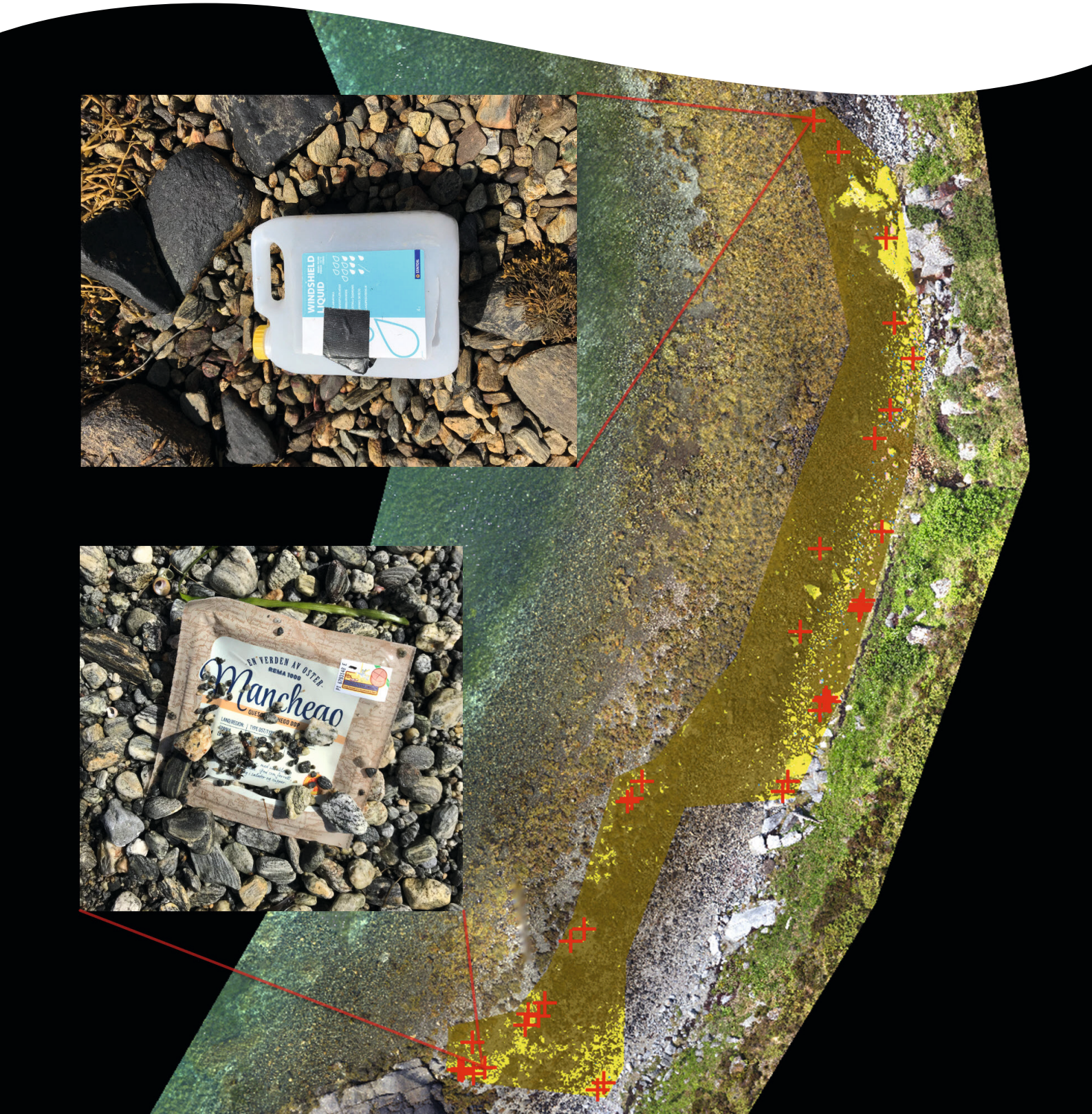
REPORT SNO. 7553-2020

Detection of macroplastic on beaches using drones and object-based image analysis



Norwegian Institute for Water Research

REPORT

Main Office
Gaustadalléen 21
NO-0349 Oslo, Norway
Phone (47) 22 18 51 00

NIVA Region South
Jon Lilletuns vei 3
NO-4879 Grimstad, Norway
Phone (47) 22 18 51 00

NIVA Region East
Sandvikaveien 59
NO-2312 Ottestad, Norway
Phone (47) 22 18 51 00

NIVA Region West
Thormøhlensgate 53 D
NO-5006 Bergen Norway
Phone (47) 22 18 51 00

NIVA Denmark
Njalsgade 76, 4th floor
DK 2300 Copenhagen S, Denmark
Phone (45) 39 17 97 33

Internet: www.niva.no

Title	Serial number	Date
Detection of macroplastic on beaches using drones and object-based image analysis	7553-2020	19.11.2020
Author(s) Bjørn Maaløe Torsvik, Robert Nøddebo Poulsen, Bert van Bavel, Hege Gundersen and Kasper Hancke	Topic group Marine biology	Distribution Open
	Geographical area Norway	Pages 47

Client(s)	Client's reference
Norwegian Institute for Water Research (NIVA)	NIVA (DigiSIS)
	Printed NIVA Project number 17004

Summary

This pilot study demonstrates UAV-based image collection being a promising technology for plastic detection in beach zones and we recommend object-based image analysis and UAV-based data collection being further developed for marine litter detection.

Objects of plastic at sizes down to 3.5 x 3.5 cm were segmented and classified successfully corresponding to approximately 4 times the pixel width of the classified images. Overall conclusions should be approached with some caution, due to the relatively few ground truth data points (13-37 geolocated plastic pieces) and their different characteristics regarding color and size.

Four keywords	Fire emneord
1. Plastic detection	1. Plastdetektering
2. Drone images	2. Dronebilder
3. Object-based image analysis (OBIA)	3. Objektbasert bildeanalyse (OBIA)
4. Random Forest, Support Vector Machine, eCognition	4. Random Forest, Support Vector Machine, eCognition

This report is quality assured in accordance with NIVA's quality system and approved by:

Kasper Hancke — Project Manager

Mats Gunnar Walday — Research Manager

ISBN 978-82-577-7288-8
NIVA-report ISSN 1894-7948

Detection of macroplastic on beaches using drones and object-based image analysis

Preface

This report is a product of a joint effort between The Norwegian Institute for Water Research (NIVA) and SpectroFly A/S. The idea and project design were initiated by Kasper Hancke and Robert Nøddebo Poulsen as a spin-off from work on using drones and image classification for coastal habitat mapping. Bjørn Maaløe Torsvik (SpectroFly A/S) performed the object-based image analysis and the project aimed to test the performance of the commercial software package eCognition for detecting marine litter in the form of macroplastic object in beach zones from drone images. The project is based on drone-captured images collected by NIVA at opportunity during multiple field studies; including the “Pilot-Møre” project, the “Strategisk instituttsatsing på digitale metoder for miljøovervåking og forskning”, and the “Frisk Oslofjord” project (www.friskoslofjord.no). We acknowledge the assistance during fieldwork and analysis from involved colleges and coworkers. Analysis and reporting were funded by NIVA’s strategic funds for New digital methods for monitoring and research (DigiSIS), and the National Research Infrastructure SeaBee - Norwegian Infrastructure for Drone-based research, mapping and monitoring in the coastal zone (www.seabee.no).

Oslo, 19.11.2020

Kasper Hancke

Table of contents

Summary

More than eight million tonnes of plastic is transported into the ocean each year, and large amounts end up in the beach zone on all continents of the globe. Here plastic debris are accumulated and buried into the sand, often consumed by animals, or washed back into the ocean. Currently, cost-efficient methods are lacking to detect and quantify macroplastic debris on beaches which is essential to determine the amount of plastic debris in beaches and to guide clean-up initiatives and monitoring.

The aim of this pilot study was to investigate the feasibility of detecting macroplastic in beach zones using high-resolution images in the Red-Green-Blue (RGB) to Near-Infrared (NIR) spectrum collected from flying drones (Unmanned Aerial Vehicles, UAV) combined with object-based image analysis (OBIA) classification techniques.

Two coastal environments (beaches) were sampled to investigate the capabilities of OBIA to detect plastic objects of 2-50 cm in size. The first, Akerøya (Oslofjord), was a sandy beach fragmented with piles of macroalgae. The second, Dymna (Møre region), was a stony beach with multiple stones and rocks of various sizes and colors. High resolution RGB or multispectral images were collected using either rotor- or fixed-winged drones at both locations. Three classification scenarios were analyzed using RGB data and for one scenario a 6-band multispectral dataset in the RGB-NIR range was used to improve detection.

Object-based image classification was used for plastic detection applying the commercial software eCognition. Analyses included multi-resolution segmentation and supervised classification, applying Support Vector Machine classifier algorithms in three scenarios and Random Trees in one. Both segmentation and classification are recognized as important steps, crucial to the detection of plastic and its resulting accuracy. In the three scenarios using RGB data, between 45% and 75% of the plastic pieces were detected (i.e. the sensitivity measure). Correspondingly, precision ranged between 3% and 21% leading to an accuracy (F-score) of 0.06 to 0.33. The classification of the fourth scenario including NIR information showed the highest accuracy of 0.67 (50% sensitivity and 87% precision), arguing for the advantages of including NIR bands for improved plastic detection.

Objects of plastic at sizes down to 3.5 x 3.5 cm were segmented and classified successfully corresponding to approximately 4 times the pixel width of the classified images. Overall conclusions should be approached with some caution, due to the relatively few ground truth data points (13-37 geolocated plastic pieces) and their different characteristics regarding color and size.

This pilot study demonstrates UAV-based image collection being a promising technology for plastic detection in beach zones and we recommend object-based image analysis and UAV-based data collection being further developed for marine litter detection.

Sammendrag

Tittel: Plastregistrering i strandsonen ved bruk av droner og objektbasert bildeanalyse
År: 2020
Forfattere: Bjørn Maaløe Torsvik, Robert Nøddebo Poulsen, Bert van Bavel, Hege Gundersen og Kasper Hancke
Utgiver: Norsk institutt for vannforskning, ISBN 978-82-577-7288-8

Mer enn åtte millioner tonn plast havner hvert år i havet, og en stor andel av dette ender opp i strandsonen på alle verdens kontinenter. Her samles plastrester som ofte begraves i sanden, blir konsumert av dyr, eller skylles tilbake i havet. Det finnes i dag ikke kostnadseffektive metoder for å registrere og kvantifisere makroplast på strender, noe som er viktig for å bestemme mengden av plast samt for å koordinere oppryddingstiltak og overvåking.

Målet med denne pilotstudien var å undersøke muligheten for å oppdage makroplast i strandsoner ved bruk av høyoppløselige bilder tatt med RGB (rød-grønn-blå) og nær-infrarød kamera fra flygende droner (såkalt UAV) kombinert med objektbasert bildeanalyse (OBIA).

Vi undersøkte to ulike kystmiljøer (strender) og evaluerte hvor effektivt vi var i stand til å oppdage plastgjenstander med størrelse 2-50 cm ved hjelp av OBIA. Det første, Akerøya i Oslofjorden, var en sandstrand med hauger av makroalger. Det andre, Dymna på Møre, var en steinstrand bestående av steiner i forskjellige størrelser og farger. Det ble tatt RGB- eller multispektrale bilder med høy oppløsning ved hjelp av både rotor- og/eller fastvinge-droner på begge lokaliteter. Tre klassifiseringsscenarier ble analysert ved bruk av RGB-data, og for ett scenario ble i tillegg et 6-bånds multispektral datasett i RGB-NIR-området anvendt for å forbedre deteksjonen.

Objektbasert bildeklassifisering ble brukt for plastdeteksjon ved bruk av den kommersielle programvaren eCognition. Analysene inkluderte «multi-resolution» segmentering og styrt klassifisering, ved bruk av algoritmer for «Support Vector Machine» (SVM) i tre scenarier og ett scenarium ved bruk av «Random Trees». Både segmentering og klassifisering er viktige trinn for påvisning av plast og nøyaktig posisjonering. I de tre scenariene der RGB-data ble brukt, ble mellom 45 og 75 % av plastbitene oppdaget (dvs. sensitivitet). Tilsvarende varierte presisjonen mellom 3 og 21 %, noe som gav en nøyaktighet (F-score) fra 0,06 til 0,33. Klassifiseringen i scenario 4, der NIR var inkludert, viste høyeste F-score på 0,67 (med 50 % sensitivitet og 87 % presisjon). Dette resultatet taler for å inkludere NIR-bånd for best plastdeteksjon.

Analysene viste at objekter i størrelser ned til 3,5 x 3,5 cm kunne segmenteres og klassifiseres som plastikkbiter. Dette tilsvarer påvisning av objekter ned i en størrelse tilsvarende fire ganger pikselbredden i de tilgjengelige bildene. Disse resultatene bør anvendes med en viss forsiktighet, på grunn av det relativt lite antall «ground truth» datapunkter (kun 13-37 geolokaliserte plastbiter), som i tillegg varierte i farge og størrelse.

Denne pilotstudien viser at bilder tatt med flygende droner i kombinasjon med objektbasert bildeanalyse er en lovende metode for registrering av plast i strandsonen, og vi anbefaler en videre utvikling av denne teknologien.

1 Introduction

1.1 Background

More than eight million tonnes of plastic are going into the ocean each year (Jambeck et al. 2015), and plastic debris in the coastal environments and in beach zones cause dramatic negative impacts on animals and human populations (UNEP, 2016). Plastic debris reaches the beach zones where it is buried into the sand and sediments or washed back out into the ocean, or in some cases consumed by animals. All of this on scales and rates that we have very limited data on, and we severely lack understanding about the dynamics of these processes (Andersen 2019).

Gathering quantitative information of the amount of plastic debris in coastal environments is fundamental to understand its sources, the pathways of transport, the temporal trends of its appearance, and for studying the impacts of plastic debris on marine ecosystems. Currently, cost-efficient methods are lacking to detect and quantify macroplastic debris on beaches which is essential to determine the amount and distribution of this global problem, and to guide monitoring programs and clean-up initiatives.

General guidelines and some operational protocols for beach litter assessments and compilations have been proposed suggesting mainly visual identification methods which require a high number of operators, is subjective, labor-intensive, time-consuming, and spatially limited (e.g. GESAMP, 2019).

To overcome these limitations new approaches are required for faster, autonomous and cost-efficient detection and mapping of plastic debris in the coastal zone. It has been proposed crucial to plan and implement routine environmental monitoring measured and standard monitoring protocols for marine litter including spatial and temporal mapping on beaches (e.g. OSPAR Commission, 2010; GESAMP, 2019).

In recent years, Unmanned Aerial Vehicles (UAV) also referred to as flying drones has become available and open new avenues to explore and monitor large areas in cost-efficient ways. Combined with sophisticated optical sensors, UAV's will enable collection of large scale, high-resolution data, and create data products that will improve research and management efforts for healthier and cleaner environments.

The scope of this pilot study was to investigate the feasibility of detecting macroplastic from regular color images (RGB) to multispectral high-resolution data collected from UAVs and evaluate the efficiency of using object-based image analysis (OBIA) classification techniques.

The following questions are sought answered:

- To what extent can macroplastic objects be detected in RGB imagery using segmentation classification?
- What is the performance of a classification configuration trained in one area for detecting macroplastic in a new area?
- How does ground surface characteristics affect plastic detection?
- Can addition of near-Infrared (NIR) band information improve macroplastic classification detection?

2 Methods

2.1 Theory

2.1.1 Image classification

Image classification of high resolution remotely sensed imagery can include segmentation, which opens the possibility to change the process from classical pixel-based approach to an object-based approach, called object-based image analysis (OBIA). OBIA includes both segmentation and classification.

2.1.1.1 Segmentation

Segmentation is the process were pixels are grouped together into segments prior to the classification, to define objects of interest. If objects are erroneously split into several segments it is called over-segmentation. Whereas under-segmentation refers to the opposite situation where objects are in segments were also surrounding pixels are included.

Multiresolution segmentation is an algorithm that combines the image pixels based on the spectral- and shape heterogeneity.

The heterogeneity threshold is defined by the “Scale” parameter, setting the tolerance of the variation inside a segment. The higher the scale parameter the higher the tolerance of variation typically resulting in larger and fewer segments.

How the heterogeneity is measured is controlled by the “Color/Shape” parameter defining the amount of weight put on color versus shape. Color refers to the standard deviation of the pixel value of the bands whereas shape measures the deviation from a shape.

Shape heterogeneity is measured according to the smoothness or the compactness of the segment (Chen et al. 2019) formulated in the two equations 1 and 2 below. A graphical illustration of the relationship between the two parameters is displayed in Figure 1.

$$Smoothness = \frac{Length\ of\ the\ border\ of\ the\ segment}{Length\ of\ the\ border\ of\ the\ bounding\ box\ of\ the\ segment} \quad \text{Eq. 1}$$

$$Compactness = \frac{Lenght\ of\ the\ border\ of\ the\ segment}{Area\ of\ the\ segment} \quad \text{Eq. 2}$$

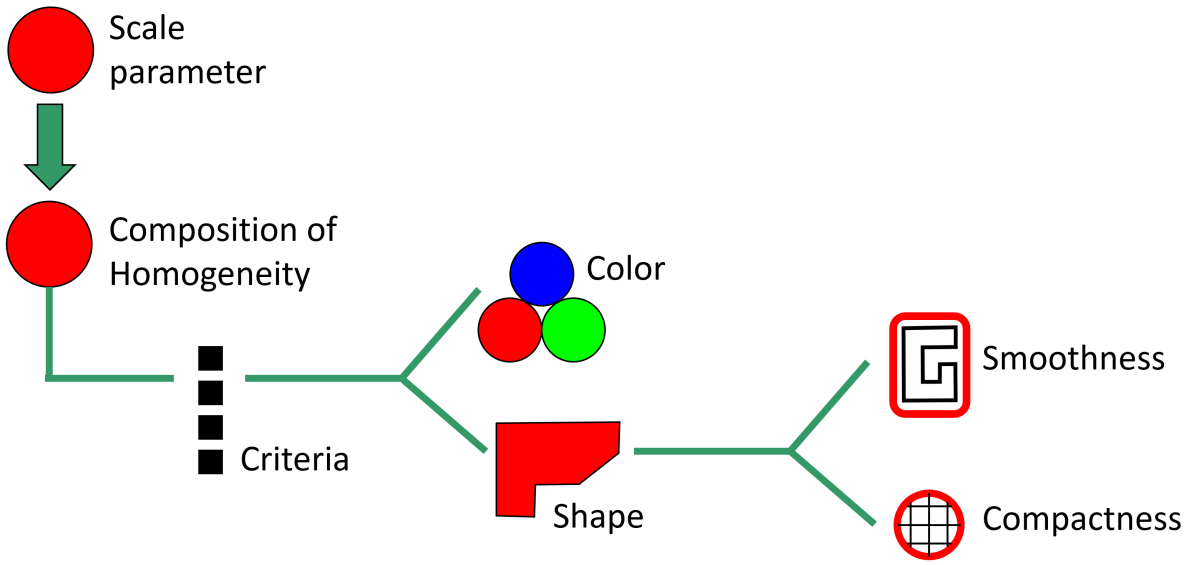


Figure 1. Relations between the parameters to adjust in the segmentation settings in eCognition (Aminipouri, 2009).

2.1.1.2 Classification

Image classification can be divided into unsupervised and supervised classification. Unsupervised classification (or clustering) is grouping pixels/segments into similar clusters without any prior knowledge of the classes. Subsequently the clusters are classified into meaningful classes. Supervised classification is classifying the pixels/segments into classes provided by the user. The user selects representative pixels/segments for each class. The algorithm uses the training data to create models that classifies the remaining pixels/segments according to the user provided classes. Many different classification algorithms exist, for instance Support Vector Machine (SVM) and Random Forest (e.g. Acuña-Ruz et al., 2018; Gonçalves et al., 2020; Qian et al., 2014).

2.1.2 Accuracy assessment

One way of assessing the accuracy of image classification is the error matrix (also called confusion matrix). When assessing two classes in an error matrix, the number of true positives (TP), false positive (FP), false negative (FN) and true negatives (TN) are counted. The error matrix is the basis for statistical measures of sensitivity (S), precision (P) and the so-called F-score (Gonçalves et al., 2020). The sensitivity is a measure of how good the classification is at detecting the ground truth, i.e. the percentage of ground truth that is right classified (Eq. 3). Precision measures how precise the classification is by measuring the percentage of the class that are TP (Eq. 4). F-score combines precision and sensitivity, measuring how well the classification performs. Results ranging from 0 to 1, with 0 as the lowest (worst) score and 1 as the highest (best) score (Eq. 5).

$$S = \frac{TP}{TP+FN} \qquad \text{Eq. 3}$$

$$P = \frac{TP}{TP+FP} \qquad \text{Eq. 4}$$

$$F = 2\,x\,\frac{P\,x\,S}{P+S} \qquad \text{Eq. 5}$$

2.2 Project Area

Two coastal study areas in Norway was included in this project. One location was Akerøya at the south east entrance to the Oslofjord, and the second site called Dymna located on Dimnøya approximately 5 km south west of Ulsteinvik on the west coast (Figure 2). Akerøya was a sandy beach location, partially covered by piles of dead seaweed and infrequent presence of small stones and shells (Figure 3). Dymna was a rocky beach, dominated by stones of multiple sizes and colors, but also areas with grass vegetation. Also, at Dymna, seaweed was sporadically distributed on the beach (Figure 4). The two locations where chosen to represent distinctly different however common backgrounds for plastic detection.

Figure 2. Location of the two study sites at the Norwegian coast.

Figure 3. Close-up image of the sandy beach at the Akerøya location, southern Norway.

Figure 4. Close-up image of the stony beach at the Dymna location, western Norway.

2.3 Data and Equipment

2.3.1 Time of data collection

Data was collected at Dymna on the 27th of June 2018 approximately 4 pm using:

- RGB sensor (Zenmuse X5) mounted on DJI Matrice 600 Pro rotor drone (Figure 5a)

Data was collected at Akerøya on the 27th of August 2019 approximately 7 pm using:

- RGB sensor (SODA) on a Sensefly eBee X fixed-wing drone (Figure 5b), and
- Multispectral (MS) sensor (Tetracam Macaw) mounted on DJI Matrice 600 Pro rotor drone (Figure 5a)

a) b)

Figure 5. The two drones used in the study, a) DJI Matrice 600 Pro rotor drone and b) Sensefly eBee X fixed-wing drone. Photos by NIVA/K Hancke.

2.3.2 Plastic pieces

At both study locations macroplastic pieces of various sizes, color, and composition were intentionally placed using a random placement design. Plastic can be categorized based on its physical size; into macro- (> 25 mm), meso- (5 mm to < 25 mm) and microplastic (1 mm to < 5 mm) (Acuña-Ruz et al., 2018). All plastic detection in this project was focused on detection of macroplastic.

2.3.3 Ground truth

At Akerøya 16 plastic pieces were geolocated with real-time kinematic (RTK) global positioning system (GPS). The plastic objects ranged in size from small pieces of a few centimeters (Figure 6A) up to around 50 cm (Figure 6B). Plastic of both soft and hard composition were included. Colors were white (8 pieces), blue (4), orange (1), red (3), pink (1), green (2), metallic (1), semitransparent (1) and transparent (4).

At Dymna 37 pieces was geolocated, ranging from 3.5 cm to a little less than half a meter in size (Figure 6C and Figure 7). Four pieces was white rectangles of 28, 14, 7 and 3.5 centimeters (Figure 6C) and nine pieces were cans (soda or beer). Materials included both soft and hard plastics. Colors included white (included in 19 pieces), blue (5), yellow (2), grey (1), green (3), red (3), brown (7), metal (9), black (2), semitransparent (4) and transparent (1). The numbers in the parentheses do not add up to the total amount of pieces since more than one color is counted for some pieces.

Figure 6. A) Close-up picture of a smaller sized plastic piece at the Akerøya study area. B) Close-up picture of a larger sized plastic piece at Akerøya study site. C) Four rectangle-shaped pieces of plastic of size 28, 14, 7 and 3.5 centimeter across, used at Dymna study site.

Figure 7. Close-up picture of a larger sized plastic piece (a plastic bag) at Dymna study site.

2.3.3.1 GPS measurements

The equipment used for geolocation was a Leica RTK GPS comprising a Leica GS16 with the CS20 tablet. It measures the position (longitude, latitude and altitude) with an accuracy of 2-3 cm.

2.3.3.2 Ground Control Points

Before initiating the drone surveys, white wooden crosses (50 cm in diameter) were placed in the survey area and positioned with RTK GPS. These were used as ground control points in the subsequent postprocessing to obtain geopositioned high accuracy orthomosaics.

2.3.4 Drones and camera specifications

Two drone systems were used for data collection: 1) a fixed-wing Sensefly Ebee X (Figure 5a) with a maximum flight time of 90 minutes (www.sensefly.com), and 2) a rotary wing drone of type DJI Matrice 600 Pro (Figure 5b) capable of carrying a payload of several kilograms (www.dji.com). Three camera systems were applied, 1) RGB Sensefly S.O.D.A. (equipped at the Ebee drone), 2) RGB DJI Zenmuse X5 (equipped at the Matrice drone) and Multispectral Tetracam Macaw-6 (equipped at the Matrice drone).

2.3.4.1 Sensefly S.O.D.A.

Sensefly S.O.D.A. collects data from three bands, red, green, and blue (RGB), with mean wavelengths of 660 nm, 520 nm and 450 nm respectively (Figure 8a). S.O.D.A. is a 20 MP RGB camera for drone mapping (Figure 8b), producing a ground sample distance of 2.5 cm/pixel at 100 m altitude.

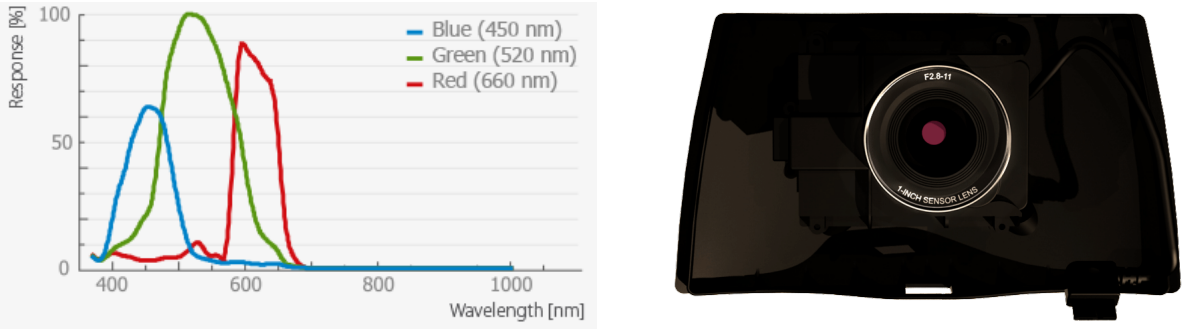


Figure 8. left) Band responses of the three bands captured by the SODA camera (www.spatialtechnologies.ca) and right) The Sensefly S.O.D.A. camera seen from the front (www.skyviv.com)

2.3.4.2 Zenmuse X5

DJI Zenmuse X5 collects data from three bands, red, green and blue (RGB), unfortunately no specifications of the average wavelengths are available.

2.3.4.3 Tetracam Macaw

The multispectral camera Macaw-6 from Tetracam is a 6 narrowband camera with a 1.3 MP resolution, corrected for incident light, wavelengths are shown in Table 1.

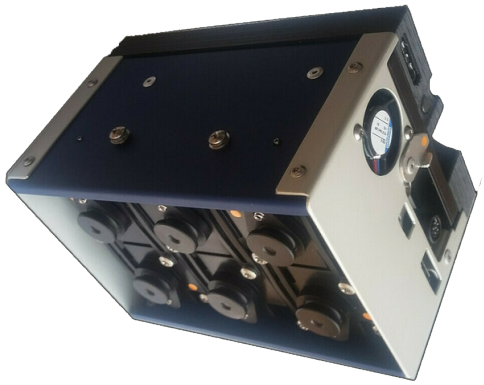


Figure 9. Tetracam Macaw-6 (www.tetracam.com)

Table 1. Wavelengths of the 6 bands of the Macaw camera. Filter shows the mean wavelength and band width, both in nm.

Channel	Filter [nm]	Band Width [nm]
Band 1	840	10
Band 2	490	10
Band 3	550	10
Band 4	670	10
Band 5	700	10
Band 6	720	10

2.3.5 Photogrammetric processing

All RGB datasets were processed in the photogrammetric software Pix4D. Details about the photogrammetric processing is available at www.pix4d.com. The multispectral dataset was processed in another photogrammetric software package – the Simactive Correlator3D. Again, details about the photogrammetric processing in Correlator3D is available at www.simactive.com.

2.3.5.1 Projection

All geographic data is in the projection Universal Transverse Mercator (UTM) zone 32 north ETRS89.

2.3.5.2 RGB Data

RGB data were available as orthomosaics with RGB bands in 8-bit unsigned (values from 0-255). The Akerøya RGB orthomosaic from 2019 has a 1.6 x 1.6 cm resolution (Figure 10), whereas the Dymna RGB orthomosaic from 2018 has 0.9 x 0.9 cm resolution (Figure 11).

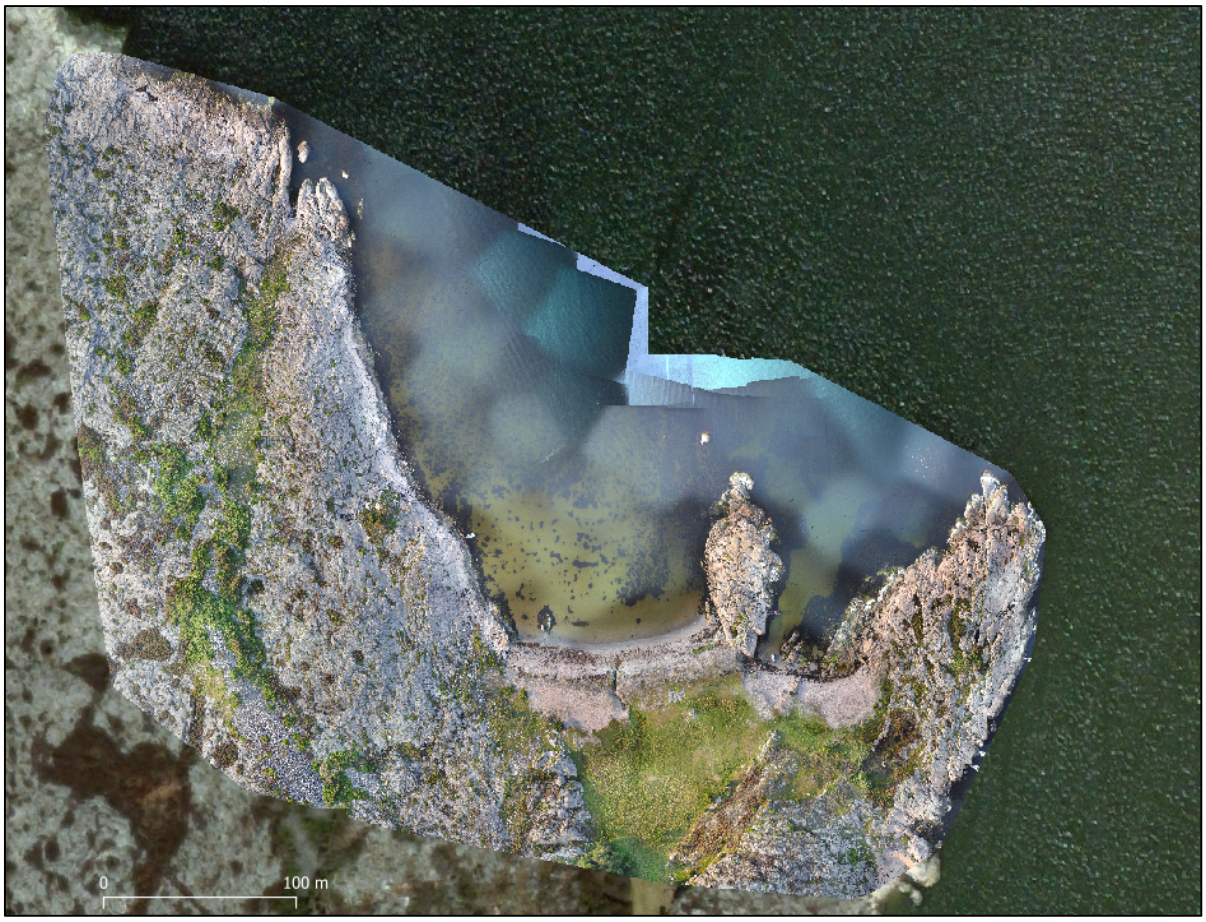


Figure 10. An RGB orthomosaic of Akerøya 2019 captured from 55 m altitude, with a 1.6 x 1.6 cm pixel resolution of the ground. Satellite imagery is used as background (Google Earth).

Figure 11. An RGB orthomosaic of Dymna 2018 captured from 35 m altitude, with 0.9 x 0.9 cm pixel resolution of the ground. Satellite imagery is used as background (Google Earth).

2.3.5.3 Multispectral Data

The multispectral dataset from Akerøya was available as an orthomosaic of the 6 bands in 8-bit unsigned (values from 0-255) in a resolution of 1.7 x 1.7 cm (Figure 12).

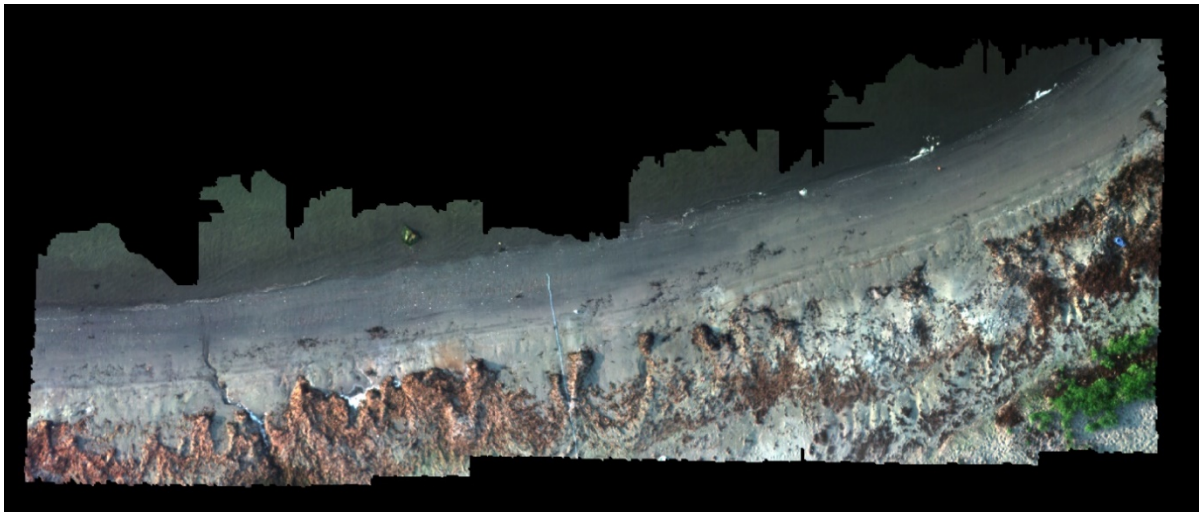


Figure 12. The orthomosaics of multispectral bands 4, 3 and 2 is here used in RGB composite visualization showing the dataset coverage of Akerøya 2019. The images were captured from 32 m altitude, with 1.7 x 1.7 cm pixel resolution of the ground.

2.4 Methodology

The present chapter describes how the analyses were performed and evaluated as well as specific settings applied. Also, a general description is given of the classification methods and accuracy assessments applied to the different study areas.

The process of classification was performed in four different scenarios, according to location, data type and approach (Table 2). The RGB-dataset from Akerøya 2019 was analyzed first by fine-tuning the settings for that area (scenario 1). The obtained configuration of this classification model was subsequently applied, without adjustments, to the Dymna 2018 RGB dataset (scenario 2). Hereby testing the performance of a preset configuration from another area with only partial matching landcover classes. Next was the RGB-dataset from Dymna 2018 where parameters were fine-tuned to meet the characteristics of the specific study area (scenario 3). This was done to test whether a site-specific classification to Dymna was an improvement compared to the previously tested preset configuration. Finally, the multispectral dataset from Akerøya 2019 was investigated in order to evaluate if the NIR band together with a site-specific adjusted classification (scenario 4) could improve the classification compared to the classification based on only the RGB dataset from Akerøya 2019.

Table 2. Parts included in the four scenarios.

Scenario number	Location (year)	Data	Approach
1	Akerøya (2019)	RGB	Finetune to the area
2	Dymna (2018)	RGB	Preset configuration (Akerøya RGB)
3	Dymna (2018)	RGB	Finetune to the area
4	Akerøya (2019)	MS (RGB+NIR)	Finetune to the area

2.4.1 Image Classification

The image classification process included segmentation of pixels and subsequent classification of segments. Both segmentation and classification were conducted using Trimble eCognition Essentials (version 1.3). Accuracy assessment was done in ArcGIS Pro. Figure 13 visualize a flowchart of the general methodology applied in scenario 1, 3 and 4. The method is based on camera, GPS data and the iterative parts of the segmentation process and subsequent classification of segments, and finally the classification accuracy assessment.

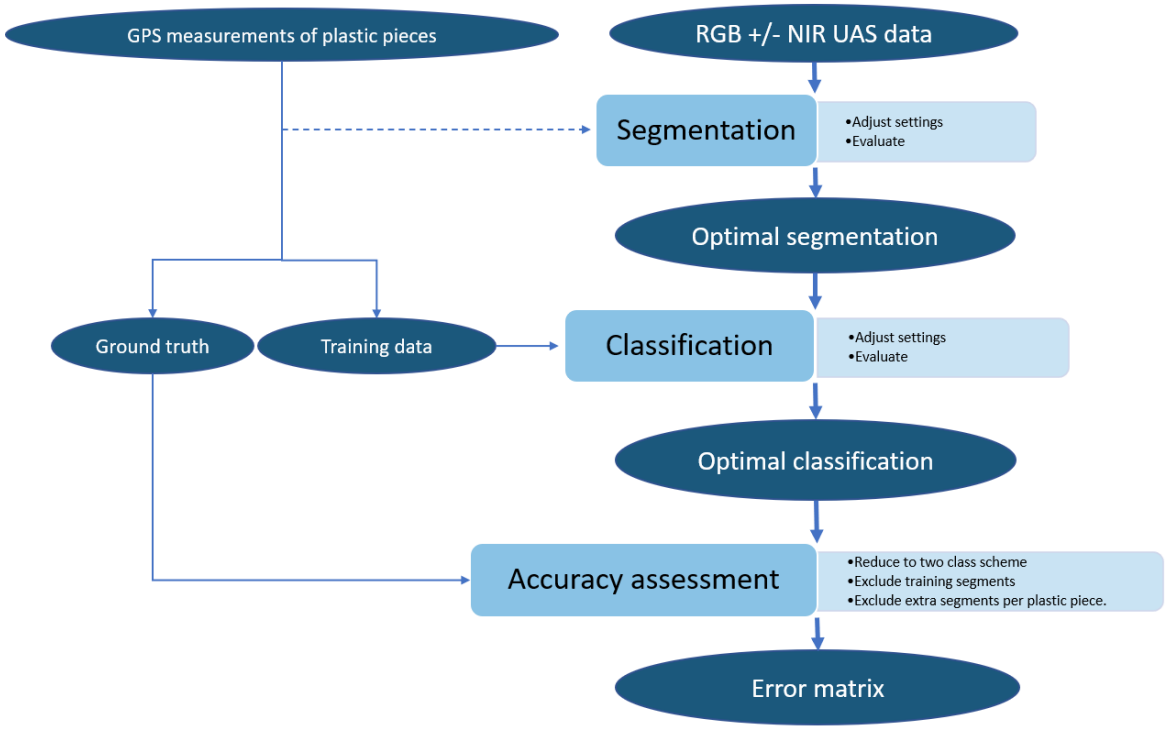


Figure 13. Flowchart of the general method applied in scenario 1, 3 and 4.

2.4.1.1 Segmentation

Imagery data comprise a high resolution and since the plastic objects consists of several pixels an object-based approach was used instead of a pixel-based. To reduce the processing time needed and exclude unnecessary landcover classes a region of interest (ROI) was selected in eCognition containing all plastic pieces.

A multiresolution segmentation was used as the segmentation method. The optimal settings were found by an iterative trial and error approach, by visually evaluating the resulting segmentation. The visual qualitative evaluation focused on three segmentation parameters 1) the purity of segments, regarding plastic piece versus surrounding landcover pixels, 2) how well small plastic pieces was captured by a single segment minimizing over- or under-segmentation of the pieces, and 3) keeping over-segmentation of larger plastic pieces at a minimum. The segmentation evaluated as the best fitting was applied in the subsequent classification.

2.4.1.2 Classification

As with the segmentation method, the optimal selections and settings were adjusted in an iterative trial and error approach. In this process adjustments were evaluated by a visual qualitative and semi-quantitative approach of the resulting classification. The evaluation focused on 1) the number of true positives, 2) the number of false positives, and 3) general fitting of the classes to the visually interpreted landcovers. The classified segments with the chosen configuration were exported from eCognition for later quantitative assessment of the accuracy of the classification. Additionally, segments used as training data were exported as points and the ROI as a polygon.

2.4.2 Accuracy Assessment

To assess the classification accuracy, the classified plastic segments were compared with the geolocated (ground truth) plastic objects. Accuracy was assessed based on geolocated plastic objects not included as training data for the classifying models. An error matrix and related percentage measures was produced to enable evaluation and comparison of the different classification results. Segments in the error matrix was counted as True Positive (TP), False Positive (FP), False Negative (FN) or True Negative (TN) as shown in Table 3. To enable this procedure the classification was transformed into a two-class system of Plastic and Other prior to the accuracy assessment. This was done by grouping all plastic classes in the Plastic class and all remaining classes into the Other class.

Table 3. Relation between error matrix term and landcover/class combination of segments.

Error matrix term	Landcover = class
True positive (TP)	Plastic = plastic
False positive (FP)	Other = plastic
False negative (FN)	Plastic = other
True negative (TN)	Other = other

The ground truth of the plastic objects was not found to be sufficient for accuracy assessment, since the geolocation only covered one position, while plastic objects and their segments covered areas. To match the number of GT plastic objects with the number of segments, each plastic object was only counted, as one segment, even though it in practice consisted of several segments. For a plastic object to be classified as a TP, only one of its segments required to be classified as plastic. Similarly, when several segments were covering a plastic object, only one plastic segment counted as a TP. All other segments not covering plastic pieces was considered as ground truth for the class “Other”. The statistical measures included sensitivity, precision, and accuracy assessment (F-score) were calculated for each scenario.

2.4.3 Settings in the four scenarios

The region of interest (ROI) for Akerøya 2019 RGB dataset was 420 m^2 including 17 ground-truth points of 16 plastic pieces (Figure 14A). The Dymna 2018 RGB dataset, used for scenario 2 and 3, had a ROI comprising 536 m^2 including 37 ground-truthed plastic objects (Figure 14B). At Akerøya, the multispectral 2019 dataset used for scenario 4, comprised a 360 m^2 ROI, including 13 ground-truthed plastic objects (Figure 14C).

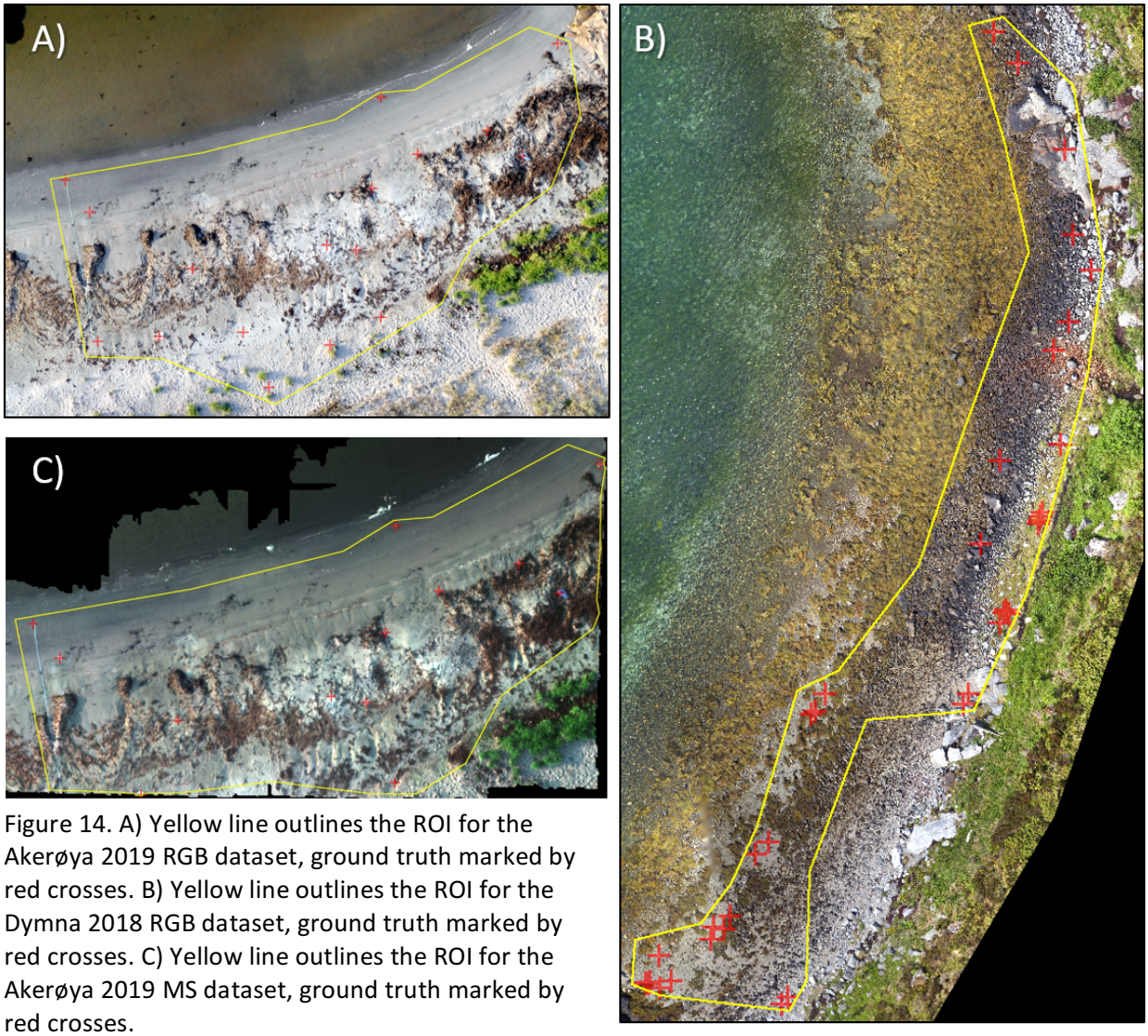


Figure 14. A) Yellow line outlines the ROI for the Akerøya 2019 RGB dataset, ground truth marked by red crosses. B) Yellow line outlines the ROI for the Dymna 2018 RGB dataset, ground truth marked by red crosses. C) Yellow line outlines the ROI for the Akerøya 2019 MS dataset, ground truth marked by red crosses.

Table 4 presents the parameters applied among the four scenarios. The "scale" parameter was set to range from 5 to 200 in all scenarios, to make the selected value comparable between the scenarios. The Color/Shape parameters ranged from 0-1, where 0 indicates maximum weight of Color and 1 maximum weight on Shape. The subdivision of the shape factor by Smoothness/Compactness was also ranging from 0 to 1, where 0 gives maximum weight to smoothness shape calculation and 1 maximum weight to compactness shape calculation method. Classification settings applied in the four different scenarios are presented in Table 5.

Table 4. Settings used in the segmentations in the different scenarios as well as size of ROI, the number of segments and the number of GT included.

Parameter	Akerøya 2019 RGB	Dymna 2018 RGB preset configuration	Dymna 2018 RGB	Akerøya 2019 MS
Scenario number	**1**	**2**	**3**	**4**
Scale	28	28	16	12
Color/Shape	0.1	0.1	0.5	0.55
Smoothness/Compactness	0.4	0.4	0.5	0.45
ROI	420 m^2	536 m^2	536 m^2	360 m^2
Segments	5,117	29,303	60,084	9,887
Ground Truth points	17	37	37	13
- ***Used for training***	*3*	*0*	*7*	*2*
- ***Excluded***	*1*	*0*	*0*	*0*
- ***Available for accuracy***	*13*	*37*	*30*	*11*

Table 5. Overview of the classes (number of training segments used per class), included features, classifier algorithm and classifier settings used in the four scenarios. No training segments are included in scenario 2 since the models from scenario 1 is used. SVM = Support Vector Machine, RT = Random Trees (a Random Forest algorithm).

	Akerøya 2019 RGB	Dymna 2018 RGB preset configuration	Dymna 2018 RGB	Akerøya 2019 MS
Scenario #	**1**	**2**	**3**	**4**
Classes (number of training segments)	- Plastic white (3) - Plastic blue (1) - Sand (28) - Sand wet (19) - Sand bright (3) - Sand overexposed (6) - Seaweed (9)	- Plastic white - Plastic blue - Sand - Sand wet - Sand bright - Sand overexposed - Seaweed	- Plastic (26) - Gravel (56) - Stone black lichen covered (41) - Stone dark (6) - Stone medium (16) - Stone bright (26) - Stone overexposed (44) - Stone miscellaneous (34) - Seaweed (42) - Seaweed dark (38) - Grass (54)	- Plastic white (2) - Plastic blue (7) - Sand wet (36) - Sand bright (33) - Overexposed (13) - Seaweed (9) - Seaweed dark (22)
Features included	- Blue - Green - Red	- Blue - Green - Red	- Blue - Green - Red	- Blue - Green - Red - NIR
Classifier	SVM	SVM	RT	SVM
Settings	Kernel type: Linear C parameter: 2	Kernel type: Linear C parameter: 2	Depth: 500 Min samples count: 0 Use surrogates: No Max categories: 16 Active variables: 0 Max tree number: 300 Forest accuracy: 0.01 Termination criteria type: both	Kernel type: Linear C parameter: 2

3 Results

Selected results from segmentation, classification, and accuracy assessment are presented in the current section. Further details about the applied segmentation and classification procedures are to be found in a work log that can be received from the authors upon request (written in Danish).

3.1 Segmentation

Selected examples of the resulting segmentation are presented in this section. Division of pixels into segments are shown in Figure 15 illustrating a segmentation nicely capturing a small plastic piece. A closeup picture of the plastic piece is presented in Figure 16. Some plastic objects were not as nicely captured by the segmentation and was consequently over-segmentated, i.e. object divided into several segments. An example on over-segmentation is shown in Figure 17. Figure 18 shows a closeup picture of the plastic object, in this case a rope. The opposite is also observed, under-segmentation, where the plastic piece is segmented together with the surrounding pixels as displayed in Figure 19. Figure 20 shows a closeup picture of the plastic piece to be identified. Hereby mixing the spectral information of the object with its surroundings in the later classification, potentially increase the risk of misclassification.

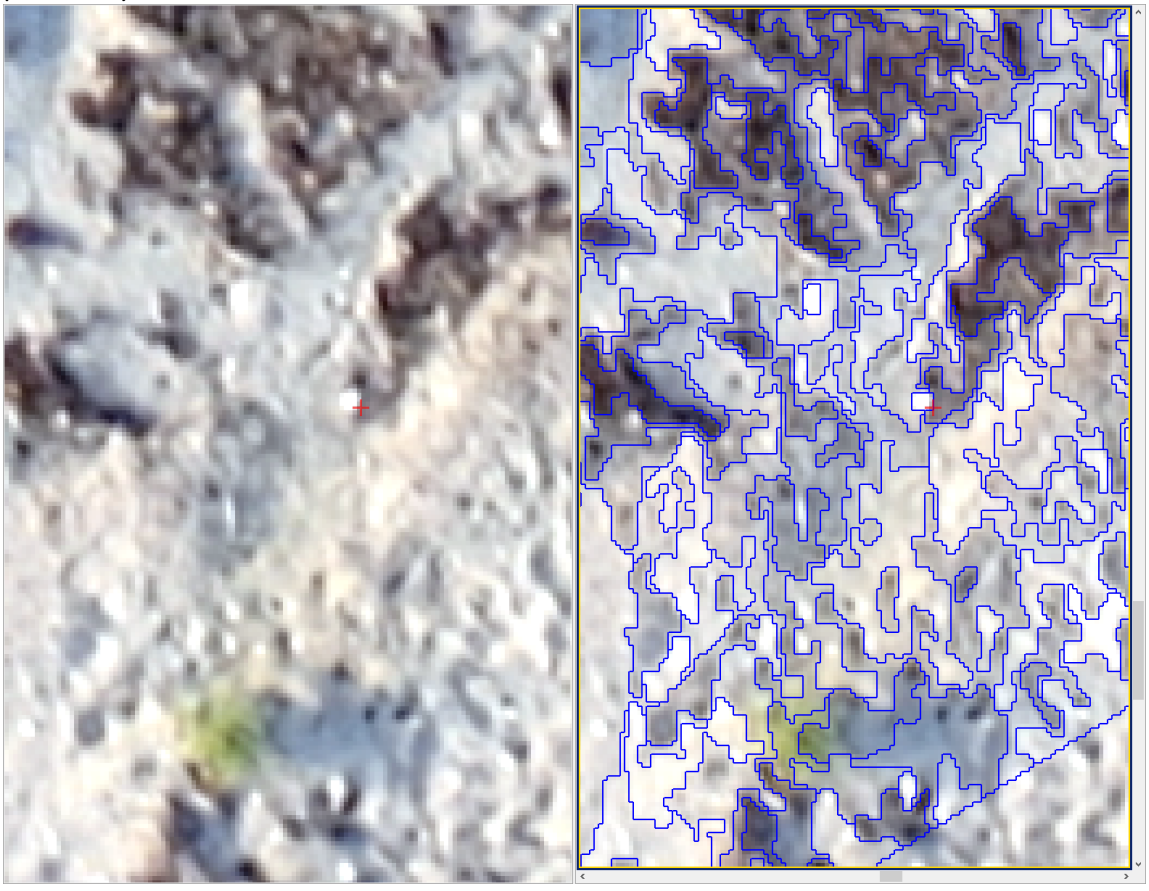


Figure 15. Screen shot from the eCognition essentials software package with an example of a small white plastic piece marked by the red cross and the segmentation capturing the plastic piece depicted as polygons with blue outlines in the right side of the picture. Example origins from RGB orthomosaic Akerøya used in scenario 1.

Figure 16. Left: Closeup picture of the plastic object in Figure 15. Right: Red line depict the extent by the closeup picture, example from RGB orthomosaic from Akerøya used in scenario 1.

Figure 17. Screenshot from the eCognition essentials software package with an example of a blue plastic piece (a rope) and the segmentation depicted as polygons with blue outlines. The right picture shows the plastic object divided into several segments, termed over-segmentation. The example originates from the RGB orthomosaic from Akerøya (scenario 1).

Figure 18. Close-up picture of the plastic object shown in Figure 17.

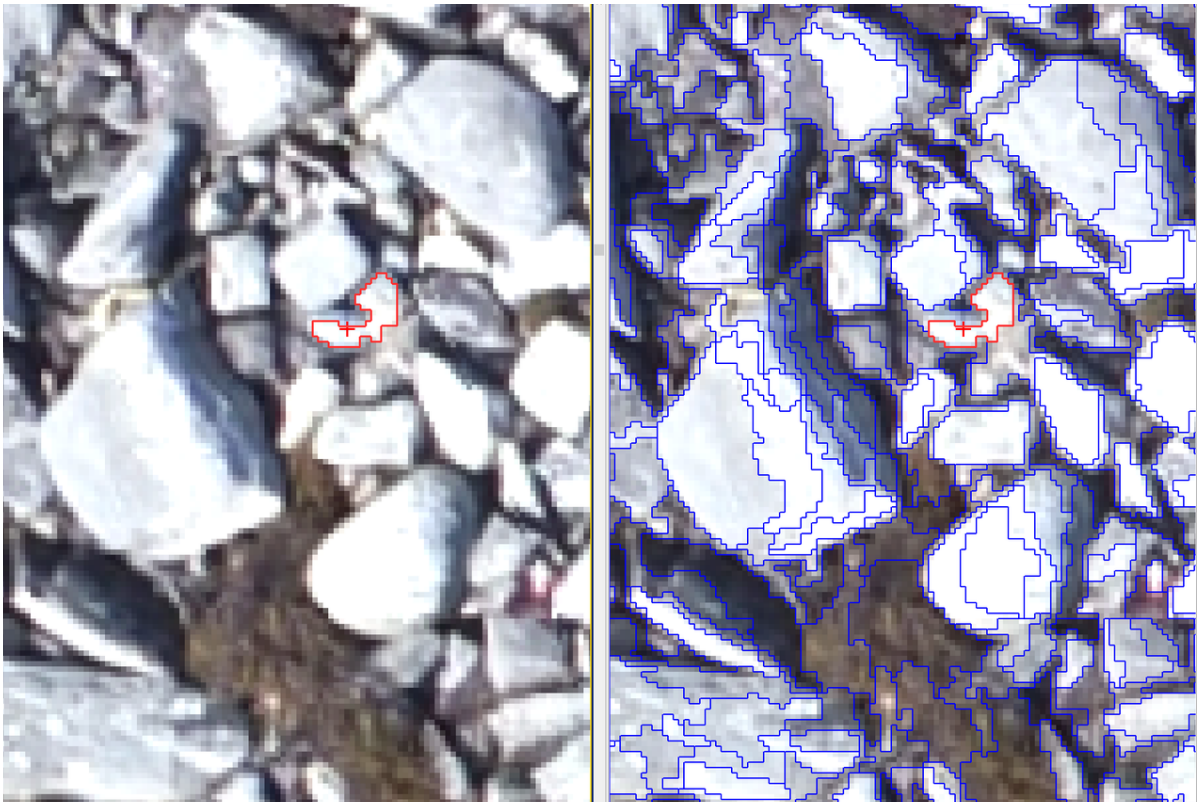


Figure 19. Screenshot from the eCognition essentials software package with an example of a small grey plastic object marked by the red cross and the segmentation depicted as polygons with red/blue outlines. The plastic object is segmented together with surrounding pixels. The example originates from the RGB orthomosaic from Dymna (scenario 2).

Figure 20. Close-up picture of the plastic object shown in Figure 19.

3.2 Classification and accuracy assessment

3.2.1 Scenario 1: Akerøya 2019 RGB

The classification of scenario 1 was divided into 7 classes (Figure 21). The best classification scheme consisted of white and blue plastic and four classes of sand mainly comprising different colors (sand, sand wet, sand light, sand overexposed) and one class for seaweed. The dominating class was bright sand highlighted in yellow, plastic classes highlighted in blue and pink (Figure 21). The error matrix for the scenario 1 classification is presented in
Table 6. In all, 12 plastic objects were included in the accuracy assessment, 4 plastic pieces was used as training data and one was excluded since the same plastic piece was covered twice. See Table 5 for information about number of segments used as training data per class. Nine plastic pieces were identified, 3 non-detected and 33 segments was misclassified as plastic. The sensitivity of identifying the plastic pieces was 75%, but due to the FP the precision was only 21% causing the resulting F-score to become 0.33 (Table 7). In other words, the classification had a plastic detection sensitivity of 75 % finding most of the plastic, but at the same time only a plastic classification precision of 21 %, meaning the plastic class was including a lot other than actual plastic. The location of the plastic objects that is identified, non-detected or excluded due to inclusion as training data is shown in Figure 22.

Table 6. Error matrix of scenario 1 classification.

		Classification		
		Plastic	Other	In total
Ground Truth	Plastic	9 TP	3 FN	12
	Other	33 FP	5110 TN	5143
		42	5113	n = 5155

Table 7. Measures of sensitivity, precision and accuracy assessment (F-score) for scenario 1.

Measure	Value
S	75% (9/12)
P	21% (9/42)
F-score	0.33

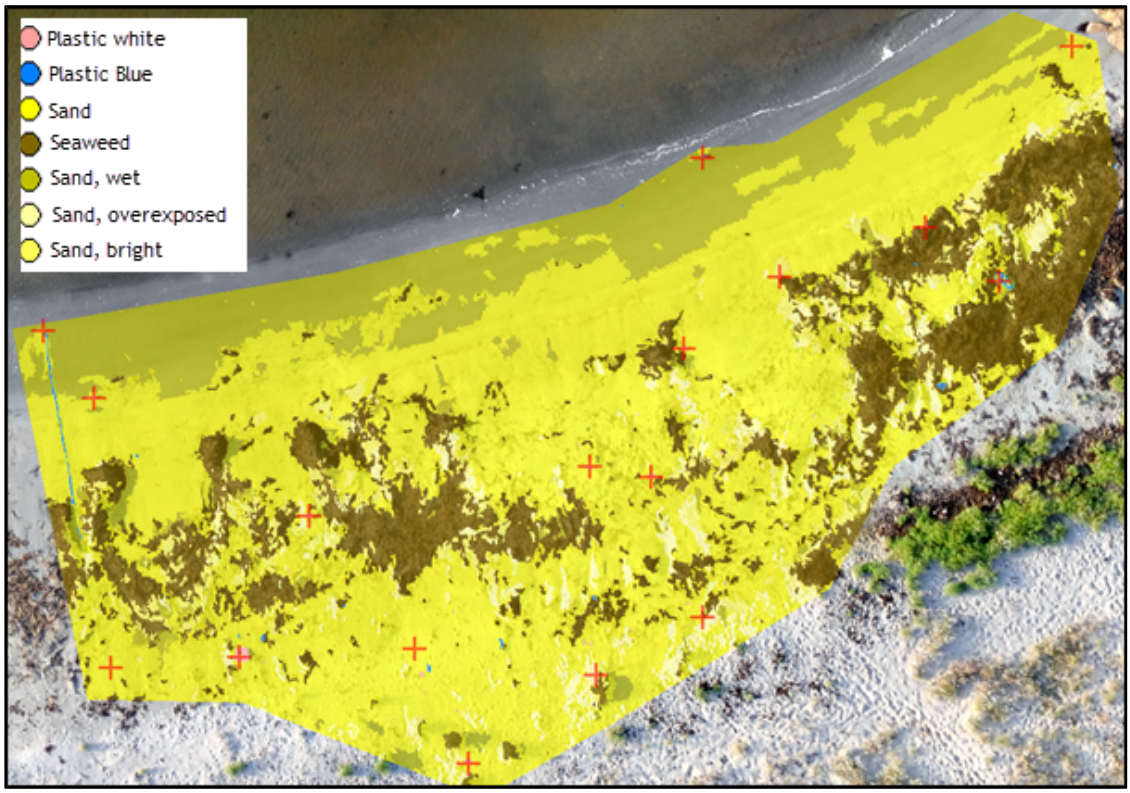


Figure 21. Classification result of scenario 1 with the location of the plastic pieces marked by red crosses.

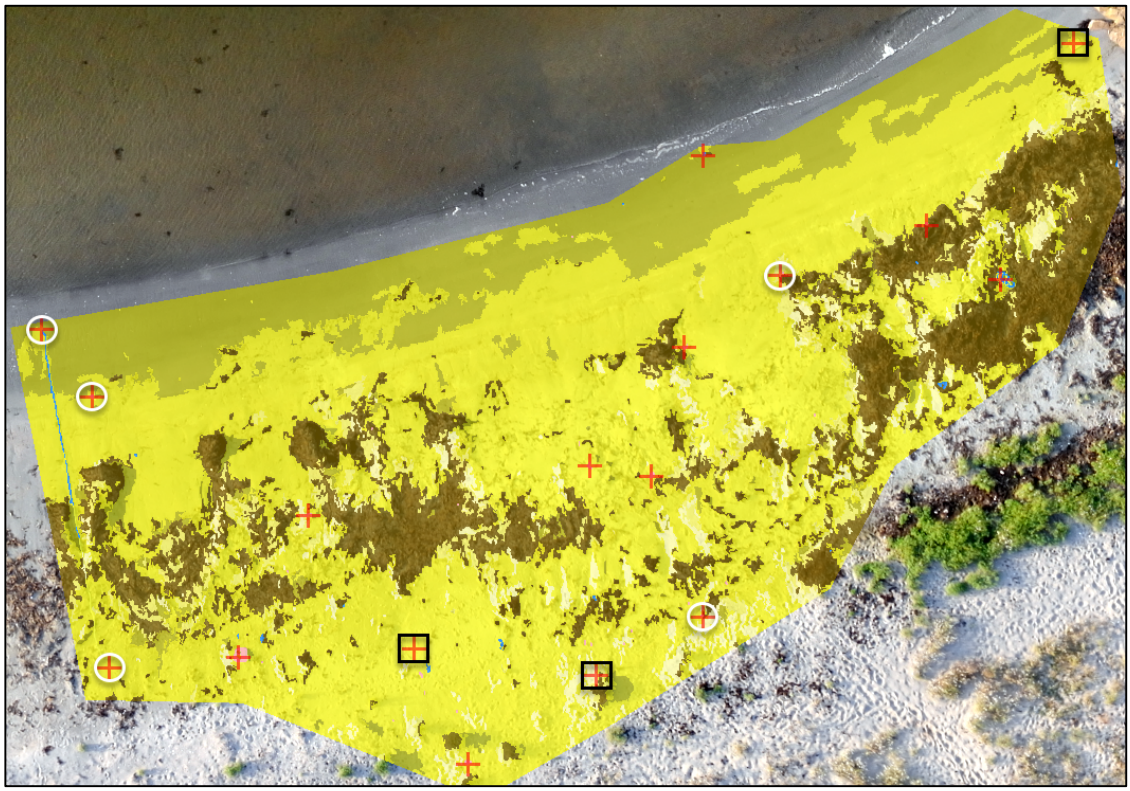


Figure 22. Identified plastic pieces marked with red crosses, non-detected marked with red crosses outlined with a black line and excluded objects with white circle around the red crosses.

3.2.2 Scenario 2: Dymna 2018 RGB with pre-set configuration

The classification of Scenario 2 was also divided into 7 classes (Figure 23). Since the configuration of the classification origin from scenario 1, the classification scheme is the same. The brown seaweed class was dominating the ROI, again segments classified as plastic was attributed blue and pink colors.

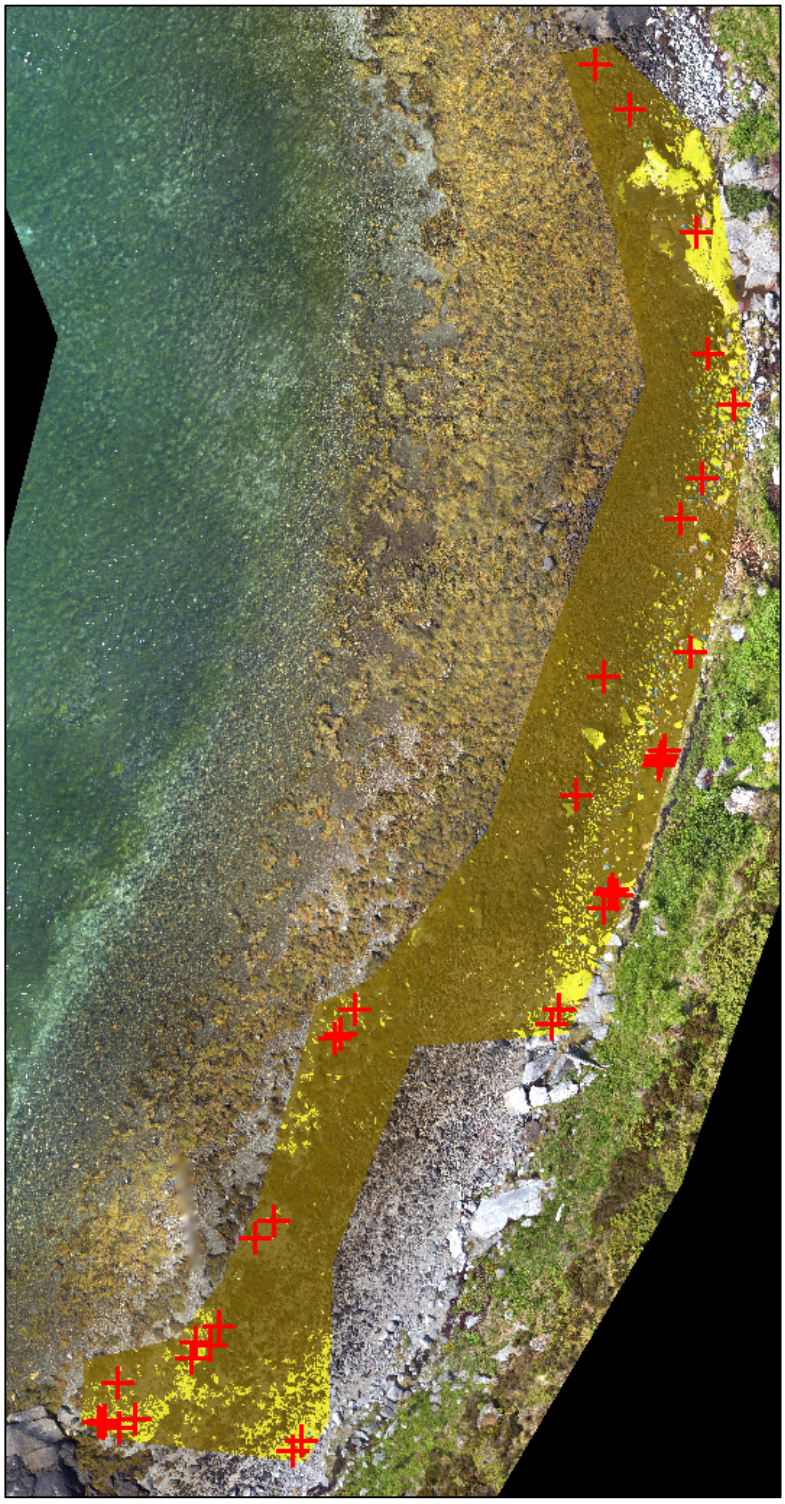


Figure 23. Classification results from scenario 2 with the plastic objects marked as red crosses.

The classification was done with the configuration from Akerøya 2019 RGB and no new training data was used. Thus all 37 plastic pieces was included in the accuracy assessment. The error matrix is displayed in Table 8, 24 plastic pieces was found, 13 non-detected and 525 segments misclassified. A sensitivity of 65% is accompanied by a precision of 4%, the F-score is 0.08 (Table 9). The location of the plastic pieces that was identified or non-detected is displayed in Figure 24 for the northern half of the ROI and Figure 25 for the southern half.

Table 8. Error matrix of the scenario 2 classification.

Classification				
		Plastic	Other	
Ground Truth	Plastic	24 TP	13 FN	37
	Other	525 FP	28721 TN	29246
		549	28734	n = 29283
Time frame:	minimum			

Table 9. Statistical measures of sensitivity, precision and F-score for scenario 2.

Measure	Value
S	65% (24/37)
P	4% (24/549)
F-score	0.08

Figure 24. Scenario 2 (northern half): Location of identified plastic pieces are marked with red crosses, non-detected pieces are marked with black squares.

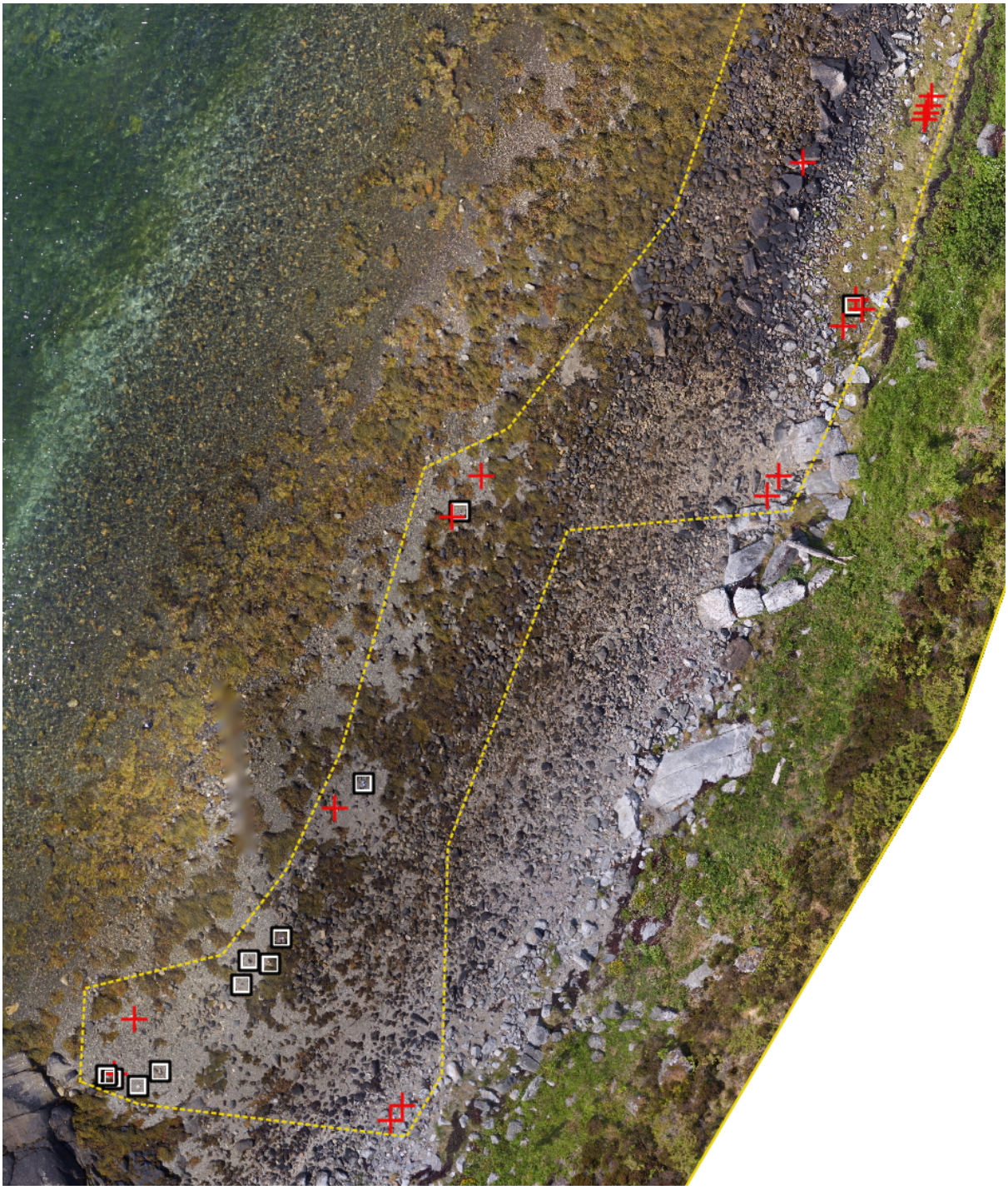


Figure 25. Scenario 2 (southern half): Location of odentified plastic pieces are marked with red crosses; non-detected pieces are marked with black squares.

3.2.3 Scenario 3: Dymna 2018 RGB

The classification of Scenario 3 was divided into 11 classes (Figure 26). The classes included plastic as one class, stone/rocks was six classes due to different colors (stone overexposed, stone light, stone medium, stone dark, stone miscellaneous and stone covered with black lichen), gravel, two classes of seaweed of different colors (seaweed and dark seaweed) and grass.

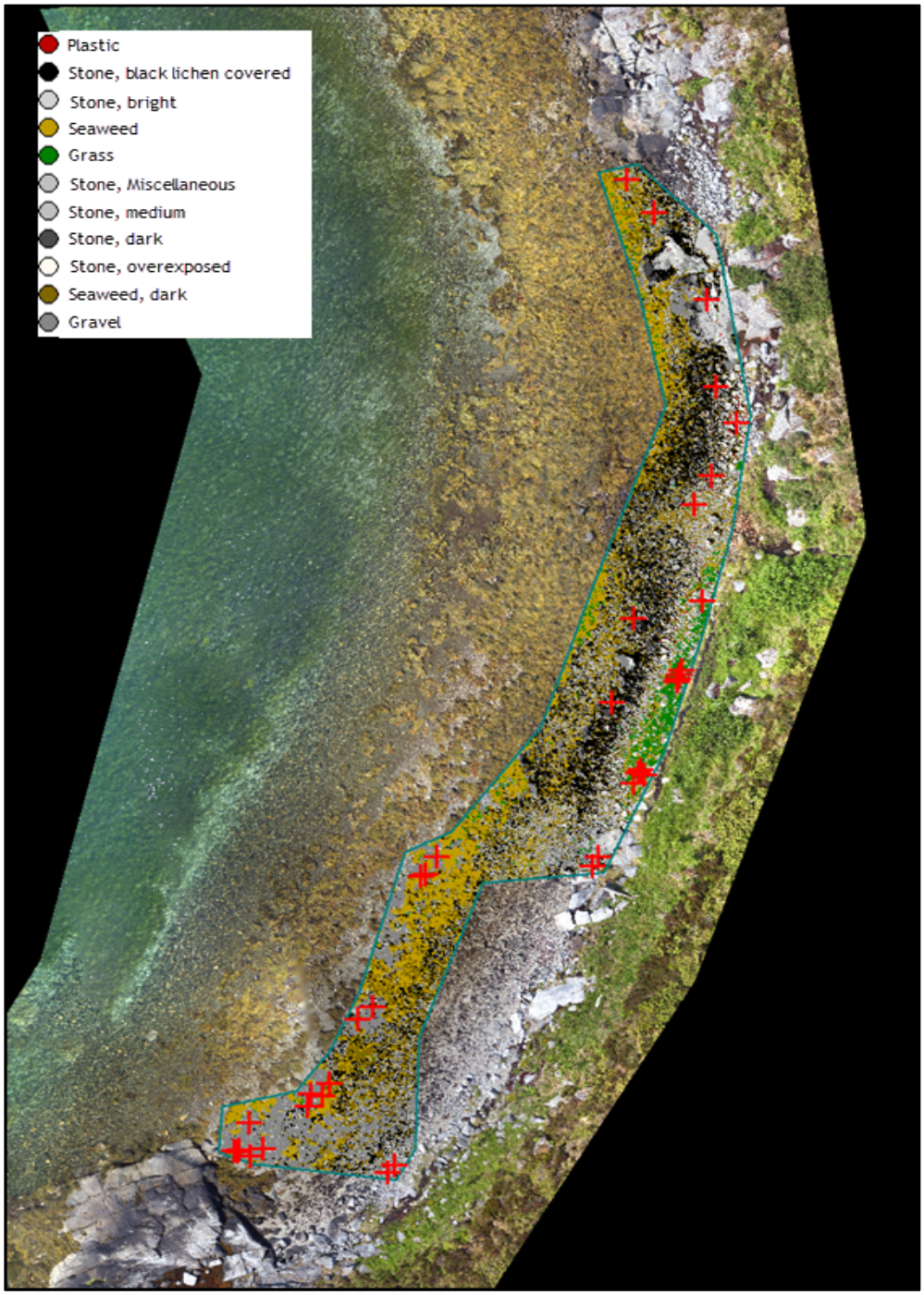


Figure 26. Classification result from scenario 3 with the location of the plastic pieces marked by red crosses. ROI marked as green outline.

Seven plastic objects were used for training, leaving 30 plastic objects to be included in the accuracy assessment. The error matrix shows 13 objects was identified, 17 non-detected and 362 misclassified (Table 10). The values above resulted in a sensitivity of 43%, precision of 3% and a accuracy (F-score) of 0.06 (Table 11). The locations of the plastic pieces that was identified or non-detected are displayed in Figure 27 for the northern half of the ROI and Figure 28 for the southern half of the investigated area.

Table 10. Error matrix of scenario 3 classification.

		Classification		
		Plastic	Other	
Ground Truth	Plastic	13 TP	17 FN	30
	Other	362 FP	59211 TN	59573
		375	59228	n =59603
Time frame:	2 hours			

Table 11. Statistical measures of sensitivity, precision and F-score for scenario 3.

Measure	Value
S	43% (13/30)
P	3% (13/375)
F-score	0.06

Figure 27. Scenario 3 northern half: Location of detected plastic pieces are marked with red crosses; non-detected pieces are marked with black squares. Plastic pieces used for training data are marked with yellow circles.

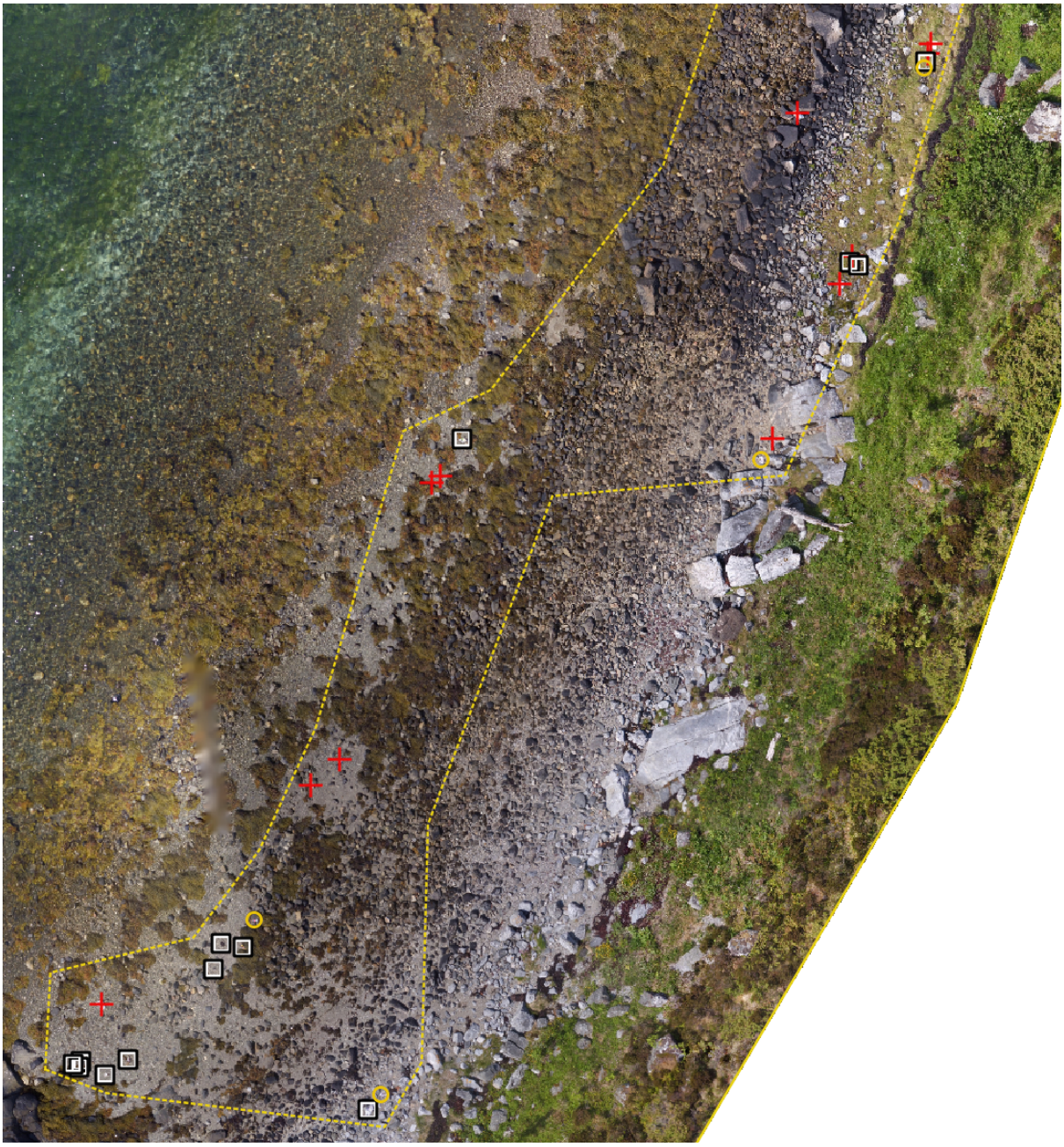


Figure 28. Scenario 3 southern half: Location of detected plastic pieces are marked with red crosses; non-detected pieces are marked with black squares. Plastic pieces used for training data are marked with yellow circles.

3.2.4 Scenario 4: Akerøya 2019 MS

The classification of scenario 4 was divided into 7 classes (Figure 29). The best classification scheme fell like in scenario 1 into; two plastic classes, one for white plastic and one for blue plastic, two classes of sand of different colors (sand wet, sand bright), two classes for seaweed (seaweed and dark seaweed) and one class for overexposed areas.

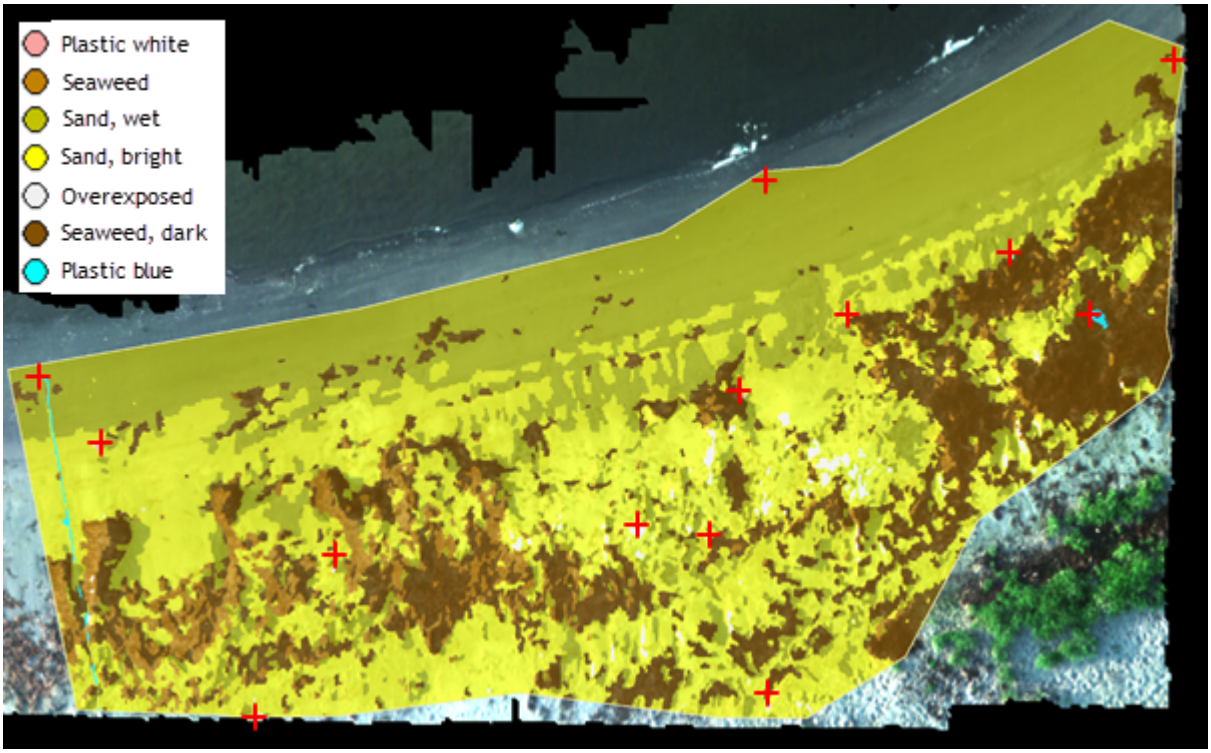


Figure 29. Classification result from scenario 4 with the location of the plastic pieces marked by red crosses.

Since the multispectral dataset of Akerøya was covering a smaller area than the RGB dataset of Akerøya, it resulted in fewer plastic objects than in Scenario 1. In total, 13 plastic pieces were included, two objects was used as training data for the plastic classes leaving 11 plastic pieces for the accuracy assessment shown in Table 12. The statistical measurements of accuracy were 55% for the sensitivity, 87% for the precision and 0.67 for the accuracy assessment (F-score, Table 13).

Table 12. Error matrix of scenario 4 classification.

		Classification		
		Plastic	Other	
Ground Truth	Plastic	6 TP	5 FN	11
	Other	1 FP	9830 TN	9831
		7	9835	n = 9842
Time frame:	4 hours			

Table 13. Statistical measures of sensitivity, precision and F-score for scenario 4.

Measure	Value
S	55% (6/11)
P	87% (6/7)
F-score	0.67

The location of the plastic pieces that was identified or non-detected as well as used as training data are displayed in Figure 30. Red crosses mark the identified objects, black squares mark the non-detected objects and the yellow circles identify objects that was used for training data.

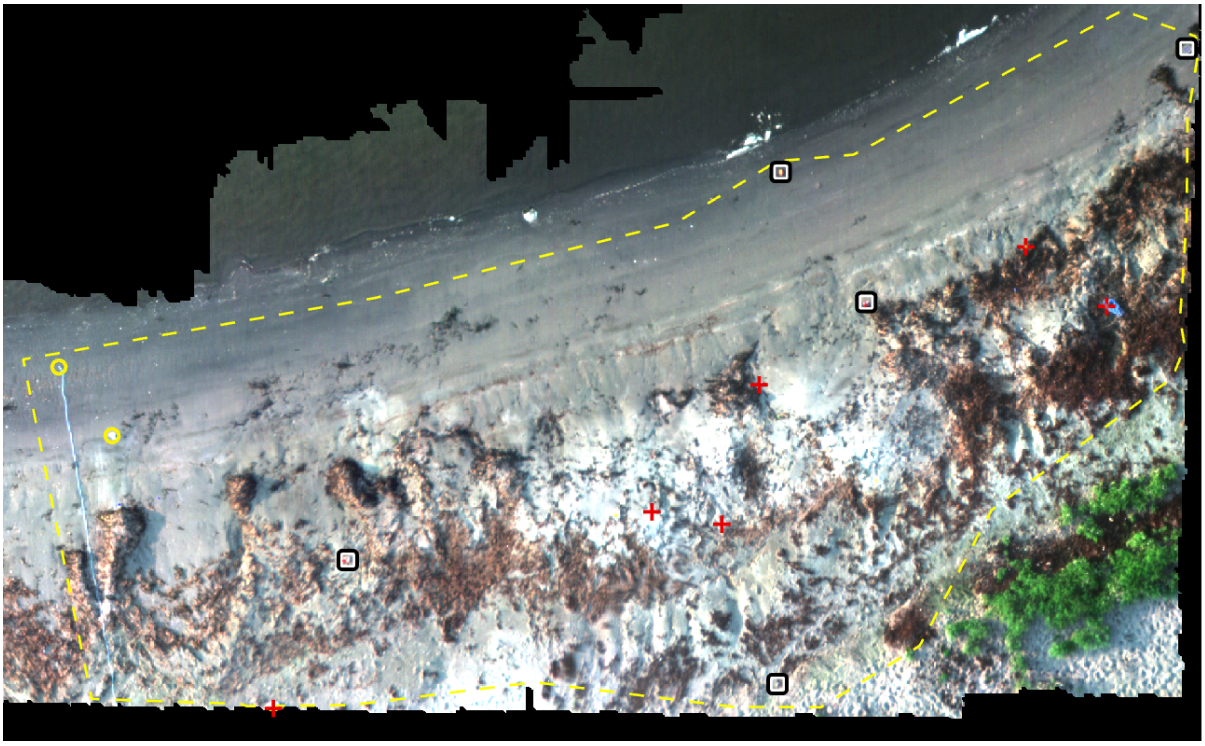


Figure 30. Scenario 4: Location of detected plastic pieces are marked with red crosses, non-detected pieces are marked with black squares. Plastic pieces used for training are marked with yellow circles.

3.2.5 Summary of the statistical measurements of accuracy

The statistical values from the accuracy assessment of the four scenarios above are displayed in Table 14. The detection accuracy of the classification from scenario 1 is performing the best and scoring the highest with a sensitivity at 75%. Both precision and F-score was highest with the classification in scenario 4 that was based on the multispectral dataset.

Table 14. Comparing the statistical measurements of accuracy from all four scenarios.

Classification	Scenario no.	Sensitivity	Precision	F-score
Akerøya 2019 RGB	1	75% (9/12)	21% (9/42)	0.33
Dymna 2018 RGB (preset configuration)	2	65% (24/37)	4% (24/549)	0.08
Dymna 2018 RGB (Fine tune)	3	43% (13/30)	3% (13/375)	0.06
Akerøya 2019 MS	4	55% (6/11)	87% (6/7)	0.67

3.3 Area of True and False positives

First having detected plastic pieces in a beach zone, it is possible from the classification output of the eCognition software package to calculate the total area of detected objects in a GIS program, which is useful when aiming to quantify the amount of detected plastic in a specific beach zone. The area of TPs, FPs and ROI for scenario 1, 3 and 4 as well as the percent coverage of plastic out of the total investigated area are presented in Table 15. In scenarios 1, 3 and 4 the sum of the area of TP and FP were 0.15%, 0.31% and 0.06% for the three scenarios, respectively.

Table 15. Area of TP, FP, ROI and the percentage the sum of the area of TP and FP is of the ROI area of scenario 1,3 and 4.

Scenario no.	TP area m^2	FP area m^2	ROI area m^2	TP + FP area percentage of ROI area
1	0.35	0.27	420	0.15%
3	0.23	1.45	536	0.31%
4	0.18	0.03	360	0.06%

3.4 Size of detectable objects

The size of the objects of interest is of relevance for the detection success and thus we performed a simple test to access the response of size on the detection success. At the Dymna study site, four rectangular objects of white plastics of 28, 14, 7 and 3.5 cm, respectively, were placed inside the ROI functioning as a scale of the detectable object sizes. All four rectangles were identified in the segmentation, and in scenario 2 also classified as plastic objects even though the main central part of the largest rectangle was classified as other than plastic (see yellow square in Figure 31). The smallest detectable object was then 3.5 x 3.5 cm, which is 12.25 cm^2, corresponding to approximately 15 times the pixel area of the RGB imagery used in scenario 2 (0.9 x 0.9 cm). Objects down to approximately 4 times the pixel width can be detected.

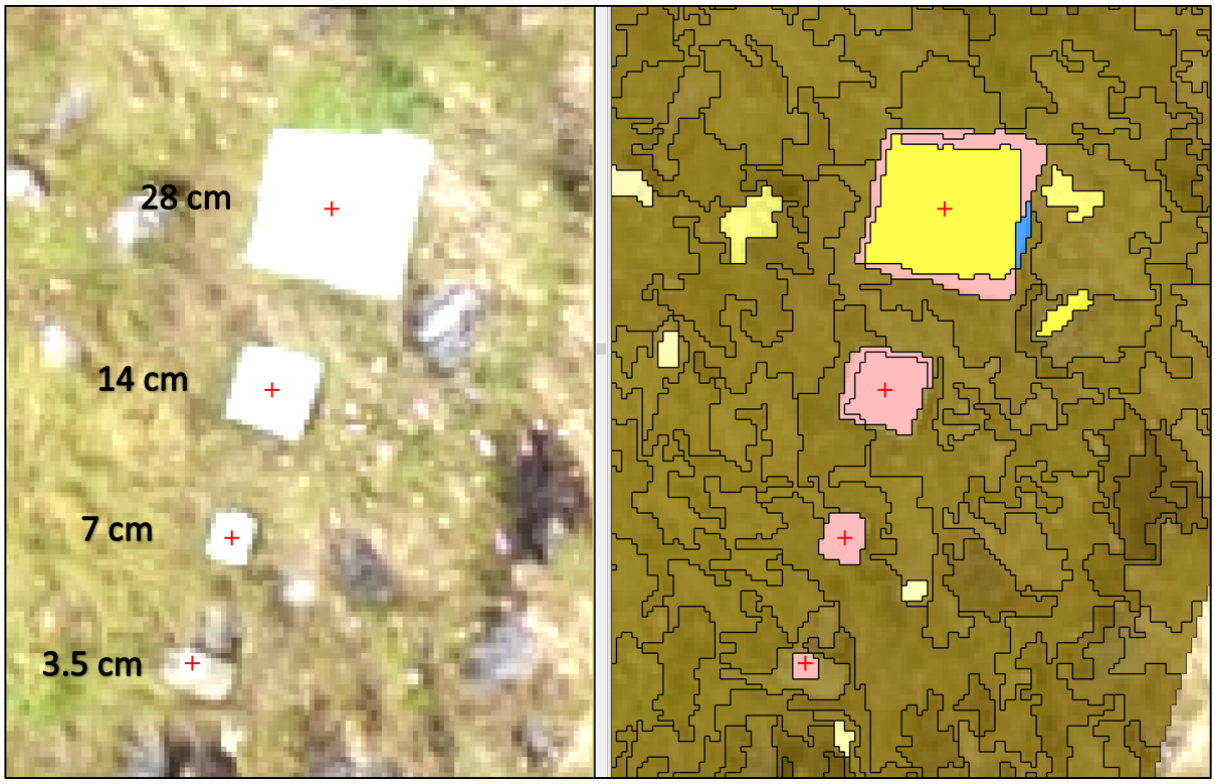


Figure 31. The four plastic rectangles placed at the Dymna study site. Left: RGB orthomosaic of the four rectangles with sizes in centimeters. Right: Classification result from scenario 2 displayed together with the used segments marked by black lines. Light red and blue color represent the two plastic classes, while the yellow class represents the class for sand.

4 Discussion

Scenario 1 performed best, regarding the plastic detection sensitivity (75%) probably due to a simpler background color and texture at Akerøya compared to Dymna. The plastic objects were easy to identify on a relatively homogenous background dominated by sand at Akerøya. Whereas the more heterogeneous background at Dymna was more challenging, comprising various stone sizes, overexposed stones, grass speckled with blooming white flowers and gravel. Somewhat surprising, the sensitivity of scenario 2 (Dymna) with the preset configuration from scenario 1 (Akerøya) performed with the second highest score of 65%, outperforming both scenario 4 and 3 with 55% and 43%, respectively. However, the relatively high sensitivity of Scenario 2 was in combination with a low precision of 4%, resulting in a low F-score of 0,08 for this scenario. Hereby having a plastic class that is detecting 65 % of the plastic pieces, but as well consist of 96 % other than plastic.

When the configuration from Scenario 1 was used in Scenario 2, it performed better on all three measures compared to Scenario 3, still, the precision and the F-score ended up in almost the same as Scenario 3. The difference in used classifier algorithm (SVM in Scenario 2 and RF in Scenario 3) possibly blur the comparison a bit, but as the best possible classification was the objective in Scenario 3, RF was evaluated in the process and found to perform better than SVM. Thereby the performance of Scenario 3 was a little less than Scenario 2 and can still be considered a valid overall result. In the more detailed comparison, it is not possible to pinpoint to what degree differences arise from the choice of classifier or the tailored adjustment of the remaining settings. Worth noting was that the number of FP decreased with 174 (almost 1/3) from 549 in Scenario 2 to 375 in Scenario 3. The concurrent decrease in TP was the main reason for the lack of improvement of the precision. This indicates a connection between the FP and TP; typically a decrease of the FP as well result in a decrease in TP and vice versa demonstrating that increased sensitivity and thus success rate for detection (TP) also follows an increasing number of False Positives. At Dymna this connectedness is between segments of plastic objects and the overexposed parts of stones and shell. An increase in the TP segments of plastic objects in the plastic class would also increase the FP segments of overexposed parts of stones included in the plastic classes. The connectedness probably exists because some segments of the plastic pieces are too similar with other segments of overexposed parts of stone and shell. The bands and features included (Table 5) in the classification process is lacking information that reveal the difference between the segments. This additional information could be from an extra band of another spectral wavelength.

The structural differences in the background and texture of the beach zone between the two locations were also expected to influence the precision. The number of potential FP at Dymna (overexposed stones) seems to exceed the potential FP at Akerøya (overexposed sand, small stones and shells). That was also reflected in the scenarios with the two highest precision measures found in Scenario 1 and 4, i.e. data from Akerøya, respectively at 21% and 87% as opposed to the Dymna data, where classifications resulted in precision values of 4% and 3% in scenario 2 and 3, respectively. This indicate the importance of the characteristics of the surroundings and the complications caused by areas with spectral similarity of plastic objects. Here again additional information making the distinction between FP segments and TP segments would probably improve the classification by decreasing FP and increasing TP.

Comparing the F-score of Scenario 1 and 4 showed the double value in Scenario 4 (0.33 compared to 0.67) even though the sensitivity was considerably higher in Scenario 1 with a difference of 20 percentage points (75% compared to 55%). A likely reason was the inclusion of the NIR band information that improved the classification by lowering the percentage of FP as the high precision of 87% reflects. However, a lowering of FP also resulted in a lowering of the TP, as seen at Dymna. The

relation between number of TP and FP showed to be relevant comparing Scenario 1 and 4. Since the sensitivity was different for the two scenarios, the uncertainty exist, whether the FP could be reduced in Scenario 1, while keeping TP high enough to achieve a similar F-score as in Scenario 4. Still the inclusion of the NIR band was considered as an advantage, referring to high precision along with the higher F-score. Another aspect of comparing Scenario 1 and 4 was the difference in ROI and thereby the included plastic objects in the two scenarios. Two of the three non-detected plastic objects from Scenario 1 was not included in Scenario 4, which compromise the direct comparability of the two scenarios classification accuracy. However, our assessment is that the high precision would not have decreased substantially if the plastic objects were to be included.

The time required for applying a preset configuration, as done in Scenario 2, enhance the speed and usability of the classification process, making it relevant to investigate how to improve this. Even though beach background was not the same in Scenario 4 and Scenario 2, an inclusion of additional information from e.g. NIR could possibly have reduced the number of FP in Scenario 2. Hereby increasing the low precision and potentially leading to a higher F-score. In a longer perspective, building up a database on classifications, grouped according to certain coastal properties, could improve re-use of classification settings and thus improve plastic detection on routine basis in terms of workhour load.

Since results from this pilot study were based on relatively few ground truthed plastic objects, it should only serve as an early demonstration of some of the methodological possibilities and challenges using remotely sensed UAV data for plastic detection. During the process of selecting optimal settings we experienced large variations in classification results, with only minor changes of settings, e.g. in the segments used as training data. The size of available ground truth points and the high variation between the plastic pieces (size, material, and color) resulted in few segments to include as training data and therefore quite possibly affecting the robustness of the findings. An earlier study by Qian et al. (2014) found that optimization of the methods of selecting best possible settings for classification algorithm together with increasing training data increased the accuracy of the resulting OBIA classification. Also automation and optimization of the selected segmentation settings with quantitative evaluation measures could possibly improve the segmentation (Chen et al., 2019) and thereby likely also the accuracy of the plastic detection. With the relatively limited extent of this project in mind, the results depicted here increase the expectations regarding potential findings in future projects. Meaning future projects could lead to cost effective UAV-based solutions for detecting and monitoring plastic objects in the coastal zone, obtaining classifications with higher F-scores. Such future studies should ideally include collection of more comprehensive and more systematically collected training datasets as well as additional information, e.g. from alternative color spaces (Gonçalves et al., 2020) or bands from the near infrared (NIR) and shortwave infrared (SWIR) domain (Garaba & Dierssen, 2018; Acuña-Ruz et al., 2018) to improve the detection performance. This will increase the usability and value of the technology and therefore support further investigation in technologies of plastic detection based on data collected from UAV.

4.1 Recommendations

When training models in order to improve classification algorithms it is essential to secure sufficient ground truth data to assess the accuracy of the classification for proper algorism validation. The numbers of ground truth objects (data set) should be carefully planned and assessed in the project planning phase prior to data collection. A minimum of 30 and preferably 50 ground truth samples per class has been proposed for the generation of high quality error matrix generation (Green et al., 2017). In smaller projects, as the present one we recommend to balance the desire for statistical strength and needed accuracy assessment with what is feasible to collect of ground truth data with the available resources.

In determination of the needed number of ground truth samples the available bands of data have to be included in the consideration. For instance, an RGB dataset forces the need for plastic pieces of similar color or alternatively adequate number of pieces per color to secure creation of classes for each color. Too few ground truth samples result in the need of a class for undefined colors, which will result in a weaker model.

Possibly the plastic color would be a less discriminating factor, if NIR and/or SWIR bands were available, either directly or for calculating specific indices with high affinity for the properties of plastic. Plastic in various colors absorbs similar wavelength in the electromagnetic spectrum as displayed in Figure 32**Feil! Fant ikke referansekilden.**. This could potentially reduce the need for collecting training data according to plastic objects color. Additionally, it could possibly improve the precision by reducing the number of the FP as indicated by the improvement of the precision of the classification in Scenario 4 compared to Scenario 1.

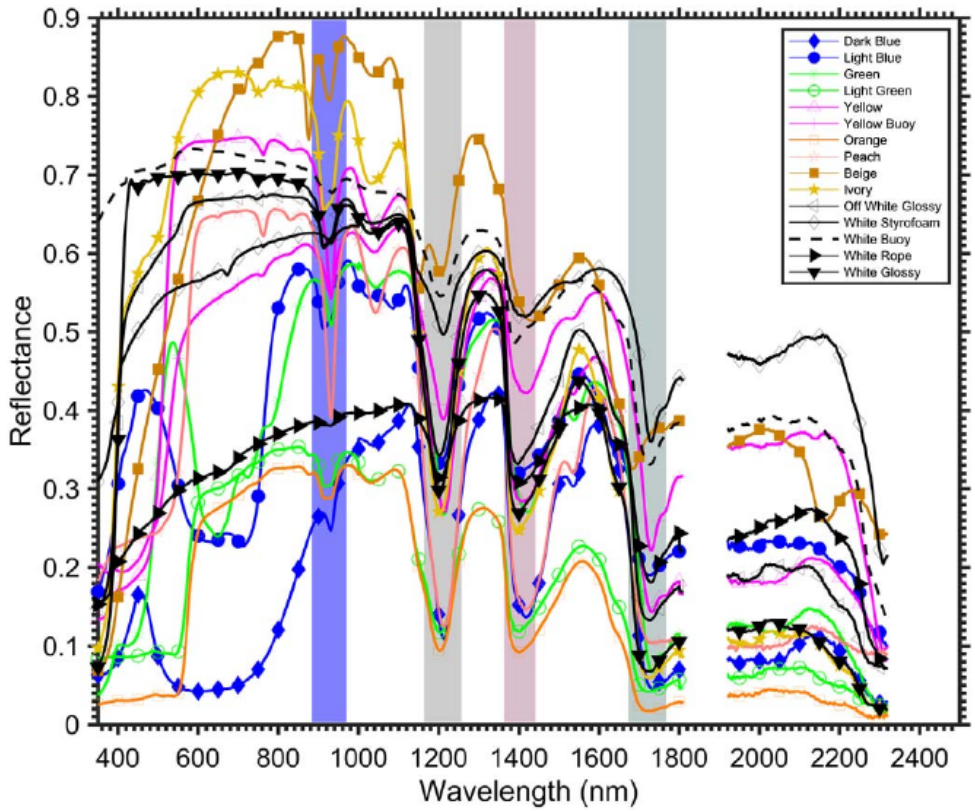


Figure 32. Reflectance at different electromagnetic wavelength of colored plastic pieces from marine litter (Source: Garaba & Dierssen, 2018).

Even with low precision due to many FP segments, results from the classification show that the ROI area can be substantially reduced, to the size of the TP and FP segments area sum (see section 3.3). In Scenario 1, 3 and 4 the sum of the TP and FP areas was lower than 0.5% of the total ROI area. This could be a relevant simplifying approach to reduce the amount of data, without losing too much information. In future studies, collecting even more data, this could become a valuable step for faster data processing with a minimal loss of information.

Using Scenario 1 as an example, the plastic objects were found with 75% sensitivity and even though the precision was low, meaning many segments were classified as FP, only 0.15% of the full ROI area comprised the TP and FP detected objects. The used approach with OBIA on RGB (or RGB+NIR) data could possibly be used as a method for reducing the total amount of information. Filtering the total dataset to the most relevant areas, and possibly adding a buffer area around detected objects could serve as an efficient way to significantly simplify and reduce the dataset size and speed up processing time for additional analyses on a targeted portion of the ROI area.

In future developments of the detection technique, it would be helpful to include a description of the included plastic objects regarding size, color, material, and contrast to the surroundings. Characteristics of the plastic objects and neighboring surroundings could lead to further insights in the classification as well as the segmentation performance and enabling more quantitative measures on the detection.

Segmentation represent a highly important step in the OBIA process since classifications are bound to the segmentation. How well the segments represent objects of varying size, color and contrast to surroundings could possibly be improved by including multiple segmentation levels. This functionality is available in the more advanced but also more expensive software package “eCognition Developers”. The “eCognition Essentials” software package used in this project is restricted on this aspect. Automized optimization of the segmentation and classification parameters are also features that can improve for detection performance and therefore relevant. The “Developers” version and other solutions enabling automized optimization will possibly result in improved detection of plastic objects of relative heterogeneous nature and is therefore recommended further tested in future projects.

The accuracy assessment was in this project based on detection without taking the object area into account. If just one small segment was classified as plastic the whole object was considered detected. If the analysis instead of the number of objects focused on area of the correctly detected objects it could in future project be relevant to include the area, as previously recommended (Radoux & Bogaert, 2017).

5 Conclusion

Object based image analysis (OBIA) classification was applied to detect plastic objects including segmentation and supervised classification. The segmentation was highly important for detecting of plastic objects since adjustment of the segmentation affected the classification accuracy. Performance of the segmentation was not tested in this project but considered highly relevant for future projects. Also, an automated process, that iteratively optimize segmentation settings would be valuable for the performance, especially for larger datasets.

With RGB imagery, macroplastic pieces (>3.5 cm) was detected with a sensitivity of up to 75% and a precision of 21% resulting in a F-score of 0.33 (Scenario 1, Akerøya). Illustrating that it is possible to detect macroplastic from RGB imagery even though the plastic class includes 79% false positive segments. The configuration from this classification detected plastic pieces with a sensitivity of 65% when applied to another area (Scenario 2, Dymna). Here the percentage of true positives decreased to 4% resulting in a F-score of 0.08. Surprisingly a finetuning of the classification to the specific area (Scenario 3, Dymna) did not improve the accuracy of the classification but instead reduced the sensitivity to 43% detected plastic pieces and the precision to 3% true positive segments in the plastic class, together giving a F-score of 0.06.

It seems like a more heterogeneous ground surface characteristic affected the classification accuracy negatively, as from comparing the results from Scenario 1 (Akerøya being relatively homogeneous) to Scenario 2 and 3 (Dymna being relatively more heterogeneous).

The highest gained accuracy assessment (F-score) was 0.67 with a sensitivity of 55% and a precision of 87% based on 11 ground truth plastic objects. This result originates from a sandy beach with RGB and NIR data (scenario 4, Akerøya), demonstrating that plastic detection from a UAV platform is possible and that the NIR band gave additional discriminating value. Including the NIR band improved the F-score from 0.33 to 0.67 (respectively Scenario 1 and 4).

Plastic objects were detected down to sizes as small as 3.5 cm x 3.5 cm corresponding to approximately 15 times the pixel area or 4 times the pixel width of the used RGB data.

This pilot study indicates the potential of plastic detection from flying drones, but results should be used with caution since the dataset was relatively small and the robustness of the conclusion likewise.

We compiled a list of recommendations and considerations for future plastic detection studies (Table 16), and recommend this list being assessed when designing new projects using unmanned aerial systems and OBIA.

Table 16. Recommendations and considerations for planning future plastic detection studies.

Checklist	
Survey conditions: ***Plastic object characteristics***	Plastic object characteristics - **Color** [consider enough similar colored plastic objects for each color, most importantly if only RGB bands are available for detection. Too few objects, to both train and classify, likely result in reduced detection performance] - **Material** [different plastic types can reflect the sunlight differently, creating additional challenges in detection] - **Shape** [differently shaped plastic objects can complicate fitting of the segmentation] - **Size** [Diverse sized plastic objects can complicate the fitting of segmentation possibly reducing the plastic detection, especially if only one level of segmentation is available] - **Number** of similar plastic pieces [30-50 objects per class are advisable if resources are available]
Survey conditions: ***Region of interest considerations***	Plastic object location: - Plan the region of interest vs. sensor **FOV** if single image analysis is desirable) - Consider **light conditions** [sunlight create shadows which complicate OBIA data processing and are ideally avoided] - **Color difference** to plastic objects [The larger the contrast between the plastic color and the background improves the object detection] - Consider the region of interest **homogeneity** compared to the plastic objects. [Increasing differences between the background and the plastic objects improves detection if the bands are spectrally separable. Heterogeneous background increase complexity and reduce detection performance.]
Survey conditions: ***Drone flight***	Flight planning: - Consider flight altitude, sensor resolution, GSD compared to plastic object sizes. [OBIA is useful when plastic objects consist of several pixels] - If image analysis on single image: Single image FOV compared to region of interest. [consider if all plastic objects are visible in a single image versus orthomosaic] - If image analysis is carried out on orthomosaic it is important to ensure sufficient overlap between images to reduce artifacts/blurry areas in the orthomosaic. [Blurry areas can cause problems in the detection of the plastic objects since the border between the plastic object and the surroundings will be less clear possibly influencing the segmentation] - Needed vs. available bands (e.g. RGB, NIR, SWIR) in sensor on the drone. - NIR and SWIR bands looks like they hold more information and gives additional information to improve detection
Software	Ideally budget for eCognition Developers or other solutions that hold functionality supporting: - Multi-level segments [Can likely improve detection performance of different sized and shaped objects] - Automatic parameter optimization [can possibly reduce time consumption and potentially improve detection]

6 References

Acuña-Ruz, T., Uribe, D., Taylor, R., Amézquita, L., Guzmán, M. C., Merrill, J., Martínez, P., Voisin, L., & Mattar B., C. (2018). Anthropogenic marine debris over beaches: Spectral characterization for remote sensing applications. Remote Sensing of Environment, 217, 309–322. DOI: 10.1016/j.rse.2018.08.008

Aminipouri, M. (2009). Object-Oriented Analysis of Very High Resolution Orthophotos for Estimating the Population of Slum Areas, Case of Dar-Es-Salaam, Tanzania. 74.

Andersen, G. S. (2019) Å se en havhest dø. Forestillinger og fakta om plast i havet. Spartacus, Oslo. 99 pp. (In Norwegian).

Chen, Y., Chen, Q., & Jing, C. (2019). Multi-resolution segmentation parameters optimization and evaluation for VHR remote sensing image based on mean NSQI and discrepancy measure. Journal of Spatial Science, 1–26. DOI: 10.1080/14498596.2019.1615011

Garaba, S. P., & Dierssen, H. M. (2018). An airborne remote sensing case study of synthetic hydrocarbon detection using short wave infrared absorption features identified from marine-harvested macro- and microplastics. Remote Sensing of Environment, 205, 224–235. DOI: 10.1016/j.rse.2017.11.023

GESAMP (2019). Guidelines or the monitoring and assessment of plastic litter and microplastics in the ocean (Kershaw P.J., Turra A. and Galgani F. editors), (IMO/FAO/UNESCO-IOC/UNIDO/WMO/IAEA/UN/UNEP/UNDP/ISA Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). Rep. Stud. GESAMP No. 99, 130 pp.

Gonçalves, G., Andriolo, U., Pinto, L., & Bessa, F. (2020). Mapping marine litter using UAS on a beach-dune system: A multidisciplinary approach. Science of The Total Environment, 706, 135742. DOI: 10.1016/j.scitotenv.2019.135742

Green, K., Congalton, R. G., & Tukman, M. (2017). Imagery and GIS: Best practices for extracting information from imagery. Esri Press.

Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R. Perryman, M., Andrady, A., Narayan, R., Law, K. L. (2015). Plastic waste inputs from land into the ocean.

OSPAR (2010). Quality Status Report 2010, OSPAR Commission, London, 176 pp.

Qian, Y., Zhou, W., Yan, J., Li, W., & Han, L. (2014). Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. Remote Sensing, 7(1), 153–168. DOI: 10.3390/rs70100153

Radoux, J., & Bogaert, P. (2017). Good Practices for Object-Based Accuracy Assessment. Remote Sensing, 9(7), 646. DOI: 10.3390/rs9070646

UNEP (2016) Marine plastic debris and microplastics – Global lessons and research to inspire action and guide policy change. United Nations Environment Programme, Nairobi.

NIVA: Norway´s leading centre of competence in aquatic environmentes

NIVA provides government, business and the public with a basis for preferred water management through its contracted research, reports and development work. A characteristic of NIVA is its broad scope of professional disciplines and extensive contact network in Norway and abroad. Our solid professionalism, interdisciplinary working methods and holistic approach are key elements that make us an excellent advisor for government and society.

Norwegian Institute for Water Research

Gaustadalléen 21 • NO-0349 Oslo, Norway
Telephone: +47 22 18 51 00
www.niva.no • post@niva.no