USING LIDAR AND SATELLITE DATA TO ESTIMATE COVER OF SUBSTRATE AND VEGETATION IN THE BALTIC SEA

Karl Florén[1], Michael Tulldahl[2] and Sofia Wikström[1]

1. Aquabiota Water Research, Stockholm, Sweden; karl.floren@aquabiota.se
2. Swedish Defence Research Agency, Stockholm, Sweden; Michael.tulldahl@foi.se

ABSTRACT

Bathymetric lidar together with high resolution satellite data was used to classify six homogenous classes (two hard bottom classes and four soft bottom classes) in a shallow archipelago area in the Baltic Sea. We utilized lidar data directly and as means for correcting the satellite data for turbidity and water depth. A maximum likelihood (ML) algorithm was used to relate biological data collected with a drop video camera with the remote sensing data. A classification accuracy of about 80 % was reached when validating against an external dataset not used for building the model. The majority of recorded video segments contained a mixture of the classes and could not be used in this analysis. To overcome this, we used the classification over several pixels to predict the cover of bladderwrack, high vegetation and hard substrate at the video segment positions. The predicted cover was validated against observed cover using the Root Mean Squared Error (RMSE). The prediction of cover for bladderwrack, high vegetation and hard substrate was not very accurate with RMSE values ranging from 29 to 46 %. The classifier was able to separate low from high cover in most cases

INTRODUCTION

Maps of underwater habitats are important tools for marine spatial planning and marine conservation, which requests cost effective ways to produce reliable maps of the aquatic environment. Remote sensing can provide high resolution information on both physical and biological features of intertidal and subtidal habitats. Commonly used techniques include aerial photography, lidar and high resolution satellite information. Previous studies have shown that it is possible to classify underwater habitats using lidar data and/or spectral information from aerial or satellite images (1,2,3,4). For instance (5) were able to classify six bottom types, with different composition of substrate and vegetation, in the north-western Baltic proper with relatively high accuracy using lidar and high resolution satellite image data. The classification was based on field sites with homogenous substrate and vegetation. However, homogenous habitats are rare in shallow archipelago areas of the Baltic Sea and little is known about the behavior of a lidar sounding when hitting mixed substrates and vegetation types. The purpose of this study is to explore the use of remote sensing data such as lidar and satellite data in a typical archipelago area in the Baltic Sea with a mosaic of different substrates and different species of algae and phanerogams. We test how the predicted cover of different vegetation and substrate types based on classification of homogenous classes corresponded to the actual cover in the field, using field data from video transects.

METHODS

The data used in this study were collected at Askö in the Södermanland Archipelago, north-western Baltic proper (Figure 1) in May 2010. With its many islands and skerries the area is representative for the archipelagos along southern Sweden's east coast. We focused the study on shallow areas where hard substrates (bedrock, stones and boulders) dominate the seabed in areas exposed to waves. In more sheltered areas finer sediments such as sand and mud are common together with rocks and boulders. On the hard substrates bladderwrack (*Fucus vesiculosus*), which can reach up to 1 m in height, form communities together with filamentous brown and red algae.

The sand and mud are usually dominated by rooted plants such as fennel pondweed (*Stuckenia pectinata*) and eelgrass (*Zostera marina*). They usually grow in patches or in mixed communities, and can reach over 1 m height.

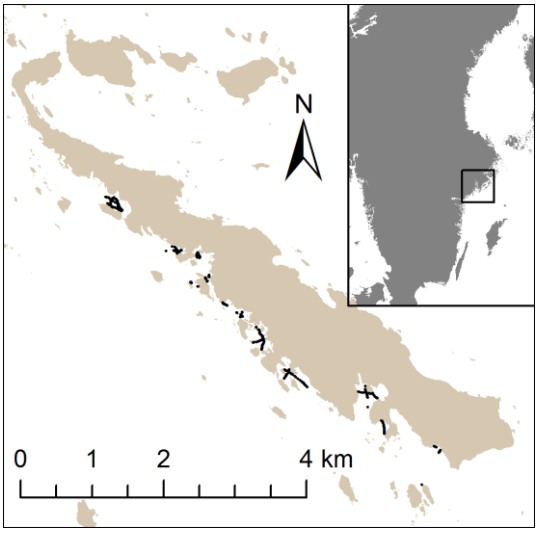

Figure 1. Study area. Black lines indicate positions for the video transects

Airborne Lidar, Satellite Image and field data

The lidar data used in the analysis were collected with the HawkEye II – Airborne coastal survey system (6). The lidar survey was performed in May 21, 2010 with an average point density of approximately 0.3 soundings per m^2 corresponding to a horizontal distance of 1.8 m between each lidar data point. In this study we use lidar data down to about 4 m depth, although the maximum depth range of the lidar data was about 10 m. Lidar classification variables were derived both form the point cloud data (depth-derived variables) and from the bottom echo part of the lidar waveform (benthic waveform variables). We used the point data for estimation of bottom slope, bottom depth standard deviation, and bottom curvature within regions around each sounding position. From the lidar benthic (sea bottom) waveform we derived the following features: width, pulse, and peak (height) of the bottom pulse. These features were subsequently corrected for depth, turbidity, and lidar system parameters before used as classification variables in combination with depth derived variables and satellite data. Details of the lidar waveform feature corrections and the lidar data and och are given in (2), and in (5), respectively.

The WorldView-2 data used in the analysis was collected in May 2, 2010 at 12h13. The pre-processing of image data involved, pansharperning e.g. (7,8) and automatic removal of sun-glints from the sea surface. The resulting image pixel size after pansharpening was 0.5 m. In the next step the lidar data and image data were geometrically aligned and the lidar variables were added to the image data as additional image layers. The lidar data layers were gridded and linearly interpolated to the image 0.5 m pixel resolution. In the final image processing step, the intensity data for the image bands Green and Yellow were corrected for depth. For the WorldView-2 data, only the Green and Yellow bands were used in the subsequent classification. Details of the image processing, the classification models, and classification steps, briefly described below, are given in (5).

Field data was collected along transect lines, between 24 and 28 May 2010, using a video hanger and an integrated (conventional) GPS positioning system. The video data were interpreted manually, by estimating the percent cover of substrate types and of plant and macroalgal species. The resulting data set of each transect line was divided into segments with a consistent cover of both substrate and vegetation. The soft substrate was often a mix between sand and mud and these substrate types were not separated in the interpretation of the video sequences. A total of 21

transects were recorded, which were divided into 733 segments that were analysed with respect to substrate and vegetation cover.

Classification of homogenous classes

We used field data from homogenous habitats for the classification using the lidar and WorldView-2 variables. To define a homogenous class we set the requirement that it should have either high or low (no) cover of one (or a few similar) substrate type(s) and/or vegetation. The resulting six classes are described in Table 1. The second column in Table 1 also shows three General classes, where the six classes are grouped into one Hard substrate class and two Soft substrate classes, one with high cover of High vegetation one without high vegetation. Table 1 also shows the number of field segments for each class. The segments were randomly divided into a training data set (67 % of the segments in each class) and an evaluation data set (33 % of the segments). For the following analysis, the classification variables were then collected at the center position of the field data segments. From each field segment center position 7 x 7 = 49 pixels were collected, corresponding to a square of 3.5 m x 3.5 m centered over each field segment (center) position.

Table 1. Description of the six classes and generalization into three classes. The number of field data segments N for training (TRAIN), evaluation (EVAL), and the total number of segments for each class is shown to the right in the table.

Class	General Class	Description, substrate and vegetation cover	*N* TRAIN	*N* EVAL	*N* Total
HardFucus	Hard	Bedrock, Boulders, or Stones 100% *Fucus vesiculosus* 100% *Pylaiella littoralis* <30%	8	4	12
HardPyla	Hard	Bedrock, Boulders, or Stones 100% *Pylaiella littoralis* >70% Red algae <50% Other vegetation <20% (mainly *Cladophora*)	5	2	7
SoftSand	Soft	Soft sediment and Sand 100% Loose algae <30%	5	3	8
SoftHiVeg	SoftHiVeg	Soft sediment and Sand 100% High Vegetation 100% (mainly *Potamogeton pectinatus*, small amounts of *Zostera marina*) Loose filamentous algae 20%-100%	9	4	13
LooseFucus	Soft	Soft sediment and Sand 100% Loose fucus >70% Loose filamentous algae <30% Other vegetation <30% (mainly *Potamogeton pectinatus*)	7	3	10
LooseFil	Soft	Soft sediment and Sand 100% Loose filamentous algae 100% Other vegetation <20% (mainly *Potamogeton* & *Chara* sp.)	7	3	10
		Total	**41**	**19**	**60**

A maximum likelihood (ML) method was used for classification tests of the homogenous classes based on the variables derived from lidar data and satellite imagery. For the ML classification, we used subsets of up to five of the available classification variables, based on internal classification accuracy evaluation (only the training data set was used). This variable optimization (choice of the most important variables) was performed using the training data with a leave-one-out cross-validation (LOOCV) technique e.g. (9). The resulting variables of highest importance were: the lidar depth standard deviation; the benthic pulse width and pulse peak from the corrected lidar

waveform data; and the Green and Yellow depth-corrected band signals from WorldView-2 imagery. An accuracy assessment was performed using the data from the field segments in the EVAL-column in Table 1.

Accuracy of mixed habitats

The difference between observed and predicted cover of *Fucus vesiculosus*, high vegetation (mainly *Stuckenia pectinata and Zostera marina*) and hard substrate was measured using Root Mean Square Error (RMSE). RMSE is a frequently used measure of the difference between predicted and observed values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{pred,i})^2}{n}}$$

To calculate predicted cover we used the 16 pixels closest to the center point of each segment, corresponding to a square of 2 x 2 m where each pixel covered 6.25 % of the square. For example 8 cells belonging to "HardFucus" would correspond to a predicted *Fucus* cover of 50 %. We also wanted to see the effect of positioning on RMSE. In conjunction with the center square eight 2 m x 2 m squares was created in different directions and in each one the cover of the responses was calculated. RMSE was calculated again and the effect of positioning on prediction accuracy was estimated by picking a random square (out of the nine positions) for each segment when RMSE was calculated. RMSE was calculated like this 10000 times and the result are shown as boxplots in Fig 3.

RESULTS

Table 2. Confusion matrix from accuracy assessment of the prediction for the six homogenous classes. Validation dataset = 19 video segments.

	Predicted							
Observed	Hard Fucus	Hard Pyla	SoftSand LoCov	Soft HiVeg	Loose Fucus	Loose Fil	Total	**PA (%)**
Hard Fucus	170	7	0	0	1	0	178	96
Hard Pyla	15	83	0	0	0	0	98	85
SoftSand LoCov	0	24	121	2	0	0	147	82
Soft HiVeg	0	49	0	96	29	22	196	49
Loose Fucus	0	0	0	20	125	2	147	85
Loose Fil	0	0	11	4	0	132	147	90
Total	185	163	132	122	155	156		
UA (%)	92	51	92	79	81	85		

Homogenous habitats in a shallow archipelago in the Baltic Sea could be classified correctly in 80 % of the cases using a ML classifier. Similar results were achieved for the general classes. However, homogenous habitats are rare in these environments so we developed a method for estimating the cover of different habitats and biotopes. The relatively high RMSE and the size of the boxes in figure 3, when comparing observed against predicted cover, tells us that the cover of all three classes was poorly estimated. The mean- and median values however show that the

classifier was able to separate high from low cover on a large scale. Mean predictions were generally higher than median which is an effect of the many zero predictions. The prevalence of the class hi vegetation was underestimated in Tulldahl's study (PA = 49 %) having an effect on the predicted cover in our analysis. In segments with an observed cover of 100 % hi vegetation mean predicted cover was below 40 %. There are a few possible reasons for the relatively low prediction accuracies. 1. Positioning problems in the field data. 2. The heterogeneous nature of the area. 3. Drifting loose algae. The heterogeneity of the seafloor places particularly high demands on position accuracy when a few meters inaccuracy often means a different substrate or a different species of macrophyte.

Table 3.The Root Mean Square Error of the difference between observed and predicted cover in the center square of the segment and as a mean value influenced by a potential positioning problem. The standard deviation indicates to what extent positioning influence the results.

	RMSE (in center square)	RMSE (mean different positions)	SD depending on position
Hard fucus	29.4	31.7	1.2
SoftHiVeg	45.8	45.5	0.9
Hard substrate	39.1	40	1.3

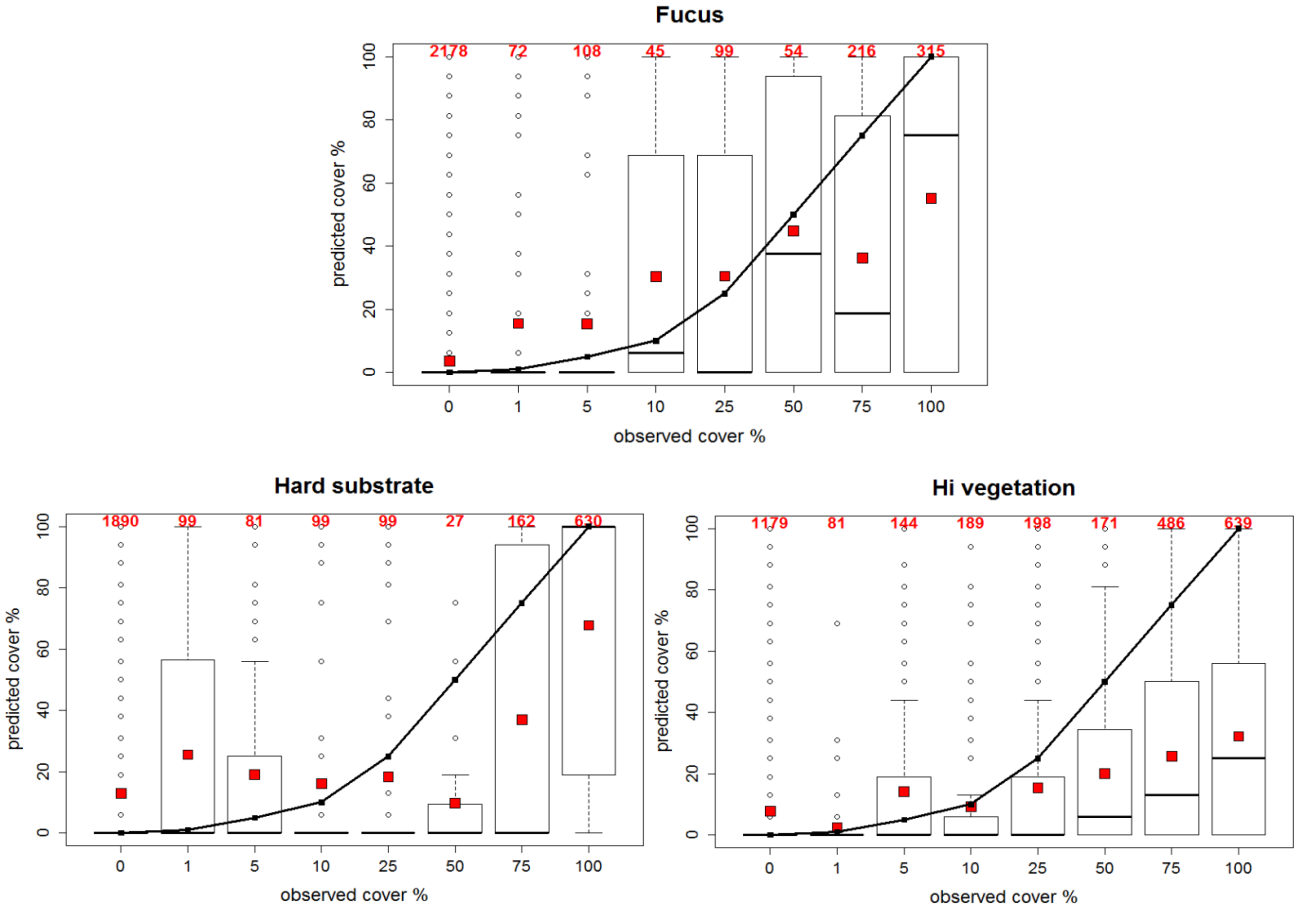

Figure 2. Predicted against observed cover of hard substrate, fucus and hi vegetation. Each of the 343segments is represented by predictions at 9 different positions (n=3087). Boxes show first quartile, median and third quartile. Whiskers show min and max values excluding outliers. Small circles show outliers more/less than 3/2 times of upper/lower quartile. Squares in red show mean. Numbers at the top in red show number of observations for each cover class. Black line show "true" cover.

CONCLUSIONS

We explored the potential of remote sensing data to map the seafloor in a typical archipelago area in the Baltic Sea by estimating cover of three common habitats. The relatively low prediction accuracies could have several explanations but it is likely that inaccurate positioning together with the small-scale heterogeneity of the habitats plays a major role. In future studies field data should be collected in accurately positioned and delimited squares to sort out some of the question marks.

ACKNOWLEDGEMENTS

This study was supported by Swedish Environmental Protection Agency (EMMA-project). Parts of the satellite data analysis were made in cooperation with the project "HISPARES - Spatial planning in archipelago waters by high spatial resolution remote sensing" financed by the European Union, Regional Development Fund, within the Central Baltic INTERREG IV A Programme 2007-2013. The authors acknowledge Airborne Hydrography AB and Brockmann Geomatics Sweden AB, respectively for cooperation concerning airborne lidar bathymetry data and satellite data.

REFERENCES

1 Collin A, B Long & P Archambault, 2012. Merging land-marine realms: Spatial patterns of seamless coastal habitats using a multispectral LiDAR. Remote Sensing Environment, 123: 390-399

2 Tulldahl M & S Wikström, 2012. Classification of aquatic macrovegetation and substrates with airborne lidar. Remote sensing environment 121: 347-357

3 Chust G, M Grande, I Galparsoro, A Uriarte & A Borja, 2010. Capabilities of the bathymetric LiDAR for coastal habitat mapping: A case study within a Basque estuary. Estuarine, Coastal and Shelf Science, 89: 200-213

4 Chust G, I Galparsoro, A Borja, J Franco & A Uriarte, 2008. Coastal and estuarine habitat mapping, using LIDAR height and intensity and multi-spectral imagery. Estuarine, Coastal and Shelf Science, 78: 633-643

5 Tulldahl M, P Philipson, H Kautsky & S Wikström, 2013. Sea Floor Classification with Satellite Data and Lidar Airborne Bathymetry (SPIE, Baltimore) Vol. 8724

6 Airborne Hydrography AB, http://www.airbornehydro.com

7 Saeedi J & K Faez, 2011. A new pan-sharpening method using multiobjective particle swarm optimization and the shiftable contourlet transform. ISPRS Journal of Photogrammetry and Remote Sensing, 66: 365-381.

8 Witharana C, D L Civco, & T H Meyer, 2013 Evaluation of pansharpening algorithms in support of earth observation based rapid-mapping workflows, Applied Geography 37: 63-87

9 Wang J J & X X Lu, 2010. Estimation of suspended sediment concentrations using Terra MODIS: An example from the Lower Yangtze River, China. Science of The Total Environment, 408: 1131-1138