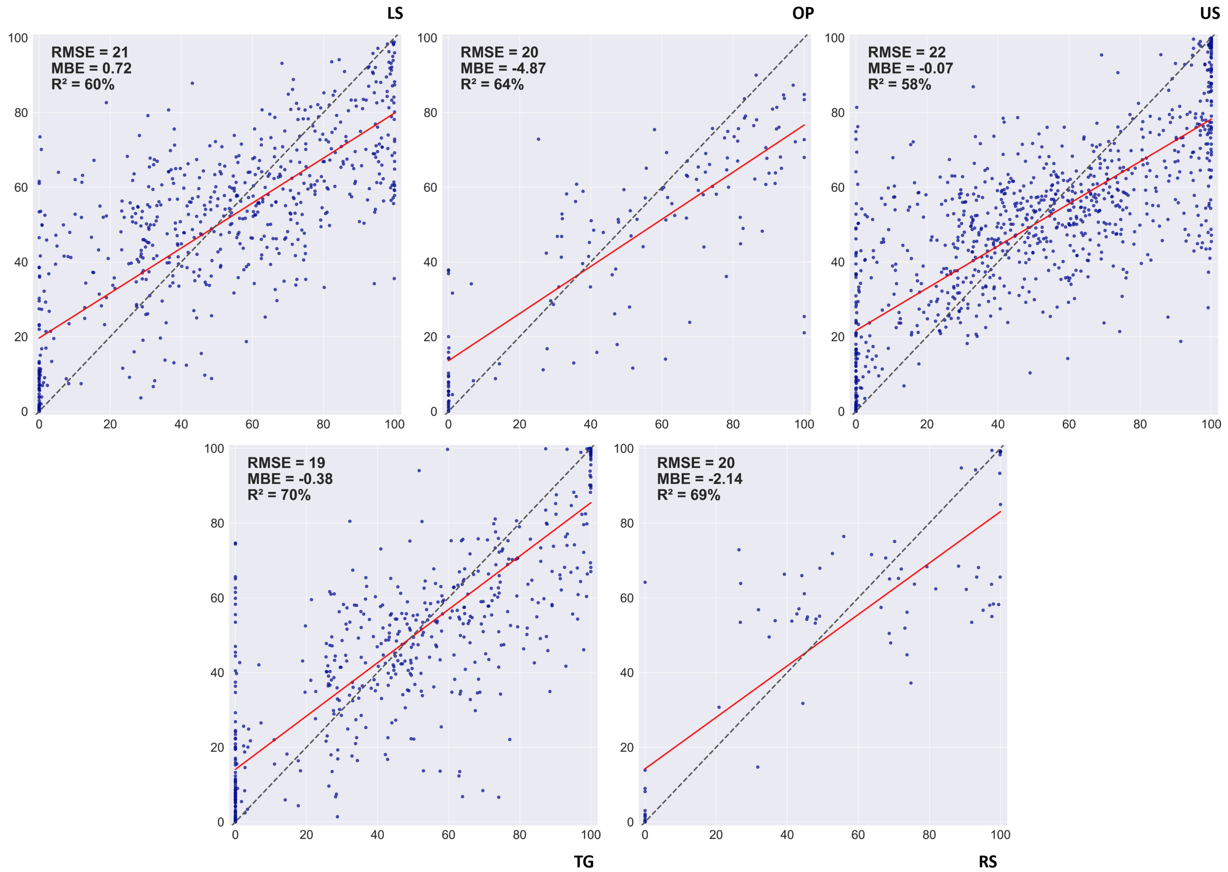

Figure 9. Prediction errors per plant community derived from Random Forest (RF) regressions with DF2. On the x axis, actual values of Plant Fractional Cover (PFC, %) and on the y axis, predicted values of PFC (%). Black dotted lines show the best fit estimated from the correlation between the predicted and measured value of the PFC (%). Red dotted lines represent the over or under estimation of the predictions with its quantification with Mean Biased Error (MBE) and Root Mean Squared Error (RMSE)

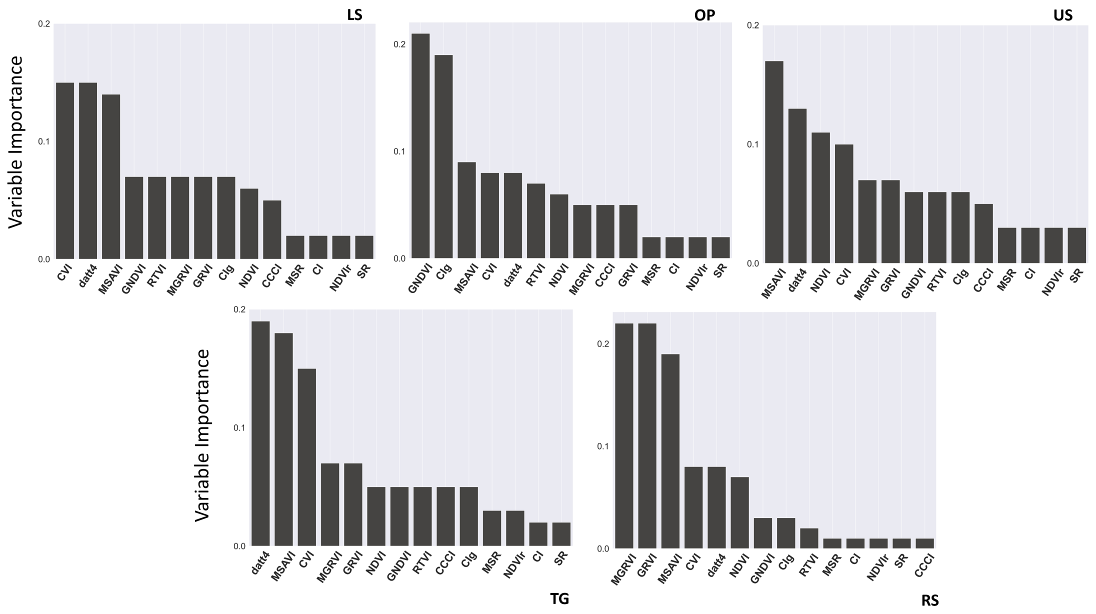

Figure 10. Variable importance retrieved by the Random Forest models derived from DF1.

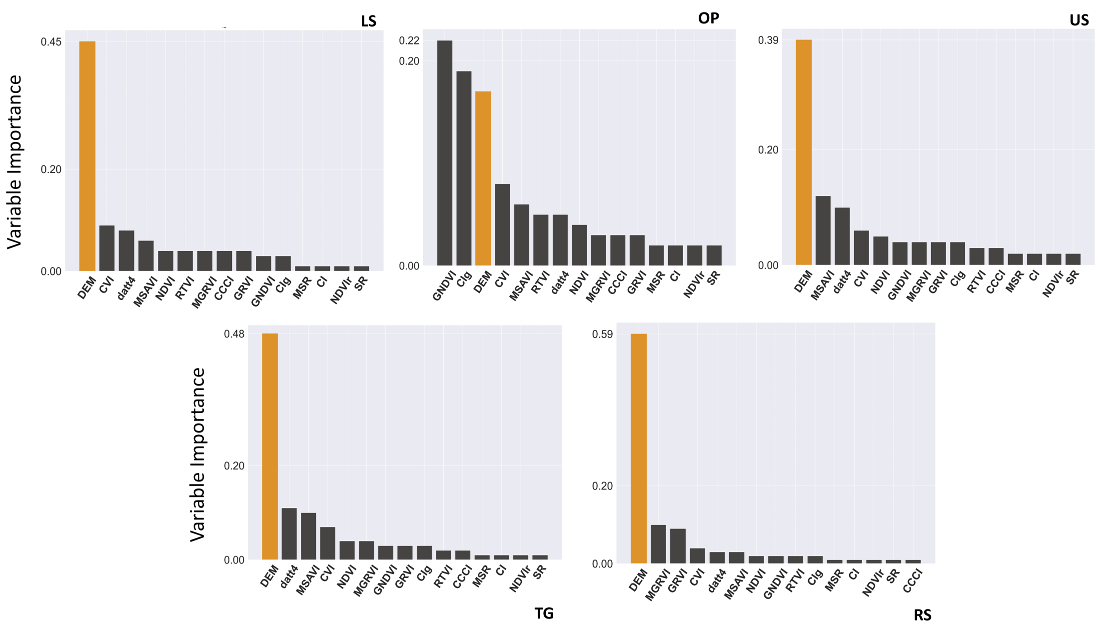

Figure 11. Variable importance retrieved by the Random Forest models derived from DF2, where the coloured bar represents the explanatory variable, Digital Elevation Model (DEM)

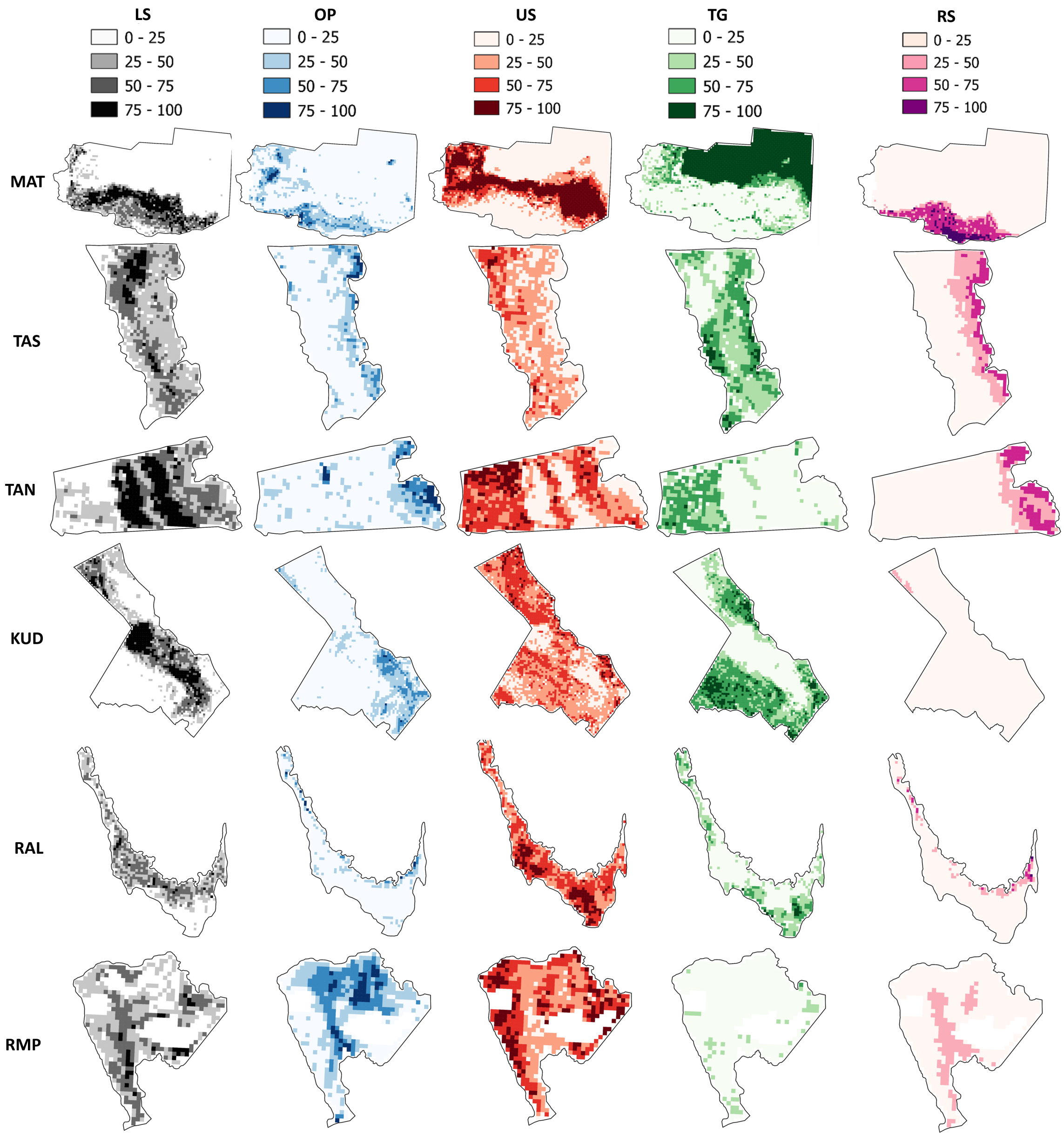

Figure 12. Maps of predicted Plant Fractional Cover (PFC ,%)for each plant community within the study areas. Matsalu (MAT), Tahu South (TAS), Tahu North (TAN), Kudani (KUD), Rälby (RAL) and Rumpo (RMP).

4 Discussion

This study is one attempt to model the distribution of five coastal wetland plant communities belonging to the formal phytosociological categorization of Burnside et al., 2007, using open data from MSI sensor on board of Sentinel-2 and the official DEM of Estonia. A fine plant community classification within the study areas derived from the spectral bands of a PS camera (Martínez Prentice et al., 2021) was the reference to calculate the distribution of each plant community.

Firstly, the spatial aggregation by average mean of the PS images from 10 cm to 10 m gave coherent similarities with the values from MSI imagery at Level 2A after finding significant relationships between PS and MSI bands with a linear fit (Figure 6), having an average of 33% of unexplained variance in the relationships. The red and green bands in MSI data display weaker linear relationships when compared to the red edge and near-infrared bands. This discrepancy can be explained by the lower reflectance values in the visible spectrum within MSI, which might be influenced by the presence of a mixture of vegetation and water within the pixel. The PS data captured more reflectance of bare soil, contributing to higher reflectance values in red and green bands. Higher relationships are observed between red edge and near infrared bands from both sensors (Figure 6). This is because these bands capture strong reflectance signals from vegetation. Similar studies compared reflectance values of PS (Díaz-Delgado et al., 2019), or another sensor on board of UAV (Padró et al., 2018), with MSI images, finding good relationships.

Secondly, the RF algorithm used in this study assessed the accuracy of non-parametric based regression models to predict the distribution of plant communities, as commonly used in Earth Observation studies (Ferreira et al., 2022), due to their robustness reported in the literature, and its use to generate distribution of plant communities at large scales (Maxwell et al., 2018; Butler and Sanderson, 2022). Its performance was assessed in this study. The models improve after the fusion of a high-resolution DEM (Figure 7). However, the mixture of reflectance signals included in one MSI pixel might have caused the deviations in the predictions (Figures 8 and 9). Overestimations of PFC in the models might be due to presence of patches where plant communities were mown or trampled, thus, retrieving VI values near to bare ground values. Underestimation, on the other hand, is due to the mixture of reflectance responses from different plant communities within the same spatial distribution of MSI pixels. Martínez Prentice et al (2021) suggest that the presence of disagreement areas is due to a mixture of radiances in transitional areas or ecotones. The effect of mixed pixels is more significant in these types of transitional areas, as the sensor receives a wide range of reflectance signals within the extent of the pixel (Muukkonen and Heiskanen, 2007). The models of OP and RS with DF1 and DF2 underestimate the PFC to a larger degree. The reflectance values retrieved from these plant communities are affected by a higher presence of water, reducing the values of VI due to lower reflectance in the Near Infrared and Red Edge spectrum.

The decrease of RMSE of predicted PFC (Figure 7) suggests that plant communities follow the elevation pattern at broader scales corresponding to the variations the microtopography in the study areas. Although the spatial resolution of the DEM in this study was 1 m and aggregated to 10 m, it showed a similar predictive efficiency as those generated with the photogrammetric point cloud derived from UAV at a resolution below 10 cm (Villoslada et al., 2020). This is also shown in the variable importances derived from RF models using DF1 and DF2 as shown in Figures 10 and 11. MSAVI was the only common variable among all the models trained and tested with DF1. Once the DEM was included as an explanatory variable in DF2, the RF models showed this as the most important variable, except for the RF model of OP, where the index GNDVI is the most important. OP is a type of plant community that is distributed over patches with a high proportion of moist and bare ground (Bergamo et al., 2022), where the visible part of the spectrum (Red and Green) retrieve greater reflectance values than the far visible part of the spectrum (Red Edge and Near Infrared). Ward et al., 2013 concluded that the presence of OP occurs at similar elevation ranges than TG by predicting with a microtopography variable. Thus, building the models with the DEM alone do not reach the metrics as with the VI, because they are indicators of vegetation corresponding to plant communities.

The final plant community distribution maps match the common patterns recognised by expert knowledge in the study areas and the maps shown in Martínez Prentice et al. 2021 at a 10 cm spatial resolution (Figure 12). According to these criteria, the PFC maps in MAT are the most representative of PFC (Figure 12). Some plant communities that were not identified in most of the study areas (Table A1) are present with high PFC values. This is the case of RS because its distribution along the coastline shows similar low values of DEM and VI due to the higher inundation levels of brackish water (Table 2). It is largely missing in KUD mostly, as this study area is relatively far from to the coastline (Figure 1). The plant community of TG is overestimated in TAN and RMP areas where it was not identified, caused by similarities in VI and DEM with other plant communities, which means similar biogeographical factors than in other study areas where it is actually present. The distribution of LS, OP and US are mostly correct according to the aformentioned criteria. OP presents a similar overestimation as RS of its PFC in RAL area only, due to the same reason concerning RS.

This study shows an acceptable empirical approximation to modelling the distribution of ESs providing units in large-scale images of coastal meadows of Estonia, this is, quantifying the prediction accuracy and error of PFC of small sized plant communities in heterogeneous ecosystems at a satellite spatial resolution. The main advantage of using the publicly available data from Sentinel-2 constellation is its short revisit time, provided by the two satellites, Sentinel-2 A and B. The methodology provides a rapid assessment of plant communities in a coastal ecosystem vulnerable to climate and land use changes using different sources of remotely sensed data. Additionally, it is shown that it is possible to reduce time and costs associated with multiple UAV flights in different areas to cover large extents by the validation of large-scale monitoring studies with open source satellite data such as Sentinel-2 and high-resolution products derived from multispectral images taken from UAV.

Upscaling remotely sensed imagery from fine to coarse resolution is necessary, although challenging. Satellite imagery provides critical information of changes needed for improved environmental management and conservation decision making at large scales. Considering this, linking UAV to Earth Observation satellites offers the opportunity for multiscale study of environmentally sensitive ecosystems such as coastal wetlands. Further work can consider using ancillary data as a co-predictor with the aggregated spectral data, such as temperature, pluviometry or distance of plant communities from the coast, to improve the prediction accuracies, as shown with DEM data in the present work. The supervised learning RF algorithm is one of the most robust ML algorithms used for ecosystem and species distribution modelling (Pichler and Hartig, 2023), however, other algorithms should be explored. Moreover, recent advances in Super-Resolution methodologies increase the spatial resolution of Sentinel-2 images by four times for all bands, at a maximum of 2.5m using Artificial Intelligence algorithms (Tarasiewicz et al., 2023). Finer scale analysis will be more suitable to study the heterogeneity of plant communities in Boreal Baltic coastal meadows.

5 Conclusions

A multiscale synergy approach between UAV and Sentinel-2 MSI was undertaken in this study to model the PFC of five plant communities in coastal meadows. Good relationships existed between both sensors, which enabled PFC to be modelled using VIs, although the fusion of DEM improved the models from 1.2 to 2 times. From this research, future studies on coastal meadows using remote sensing from satellites should be focused on finding methods to achieve local calibration of the image based on values retrieved from UAV mounted multispectral cameras and thus, achieve stronger synergies between both sensors (Emilien et al, 2021). Due to the high repeat time and long duration of data collection, upscaling from UAV to satellite imagery provides an excellent resource for monitoring and assessment of the response of coastal ecosystems to loss and degradation as a result of climate change or other anthropogenic stressors. This will allow land users and managers to appropriately assess conservation priorities and implement and monitor responses.

Appendix A

A1 Additional figures and tables

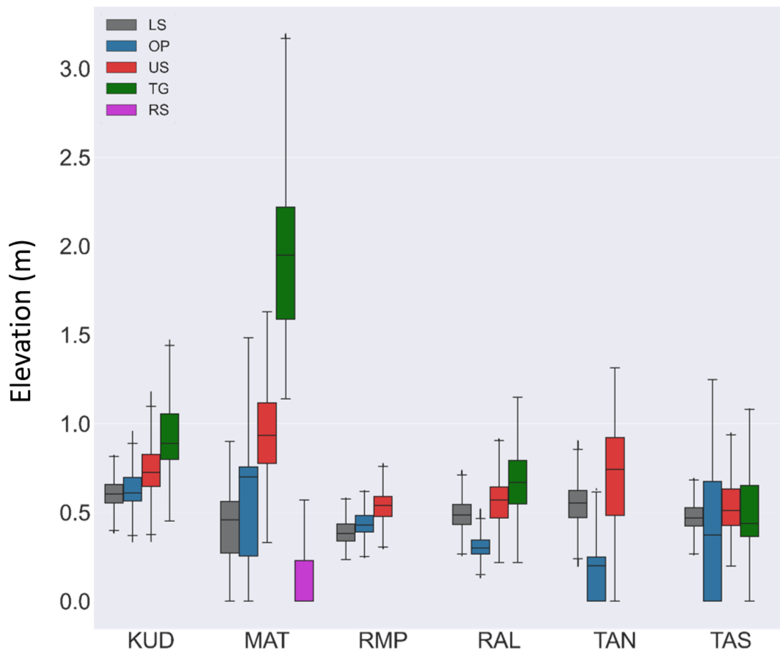

Figure A1. Boxplot of elevation ranges (m) per category of plant communities in each study area, showing the microtopography gradient within the established plant communities.

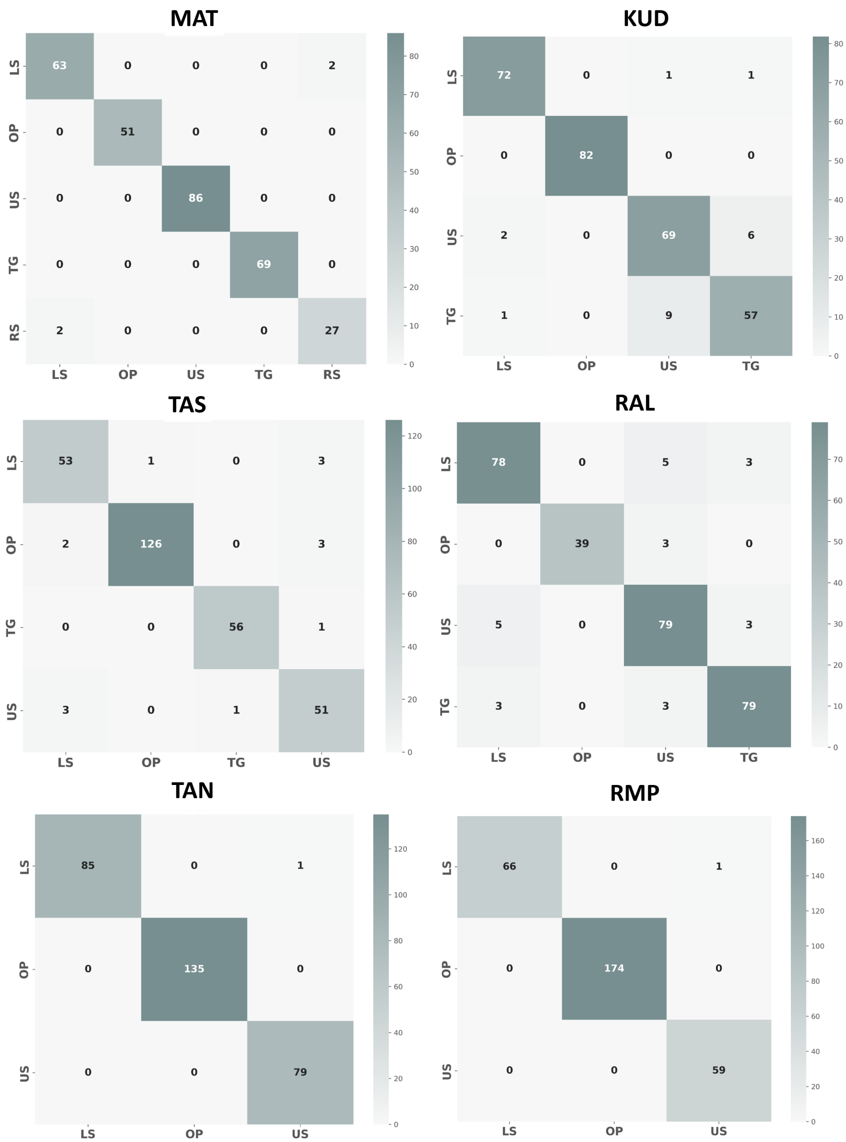

Figure A2. Confusion Matrix from the results of Random Forest pixel classification in Martinez Prentice et al., 2021. MAT (Matsalu), KUD (Kudani), TAS (Tahu South), RAL (Rälby), TAN (Tahu North) and Rumpo (RMP). Kappa values are MAT: 0.98, KUD: 0.92, TAS: 0.93, RAL: 0.89, TAN: 0.99 and RMP: 0.99. Each class of Predicted and Actual Plant Communities are LS (Lower Shore), OP (Open Pioneer), US (Upper Shore), TG (Tall Grassland) and RS (Reed Swamp). Numbers are the pixels classified in each quadrat

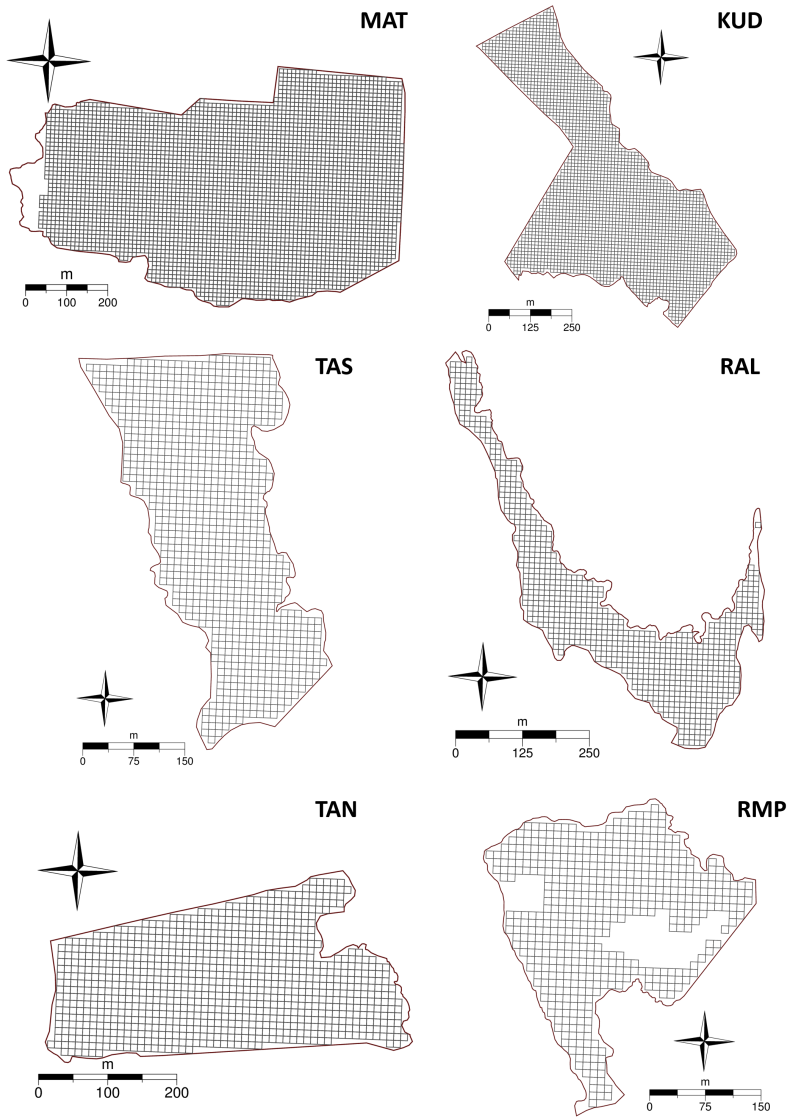

Figure A3. Polygon grids from Sentinel-2 MultiSpectral Instrument (MSI) pixels (9766) covering the six study areas. 1. Matsalu (MAT), 2. Tahu South (TAS), 3. Tahu North (TAN), 4. Kudani (KUD), 5. Rälby (RAL), 6. Rumpo (RMP)

Table A1. Plant communities sampled in each study area. Matsalu (MAT), Tahu South (TAS), Tahu North (TAN), Kudani (KUD), Rälby (RAL) and Rumpo (RMP); Lower Shore (LS), Open Pioneer (OP), Upper Shore (US), Tall Grass (TG), Reed Swamp (RS).

Study Area	Plant Communities
MAT	LS, OP, US, TG, RS
TAS	LS, OP, TG, US
TAN	LS, OP, US
KUD	LS, OP, TG, US
RAL	LS, OP, TG, US
RMP	LS, OP, US

Author contributions. Conceptualization, R.M.P., M.V.P. and R.D.W.; methodology, R.M.P., C.B.J., M.V.P., T.F.B. and R.D.W.; software, R.M.P. and M.V.P.; validation, R.M.P., M.V.P. and R.D.W.; formal analysis, R.M.P.; investigation, R.M.P., M.V.P., T.F.B., C.B.J. and R.D.W.; resources, R.D.W. and K.S.; data curation, R.M.P., M.V.P. and R.D.W.; writing—original draft preparation, R.M.P., M.V.P. and R.D.W.; writing—review and editing, R.M.P., M.V.P., R.D.W. and C.B.J.; visualization, R.M.P.; supervision, M.V.P., R.D.W. and K.S.; project administration, R.D.W. and K.S.; funding acquisition, R.D.W. and K.S. All authors have read and agreed to the published version of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. This research was funded by the Doctoral School of Earth Sciences and Ecology, financed by the European Union, European Regional Development Fund (Estonian University of Life Sciences ASTRA project “Value-chain based bio-economy”.

Moreover, the authors would like to thank the anonymous reviewers for their valuable comments and critical questions, which helped to improve the quality of the present manuscript

References

Adam, E., Mutanga, O., and Rugege, D.: Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review, Wetlands Ecology and Management, 18, 281–296, https://doi.org/10.1007/s11273-009-9169-z, 2010.

Andreatta, D., Gianelle, D., Scotton, M., and Dalponte, M.: Estimating grassland vegetation cover with remote sensing: A comparison between Landsat-8, Sentinel-2 and PlanetScope imagery, Ecological Indicators, 141, 109 102, https://doi.org/https://doi.org/10.1016/j.ecolind.2022.109102, 2022.

Barnes, E., Clarke, T., Richards, S., Colaizzi, P., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., and Thompson, T.: Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data, in: Proceedings of the 5th International Conference on Precision Agriculture and other resource management, Bloomington, MN USA, 2000.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., and Bareth, G.: Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley, International Journal of Applied Earth Observation and Geoinformation, 39, 79–87, https://doi.org/https://doi.org/10.1016/j.jag.2015.02.012, 2015.

Berg, M., Joyce, C., and Burnside, N.: Differential responses of abandoned wet grassland plant communities to reinstated cutting management, Hydrobiologia, 692, 83–97, https://doi.org/10.1007/s10750-011-0826-x, 2012.

Berg, M. J.: Abandonment and Reinstated Management upon Coastal Wet Grasslands in Estonia, Phd thesis, University of Brighton, 2008.

Bergamo, T. F., Ward, R. D., Joyce, C. B., Villoslada, M., and Sepp, K.: Experimental climate change impacts on Baltic coastal wetland plant communities, Scientific Reports, 12, 20 362, https://doi.org/10.1038/s41598-022-24913-z, 2022.

Bergamo, T. F., de Lima, R. S., Kull, T., Ward, R. D., Sepp, K., and Villoslada, M.: From UAV to PlanetScope: Upscaling fractional cover of an invasive species Rosa rugosa, Journal of Environmental Management, 336, 117 693, https://doi.org/https://doi.org/10.1016/j.jenvman.2023.117693, 2023.

Blan, L. and Butler, R.: Comparing Effects of Aggregation Methods on Statistical and Spatial Properties of Simulated Spatial Data, Photogrammetric Engineering and Remote Sensing, 65, 73–84, https://api.semanticscholar.org/CorpusID:131331684, 1999.

Brodu, N.: Super-Resolving Multiresolution Images With Band-Independent Geometry of Multispectral Pixels, IEEE Transactions on Geoscience and Remote Sensing, 55, 4610–4617, https://doi.org/10.1109/TGRS.2017.2694881, 2017.

Burkhard, B., Kroll, F., Nedkov, S., and Müller, F.: Mapping ecosystem service supply, demand and budgets, Ecological Indicators, 21, 17–29, https://doi.org/https://doi.org/10.1016/j.ecolind.2011.06.019, challenges of sustaining natural capital and ecosystem services, 2012.

Burnside, N., Joyce, C., Puurmann, E., and Scott, D.: Use of vegetation classification and plant indicators to assess grazing abandonment in Estonian coastal wetlands, Journal of Vegetation Science, 18, 645–654, https://doi.org/10.1111/j.1654-1103.2007.tb02578.x, 2007.

Butler, L. and Sanderson, R. A.: National-scale predictions of plant assemblages via community distribution models: Leveraging published data to guide future surveys, Journal of Applied Ecology, 59, 1559–1571, https://doi.org/https://doi.org/10.1111/1365-2664.14166, 2022.

Chen, A., Orlov-Levin, V., and Meron, M.: Applying high-resolution visible-channel aerial imaging of crop canopy to precision irrigation management, Agricultural Water Management, 216, 196–205, https://doi.org/https://doi.org/10.1016/j.agwat.2019.02.017, 2019.

Chen, P.-F., Nicolas, T., Wang, J.-H., Philippe, V., Huang, W.-J., and Li, B.-G.: New index for crop canopy fresh biomass estimation, Spectroscopy and Spectral Analysis, 30, 512–517, 2010.

Colomina, I. and Molina, P.: Unmanned aerial systems for photogrammetry and remote sensing: A review, ISPRS Journal of Photogrammetry and Remote Sensing, 92, 79–97, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2014.02.013, 2014.

Copernicus Hub: European Space Agency (ESA). Copernicus Open Access Hub., https://scihub.copernicus.eu/dhus/#/home, accessed: 2022-01-30.

Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V. E., Strasser, T., Vanden Borre, J., Toon, S., and Michael, F.: Remote sensing for mapping natural habitats and their conservation status – New opportunities and challenges, International Journal of Applied Earth Observation and Geoinformation, 37, 7–16, https://doi.org/https://doi.org/10.1016/j.jag.2014.11.005, special Issue on Earth observation for habitat mapping and biodiversity monitoring, 2015.

Courouble, M.: Global Wetland Outlook: Special Edition 2021, Secretariat of the Convention on Wetlands, Gland, Switzerland, 2021.

Cracknell, A. P.: UAVs: regulations and law enforcement, International Journal of Remote Sensing, 38, 3054–3067, https://doi.org/10.1080/01431161.2017.1302115, 2017.

Datt, B.: Remote Sensing of Chlorophyll a, Chlorophyll b, Chlorophyll a+b, and Total Carotenoid Content in Eucalyptus Leaves, Remote Sensing of Environment, 66, 111–121, https://doi.org/https://doi.org/10.1016/S0034-4257(98)00046-7, 1998.

Davidson, N.: How much wetland has the world lost? Long-term and recent trends in global wetland area, Marine and Freshwater Research, 65, 936–941, https://doi.org/10.1071/MF14173, 2014.

De Simone, W., Allegrezza, M., Frattaroli, A. R., Montecchiari, S., Tesei, G., Zuccarello, V., and Di Musciano, M.: From Remote Sensing to Species Distribution Modelling: An Integrated Workflow to Monitor Spreading Species in Key Grassland Habitats, Remote Sensing, 13, https://doi.org/10.3390/rs13101904, 2021.

Diekmann, M.: Species indicator values as an important tool in applied plant ecology – a review, Basic and Applied Ecology, 4, 493–506, https://doi.org/https://doi.org/10.1078/1439-1791-00185, 2003.

Díaz-Delgado, R., Cazacu, C., and Adamescu, M.: Rapid Assessment of Ecological Integrity for LTER Wetland Sites by Using UAV Multispectral Mapping, Drones, 3, https://doi.org/10.3390/drones3010003, 2019.

Emilien, A.-V., Thomas, C., and Houet, T.: UAV & Satellite synergies for optical remote sensing applications: a literature review, Science of Remote Sensing, 3, 100 019, https://doi.org/10.1016/j.srs.2021.100019, 2021.

ESA: SNAP.ESA Sentinel Application Platform v9, http://step.esa.int., 2014.

Evans, J. S., Murphy, M. A., Holden, Z. A., and Cushman, S. A.: Modeling Species Distribution and Change Using Random Forest, pp. 139–159, Springer New York, New York, NY, https://doi.org/10.1007/978-1-4419-7390-0_8, 2011.

E Thessen, A.: Adoption of Machine Learning Techniques in Ecology and Earth Science, One Ecosystem, 1, e8621, https://doi.org/10.3897/oneeco.1.e8621, 2016.

Fawcett, D., Panigada, C., Tagliabue, G., Boschetti, M., Celesti, M., Evdokimov, A., Biriukova, K., Colombo, R., Miglietta, F., Rascher, U., and Anderson, K.: Multi-Scale Evaluation of Drone-Based Multispectral Surface Reflectance and Vegetation Indices in Operational Conditions, Remote Sensing, 12, https://doi.org/10.3390/rs12030514, 2020.

Fernández-Guisuraga, J. M., Sanz-Ablanedo, E., Suárez-Seoane, S., and Calvo, L.: Using Unmanned Aerial Vehicles in Post-fire Vegetation Survey Campaigns through Large and Heterogeneous Areas: Opportunities and Challenges, Sensors, 18, https://doi.org/10.3390/s18020586, 2018.

Ferreira, B., Silva, R. G., and Iten, M.: Earth Observation Satellite Imagery Information Based Decision Support Using Machine Learning, Remote Sensing, 14, https://doi.org/10.3390/rs14153776, 2022.

Findell, K. L., Berg, A., Gentine, P., Krasting, J. P., Lintner, B. R., Malyshev, S., Santanello, J. A., and Shevliakova, E.: The impact of anthropogenic land use and land cover change on regional climate extremes, Nature Communications, 8, 989, https://doi.org/10.1038/s41467-017-01038-w, 2017.

Gardner, T. A., Barlow, J., Chazdon, R., Ewers, R. M., Harvey, C. A., Peres, C. A., and Sodhi, N. S.: Prospects for tropical forest biodiversity in a human-modified world, Ecol Lett, 12, 561–582, 2009.

Gedan, K., Kirwan, M., Wolanski, E., Barbier, E., and Silliman, B.: The Present and Future Role of Coastal Wetland Vegetation in Protecting Shorelines: Answering Recent Challenges to the Paradigm, Climatic Change, 106, 7–29, https://doi.org/10.1007/s10584-010-0003-7, 2011.

Gillies, S.: Rasterio: geospatial raster I/O for Python programmers, https://github.com/mapbox/rasterio, 2013.

Gitelson, A. and Merzlyak, M. N.: Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation, Journal of Plant Physiology, 143, 286–292, https://doi.org/https://doi.org/10.1016/S0176-1617(11)81633-0, 1994.

Gitelson, A., Viña, A., Arkebauer, T., Rundquist, D., Keydan, G., Leavitt, B., and Keydan, G.: Remote estimation of leaf area index and green leaf biomass in maize canopies, Geophysical Research Letters, 30, https://doi.org/10.1029/2002GL016450, 2003.

Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N.: Use of a green channel in remote sensing of global vegetation from EOS-MODIS, Remote Sensing of Environment, 58, 289–298, https://doi.org/https://doi.org/10.1016/S0034-4257(96)00072-7, 1996.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, Nature, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Heil, J., Jörges, C., and Stumpe, B.: Fine-Scale Mapping of Soil Organic Matter in Agricultural Soils Using UAVs and Machine Learning, Remote Sensing, 14, https://doi.org/10.3390/rs14143349, 2022.

Hopkinson, C. S., Cai, W.-J., and Hu, X.: Carbon sequestration in wetland dominated coastal systems—a global sink of rapidly diminishing magnitude, Current Opinion in Environmental Sustainability, 4, 186–194, https://doi.org/https://doi.org/10.1016/j.cosust.2012.03.005, 2012.

Isenburg, M., Liu, Y., Shewchuk, J., Snoeyink, J., and Thirion, T.: Generating Raster DEM from Mass Points Via TIN Streaming, pp. 186–198, https://doi.org/10.1007/11863939_13, 2006.

Isgró, M. A., Basallote, M. D., Caballero, I., and Barbero, L.: Comparison of UAS and Sentinel-2 Multispectral Imagery for Water Quality Monitoring: A Case Study for Acid Mine Drainage Affected Areas (SW Spain), Remote Sensing, 14, https://doi.org/10.3390/rs14164053, 2022.

Jiang, J., Johansen, K., Tu, Y.-H., and McCabe, M.: Multi-sensor and multi-platform consistency and interoperability between UAV, Planet CubeSat, Sentinel-2, and Landsat reflectance data, GIScience & Remote Sensing, 59, 936–958, https://doi.org/10.1080/15481603.2022.2083791, 2022.

Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L. J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, and Leblanc, F.: geopandas/geopandas: v0.8.1, https://doi.org/10.5281/zenodo.3946761, 2020.

Knight, J. F., Lunetta, R. S., Ediriwickrema, J., and Khorram, S.: Regional Scale Land Cover Characterization Using MODIS-NDVI 250 m Multi-Temporal Imagery: A Phenology-Based Approach, GIScience & Remote Sensing, 43, 1–23, https://doi.org/10.2747/1548-1603.43.1.1, 2006.

Kont, A., Jaagus, J., and Aunap, R.: Climate change scenarios and the effect of sea-level rise for Estonia, Global and Planetary Change, 36, 1–15, https://doi.org/https://doi.org/10.1016/S0921-8181(02)00149-2, 2003.

Laliberte, A. S., Goforth, M. A., Steele, C. M., and Rango, A.: Multispectral Remote Sensing from Unmanned Aircraft: Image Processing Workflows and Applications for Rangeland Environments, Remote Sensing, 3, 2529–2551, https://doi.org/10.3390/rs3112529, 2011.

Lima-Cueto, F. J., Blanco-Sepúlveda, R., Gómez-Moreno, M. L., and Galacho-Jiménez, F. B.: Using Vegetation Indices and a UAV Imaging Platform to Quantify the Density of Vegetation Ground Cover in Olive Groves (Olea Europaea L.) in Southern Spain, Remote Sensing, 11, https://doi.org/10.3390/rs11212564, 2019.

Luck, G., Daily, G., and Ehrlich, P.: Population diversity and ecosystem services, Trends in Ecology & Evolution, 18, 331–336, https://doi.org/10.1016/S0169-5347(03)00100-9, 2003.

Maa-amet geoportaal: Eesti Maa-amet (Estonian Land Board), https://geoportaal.maaamet.ee/, accessed: 2022-03-21.

Maasing, V. and Paal, J.: Estonian wetlands- classification and biodiversity, Svensk Botanisk Tidskrift, 92, 147–161, 1998.

Magurran, A. E.: Ecological Diversity and Its Measurement, Princeton University Press, Princeton, New Jersey, 1988.

Mahdianpari, M., Granger, J. E., Mohammadimanesh, F., Salehi, B., Brisco, B., Homayouni, S., Gill, E., Huberty, B., and Lang, M.: Meta-Analysis of Wetland Classification Using Remote Sensing: A Systematic Review of a 40-Year Trend in North America, Remote Sensing, 12, https://doi.org/10.3390/rs12111882, 2020.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., and Gascon, F.: Sen2Cor for Sentinel-2, in: Image and Signal Processing for Remote Sensing XXIII, edited by Bruzzone, L., vol. 10427, p. 1042704, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.2278218, 2017.

Mao, P., Ding, J., Jiang, B., Qin, L., and Qiu, G. Y.: How can UAV bridge the gap between ground and satellite observations for quantifying the biomass of desert shrub community?, ISPRS Journal of Photogrammetry and Remote Sensing, 192, 361–376, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2022.08.021, 2022.

Martínez Prentice, R., Villoslada Peciña, M., Ward, R. D., Bergamo, T. F., Joyce, C. B., and Sepp, K.: Machine Learning Classification and Accuracy Assessment from High-Resolution Images of Coastal Wetlands, Remote Sensing, 13, https://doi.org/10.3390/rs13183669, 2021.

Maurya, A., Nadeem, M., Singh, D., Singh, K., and Rajput, N.: Critical Analysis of Machine Learning Approaches for Vegetation Fractional Cover Estimation Using Drone and Sentinel-2 Data, pp. 343–346, https://doi.org/10.1109/IGARSS47720.2021.9554422, 2021.

Maxwell, A. E., Warner, T. A., and Fang, F.: Implementation of machine-learning classification in remote sensing: an applied review, International Journal of Remote Sensing, 39, 2784–2817, https://doi.org/10.1080/01431161.2018.1433343, 2018.

McKinney, W.: Data Structures for Statistical Computing in Python, pp. 56–61, https://doi.org/10.25080/Majora-92bf1922-00a, 2010.

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., Solovchenko, A. E., and Pogosyan, S. I.: Application of Reflectance Spectroscopy for Analysis of Higher Plant Pigments, Russian Journal of Plant Physiology, 50, 704–710, https://doi.org/10.1023/A:1025608728405, 2003.

Muukkonen, P. and Heiskanen, J.: Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories, Remote Sensing of Environment, 107, 617–624, https://doi.org/https://doi.org/10.1016/j.rse.2006.10.011, 2007.

Okolie, C. J. and Smit, J. L.: A systematic review and meta-analysis of Digital elevation model (DEM) fusion: pre-processing, methods and applications, ISPRS Journal of Photogrammetry and Remote Sensing, 188, 1–29, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2022.03.016, 2022.

Olden, J. D., Lawler, J. J., and Poff, N. L.: Machine learning methods without tears: a primer for ecologists, Q Rev Biol, 83, 171–193, 2008.

Padró, J.-C., Muñoz, F.-J., Avila, L. A., Pesquer, L., and Pons, X.: Radiometric Correction of Landsat-8 and Sentinel-2A Scenes Using Drone Imagery in Synergy with Field Spectroradiometry, Remote Sensing, 10, https://doi.org/10.3390/rs10111687, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Édouard Duchesnay: Scikit-learn: Machine Learning in Python, 2018.

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., and Turner, W.: Satellite remote sensing for applied ecologists: opportunities and challenges, Journal of Applied Ecology, 51, 839–848, https://doi.org/https://doi.org/10.1111/1365-2664.12261, 2014.

Pichler, M. and Hartig, F.: Machine learning and deep learning—A review for ecologists, Methods in Ecology and Evolution, 14, 994–1016, https://doi.org/10.1111/2041-210x.14061, 2023.

Probst, P., Wright, M. N., and Boulesteix, A.-L.: Hyperparameters and tuning strategies for random forest, WIREs Data Mining and Knowledge Discovery, 9, https://doi.org/10.1002/widm.1301, 2019.

Qi, J., Chehbouni, A., Huete, A., Kerr, Y., and Sorooshian, S.: A modified soil adjusted vegetation index, Remote Sensing of Environment, 48, 119–126, https://doi.org/https://doi.org/10.1016/0034-4257(94)90134-1, 1994.

Rannap, R., Briggs, L., Lotman, K., Lepik, I., Rannap, V., and Põdra, P.: Coastal Meadow Management—Best Practice Guidelines, vol. 4, Ministry of the Environment of the Republic of Estonia, 2004.

Riihimäki, H., Luoto, M., and Heiskanen, J.: Estimating fractional cover of tundra vegetation at multiple scales using unmanned aerial systems and optical satellite data, Remote Sensing of Environment, 224, 119–132, https://doi.org/https://doi.org/10.1016/j.rse.2019.01.030, 2019.

Rivis, R., Kont, A., Ratas, U., Palginomm, V., Antso, K., and Tõnisson, H.: Trends in the development of Estonian coastal land cover and landscapes caused by natural changes and human impact, Journal of Coastal Conservation, 20, https://doi.org/10.1007/s11852-016-0430-3, 2016.

Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J.: An assessment of the effectiveness of a random forest classifier for land-cover classification, ISPRS Journal of Photogrammetry and Remote Sensing, 67, 93–104, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2011.11.002, 2012.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W.: Monitoring vegetation systems in the great plains with ERTS, 1973.

Schuster, C., Förster, M., and Kleinschmit, B.: Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data, International Journal of Remote Sensing, 33, 5583–5599, https://doi.org/10.1080/01431161.2012.666812, 2012.

Shiferaw, H., Bewket, W., and Eckert, S.: Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem, Ecol Evol, 9, 2562–2574, 2019.

Spiegelberger, T., Gillet, F., Amiaud, B., Thébault, A., Mariotte, P., and Buttler, A.: How do plant community ecologists consider the complementarity of observational, experimental and theoretical modelling approaches?, Plant Ecology and Evolution, 145, 4–12, https://doi.org/10.5091/plecevo.2012.699, 2012.

Sutton-Grier, A. E. and Sandifer, P. A.: Conservation of Wetlands and Other Coastal Ecosystems: a Commentary on their Value to Protect Biodiversity, Reduce Disaster Impacts, and Promote Human Health and Well-Being, Wetlands, 39, 1295–1302, https://doi.org/10.1007/s13157-018-1039-0, 2019.

Tang, B., Frye, H. A., Gelfand, A. E., and Silander, J. A.: Zero-Inflated Beta Distribution Regression Modeling, Journal of Agricultural, Biological and Environmental Statistics, 28, 117–137, https://doi.org/https://doi.org/10.1007/s13253-022-00516-z, 2023.

Tarasiewicz, T., Nalepa, J., Farrugia, R. A., Valentino, G., Chen, M., Briffa, J. A., and Kawulok, M.: Multitemporal and Multispectral Data Fusion for Super-Resolution of Sentinel-2 Images, IEEE Transactions on Geoscience and Remote Sensing, 61, 1–19, https://doi.org/10.1109/TGRS.2023.3311622, 2023.

Turpie, K.: Explaining the Spectral Red-Edge Features of Inundated Marsh Vegetation, Journal of Coastal Research, 29, 1111–1117, https://doi.org/10.2112/JCOASTRES-D-12-00209.1, 2013.

Valavi, R., Elith, J., Lahoz-Monfort, J., and Guillera-Arroita, G.: Modelling species presence-only data with random forests, Ecography, 44, https://doi.org/10.1111/ecog.05615, 2021.

Van der Maarel, E.: Vegetation ecology, Blackwell Science Ltd, Oxford, 2005.

Villoslada, M., Bergamo, T., Ward, R., Burnside, N., Joyce, C., Bunce, R., and Sepp, K.: Fine scale plant community assessment in coastal meadows using UAV based multispectral data, Ecological Indicators, 111, 105 979, https://doi.org/https://doi.org/10.1016/j.ecolind.2019.105979, 2020.

Villoslada Peciña, M., Bergamo, T., Ward, R., Joyce, C., and Sepp, K.: A novel UAV-based approach for biomass prediction and grassland structure assessment in coastal meadows, Ecological Indicators, 122, 107 227, https://doi.org/https://doi.org/10.1016/j.ecolind.2020.107227, 2021.

Vincini, M., Frazzi, E., and D'Alessio, P.: A broad-band leaf chlorophyll vegetation index at the canopy scale, Precision Agriculture, 9, 303–319, https://doi.org/10.1007/s11119-008-9075-z, 2008.

Ward, R., Burnside, N., Joyce, C., and Sepp, K.: The use of medium point density LiDAR elevation data to determine plant community types in Baltic coastal wetlands, Ecological Indicators, 33, 96–104, https://doi.org/10.1016/j.ecolind.2012.08.016, 2013.

Ward, R., Burnside, N., Joyce, C., and Sepp, K.: Importance of Microtopography in Determining Plant Community Distribution in Baltic Coastal Wetlands, Journal of Coastal Research, 32, 1062–1070, https://doi.org/10.2112/JCOASTRES-D-15-00065.1, 2016.

Ward, R. D.: Landscape and ecological modelling: Development of a plant community prediction tool for Estonian coastal wetlands, Phd thesis, University of Brighton, 2012.

Ward, R. D.: Carbon sequestration and storage in Norwegian Arctic coastal wetlands: Impacts of climate change, Science of The Total Environment, 748, 141 343, https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.141343, 2020.

Wu, C., Niu, Z., Tang, Q., and Huang, W.: Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation, Agricultural and Forest Meteorology, 148, 1230–1241, https://doi.org/https://doi.org/10.1016/j.agrformet.2008.03.005, 2008.

Zabala, S.: Comparison of Multi-Temporal and Multispectral Sentinel-2 and Unmanned Aerial Vehicle Imagery for Crop Type Mapping, Master's thesis, Lund University, Lund, Sweden, 2017.

Zhang, L., Huettmann, F., Zhang, X., Liu, S., Sun, P., Yu, Z., and Mi, C.: The use of classification and regression algorithms using the random forests method with presence-only data to model species' distribution, MethodsX, 6, 2281–2292, https://doi.org/https://doi.org/10.1016/j.mex.2019.09.035, 2019.

Zhu, W., Rezaei, E. E., Nouri, H., Yang, T., Li, B., Gong, H., Lyu, Y., Peng, J., and Sun, Z.: Quick Detection of Field-Scale Soil Comprehensive Attributes via the Integration of UAV and Sentinel-2B Remote Sensing Data, Remote Sensing, 13, https://doi.org/10.3390/rs13224716, 2021.