Environmental Protection in the Arctic — support of German activities in the Arctic Council in terms of a pilot study on monitoring plastic litter on arctic coastlines applying remote sensing techniques

Final report

German Environment Agency

Umwelt Bundesamt

Environmental Protection in the Arctic – support of German activities in the Arctic Council in terms of a pilot study on monitoring plastic litter on arctic coastlines applying remote sensing techniques

Final report

by

Schnurawa, Marc; Vilela, Raul; Kersten, Anna
BioConsult SH GmbH & Co. KG, Husum

Schulz, Marcus
AquaEcology GmbH & Co. KG, Oldenburg

On behalf of the German Environment Agency

Imprint

Abstract: Environmental Protection in the Arctic – support of German activities in the Arctic Council in terms of a pilot study on monitoring plastic litter on arctic coastlines applying remote sensing techniques

A consistent and Arctic-wide quantification of beach litter is an essential step for the development of a Regional Action Plan (RAP) as intended by the Protection of the Arctic Environment working group from the arctic council (PAME) for 2021. The quantification of beach litter not only gives information on the actual state of pollution, but also allows for setting a baseline, which can be used to validate the possible success of measures implemented in an action plan.

To achieve such a baseline, a uniform methodology is necessary which allows for a cost- and time-efficient monitoring of the Arctic coastlines. Traditional beach litter monitoring should be supported and spatially extended by applying remote sensing techniques, such as satellite imagery or drone surveys. The derivation of a suitable methodology is the major aim of this study. The new methods were tested on several hot spots of beach litter. For this purpose, hot spots were identified before, by analysing existing beach litter data, as well as information on sources and influencing factors (e.g. population density, sewage treatment, etc.). Statistical and modelling approaches were applied, suggesting potential hot spots of beach litter accumulation.

To select a suitable methodology, a literature study was performed comparing the potential of different remote sensing techniques regarding spatial and spectral resolution, time and cost efficiency and their potential for a (semi-) automatic classification. Considering the restrictions of spatial resolution for all remote sensing techniques, this study focuses on macro litter (>2.5 cm). A combination of several methodologies was recommended and performed on beach sections on Greenland and Svalbard, i.e. satellite imagery with high spatial resolution data from drone surveys and results from conventional beach litter surveys applying the OSPAR method.

The drone surveys were evaluated by a manual screening of the drone imagery and the application of machine learning approaches. During the manual screening of the drone imagery, up to 17.5% of the plastic items, identified during the OSPAR (Convention for the Protection of the North Sea and the North-East Atlantic) monitoring, could be detected. The low recovery rate is the result of the litter size distribution with most items <10 cm, the heterogenic beach environment, and the ground sample distance (GSD) (1.4 cm and 3.4 cm for the RGB (Red-Green-Blue) and the visible infrared (VIR) sensor, respectively). A (semi-) automatic classification of beach litter was applied testing machine learning approaches like Random Forest, Support Vector Machine, and Neural Network. The highest overall accuracy (OA) was reached applying Random Forest on VIR imagery with an OA of 90.6%, showing the potential of VIR imagery. Misclassifications of plastic objects applying machine learning algorithms occurred mainly with rocks and wood. As the occurrence of plastic objects was much smaller compared to rocks and wood, wrong assignments had a great impact on the classification maps and plastic was strongly overestimated. To meet up with the complex environmental conditions and the litter size distribution in the Arctic, a sub-centimetre GSD is recommended for future studies in the Arctic. For large areas, the drone-based beach litter monitoring was up to 22-times faster compared to the OSPAR monitoring. For small areas, an OSPAR monitoring could be more time efficient. Anyhow, drone surveys can also be beneficial for small areas, as the beaches in the Arctic are often remote and the time at the beaches is often limited. The results of the WorldView 3 (WV3) imagery showed that beach litter detection with satellite imagery is still limited by its spatial resolution and only large accumulations of litter seem to be detectable. The litter size and the spatial distribution of litter on the monitored Arctic beaches were too low to be identified on the WV3 imagery. A higher spatial resolution or a larger spectral coverage would be required to also detect lower litter accumulations on beaches from satellites. In future studies, sub-pixel detection approaches for beach litter detection should be tested for survey areas with very high litter abundances to determine the minimum coverage of plastics that is required for a successful detection. The results of such an evaluation can further be used to determine the

resolution of satellite imagery needed to detect litter accumulations as found on Greenland and Svalbard.

Kurzbeschreibung: Umweltschutz in der Arktis – Unterstützung der deutschen Aktivitäten im Arktischen Rat durch eine Pilotstudie zum Monitoring von Kunststoffmüll an arktischen Küsten mittels fernerkundlicher oder luftgestützter Methoden

Eine arktisweit einheitliche Erfassung von Strandmüll ist die grundlegende Voraussetzung für die Entwicklung eines regionalen Aktionsplans (RAP) wie von der Protection of the Arctic Marine Environment Arbeitsgruppe des Arktischen Rats (PAME) für 2021 angestrebt. Die Erfassung von Strandmüll gibt dabei sowohl Informationen über den aktuellen Verschmutzungszustand als auch eine Grundlage für die Bewertung des Erfolgs möglicher Aktionspläne.

Um dies zu erreichen, ist eine einheitliche Methodik erforderlich, die eine kosten- und zeiteffiziente Erfassung von arktischen Stränden ermöglicht. Die traditionelle Erfassung von Strandmüll wurde in diesem Projekt durch fernerkundliche Methoden wie Satellitenbilder oder Drohnenerfassungen unterstützt und erweitert. Die Entwicklung einer geeigneten Methodik war das Hauptziel dieser Studie, und die Methodik wurde an ausgewählten Hotspots von Strandmüll auf Grönland und Spitzbergen evaluiert. Zu diesem Zweck wurden zuvor potenzielle Hotspots identifiziert, indem Daten von bereits existierenden Standmüllkartierungen sowie Informationen zu potenziellen Quellen und weiterer geografischer Faktoren (z. B. Bevölkerungsdichte, Abwasserbehandlung usw.) ausgewertet wurden. Zur Auswertung wurden ein statistischer und ein Modell-basierter Ansatz gewählt.

Für die Entwicklung einer geeigneten Methodik wurde eine Literaturstudie durchgeführt, in der das Potenzial verschiedener Fernerkundungsmethoden hinsichtlich räumlicher und spektraler Auflösung, Zeit- und Kosteneffizienz und der Anwendbarkeit einer (halb-) automatischen Klassifizierung verglichen wurde. In Anbetracht der Einschränkungen der räumlichen Auflösung aller Fernerkundungstechniken konzentriert sich diese Studie auf Makromüll (>2,5 cm). Eine Kombination von Satellitenbildern mit Drohnenerfassungen und konventionellen Strandmüllkartierungen wird empfohlen.

Für die Drohnenerfassung wurde eine manuelle Sichtung der Drohnenbilder und eine Anwendung Maschinellen Lernens getestet. Bei der manuellen Sichtung konnten bis zu 17,5% der Plastikobjekte im Vergleich zur OSPAR (Vertrag zum Schutz der Nordsee und des Nordostatlantiks) Erfassung gefunden werden. Die niedrige Erfassungsrate resultierte aus der Größenverteilung der Plastikobjekte mit der Mehrzahl der Objekte <10 cm, der heterogenen Strandumgebung und der Bodenauflösung (ground sample distance: GSD) von 1,4 cm und 3,4 cm für den RGB (Rot-Grün-Blau) - beziehungsweise den VIR (sichtbar – Infrarot) - Sensor. Eine (halb-) automatische Auswertung der Drohnenbilder wurde anhand verschiedener Anwendungen Maschinellen Lernens (*Random Forest, Support Vector Machine*, *Neuronale Netzwerke*) getestet. Die höchste Gesamtgenauigkeit (overall accuracy: OA) wurde unter Anwendung von Random Forest für VIR-Bilder erreicht mit einer OA von 90,6%. Die häufigste Fehlklassifikation von Plastikobjekten fand mit Steinen und Holz statt. Auf Grund des deutlich geringeren Müllvorkommens im Vergleich zum Vorkommen von Steinen und Holz, führte schon ein geringer Prozentsatz an falsch bestimmten Steinen und Holz zu einer erheblichen Überschätzung des Plastikmülls. Um den komplexen Umweltbedingungen und der Größenverteilung der Plastikobjekte in der Arktis gerecht zu werden, wird für zukünftige Studien eine GSD im Subzentimeter-Bereich empfohlen. Für große Flächen war das drohnenbasierte Strandmüll-Monitoring im Vergleich zum OSPAR-Monitoring bis zu 22-mal schneller, wohingegen für kleinere Flächen das OSPAR-Monitoring effizienter war. Trotzdem können Drohnenerfassungen auch für kleine Gebiete von Vorteil sein, da die Strände in der Arktis oft abgelegen sind und die Zeit an den Stränden begrenzt sein kann. Die Ergebnisse der WorldView 3 (WV3) -Bilder zeigen, dass die Erkennung von Strandmüll mit Satellitenbildern immer noch durch die räumliche Auflösung begrenzt ist und nur große Müllansammlungen

erkennbar sind. Eine höhere räumliche Auflösung oder eine größere spektrale Abdeckung wäre erforderlich, um auch geringere Müllansammlungen an Stränden von Satelliten aus zu erkennen. In zukünftigen Studien sollten Anwendung der Sub-Pixel Klassifikation in Untersuchungsgebieten mit sehr hohem Müllaufkommen getestet werden, um die Mindestpixelabdeckung durch Plastikobjekte zu bestimmen, die für eine erfolgreiche Erkennung erforderlich ist. Die Ergebnisse einer solchen Auswertung können verwendet werden, um die Mindestauflösung von Satellitenbildern zu bestimmen, um Müllansammlungen, wie sie auf Grönland und Spitzbergen vorgefunden wurden, erkennen zu können.

Table of content

List of figures

14

List of tables

List of abbreviations

AE	AquaEcology GmbH & Co. KG
ASTD	Arctic Ship Traffic Data
AVIRIS	Airborne Visible/ Infrared Imaging Spectrometer
BCSH	BioConsult SH GmbH & Co. KG
CFD	Computational Fluid Dynamics
CLIFRO	R package for download and visualization of the climate database of New Zealand
DLR	Deutsche Zentrum für Luft- und Raumfahrt e. V.
EMODNET	European Marine Observation and Data Network
EMS	Electromagnetic spectrum
EnMap	Environmental Mapping and Analysis Program
FDI	Floating Debris Index
GNSS	Global Navigation Satellite System
GSD	Ground sample distance
HELCOM	Baltic Marine Environment Protection Commission (Helsinki Commission, HELCOM
HoG	Histogram of orientated Gradients
HR	High resolution
ICG ML	Intersessional Correspondence Group Marine Litter (OSPAR Working Group)
IR	Infrared
ISO	Film speed
ITRES	ITRES Research Limited
KNN	K-nearest neighbors
LDPE	Low Density Polyethylen
LR	Low resolution
ML	Maximum Likelihood classifier
MSFD	Marine Strategy Framework Directive
MTOW	Maximum Takeoff Weight
NetCDF	Network Common Data Form
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NP	Norsk Polar Institute
OA	Orbicon Arctic A/S
OSPAR	Convention for the Protection of the Marine Environment of the North-East Atlantic (named after the predecessors Oslo Convention and Paris Convention)
PA	Producer's Accuracy
PAME	Protection of the Arctic Marine Environment (Working Group of the Arctic Council)

PERMANCOVA	Permutational analysis of covariance
PET	Polyethylene terephthalate
R	Statistical software
RAP	Regional Action Plan
RF	Random Forest
RGB	Red-Green-Blue
ROMS	Regional Ocean Modelling System
SDMTools	R package: Species Distribution Modelling Tools
SWIR	Shortwave Infrared
SVM	Support Vector Machine
UA	User's Accuracy
UAV	Unmanned Aerial Vehicle/ Drone
UBA	German Environmental Agency (Umweltbundesamt)
VIR	Visible Infrared
VIS	Visible spectrum
VNIR	Visible and near Infrared
WSP	WSP Global
WV3	WorldView 3
WWF	World Wide Fund For Nature

Summary

This study aims to develop a large-scale application for an Arctic-wide consistent and comparable data acquisition of beach litter. Therefore, traditional applications of beach litter monitoring were combined and extended by remote sensing techniques. The developed applications were tested on selected beaches on Svalbard and Greenland and validated by comparing it to a traditional beach litter monitoring approach on the same test sites. Methods of traditional beach litter monitoring were evaluated to select the most suitable approach for this study. The test sites on Svalbard and Greenland were chosen after the identification of potential hotspots of beach litter accumulation applying statistical and modelling approaches. Potential test sites were further evaluated by their remoteness and by the presence of regional logistics facilities.

In work package 1, AquaEcology GmbH & Co. KG (AE) performed a literature study on existing methods of beach litter monitoring. Distinct monitoring methods, such as those by Alkalay et al. (2007), Bravo et al. (2009), Cheshire et al. (2009), Opfer et al. (2012) and OSPAR (2010), were compared. The OSPAR method is applied by the European countries and Greenland, while the protocols of Opfer et al. (2012) and Cheshire et al. (2009) are applied in the United States and Canada, respectively. For this study, the OSPAR method (2010) was applied, as it was found to be the most detailed one among the methods described above. During OSPAR surveys, standard 100 m sections of beach along the coastline are monitored, recording the abundance of macroscopic beach litter. The recorded litter is then collected and removed from the survey sites. The mean total abundance of beach litter on OSPAR beaches on Greenland and Svalbard was analysed for the years 2016 to 2019. On Greenland values ranged from one item at the eastern coast to over 800 items at the western coast. On Svalbard, the mean total abundances of beach litter approximated to 200 items. An analysis of the spatial distribution of beach litter shows the occurrence of high amounts of beach litter on sections of beaches highly exposed to wind and currents indicating that most beach litter is sea-based. Sea-based sources contain mostly shipping and fishing. Several studies confirm shipping and especially fishing as the dominating sources of marine litter in the Arctic (Bergmann et al., 2017; Buhl-Mortensen and Buhl-Mortensen, 2017; Tekmann et al., 2017).

BioConsult SH GmbH & Co. KG (BCSH) conducted a literature study on existing remote sensing techniques for the detection of plastic litter on Arctic beaches in order to develop an application for an Arctic-wide consistent and comparable data acquisition. The literature study focused on three remote sensing platforms: Satellite, airborne and drone (UAV), which were evaluated in terms of spectral and spatial resolution, area coverage, acquisition limitations and costs. Satellite platforms showed high potential for plastic beach litter detection on Arctic beaches permitting the acquisition of large-scale areas with a high spectral coverage. Due to the remoteness of the Arctic and the resulting time and cost intensive logistics of field works, satellite platforms are predestinated for an Arctic-wide monitoring. Several studies (Topouzelisa et al. 2019, Biermann et al. 2020, Acuña-Ruz et al. 2018 and Maximenko et al. 2019) have shown the potential of satellite systems for the detection of spectral characteristics of plastics on pixel level. Even though the main limitation remains the relatively coarse spatial resolution of >1 m for multispectral imagery resulting in a potential detectability of objects >1 m^2. Sub-pixel classification could further increase the object detectability, but this was only tested over sea so far. A further challenge of satellite-based beach litter detection is the high cloud coverage in the Arctic which requires a broad acquisition timeframe and high revisit time. Airborne-based digital surveys for litter detection were conducted in various studies, showing its potential for a visual census of plastic litter as well as for (semi-) automatic classification approaches (Garaba et al. 2018, Garaba et al. 2018b, Moy et al. 2018, Pichel et al. 2012 and Garcia-Garin et al. 2019).

The visual analysis permitted an identification of objects with a size of ten times the spatial resolution. For a (semi-) automatic classification hyperspectral sensors were applied. Garaba et al. (2018b) described a potential detection of objects at sub-pixel scale of 5% of a pixel over sea, resulting in an object size of around 0.025 x 0.06 m. Airborne-based systems show great potential regarding spatial and spectral resolution but are limited due to their relatively high cost, especially for an Arctic-wide acquisition. Costs include instrument-running costs, aircraft/ fuel overheads, mobilization and demobilization and personnel and living expenses (Coulter et al. 2007). A rough estimation of area coverage is based on aerial survey experience of BCSH and resulted in an approximate coverage of around 800 km of coastline per day at a flight altitude of 550 m but can vary significantly depending on coastal characteristics. UAVs were successfully applied in various studies of beach litter monitoring for a visual analysis and for (semi-) automatic detection approaches (Martin et al. 2018, Bao et al. 2018). UAVs permit the highest spatial resolution of the three platforms with sub-centimetre acquisition in RGB and 3.4 cm for multispectral sensors. The application of hyperspectral sensors is possible but still very costive. Several studies showed the potential of (semi-) automatic classification over sandy beaches (Martine et al. 2018, Bao et al. 2018) which has still to be confirmed on different beach types. The relatively low costs and the fast acquisition time make UAV surveys particularly interesting on a locale scale and as a support to traditional beach litter monitoring. The main limitation of UAV surveys for an Arctic-wide monitoring of beach litter are a large-scale application and the accessibility of remote areas as well as the climatic limitations (rain or strong winds >30 km/h). In addition to the literature review of the three remote sensing platforms, the applicability of (semi-) automatic classification methods was investigated. For a large-scale application in the Arctic, satellite data are of high interest and thus provide an appropriate classification on pixel level using the spectral characteristics of plastic. Garaba et al. (2020) presented a continuous spectrum of plastic litter identifying wavelength ranges that are characteristic for the presence of plastics (around 931 nm, 1215 nm, 1417 nm, 1732 nm). Sensors covering these wavelength ranges are particularly well-suited for (semi-) automatic detection of plastic litter. Comparing the three platforms with their advantages and limitations, BCSH suggests a combination of at least two platforms to profit from synergistic effects. For this study, BCSH recommends the application of WorldView3 (WV3) satellite imagery and UAV imagery from the multispectral MicaSense Altum sensor. Combining satellite imagery with large aerial coverage but relatively low spatial resolution, with high spatial resolution data from UAV surveys is very promising.

In work package 2, AE and BCSH summarized and acquired the data, necessary for the identification of accumulations of beach litter on Greenland and Svalbard. Orbicon Arctic A/S (OA) gathered and provided data regarding settlements, tourism, harbours, landfills, and mining activities on Greenland. Data availability significantly differed between Svalbard and Greenland. Therefore, data evaluations were different, as well. While hydrodynamic and Lagrangian transport models were available for the waters around Svalbard only, information on shipping, fishing, and tourism could be acquired for Greenland only. The required data included geo-referenced data on settlements, harbours, currents, and wind as well as information on landfills and incineration. AE created geo-referenced shape files of meteorological stations, harbours, and settlements and aggregated ASTD data on shipping and fishing activities to total abundances in the waters of Greenland. Afterwards, AE carried out a statistical analysis to identify drivers of beach litter abundances on Greenland. A permutational analysis of covariance (PERMANCOVA) was calculated with mean abundances of fishing vessels as factor, and numbers of inhabitants, distance to next settlement, distance to next harbour, and overnight stays as covariates. Mean total abundances of beach litter on Greenland, taken from the OSPAR beach litter database, were used as dependent variable. Normality (Kolmogorov-Smirnov test) and homogeneity of variance (Levene test) were given, but autocorrelation could not be excluded definitely. Therefore, the model is assumed to be not or barely biased. The model fit was good

with a measure of determination of 0.83. Fishing and overnight stays could be identified as significant factor and covariate, respectively, which agrees with previous studies on beach litter in the Arctic. Particularly, sea-based sources are often highlighted as dominant input parameter (Bergmann et al. 2017, Kirkfeldt 2016, Tekmann et al. 2017). Further evidence for sea-based sources are wind and data on currents. Abundances of beach litter at the west coast of Greenland are highest where strong westerly or southwesterly winds are dominating. In contrast, beaches with less westerly wind reflect lower abundances. The proximity to larger cities is an additional factor for beach litter accumulation. Evaluating all data, hotspots of beach litter accumulation were expected to be found close to Nuuk and Sisimiut, making them potential beaches to be surveyed in this project. At Svalbard, easterly winds are predominating, indicating lower abundances at the western coast of Svalbard. Highest abundances were found on beach sections exposed to wind and currents in the north of Svalbard (Bergmann et al., 2017).

In addition to the statistical approach described above, BCSH applied a modelling approach to predict locations around Svalbard with higher probabilities of beaching of macro-plastics. The model approach was only applied on Svalbard, as no high-resolution dataset of sea current was available for Greenland. Therefore, two sets of data on sea currents with different resolutions were applied and the outputs from the models were assessed. Monthly fishing data were used to create input data for the model calculating monthly densities, based on which plastic particles were deployed randomly at sea every 100 km of route. Each monthly dataset was loaded in the current model. Furthermore, two current datasets were used as an input. From the Marine Copernicus website (https://marine.copernicus.eu), a dataset with low spatial resolution but large spatial coverage was downloaded and used for simulation of trajectories of plastic particles released according to the monthly fishing intensity. Besides, it was used to feed particles to the higher resolution model. The dataset had a spatial resolution of 0.025°, a spatial extension from -10°W to 60°E and from 55°N to 90°N and a temporal coverage over two years, from August 2008 to August 2010. The second dataset had higher spatial resolution but less spatial coverage and is available on the Norwegian Meteorological Service website (https:\\api.npolar.no). The dataset is available for the years 2005 to 2010 at 800 m resolution but is not covering the full fishing area around Svalbard. Therefore, plastic particles were fed by the lower resolution model. The temporal range was set between August 2009 and August 2010. Wind and wave stoke were not available for the considered area and time frame and therefore not included in this study. The model was performed with OpenDrift v1.0, an open-source Python based framework for Lagrangian particle modelling that is commonly used to predict pathways of floating objects in the ocean (Dagestad et al. 2018). As wind and wave stoke were not available, a 10% uncertainty factor was included in the current. Litter particles of both models that stranded at the end of the study period were used for further statistical analysis. Both models showed similar stranding patterns with highest stranding probability at the Southern- and Western coast. A linear regression model confirmed a significant relation with a measure of determination (R^2) of 0.758 but also showed higher variations among predicted standings in the southern coast. A spatial analysis on significant difference confirmed the results from the linear regression. Both models indicated potential litter accumulation on the West coast of Svalbard, a finding which contrasts with the above-mentioned assumption of low abundances due to easterly winds. One explanation might be that wind and wave data were not incorporated in this study. Abundance data gathered by this studies' fieldwork can improve the model output and should be included in future approaches. For future approaches, BCSH suggests the incorporation of the Leeway configuration which is a more accurate model configuration for big size plastics compared to PlasticDrift. For the application of the Leeway configuration, wind and wave data are required. BCSH recommends the application of the

Copernicus model, which shows similar results and is not limited to the years 2005 to 2010 as is the API model. Additional input parameter such as the type of coast (beach or cliff), the slope of the beach or wave angle incidence could further improve the model.

The expedition to Svalbard lasted from June 25, 2021 to July 17, 2021. A total of two 50 m sections of beach were completely surveyed applying the OSPAR method. Coordinates were determined by means of a high-resolution GPS. The coordinates of single pieces of litter could not be measured because of lacking access to internet and radio. During the stay on Greenland, beach litter surveys were performed near Nuuk and Sisimiut with up to four replicate sections of beach. Travelling to the selected beaches was done by boat with a boat guide. Conventional surveys were carried out applying the OSPAR method (OSPAR, 2010), where the most recent categorization of 2021 was used. Categorization was amended by beach litter types, which are characteristic of the western coast of Greenland. These new litter types were derived from the Strietmann et al. (2021) and are based on the splitting of OSPAR beach litter types, as was similarly done by the Joint Category list of the MSFD. Power analyses on one-sided one-sample Wilcoxon rank sum tests were calculated to estimate how long beach sections have to be for a significant detection of beach litter reduction by as much as 10%, 20%, 30%, 40% and 50%. For this purpose, a significance level of 0.05 was used.

On Svalbard, mean total abundance was as low as 28.5 pieces of litter per 50 m beach. Beach litter consisted of 57.9% plastic, 40.4% wood and 1.7% rubber. The high percentage of plastics agrees well with the results of previous beach litter studies on Svalbard (Bergmann et al., 2017). Among the top-14 litter categories, six are definitely attributable to fishing and shipping. Thereby, the importance of sea-based sources is highlighted, which is in line with the results of previous studies (Bergmann et al., 2017; Weslawski and Kotwicki, 2018).

Near Nuuk and Sisimiut, mean total abundances of beach litter amounted to 116.4±111.9 pieces and 87,3±33,5 pieces on a 50 m section of beach, respectively. This was in the same order of magnitude as Strietmann et al. (2021) found on the western coast of Greenland. Material composition was dominated by plastics and processed wood, which again is in line with the results of Strietmann et al. (2021). Beach litter types, which can definitely be assigned to sea-based sources, were the most common among the top-ten litter types. This finding is partly contradictory to the results of Strietmann et al. (2021), who found that domestic waste was predominant. These differences are likely due to the fact that the latter authors partly surveyed beaches in closer vicinity to settlements than in the study at hand. However, the precise distances to settlements are not given in the study of Strietmann et al. (2021). Power analyses gave statistical powers <50% and >80% for beaches near Nuuk and Sisimiut, respectively, for beach litter reductions of 50%. At lower reduction rates, statistical power was even lower and thus not sufficient. Therefore, surveyed beach sections should have a minimum length of 200 m, because near Sisimiut, replicate data of four beach sections of 50 m length were used for power analyses. In turn from a logistic point of view, this supports the application of UAVs for beach litter surveys on Greenland.

To evaluate the potential for remote sensing techniques in the Arctic, drone and satellite imagery were acquired.

The drone images were recorded with a WingtraOne drone and two sensors, the Sony QX1 (RGB) and the Micansense Altum (VIR) with a ground sample distance (GSD) of 1.4 cm and 3.4 cm respectively. The drone surveys were conducted before performing the OSPAR survey, to make sure that litter items were not accidently moved or buried. A total of eight study sites were

monitored, two on Svalbard and six on Greenland. The area coverage varied between 1.31 ha and 33.85 ha with a minimum of two flights per area, one for each sensor. The required flight time per study site was between 06:56 and 49:26 minutes in which 108 to 1877 images were recorded.

The single images were georeferenced with a post-processed kinematics (PPK) approach and used as an input in the photogrammetry software Pix4D mapper version 4.6.4 to produce elevation models, orthophotos and index maps. The orthophotos of the RGB imagery were then screened manually for litter objects within a GIS environment (QGIS version 3.16.6.) with a scale of 1:20. Between 0 and 63 pieces could be detected which is compared to the OSPAR data at maximum 24% and a mean of 14.8% of all found litter objects. Previous studies described a detection of litter objects by manual screening with a detectability between 18% (Merlino et al., 2020) to almost 100% of litter objects (Andriolo et al., 2020; Escobar Sánchez et al., 2021; Gonçalves et al., 2020b). Three factors were identified to explain the variation in detectability between previous studies, which could have led to the relative low recovery rates in this study: The litter size-distribution, the beach environment, and the GSD of the drone footage. The smaller the objects are, the more difficult they are to spot. Therefore, depending on the litter size distribution on the study area, the detectability of objects via manual screening can vary significantly (Martin et al., 2018; Merlino et al. 2020; Lo et al., 2020). On the surveyed beaches in the Arctic, over 70% of the plastic objects detected on drone footage were larger than >10 cm, reflecting the difficulties of detecting smaller objects. Furthermore, the substrate of the beach (sand or gravel) and its heterogeneity (e.g., occurrence of vegetation, wood, or rocks) influence the manual screening. The monitored beaches in the Arctic had a heterogenic beach background with coloured pebbles, vegetation and driftwood which could be mistaken for litter objects and made the manual screening difficult. Merlino et al. (2020) described similar difficulties with small objects and complex beach backgrounds even having a GSD of 0.18 cm. The third parameter, the GSD, is the only one of the above-mentioned parameters that can be directly defined by the operator as it depends on the chosen sensor and the flight altitude. A lower GSD decreases the minimum size of an object sufficient to be identified, but is limited by the applied sensor and may lead to a decrease in area coverage. The detected objects of the manual screening were used to create density maps for each region which could be used as reference for WV3 satellite imagery later. Furthermore, they were used as reference data to test several approaches of (semi-) automatic classification (Random Forest, Support Vector Machine, and Neural Network). The best statistical results for (semi-) automatic classifications were reached applying Random Forest on VIR imagery with an overall accuracy of 90.6% and a F-score of 77%. The corresponding User's Accuracy was 81.5%. Misclassifications of plastics occurred mainly with the classes *rocks* and *wood*. As the plastic object occurrence was much smaller compared to those classes, wrong assignments had a great impact on the classification maps where plastic was strongly overestimated. The difficulty to spectrally differentiate between plastic, wood and rocks could be confirmed by comparing their spectral profiles. Anyhow, the statistic classification accuracy was comparable to previous studies which used better GSDs and had more homogeneous beach backgrounds (Falatti et al., 2019; Gonçalves et al., 2020a,b,c; Papakonstantinou et al., 2021; Wolf et al., 2020).

This study confirms the great potential of (semi-) automatic classification approaches offering a method for a time- and cost-efficient large-scale detection of beach litter using drone imagery. However, the promising statistical results must always be confirmed by the resulting classification maps. In this study, the results showed that a strong overestimation of plastic occurs applying classification approaches that focus on the spectral characteristics alone. The feature spaces of the applied sensors were not able to distinguish between all occurring land cover classes (e.g. plastics, wood and rocks). Therefore, the spectral coverage of the sensors

must be extended, or other structural parameters must be included. The application of an object-based classification could reduce the misclassification of the wood class but would require a higher GSD. Image recognition methods are also promising but require a higher GSD and a large amount of training data. So far, most studies only focused on litter detection and not litter type identification as performed with the OSPAR method. Only Wolf et al. (2020) presented an approach for plastic type identification using image recognition techniques. For a future implementation into ongoing monitoring programmes, this is an essential step, which must be further investigated and developed, considering the required area coverage (limiting the GSD), the beach environment and the litter size distribution for the Arctic.

The WV3 satellite imageries of the study areas were acquired two to eight days before the actual field work covering up to 7500 ha and 23 km of coastline. The imagery included eight spectral bands from visible to NIR with 1.2 m resolution and eight bands in SWIR with 3.7 m resolution. As the occurring plastic objects were of sub-pixel size, a manual screening was not possible. Instead, this study focused on the question whether satellite imagery can identify areas of high beach litter accumulations. Therefore, the density maps of the manual drone screening were used to create a reference data set from which the spectral signal of different land cover classes was extracted. The density maps also helped to identify the plastic coverage per WV3 pixel. Considering all pixels at 1.2 m resolution which were covered by plastic, 95% and 74% of them had a plastic coverage of less than 5% for Svalbard and Greenland respectively. Only a total of 13 pixels were covered by more than 30% of plastic (1x Svalbard and 12x Greenland). For SWIR imagery at 3.7 m resolution, only seven pixels could be identified that were covered by more than 10% of plastic with a maximum between 20% and 30% coverage. Studies of floating marine litter showed that litter could be detected on Sentinel 2 satellite imagery with a plastic coverage in sub-pixel size of 30 – 55% (Biermann et al., 2020; Topouzelis et al., 2019). Sentinel 2 has a similar spectral coverage as WV3 imagery with 12 bands from visible to SWIR. A sub-pixel detection of plastics over open water for objects covering down to 5% of a pixel was described by Garaba et al. (2018b), but only by applying a higher spectral coverage using hyperspectral imagery. The only study applying satellite imagery for beach litter detection was conducted by Acuña-Ruz et al. (2018) using WV3. These authors applied a semi-automatic classification approach on pixel level and could detect litter objects or patches of litter with a minimum size of 1 m² with an overall accuracy up to 88%.

A detection of beach litter on a sub-pixel scale by spectral unmixing of the single pixels of WV3 imagery was not successful in this study. Spectral unmixing intents to identify the percentage of each reference material within a pixel by identifying the proportion of the respective signals. In difference to floating marine litter the background of beach environments was more complex (e.g. vegetation, sand, rocks, wood) and the litter size and spatial distribution were too low. When including SWIR imagery, the sub-pixel size of 30 – 55% as described in Biermann et al. (2020) and Topouzelis et al. (2019), was not reached for any pixel. Compared to Garaba et al. (2018b), the spectral coverage seemed to be insufficient for a sub-pixel detection down to 5%, probably also being influenced by the complex background signals. A classification at pixel level as described by Acuña-Ruz et al. (2018) was not successful as the litter density and the litter size of single objects or patches was not sufficient. Even though high amounts of plastics were observed on Arctic beaches, their spatial distribution was too sparse and the object size to small.

This study investigated the potential application of remote sensing techniques for beach litter monitoring in the Arctic. The results showed the great potential for drone surveys and compared a manual and a (semi-) automatic detection of plastic objects. The integration of VIR sensors could improve the detection accuracy of (semi-) automatic detection, even though the GSD was coarser compared to RGB imagery. Anyhow, the litter size distribution, the complex background of the beaches, as well as the required high area coverage still pose challenges for beach litter

detection. Furthermore, the identification of the detected litter must be further investigated and requires both, a very high GSD and large amount of training data. Satellite imagery is still limited by its spatial resolution. The litter size and its spatial distribution on Arctic beaches were too low to be identified on WV3 imagery. Therefore, a higher spatial resolution or a larger spectral coverage is required to also detect lower litter accumulations on beaches from Satellites.

Applying drone surveys on Svalbard and Greenland, the monitoring time of large areas can be reduced by a factor of 22 compared to an OSPAR monitoring. However, for small areas as a single 50 m transect, a OSPAR monitoring might be faster because of the time required for the data processing of drone imagery. But even when monitoring small areas, drone surveys can be beneficial, as the beaches in the Arctic are often remote and the time at the beaches can be limited. Satellite imagery permit a high area coverage, but the costs of the data acquisition must be further reduced to permit a large-scale application in future.

Zusammenfassung

Das Ziel dieser Studie ist die Entwicklung einer Methode für eine einheitliche und vergleichbare Erfassung von Plastikmüll an arktischen Stränden. Dafür soll die konventionelle Strandmüllerfassung durch Fernerkundung unterstützt und erweitert werden. Die entwickelte Methode wurde an ausgewählten Stränden auf Spitzbergen und Grönland getestet. Die Untersuchungsgebiete stellten Strände dar, an denen ein erhöhtes Müllvorkommen aufgrund statistischer Auswertungen und Modellergebnissen erwartet wurde.

In Arbeitspaket 1 hat AquaEcology GmbH & Co. KG (AE) eine Literatursuche zu bestehenden Methoden der Strandmüllerfassung durchgeführt. Verschiedene Methoden, wie die von Alkalay et al. (2007), Bravo et al. (2009), Cheshire et al. (2009), Opfer et al. (2012) und OSPAR (2010) wurden miteinander verglichen. Die OSPAR Methode wird von den europäischen Ländern und Grönland verwendet, das Protokoll nach Opfer et al. (2012) von den USA und das Protokoll nach Cheshire et al. (2009) von Kanada. Für diese Studie wurde die OSPAR Methode (2010) ausgewählt, weil sie die detaillierteste und ausgereifteste Methode darstellt. Im Zuge einer OSPAR Erfassung wurden 100 m lange Strandsegmente entlang der Küste nach Müll untersucht und die Anzahl der Objekte erfasst. Der erfasste Müll wurde gesammelt und vom Untersuchungsgebiet entfernt. Die durchschnittliche Gesamtanzahl von Müll an OSPAR Stränden auf Grönland und Spitzbergen wurde für die Jahre 2016 bis 2019 ausgewertet. Auf Grönland reichte die Anzahl von einzelnen Müllteilen bis zu über 800 Fundstücken, die an einzelnen Strandabschnitten der Westküste gefunden wurden. Auf Spitzbergen lag der Wert der durchschnittlichen Gesamtzahl bei etwa 200 Fundstücken. Eine räumliche Auswertung ergab eine hohe Ansammlung von Strandmüll an Orten, die exponiert für Wind und Strömung waren. Dies deutet auf einen großen Einfluss von Müll aus seebasierten Quellen hin, der zu großem Teil von der Schifffahrt und der Fischerei stammt. Die Bedeutung der Fischerei als Quelle von Plastikmüll in der Arktis wurde in anderen Studien bestätigt (Bergmann et al., 2017; Buhl-Mortensen and Buhl-Mortensen, 2017; Tekmann et al., 2017).

BioConsult SH GmbH & Co. KG (BCSH) hat eine Literatursuche zu bestehenden Anwendungen der Fernerkundung zur Erfassung von Strandmüll durchgeführt. Die Literatursuche hat einen Schwerpunkt auf die drei Fernerkundungsplattformen Satellit, Flugzeug und Drohne (UAV) gelegt und sie anhand ihrer spektralen und räumlichen Auflösung, der abdeckenden Fläche, Einschränkungen bei der Erfassung und ihrer Kosten verglichen. Die Ergebnisse der Literatursuche ergaben ein großes Potential von Satelliten für die Detektion von Strandmüll in der Arktis, da sie eine großflächige Erfassung bei einer breiten spektralen Abdeckung ermöglichen. Satellitenerfassungen sind prädestiniert für eine Kartierung in unzugänglichem Gelände, die ansonsten hohe Kosten und einen großen Zeitaufwand an Feldarbeit mit sich bringen würde. Zahlreiche Studien haben bereits das Potential von Satellitenerfassung für die Erfassung von spektralen Eigenschaften von Plastik gezeigt (Topouzelisa et al. 2019, Biermann et al. 2020, Acuña-Ruz et al. 2018 and Maximenko et al. 2019). Die größte Einschränkung stellt die relativ grobe räumliche Auflösung dar. Im Multispektralbereich liegt sie bei >1 m und somit bei einer Pixelgröße von 1 m². Dadurch sind selbst auf Pixelbasis nur Objekte ab einer Größe von 1 m² erkennbar. Eine Klassifikation auf Sub-Pixel Ebene wurde bislang nur über Wasser getestet. Eine weitere Einschränkung ist die hohe Wolkenbedeckung in der Arktis, die ein grobes Erfassungszeitfenster und eine hohe Wiederbesuchsrate des Satelliten nötig macht.

Auch Flugzeugerfassungen eignen sich für die Detektion von Plastikmüll. Zahlreiche Studien haben Strand-/ und Meeresmüll mit Flugzeugen erfasst und die Daten sowohl visuell als auch (halb-) automatisch ausgewertet (Garaba et al. 2018, Garaba et al. 2018b, Moy et al. 2018, Pichel et al. 2012 and Garcia-Garin et al. 2019). Die visuelle Auswertung ermöglichte die Erfassung von Objekten der zehnfachen Pixelgröße. Garaba et al. 2018b beschreiben zudem eine mögliche

Erfassung von Plastikmüll über Wasser in Hyperspektralbildern. Eine (teil-) automatische Klassifikation ermöglicht eine Erfassung von Objekten im Sub-Pixel Bereich, die 5% eines Pixels abdecken und damit eine Erfassung von Objekten in der Größenordnung von 0.025 m x 0.06 m. Flugzeugerfassungen ermöglichen eine hohe räumliche Auflösung bei einer großen spektralen Abdeckung. Die größte Einschränkung sind jedoch die relativ hohen Kosten insbesondere für eine Arktis-weite Erfassung. Die Kosten beinhalten Instrumente, Flugzeug und Treibstoff, Mobilisierung und Personalkosten (Coulter et al. 2007). BCSH schätzt die räumliche Abdeckung einer Flugzeugerfassung pro Tag auf etwa 800 km Küstenlinie ein bei einer Flughöhe von 550 m. Dieser Wert ist als grobe Einschätzung zu verstehen und kann abhängig von den Küsteneigenschaften signifikant variieren.

Auch UAVs wurden bereits erfolgreich für die Erfassung von Strandmüll verwendet. Sowohl visuelle als auch (teil-) automatische Methoden kamen zum Einsatz, wobei (teil-) automatischen Klassifikationen bisher nur über Sandstränden durchgeführt wurden (Martin et al. 2018, Bao et al. 2018). UAVs ermöglichen die höchste räumliche Auflösung der drei Methoden mit Auflösungen <1 cm im RGB und 3.4 cm für Multispektral Sensoren. Der Einsatz von Hyperspektral-Sensoren ist möglich, jedoch noch sehr kostenintensiv. Die geringen Kosten und die schnelle Durchführbarkeit einer Befliegung machen UAVs vor allem für kleinräumige Erfassungen und als Unterstützung konventioneller Strandmüllerfassung interessant. Die größte Einschränkung sind die Kosten einer großflächigen Anwendung und die schwierige Anwendung in unzugänglichem Gelände. Weitere Einschränkungen sind Regen und starker Wind (>30 km/h).

Zusätzlich zu den drei Fernerkundungsplattformen wurde die Möglichkeit einer (teil-) automatischen Erfassung untersucht. Für eine Arktis-weite Erfassung von Strandmüll bieten sich Satellitendaten an und somit eine (teil-) automatische Erkennung auf Pixelebene anhand der spektralen Eigenschaften von Plastik. Garaba et al. (2020) haben die spektralen Eigenschaften von Plastik untersucht und charakteristische Wellenlängenbereiche identifiziert (um 931 nm, 1215 nm, 1417 nm und 1732 nm). Sensoren, die diese Bereiche abdecken, sind für eine (teil-) automatische Erkennung von Plastikmüll sehr geeignet. BCSH spricht sich nach vorausgegangener Literatursuche für die Kombination von mindestens zwei Plattformen aus, um von Synergieeffekten zu profitieren. Für diese Studie empfiehlt BCSH eine Kombination aus WorldView3 (WV3) Satelliten Daten und UAV-Aufnahmen eines MicaSense Altum-Sensors. Die Kombination aus flächendeckenden Satellitendaten mit hoher spektraler Abdeckung und UAV-Daten mit sehr hoher räumlicher Auflösung wird als vielversprechend angesehen.

In Arbeitspaket 2 haben AE und BCSH Daten zusammengestellt, die nötig sind, um potenzielle Ansammlungen von Strandmüll auf Grönland und Spitzbergen zu erkennen. Die benötigten Daten umfassen georeferenzierte Informationen zu Siedlungen, Häfen, Strömungen und Wind, sowie von Deponiegeländen und Müllverbrennungsanlagen. AE hat zudem georeferenzierte Shape Files zu Meteorologischen Stationen, Häfen und Siedlungen erstellt und die Daten zur Schifffahrt und Fischerei in der Arktis (ASTD) Daten zur Schifffahrt und Fischerei zu Gesamthäufigkeiten für West- und Ost-Grönland aggregiert. Im Anschluss hat AE eine statistische Analyse durchgeführt, um die Ursachen von Strandmüll Ansammlungen auf Grönland zu untersuchen.

Dafür wurde eine permutationale Analyse der Kovarianz (PERMANCOVA) berechnet. Als unabhängige Variablen wurden Fischerei, Einwohnerzahl, Entfernung zu Siedlungen, Entfernung zum nächsten Hafen und Übernachtungen gewählt. Die durchschnittliche Gesamtanzahl an Müll von OSPAR-Stränden auf Grönland wurde als abhängige Variable festgelegt. Die Ergebnisse zeigen, dass Normalität und Varianzhomogenität gegeben waren, jedoch eine Autokorrelation nicht endgültig ausgeschlossen werden konnte. Das Modell gilt somit als nicht, beziehungsweise

wenig verzerrt („biased"). Das Bestimmtheitsmaß nimmt mit 0.83 einen sehr guten Wert an. Die Analyse ergibt einen signifikanten Einfluss von Fischerei und Tourismus auf das Aufkommen von Strandmüll. Vor allem die Ergebnisse bezüglich Fischerei stehen in Übereinstimmung mit zahlreichen Studien, die auf den Einfluss von seebasierten Quellen auf Strandmüllansammlungen hinweisen (Bergmann et al. 2017, Kirkfeldt 2016, Tekmann et al. 2017). Weitere Hinweise auf erhöhte Mengen an Strandmüll geben Wind- und Strömungsdaten. Strände, die besonders exponiert gegenüber Strömungen und auflandigem Wind sind, weisen die höchsten Aufkommen von Strandmüll auf. Zudem scheint die Nähe zu größeren Städten einen Einfluss zu haben. Die Analyse ergibt, dass hohe Ansammlungen in der Nähe von Nuuk und Upernavik zu erwarten waren. Auf Spitzbergen dominieren östliche Winde. Somit wurden geringe Abundanzen an der Westküste erwartet. Große Mengen an Strandmüll wurden im Norden Spitzbergens festgestellt, an Stränden, die sehr gegenüber Wind und Strömung exponiert sind. (Bergmann et al., 2017).

BCSH hat zudem einen Modell-basierten Ansatz genutzt, um Küstenabschnitte auf Spitzbergen zu identifizieren, die eine hohe Wahrscheinlichkeit für Ansammlungen von Strandmüll aufweisen. Dafür wurden zwei Strömungsdatensätze mit unterschiedlicher räumlicher Auflösung verwendet. Monatliche Dichteverteilungen von Fischerei wurden als Grundlage genommen, um einen Datensatz von Plastikpartikeln auf See zu erstellen. Dabei wurde für jeden Monat eine zufällige Verteilung von Plastikpartikeln für Abschnitte je 100 km erstellt. Die monatlichen Datensätze dienten als Eingabedaten der Modelle. Ein Datensatz mit geringer räumlicher Auflösung, aber großer Flächenabdeckung wurde von der Marine Copernicus Website (https://marine.copernicus.eu) heruntergeladen und für eine Simulation der Verbreitungswege der Plastik Partikel benutzt. Des Weiteren dienten die Ergebnisse als Grundlage für das Modell mit einer hohen räumlichen Auflösung. Der Datensatz der Marine Copernicus Website hat eine räumliche Auflösung von 0.025°, eine räumliche Ausdehnung zwischen -10°W bis 60°E und von 55°N bis 90°N, und eine zeitliche Abdeckung von zwei Jahren, von August 2008 bis August 2010. Der zweite Datensatz stammt von der Homepage des Norwegian Meteorological Service (https://api.npolar.no) und hat eine hohe räumliche Auflösung, jedoch eine geringe Flächenabdeckung. Der Datensatz ist für die Jahre 2005 bis 2010 verfügbar und hat eine räumliche Auflösung von 800 m, deckt jedoch nicht das gesamte Fischereigebiet rund um Svalbard ab. Als Simulationszeit der Anwendung wurde für beide Modelle der Zeitraum zwischen August 2009 und August 2010 gewählt. Daten zu Wind und Wellen waren für das Untersuchungsgebiet in dem Untersuchungszeitraum nicht verfügbar und konnten deshalb nicht in diese Studie integriert werden.

Das Modell wurde mit OpenDrift v1.0 gerechnet, einer Open-Source-Python-basierten Anwendung für Lagrange-Partikelmodellierung, die oft zur Bestimmung von Verbreitungswegen schwimmender Objekte im Ozean verwendet wird (Dagestad et al. 2018). Auf Grund der fehlenden Wind- und Wellen-Daten wurde ein Unsicherheitsfaktor von 10% verwendet. Beide Modelle zeigten eine ähnliche Verteilung der gestrandeten Partikel mit einer hohen Strandungswahrscheinlichkeit an der südlichen und westlichen Küste von Spitzbergen. Ein lineares Regressionsmodell bestätigte die Übereinstimmungen der beiden Modelle mit einem Bestimmtheitsgrad (R^2) von 0.758., zeigte jedoch große Unterschiede an der südlichen Küste. Beide Modelle deuten auf eine hohe Wahrscheinlichkeit von Strandmüll-Ansammlungen an der Westküste Spitzbergens hin, was im Gegensatz zur Annahme von niedrigen Ansammlungen auf Grund der dominierenden Ost-Winde steht. Eine Erklärung kann die fehlende Windkomponente des Models geben. Das im Zuge der Feldarbeit ermittelte Strandmüllvorkommen kann dazu dienen, die Modellergebnisse weiter zu verbessern und sollte in zukünftigen Anwendungen mit

einbezogen werden. Für zukünftige Modelle empfiehlt BCSH zudem die Einbeziehung der Leeway-Konfiguration, die im Vergleich zu PlasticDrift eine genauere Modellkonfiguration für große Plastikpartikel darstellt. Für die Anwendung der Leeway-Konfiguration werden jedoch Wind- und Wellendaten benötigt. BCSH empfiehlt zudem die Anwendung des Copernicus-Modells, das verglichen mit dem API-Modell zu ähnlichen Ergebnissen kommt und nicht auf die Jahre 2005 bis 2010 beschränkt ist. Zusätzliche Eingabeparameter wie Küsteneigenschaften (Strand oder Klippe), die Neigung des Strandes oder der Welleneinfallswinkel können das Modell weiter verbessern, übersteigen jedoch den Rahmen dieser Studie.

Die Expedition nach Spitzbergen dauerte vom 25. Juni, 2021 bis zum 17. Juli, 2021. Insgesamt wurden zwei 50 m-Abschnitte eines Strandes bei Kapp Mitra vollständig nach der OSPAR-Methode kartiert. Die Koordinaten wurden mittels eines hochauflösenden GPS bestimmt. Die Koordinaten einzelner Müllteile konnten mangels Zugangs zu Internet und Funk nicht gemessen werden. Während des Aufenthaltes auf Grönland wurden in der Nähe von Nuuk und Sisimiut Strandmüllerhebungen durchgeführt mit bis zu vier Replikat-Strandabschnitten. Die Anreise zu den ausgewählten Stränden erfolgte per Boot mit einem Bootsführer. Herkömmliche Erhebungen wurden nach der OSPAR-Methode (OSPAR, 2010) durchgeführt, wobei die aktuelle Kategorisierung von 2021 verwendet wurde. Die Kategorisierung wurde um Strandmülltypen ergänzt, die für die Westküste Grönlands charakteristisch sind. Diese neuen Mülltypen wurden aus dem Strietmann et al. (2021) abgeleitet und basieren auf der Aufteilung der OSPAR-Strandmülltypen, wie dies in ähnlicher Weise durch die Joint Category-List der Meeres-Strategie-Rahmenrichtlinie erfolgt ist. Power-Analysen für einseitige Wilcoxon-Rangsummentests mit einer Stichprobe wurden gerechnet, um abzuschätzen, wie lang Strandabschnitte sein müssen, um eine signifikante Verringerung der Strandverschmutzung um bis zu 10%, 20%, 30%, 40% und 50% zu erkennen. Hierfür wurde ein Signifikanzniveau von 0,05 verwendet.

Auf Spitzbergen lag die durchschnittliche Gesamtabundanz bei 28,5 Müllteilen pro 50 m Strand. Der Strandmüll bestand zu 57,9% aus Plastik, zu 40,4% aus Holz und zu 1,7% aus Gummi. Der hohe Plastikanteil stimmt gut mit den Ergebnissen früherer Strandmülluntersuchungen auf Spitzbergen überein (Bergmann et al., 2017). Unter den Top-14-Müllkategorien waren sechs eindeutig der Fischerei und der Schifffahrt zuzuordnen. Dadurch wird die Bedeutung von seebasierten Quellen hervorgehoben, was mit den Ergebnissen früherer Studien übereinstimmt (Bergmann et al., 2017; Weslawski und Kotwicki, 2018). In der Nähe von Nuuk und Sisimiut betrug die durchschnittliche Gesamtabundanz an Strandmüll 116,4 ± 111,9 Teile bzw. 87,3 ± 33,5 Teile auf einem 50 m langen Strandabschnitt. Dies lag in der gleichen Größenordnung, wie von Strietmann et al. (2021) an der Westküste Grönlands gefunden wurde. Die Materialzusammensetzung wurde von Plastik und verarbeitetem Holz dominiert, was wiederum mit den Ergebnissen von Strietmann et al. (2021) übereinstimmt. Unter den Top-Ten-Mülltypen waren solche am häufigsten, die eindeutig seebasierten Quellen zugeordnet werden können. Dieser Befund steht teilweise im Widerspruch zu den Ergebnissen von Strietmann et al. (2021), die feststellten, dass Hausmüll an der Westküste von Grönland überwiegt. Diese Unterschiede sind eventuell darauf zurückzuführen, dass die letztgenannten Autoren zum Teil in größerer Nähe zu Siedlungen kartierten als in der vorliegenden Studie. Die genauen Entfernungen zu Siedlungen sind in der Studie von Strietmann et al., 2021 jedoch nicht angegeben. Die Power-Analysen ergaben Teststärken <50% und >80% für Strände in der Nähe von Nuuk bzw. Sisimiut für eine Verringerung des Strandmülls um 50%. Bei niedrigeren Reduktionsraten war die statistische Teststärke noch geringer und damit nicht ausreichend. Kartierte Strandabschnitte

sollten daher eine Mindestlänge von 200 m aufweisen, wie anhand der Poweranalysen an Daten von vier 50 m-langen Strandreplikaten in der Nähe von Sisimiut gezeigt werden konnte. Aus logistischer Sicht unterstützt dies wiederum den Einsatz von Fernerkundungsdaten für Strandmüllerhebungen auf Grönland.

Um das Potenzial von Fernerkundungstechniken für Strandmüllerfassungen in der Arktis zu untersuchen, wurden Drohnen- und Satellitenbilder erfasst bzw. erworben.

Für die Drohnenerfassung wurde eine WingtraOne Drohne mit zwei Sensoren benutzt, die Sony QX1 (RGB) und die Micasense Altum (VIR) mit einer *ground sample distance* GSD von 1,4 cm bzw. 3,4 cm. Die Drohnenflüge fanden im Vorfeld des OSPAR Monitorings statt, damit Müllobjekte nicht versehentlich bewegt oder verschüttet werden. Insgesamt wurden acht Untersuchungsgebiete beflogen, zwei auf Spitzbergen und sechs auf Grönland. Die dabei erfassten Flächen umfassten zwischen 1,31 ha und 33,85 ha. Es fanden pro Gebiet mindestens zwei Flüge, eine pro Sensor, statt. Die Flugzeit pro Untersuchungsgebiet lag zwischen 06:56 und 59:26 Minuten, wobei zwischen 108 und 1877 Bilder aufgenommen wurden. Die einzelnen Bilder wurden mit Hilfe von *post-processed kinematics* (PPK) zentimetergenau georeferenziert. Anschließend wurden in der Photogrammetrie-Software Pix4D mapper Version 4.6.4 Höhenmodelle, Orthofotos und Indexkarten erstellt. Die Orthofotos des RGB Sensors wurden in einer GIS-Umgebung (QGIS version 3.16.6.) mit einem Maßstab von 1:20 manuell gesichtet, wobei alle gefundenen Müllobjekte als Polygone markiert wurden. Insgesamt wurden zwischen 0 und 63 Müllobjekte erfasst. Im Vergleich zur OSPAR Erfassung wurden im Mittel 14,8% der Müllobjekte gefunden mit einem Maximalwert von 24%. Frühere Studien zur Strandmüllerfassung mit Drohnen konnten zwischen 18% (Merlino et al., 2020) und nahezu 100% der Müllobjekte detektieren (Andriolo et al., 2020; Escobar Sánchez et al., 2021; Gonçalves et al., 2020b). Für die große Variabilität wurden drei Faktoren identifiziert, die auch für die vergleichsweise niedrige Erfassungsrate in dieser Studie verantwortlich sein können: Die Größenverteilung der Müllobjekte, Strand- und Umgebungsparameter und die GSD der Drohnenbilder. Je kleiner Müllobjekte sind, desto schwieriger ist es, sie auf den Drohnenbildern zu erkennen. Deshalb ist die Erfassungsrate von Müllobjekten stark abhängig von der Größenverteilung der Müllobjekte vor Ort (Martin et al., 2018; Merlino et al. 2020; Lo et al., 2020). Auf den Drohnenbilder der erfassten Strände in der Arktis wurden zu einem Großteil (> 70%) Müllobjekte gefunden, die größer als 10 cm waren, was die Schwierigkeit der Erfassung kleinerer Objekte bestätigt. Des Weiteren wird die manuelle Sichtung erheblich durch das Strandsubstrat (sandig oder steinig) und die Heterogenität (Vorkommen von Vegetation, Holz oder Steinen) beeinflusst. Die Strände auf Grönland und Spitzbergen waren zum Großteil heterogen mit farbigen Steinen, Vegetation und Treibholz, die alle mit Müllobjekten verwechselt werden konnten. Merlino et al. (2020) beschrieb ähnliche Schwierigkeiten bei der Erfassung kleiner Müllobjekte auf heterogenen Stränden, selbst bei einer GSD von 0,18 cm. Der dritte Parameter, die GSD ist der Einzige, der direkt beeinflusst werden kann und hängt maßgeblich von dem verwendeten Sensortyp und der Flughöhe ab. Ein niedriger GSD-Wert verringert die Minimalgröße eines Objektes, die notwendig zur deren Erfassung ist, kann aber auch zu einer niedrigeren Flächenabdeckung führen.

Die während der manuellen Sichtung erfassten Objekte wurden zur Erstellung einer Dichtekarte verwendet und dienten als Referenz für die Auswertung der WV3 Satellitenbilder. Außerdem wurden sie benutzt, um verschiedene Anwendungen des Maschinellen Lernens (*Random Forest, Support Vector Machine, Neural Networks*) zu trainieren und zu validieren. Die besten statistischen Ergebnisse wurden mit Random Forest und VIR-Drohnenbildern erzielt und ergaben eine Gesamtgenauigkeit (OA) von 90,6% bei einem F-score von 77%. Die dazugehörige Benutzergenauigkeit (UA) betrug 81,5%. Vor allem Steine und Holz wurden häufig

fälschlicherweise als Plastik klassifiziert. Da das Müllvorkommen deutlich kleiner als das Vorkommen von Steinen und Holz war, führte schon ein geringer Prozentsatz an falsch bestimmten Steinen und Holz zu einer erheblichen Überschätzung von Plastikmüll. Die Schwierigkeit der Algorithmen, Plastik von Steinen und Holz auf Basis der spektralen Eigenschaften zu unterscheiden, wurde durch einen Vergleich der jeweiligen spektralen Kurven bestätigt. Vergleicht man nur die statistischen Ergebnisse der Klassifikationsgenauigkeit mit früheren Studien, konnten trotz gröberer GSD und heterogeneren Stränden, vergleichbare Ergebnisse erzielt werden (Falatti et al., 2019; Gonçalves et al., 2020a,b,c; Papakonstantinou et al., 2021; Wolf et al., 2020).

Diese Studie bestätigt das große Potential (halb-) automatischer Klassifikationsmethoden für eine zeit- und kosteneffiziente, großflächige Erfassung von Strandmüll in der Arktis mit Hilfe von Drohnenbildern. Trotz der vielversprechenden statistischen Ergebnisse müssen diese jedoch stets durch die Ergebnisse der Klassifikationskarten bestätigt werden. Die Ergebnisse der Drohnen-gestützten Klassifikation in der Arktis zeigen dabei eine starke Überschätzung des Plastikvorkommens. Anhand der spektralen Eigenschaften des Drohnensensors war es nicht möglich, alle vorkommenden Materialien voneinander zu unterscheiden (u.a. Plastik, Holz und Steine). Daher muss die spektrale Abdeckung der Sensoren erweitert, oder andere Strukturparameter mit einbezogen werden. Die Anwendung von Objekt-basierten Klassifikationsmethoden könnte die Unterscheidbarkeit von Holzobjekten und Plastik verbessern, benötigt jedoch eine bessere GSD. Auch Verfahren der automatischen Bildererkennung sind vielversprechend, sie benötigen jedoch auch eine niedrigere GSD und eine sehr große Anzahl an Trainingsdaten. Neben einer Mülldetektion sollte langfristig auch eine Identifikation der Strandmüllobjekte angestrebt werden. Bisher wurde eine solche nur durch Wolf et al. (2020) mit Hilfe von Bilderkennungstechniken erprobt. Für eine Integration in bestehende Monitoring Programme ist die Identifikation der erfassten Müllobjekte jedoch essenziell und sollte mit einem Fokus auf die Arktis-spezifischen Anforderungen (die benötigte Flächenabdeckung und dadurch limitierte GSD, die Strandparameter und die Größenverteilung der Müllobjekte) weiter erprobt werden.

Für die Erprobung von Satellitenbildern wurden WV3-Bilder zwei bis acht Tage vor der Feldarbeit aufgezeichnet. Die erfassten Satellitenbilder erstrecken sich über eine Fläche von bis zu 7500 ha bei einer Küstenlänge von bis zu 23 km. Die WV3 Bilder deckten dabei acht spektrale Bänder im sichtbaren bis NIR Bereich mit einer Auflösung von 1,2 m und acht Bänder im SWIR Bereich mit 3,7 m Auflösung ab. Da die vorkommenden Plastikobjekte im Sub-Pixel Bereich lagen, war eine manuelle Sichtung nicht möglich. Stattdessen wurde untersucht, ob Müllansammlungen mit Hilfe der spektralen Eigenschaften detektiert werden können. Die Dichtekarten der manuellen Drohnensichtung wurden als Referenzdaten genutzt, und die spektralen Eigenschaften der vorkommenden Materialien wurden untersucht. Die Dichtekarten dienten des Weiteren dazu die Plastikabdeckung der einzelnen Pixel zu bestimmen. Unter Einbezug aller Pixel mit einer Auflösung von 1,2 m, die ein Plastikvorkommen aufwiesen, wurden 95% bzw. 74% (Spitzbergen bzw. Grönland) dieser von weniger als 5% Plastik abgedeckt. Insgesamt wurden lediglich 13 Pixel identifiziert, die eine Plastikabdeckung von mehr als 30% aufweisen (1x Spitzbergen und 12x Grönland). In den WV3 SWIR Bildern (Auflösung von 3,4 m) wurden sieben Pixel identifiziert, die mehr als 10% Plastik enthielten bei einer maximalen Abdeckung zwischen 20% und 30%. Vorherige Studien zu schwimmendem Meeresmüll zeigten, dass Müllobjekte, die bis zu 30% - 55% eines Pixels abdeckten, auf Sentinel 2-Satellitenbildern erkannt werden konnten (Biermann et al., 2020; Topouzelis et al., 2019). Sentinel 2 weist dabei eine ähnliche spektrale Abdeckung wie WV3 auf mit 12 Bändern vom sichtbaren bis zum kurzwelligen-Infrarot (SWIR) Bereich. Garaba et al. (2018b) beschrieben zudem eine Sub-Pixel Erfassung von Plastikmüll über offenem Wasser für Objekte,

die bis zu 5% eines Pixels abdecken. Dies erfolgte jedoch unter Verwendung einer höheren spektralen Abdeckung mit Hilfe eines Hyperspektralsensors. Die bisher einzige publizierte Erfassung von Strandmüll mit Hilfe von Satellitenbildern wurde von Acuña-Ruz et al. (2018) durchgeführt. Die Autoren verwendeten WV3 Bilder für eine (halb-) automatische Erfassung von Plastikmüll auf Pixelebene und konnten Müllobjekte in einer Größenordnung von ca. 1 m² mit einer Gesamtgenauigkeit (OA) von 88% erfassen.

Die Erfassung von Strandmüll auf Sub-Pixel Ebene durch spektrales Entmischen der WV3 Bilder war in dieser Studie nicht erfolgreich. Beim spektralen Entmischen soll der Anteil der einzelnen Referenzmaterialien innerhalb eines Pixels identifiziert werden, indem der Prozentsatz der jeweiligen spektralen Signale am Gesamtsignal des Pixels identifiziert wird. Im Gegensatz zu Studien, die Meeresmüll erfasst haben, waren die Hintergrundsignale der Strände zu komplex (u.a. Vegetation, Sand, Steine und Holz), und die Größe der Müllobjekte sowie die Mülldichte waren zu gering. Bei Einbezug der SWIR Bilder konnte eine Sub-Pixel Abdeckung von 30% - 55% wie in Biermann et al. (2020) and Topouzelis et al. (2019) nicht erreicht werden. Verglichen zu Garaba et al. (2018b) scheint die spektrale Abdeckung nicht groß genug gewesen zu sein, um Sub-Pixel mit bis zu 5% Plastikabdeckung zu erkennen. Für eine Klassifizierung auf Pixelebene, wie in Acuña-Ruz et al. (2018) beschrieben, waren die Größe der Müllobjekte sowie die Mülldichte zu gering.

In dieser Studie wurde die Anwendung von Fernerkundungsmethoden zur Erfassung von Strandmüll in der Arktis untersucht. Die Ergebnisse zeigen dabei ein großes Potenzial für Drohnenerfassungen. Die Anwendung eines VIR-Sensors wurde erfolgreich erprobt und es konnte damit eine höhere Gesamtgenauigkeit (OA) im Vergleich zur Anwendung von RGB Bildern erreicht werden, obwohl die GSD der RGB-Bilder niedriger war. Dennoch stellen die Größenverteilung der vorkommenden Müllobjekte, die heterogene Strandumgebung, sowie die geforderte hohe Flächenabdeckung eine große Herausforderung dar. Neben der reinen Erfassung von Müllobjekten muss in Zukunft zudem ein Fokus auf der Identifikation der Objekte liegen. Diese benötigt jedoch eine stark verbesserte GSD und eine große Anzahl an Trainingsdaten. Eine Erfassung von Strandmüllansammlungen mit Hilfe von Satellitenbildern ist auf Grund der räumlichen Auflösung stark eingeschränkt. Die Größe der in Grönland und Spitzbergen vorkommenden Müllobjekte sowie die Mülldichte waren zu gering, um diese auf WV3 Satellitenbildern zu erkennen. Eine höhere räumliche Auflösung oder eine größere spektrale Abdeckung wären erforderlich, um auch geringere Müllansammlungen an Stränden mit Hilfe von Satelliten zu erkennen.

Die Auswertung der Drohnendaten auf Spitzbergen und Grönland haben gezeigt, dass die Zeit, die für eine großflächige Erfassung von Strandmüll benötigt wird, durch den Einsatz von Drohnen um bis zu 22-fach reduziert werden kann im Vergleich zu einer Strandmüllerfassung nach OSPAR. Für kleine Flächen, wie ein einzelnes 50 m Transekt, kann eine Erfassung nach OSPAR jedoch schneller sein. Grund dafür ist die Zeit, die für die Datenprozessierung benötigt wird. Trotzdem können Drohnenaufnahmen auch zur Erfassung kleiner Gebiete von Vorteil sein, da die Strände in der Arktis oft abgelegen sind und die Zeit an den Stränden begrenzt sein kann. Satellitenbilder ermöglichen eine hohe Flächenabdeckung, wie sie in der Arktis benötigt wird. Jedoch müssen die Kosten der Datenerfassung weiter gesenkt werden, um eine kosteneffiziente und großflächige Anwendung in Zukunft zu ermöglichen.

1 Introduction

During the last decades, marine pollution with anthropogenic litter has become a major environmental problem worldwide. Numerous studies carried out in marine compartments have focused on temporal and spatial trends of anthropogenic marine litter, especially on that made of plastic (Browne et al., 2010; Galgani et al., 2000; Howell et al., 2012; Ribic et al., 2010). Marine litter has adverse impacts on marine animals through entanglement of marine vertebrates and invertebrates in discarded and lost fishing gear and other plastic litter items (Bullimore et al., 2001; Fowler, 1987; Gregory, 2009; Matsuoka et al., 2005; Pichel et al., 2012; Votier et al., 2011), as well as through ingestion, especially of micro- and mesoplastics (Browne et al., 2008; van Franeker et al., 2011). Because of their durability and associated long lifespan, floating debris can also act as a vector for invasive species (Barnes and Fraser, 2003; Barnes and Milner, 2005; Majer et al., 2012). In several international treaties on the protection of the seas, beach litter has become an indicator of the overall pollution of marine waters with artificial debris (OSPAR, 1992, MSFD, 2008, 2010). In the past, several methods have been developed to quantify the amount of litter washed ashore (Alkalay et al., 2007; Bravo et al., 2009; Cheshire et al., 2009; Opfer et al., 2012; OSPAR, 2010). These methods require the presence of surveyors on the beach to be surveyed for beach litter, and therefore, they are not efficient or applicable in remote areas, such as Arctic coastlines, which are barely accessible to beach goers. The project at hand was launched to handle these problems and to develop an Arctic-wide monitoring of beach litter applying remote sensing techniques.

This project is structured into four work packages:

- Work Package 1: Literature research on possible methods of beach litter detection, including remote sensing-based techniques
- Work Package 2: Identification of potential hotspots of beach litter accumulation
- Work Package 3: Test and validation of the developed methodology on selected beaches with high abundances of beach litter
- Work Package 4: Participation in scientific conferences

Hence, in this project, innovative remote sensing techniques for the quantification of beach litter have been developed and tested on sections of beaches on Greenland and Svalbard. For this purpose, initially, beaches with high abundances of beach litter had to be identified and selected to obtain a sufficiently high number of litter items, necessary for the training and testing of algorithms of automatic image recognition and land use classification.

In work package 1 of the study, BioConsult SH & Co. KG (BCSH) conducted literature research on remote sensing techniques for the detection of plastic litter on Arctic beaches, summarizing and comparing the three main platforms, satellite, aircraft, and drone (UAV). Finally, BCSH gives a recommendation for the method to being applied in this study.

In work package 1 and 2, AquaEcology GmbH & Co. KG (AE) has taken responsibility to identify appropriate beach sections on Greenland and Svalbard. Data on amounts of beach litter from previous studies, as well as information on driving forces, linked to potential enrichments of beach litter, such as data on wind, currents, settlements, harbours, tourism, fishing, and shipping activities, have been acquired and used for this purpose. This data was used as input for a statistical model explaining the variation of total abundances of beach litter on the coast of Greenland. BCSH performed a modelling approach as an additional indicator for areas of potential beach litter accumulation; two models were run on the area around Svalbard. In

addition to the main goal of the identification of potential hotspots, the model aimed to identify the main input parameters required to predict litter accumulation and to compare how different model resolutions influence the model output.

In work package 3, on the selected beach sections, AE and BCSH performed beach litter surveys with the support of Orbicon Arctic (OA) and the Norsk Polar Institute (NP). In parallel, BSCH applied remote sensing techniques, namely automatic imaging from a drone, and satellite images were acquired for the same beach section and survey date. Finally, the remote sensing techniques were calibrated and validated by means of results of the beach litter surveys. Due to the prevalent Covid-19 situation, the planned fieldwork season 2020 was postponed into the summer 2021.

2 Work package 1

2.1 Traditional beach litter monitoring

2.1.1 Methods

AE has performed a literature study on methods of beach litter monitoring worldwide and in the Arctic. Beach litter data were taken from Bergmann et al. (2017), as well as from Weslawski and Kotwicki (2018). Geo-referenced shape files were generated from these data. In addition, OSPAR beach litter data from Iceland, Svalbard, and Greenland were downloaded from the online OSPAR beach litter database (https://www.mcsuk.org/ospar/). Descriptive statistics on OSPAR beach litter data were calculated using the open source software R (https://cran.r-project.org). While beach litter data were downloaded from the web and taken from publications, information on potential drivers of beach litter was partly provided by the project partners Orbicon Arctic A/S (OA) and the Norsk Polar Institute (NP).

2.1.2 Results and Discussion

2.1.2.1 Available monitoring protocols of beach litter

Worldwide, there has been a number of different beach litter monitoring methods, such as those by Alkalay et al. (2007), Bravo et al. (2009), Cheshire et al. (2009), Opfer et al. (2012) and OSPAR (2010).

Alkalay et al. (2007) ignored waste other than plastic items and monitored transects perpendicular to the strandline. Abundances of plastic particles larger than 2 cm were chosen as the basis for a simple beach pollution index. Surveys are carried out every two weeks. In order to achieve optimal characterization of beach cleanliness, transects of 10 m in width are surveyed. 10 m transects are assumed to reliably represent the cleanliness of the beach. To facilitate the counting procedure, transects are divided into 5 strips of 2 m in width. Beaches are morphologically defined as segments, each characterized by the same coastal conditions (sandy/gravelly, narrow/wide, open/bordered by cliffs, etc.). To eliminate bias, the exact measurement location point is not defined, and the surveyors choose the measurement location randomly before entering the beach (for example by choosing a number of footsteps), prior to knowledge of its cleanliness. Once reaching the location, the exact point is taken as representative of the beach cleanliness. Beach litter density is given in [parts/m^2]. Results for appearance of litter on the coasts are graded as follows:

► 0–0.1 parts/m^2—very clean—no litter is seen,

► 0.1–0.25 parts/m^2—clean—no litter is seen over a large area,

► 0.25–0.5 parts/m^2—moderate—a few pieces of litter can be detected,

► 0.5–1 parts/m^2—dirty—a lot of waste on the shore,

► More than 1 part/m^2—extremely dirty—most of the shore is covered with plastic debris.

Bravo et al. (2009) randomly selected beaches along the Chilean coast. These authors used densities per square meter as a pollution indicator and found plastic items and cigarette butts to be the most abundant beach litter items. Most of the sites surveyed so far were sandy beaches, but some of them were pebble beaches; rocky shores were not surveyed. Several transects are surveyed. These transects are perpendicular to the coastline, i.e. from the low tide line (low

station) to the base of dunes (high station). On each transect, a minimum of two stations are surveyed (between two and six stations, depending on the width of the beach). Each station covers an area of 3 m x 3 m, which is delimited by ropes or measuring tape. All types of beach litter within these 9 m² are counted and classified. In Bravo et al. (2009), no information on the categorization of litter is given. Amounts of beach litter are given as densities [m^{-2}].

Similar to the OSPAR method (2010), in Cheshire et al. (2009), beach selection is made according to criteria such as accessibility, beach cleaning activities and neighbourhood to potential sources. The basic sampling unit for beach litter surveys is a fixed section (length) of beach. Survey teams re-survey the same sampling units over an extended period of time (e.g. every three months for a period of five or more years). Sampling units of 100 – 1000 m are recommended. Smaller sampling sub-units may also be employed for ubiquitous items such as cigarette butts but these do not form part of the standard methodology. These sub-units should be 10 m wide strips from the water line to the back of the beach. The minimum sampling frequency should be annually. Ideally, it is recommended that locations are surveyed every three months. Cheshire et al. (2009) recommend using flux rates (i.e. net deposition rates [$d^{-1} m^{-2}$]) of 77 litter types as a measure of pollution rather than their abundance.

The method of Opfer et al. (2012) relies on a random selection of beach transects and distinguishes a total of 43 litter types. This method requires several criteria for beach selection, such as substrate, accessibility, minimum length, and cleaning activities. Opfer et al. (2012) distinguish between accumulation surveys and standing stock surveys. For accumulation surveys, it is recommended making the survey during low tide. To cover the entire site from water's edge to the back of the shoreline, surveyors should traverse the survey area parallel or perpendicular to the water. Surveyors should traverse the survey area in a pre-determined walking pattern until the entire site is cleaned of marine debris. Litter density counts consider items that measure over 2.5 cm in the longest dimension. For standing stock surveys, it is prescribed to divide sections of 100 m length into 5 m segments. There should be 20 of them. Each 5 m segment should run from the water's edge to the back of the shoreline. Among the 20 transects, four are randomly chosen for field surveys. For standing stock surveys, litter is not removed from the beach.

The most detailed guidelines and categorizations of beach litter monitoring are given in OSPAR (2010). Since 2001, on 129 beach sections of 100 m, bordering on the North Sea and the North-East Atlantic, beach litter surveys have been carried out in regular time intervals of three months, applying a category list of 121 litter types. Part of the huge OSPAR beach litter database was statistically analysed by Schulz et al. (2013, 2015a, 2015b, 2017, 2019). These authors made a first proposal for an evaluation system, based on both abundances and trends of beach litter types and general categories. In the OSPAR Intersessional Correspondence Group Marine Litter (ICG ML), discussion on an appropriate assessment method led to the method described by Schulz et al. (2017, 2019) as standard for the OSPAR beach litter database.

In detail, OSPAR beach litter surveys are performed as follows, with beaches included in the monitoring programme selected according to the criteria listed below:

► composed of sand or gravel and exposed to the open sea,

► accessible to surveyors all year round,

► accessible for ease of marine litter removal,

► a minimum length of 100 meters and if possible, over 1 km in length,

► free of 'buildings' all year round,

► ideally not subject to any other litter collection activities.

Note that in arctic regions, the accessibility of beaches is limited to the summer season. Ideally, surveys are carried out at intervals of about three months in winter (mid-December–mid-January), spring (April), summer (mid-June–mid-July), and autumn (mid-September–mid-October). During surveys, the abundance of macroscopic beach litter (items >2.5 cm in their longest dimension) is recorded on standard 100 m sections of beach along the coastline. All litter is collected and removed from the monitored section during the survey.

Data on the amount of litter on a given section of coastline are recorded at the level of litter types. Litter types are identifiable pieces of litter such as plastic bottles, Tetrapak containers, metal drinks cans, rubber gloves, etc. Each piece of litter is assigned to one of 121 different litter types. Moreover, litter types are assigned to different categories according to the material they are made of (i.e. plastic/ polystyrene, rubber, cloth/ textile, paper/ cardboard, wood, metal, glass, and ceramic/ pottery), or their use (sanitary and medical waste). In addition, each litter type is assigned to a given purpose, such as packaging, user item, consumer, and professional.

Overall, there is no method of beach litter surveys applied worldwide. European countries and Greenland apply the OSPAR method, while in the United States the protocol of Opfer et al. (2012) is applied. In Canada, the protocol of Cheshire et al. (2009) is used.

The main advantage of the protocol of Alkalay (2007) is its simplicity, which allows for its application within citizen science projects. This also holds for the protocol of Bravo et al. (2009). However, the lack of any detailed categorization including information on purpose and source makes these two protocols highly disadvantageous. The main advantage of the protocol of Opfer et al. (2012) is its subdivision into standing stock surveys and accumulation surveys. However, no information is given on the frequency of surveys, and categorizations lacks attribution to sources and purposes. Among the five presented protocols, those of Cheshire et al. (2009) and OSPAR (2010) have the most detailed categorizations allowing for source assignments of beach litter. Moreover, these two protocols include detailed prescriptions of the selections of survey sites and frequencies of surveys.

The project partners agreed on using the OSPAR method (2010), which among the methods described above, has the most detailed categorization. The OSPAR method is well standardized, and in the recent past, assessment of OSPAR beach litter data has been standardized as well (Schulz et al., 2017, 2019). The OSPAR protocol is widely used and has also been adapted by other Regional Seas conventions, such as the Baltic Marine Environment Protection Commission (Helsinki Commission, HELCOM) and the Barcelona Convention. Within the EU, further refinement of beach litter categorization to a joint category list is based on the OSPAR categorization. Finally, the OSPAR protocol has been applied on Greenland and Svalbard in the recent past. Thus, results of the envisaged surveys are comparable to historical data from these islands. Table 1 gives a comparative overview of existing methods of beach litter monitoring.

Table 1: Overview of existing protocols of beach litter monitoring.

Method	Selection of beaches	Categorisation	Measured parameter	Data assessment
Alkalay et al. (2007)	No information given	Plastics and non-plastics	Density per square meter	Beach pollution index
Bravo et al. (2009)	randomly	Simple categorization	Densities per square meter	-
Cheshire et al. (2009)	Stratified random selection of beaches	77 categories	Net deposition rates	-
Opfer et al., (2012)	Stratified random selection of transects	43 categories	Abundance	-
OSPAR (2010)	Stratified random selection of beach sections	121 single categories, 8 material categories and further categories of purpose and use	Abundance	Schulz et al. (2017, 2019)

2.1.2.2 Identification of hot spots of beach litter on Greenland and Svalbard

Between 2016 and 2019, on 17 OSPAR beaches on Greenland, mean total abundances of beach litter ranged from one item in remote areas at the eastern coast to more than 800 items at the beach of Upernavik, located at the western coast (Figure 1, Figure 2, A.1). In the same period on the two OSPAR beaches on Svalbard, mean total abundances of beach litter were less variable than on Greenland and were in the order of magnitude of approximately 200 items on a 100 m-long beach section.

On Svalbard, total abundances of beach litter, surveyed by Weslawski and Kotwicki (2018), are given in [items ha^{-1}] and were as low as eleven items ha^{-1} at maximum (Figure 3). Amounts of beach litter taken from Bergmann et al. (2017) range from nine to 524 g m^2 (Figure 3). The different dimensions of amounts of beach litter significantly impede any comparison between datasets. Standardisations of amounts of beach litter could only be made for each dataset separately (Figure 3). However, amounts of beach litter tended to be higher at the northern coast than on the western coast of Svalbard. No information is given on amounts of beach litter in the southern and eastern part of Svalbard. Therefore, interpretations of beach litter data have to be made cautiously.

In general, amounts of beach litter were high on sections of beaches highly exposed to easterly wind and currents (Figure 3, Bergmann et al., 2017). In contrast, amounts of beach litter were low on coastlines sheltered from easterly wind and currents at the western coast (Weslawski and Kotwicki, 2018). This is in line with other studies on marine litter in the Arctic (Bergmann et al., 2017; Buhl-Mortensen and Buhl-Mortensen, 2017; Tekmann et al., 2017). However, according to the vicinity to potential sea-based sources, amounts of beach litter are supposed to be high at the southern coast of Svalbard. This could not be shown, because of lacking beach litter data.

Figure 1: **Box-Whisker-Plots of total abundances of beach litter on Greenland, Iceland, and Svalbard**
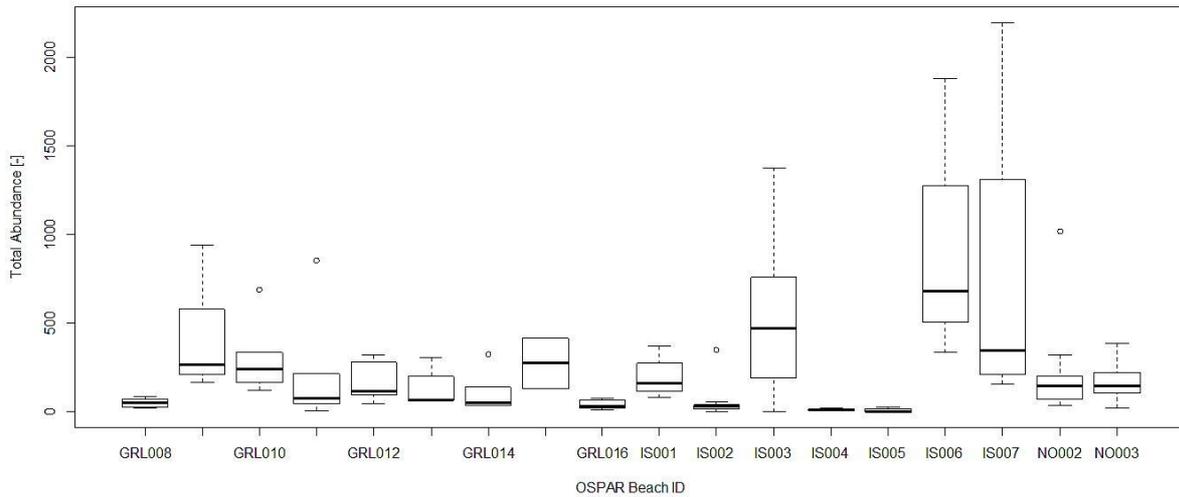
Acronyms on the x-axis are OSPAR beach IDs (GRL = Greenland, IS = Iceland, NO = Norway -here Svalbard). In the boxes, horizontal lines mark the three quartiles, error bars give standard deviations, and empty circles show outlying values. The diagram was created with the open source software R (https://cran.r-project.org). Beaches with too low numbers of surveys (i.e. <4) were omitted from the boxplot. Source: Compiled by AE

Figure 2: **Locations of beach litter surveys, harbours, settlements, and meteorological stations on Greenland**
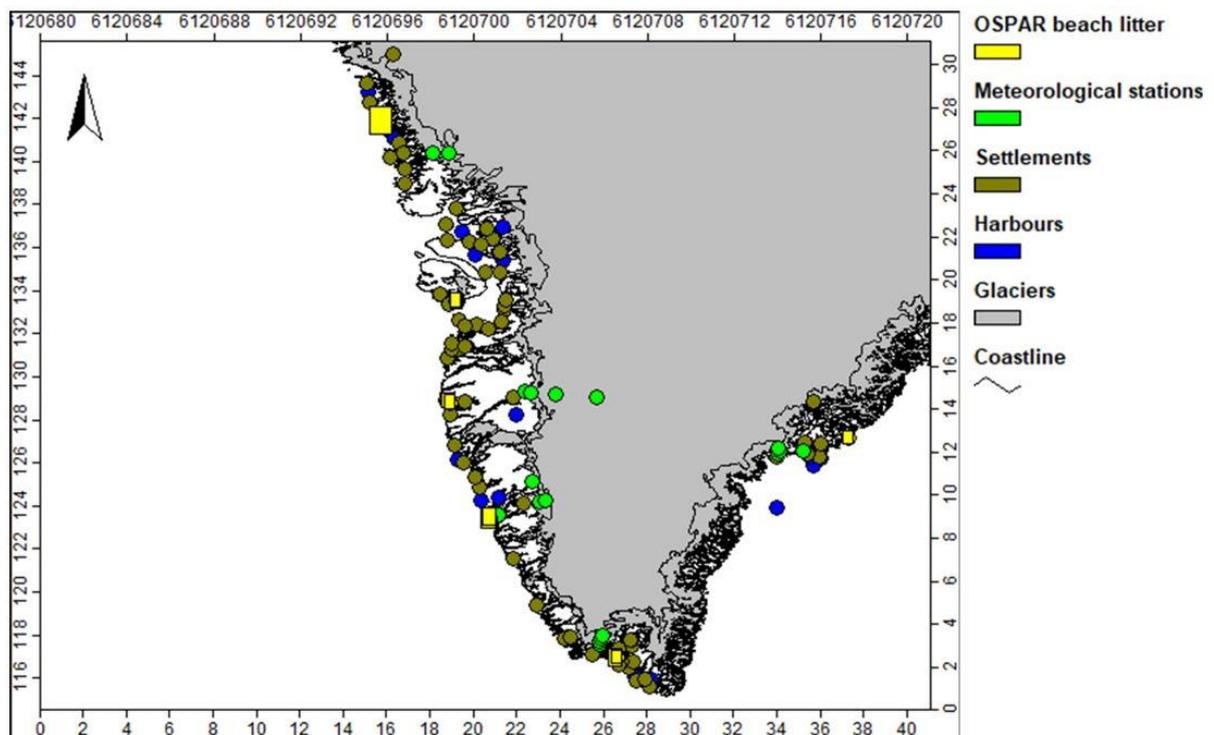
Easting and Northing are UTM coordinates. The size of bars at monitoring locations of beach litter corresponds to the mean total abundance of beach litter, standardized to values from 10 to 20. Source: Compiled by AE

Figure 3: Locations of beach litter surveys, harbours, and meteorological stations on Svalbard.
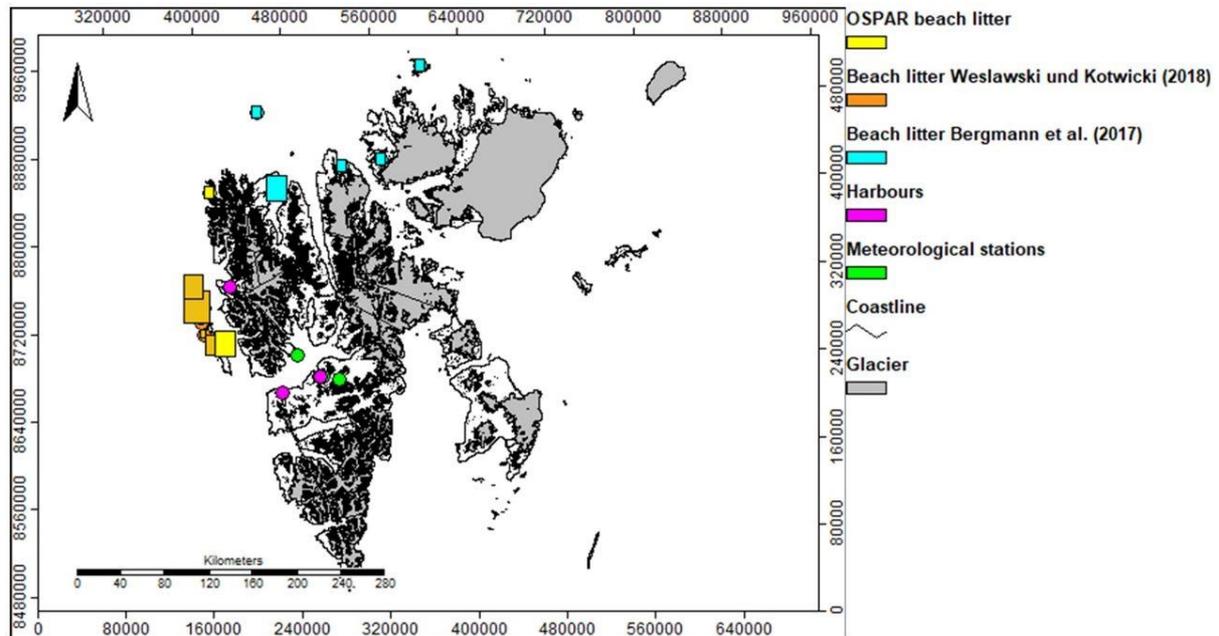
Easting and Northing are UTM coordinates. The size of bars at monitoring locations of beach litter corresponds to the mean total amount of beach litter, standardized to values from 10 to 20. Source: Compiled by AE

2.2 Remote sensing-based beach litter monitoring

2.2.1 Methods

BCSH has conducted a literature study targeted at summarizing and comparing distinct remote sensing techniques for the detection of plastic litter on Arctic beaches. The main objective was the development of a large-scale application for Arctic-wide consistent and comparable data acquisition. The developed methodology should be evaluated in terms of cost efficiency, accessibility limitations and data quality. BCSH distinguished between three major acquisition platforms: satellite, aircraft, and drone (UAVs). State of the art application areas for each approach were summarized, and spectral and spatial resolution, area coverage, acquisition limitations and costs were compared (Table 2). Published studies applying these methods in litter detection were also examined. As the number of studies applying remote sensing techniques for beach litter detection is still limited, studies focussing on marine floating plastics were also included.

The literature study was conducted through a detailed reference search of recent publications, with the online portal Litterbase from the Alfred Wegener Institute of Polar and Marine Research (AWI) serving as a starting point. Litterbase is an online portal for marine litter, which has summarized results from 2044 scientific studies. A second source of information used was the PAME Desktop study on Marine Litter form 2019 (PAME 2019). In addition, discussions with leading scientists (e.g. Dr. Melanie Bergmann, Tomás Acuña Ruz) on the subject of marine plastic litter were conducted.

2.2.2 Introduction to Remote Sensing

For a better understanding of the investigated methods, a short introduction detailing the electromagnetic spectrum and the applied terminology of remote sensing platforms is given below.

2.2.2.1   Electromagnetic spectrum

The electromagnetic spectrum (EMS) is the range of frequencies of all types of electromagnetic radiation from Gamma-rays (<0.01 nm) to Radio waves (>30 cm), typically characterized by their respective wavelengths (Campbell et al. 2011). In remote sensing, most sensors operate within the visible spectrum (VIS) (380 – 780 nm), Infrared (IR) (780 nm – 0.1 mm; including near Infrared (NIR): 780 nm to 1400 nm and shortwave Infrared (SWIR) :1.4 to 3 µm) and microwave (0.1 mm – 1 m). Gamma rays, X- rays, ultraviolet, most IR and long-wavelength radio waves are blocked or absorbed by the atmosphere and are therefore not used by passive remote sensing sensors (Figure 4).

Figure 4:   Electromagnetic spectrum
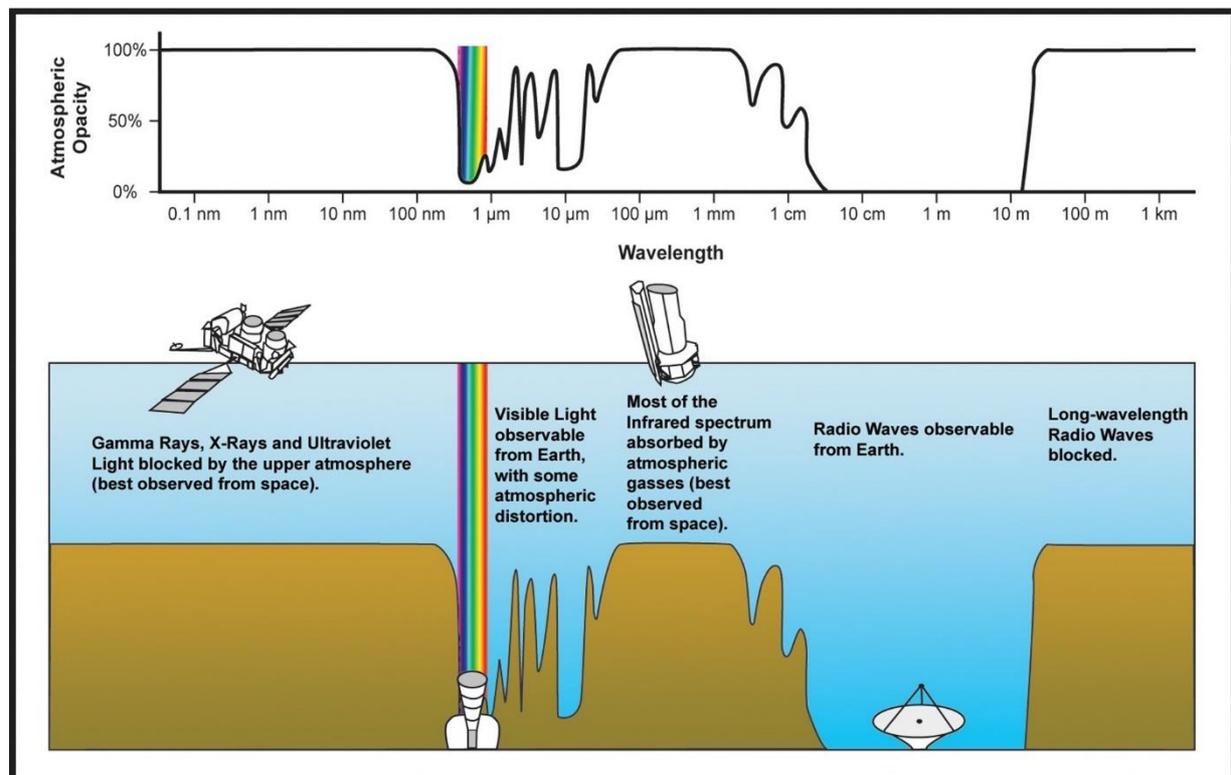
Atmospheric electromagnetic transmittance and opacity. Source: NASA

2.2.2.2   Active and Passive sensors

The properties of the EMS are used by different kinds of sensors, generally divided into active and passive. Active sensors transmit an electromagnetic pulse and measure the reflected or scattered signal. Passive sensors measure the natural emissions of the Earth's surface and the atmosphere depending on their spectral resolution and coverage (Erdle et al. 2011). As all earth surface materials absorb incident sunlight differently according to wavelength, the detected reflectance signal is characteristic of each material and can be used as a determinant of classification approaches (Richards 1999).

2.2.2.3   Spectral resolution and coverage

The spectral resolution of a sensor describes the width of the spectral bands and therefore displays its capacity to resolve features in the EMS. As an example, the MicaSense Altum has a spectral resolution of 10 nm for the green, red and redEdge band, 20 nm for the blue band and 40 nm for the NIR band (MicaSense Inc 2020).

The spectral coverage describes the area of the EMS covered by the spectral bands of a sensor. The spectral coverage of a sensor can vary between multispectral (generally 2 - 10 bands: e.g. RGB with the three bands for red, green, and blue) and hyperspectral sensors (100 - 1000 spectral bands) (Pettorelli, 2019).

2.2.2.4   Spatial resolution and area coverage

The spatial resolution describes the pixel size of the imagery acquired by a remote sensing sensor. The pixel size of satellite platforms is fixed for its single bands but can vary for airborne or UAV acquisitions depending on the flight altitude (Richards 1999). Single values of the spatial resolution indicate the same value along and across track. In the case of a difference along and across track, two values are given. The area coverage refers to the image area covered by a sensor and discriminates between small and large-scale acquisitions.

2.2.2.5   Spectral signature

The reflected radiation of a surface material is a function of wavelength and is called the spectral signature. The spectral signature is characteristic of different materials and can be used for classification approaches (Richards 1999). Figure 5 shows an example for the spectral signature of plastics.

Figure 5:   Electromagnetic spectrum of plastic
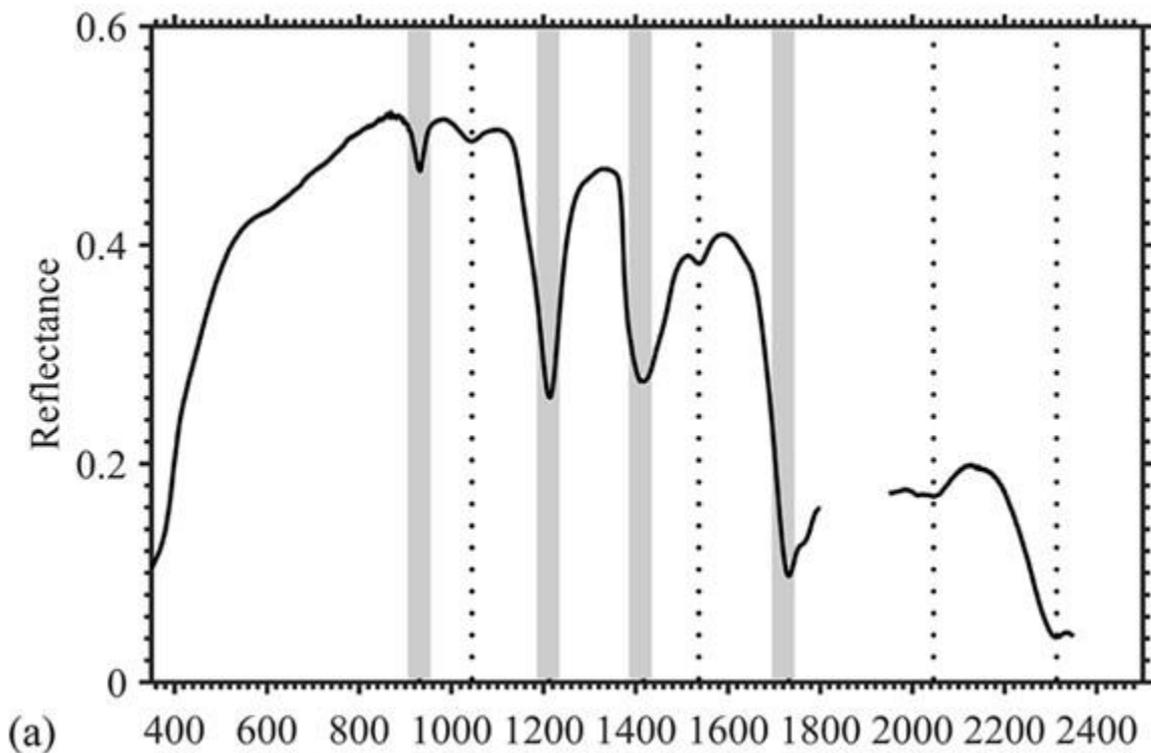
Example of a reflectance spectrum of plastic litter with the four characteristic absorption areas marked in grey
Source: Garaba et al. 2020 licensed under CC BY 4.0

2.2.3   Results and Discussion

2.2.3.1   Review of satellite imagery for plastic litter detection

Satellite platforms permit acquisition of large areas with spatial resolutions down to a few decimetres and with a wide range of sensors. Maximenko et al. (2019) describes potential applications for the detection of marine plastic litter for active and passive sensors. Radar Sensors (active sensors) allow for a description of the dynamics of a floating object as drift velocity or the generated wake. This applies particularly to floating plastics. Radar sensors can operate down to sub-metre spatial resolution. Passive sensors can be used for RGB composites to detect plastic litter and can potentially identify the particular type by giving information on colour and shape of the detected litter. A key requirement is a high spatial resolution (Maximenko et al. 2019). Satellite platforms with passive sensors cover a wide range of spatial resolution from sub-metre for the panchromatic band (e.g. WorldView-3 (WV3), QuickBird) to a few metres (WV3: 1.24 m; Skysat: 2 m; QuickBird: 2.4 m), several tenths of metre (Landsat 7/8: 30 m); Sentinel-2A/2B: 10 - 20 m) and hundreds of metres (MODIS: 250 - 1000 m) in the multispectral bands. Next to the spatial resolution, important characteristics are the spectral resolution and the spectral coverage of the sensor and the revisit time of the platform.

To date only a handful of studies have been published using satellite imagery for the detection of plastic litter on beaches or in the sea. Topouzelis et al. (2019) and Biermann et al. (2020) examined the potential of Sentinel-2 imagery to detect floating plastics. Topouzelis et al. (2019) used three 100 m^2 artificial floating plastic targets consisting of PET-1 1.5 l water bottles, LDPE plastic bags and nylon fishing ghost nets to demonstrate their discrimination capability from natural sea-water. Biermann et al. (2020) successfully classified floating plastics on sub-pixel scale with an accuracy up to 86% developing the Normalized Difference Vegetation Index (NDVI) and a Floating Debris Index (FDI). The authors detected spectral characteristics of plastic in pixel filled by at least 30% of bottles or bags or by 50% of fishing nets. Both studies take advantage of the characteristic spectral response of seawater being the main background signal, even though Biermann et al. (2020) also describes a discrimination between plastics and seaweed, spume, and timber (Figure 6). The main restriction of Sentinel-2 data for beach litter detection remains its spatial resolution of 10 - 20 m. Even by assuming a detectability at a sub-pixel scale of 30%, plastic patches of 30 - 120 m^2 would be required.

The only study applying satellite imagery to beach litter so far was conducted by Acuña-Ruz et al. (2018), using WV3 imagery for the detection of plastic litter on beaches of Chiloé, Chile. The authors applied a semi-automatic approach of land use classification on a pixel level using the spectral data acquired by WV3. Litter objects or patches of litter with a minimum size of 1 m^2 could be detected with an overall classification accuracy up to 88%.

Figure 6: Classification of floating materials using NDVI and FDI
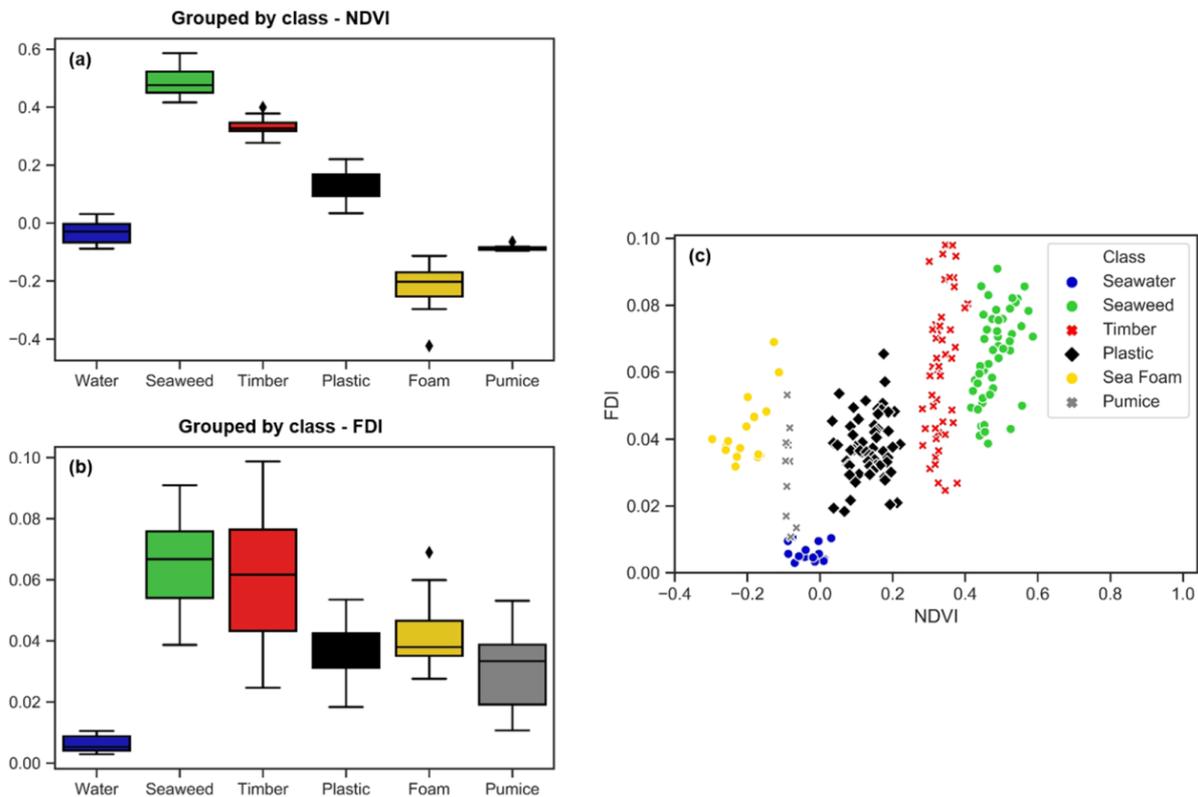
Known floating materials in the environment. Discrimination using NDVI alone (a), FDI (b) and a combination of NDVI and FDI (c). The combination of NDVI and FDI reflects distinct clustering of materials but also NDVI alone shows great potential for the discrimination of plastics. Source: Biermann et al. 2020 licensed under CC BY 4.0

On April 1, 2022, the hyperspectral satellite mission EnMAP (Environmental Mapping and Analysis Program) by DLR (Deutsche Zentrum für Luft- und Raumfahrt e. V.) was launched. EnMAP provides a broad spectral coverage within the visible near-Infrared (VNIR) 420 - 1000 nm and the SWIR 900 - 2450 nm. The spatial resolution is 30 m. An overview of sensors already used for detection of plastic debris was given by Maximenko et al. (2019). The authors summarize the platforms used and the respective spatial and spectral resolutions (Figure 7).

Satellite imagery shows high potential for plastic beach litter detection on Arctic beaches due to its area coverage and spectral resolution, facilitating data acquisition in remote areas and permitting the detection of spectral characteristics of plastics as demonstrated in various studies (Topouzelis et al. 2019, Biermann et al. 2020, Acuña-Ruz et al. 2018 and Maximenko et al. 2019). The main limitation of satellite acquisition is still the relatively coarse spatial resolution of multispectral imagery starting from 1 m upwards, as well as the potential costs for large-scale applications. Even though most high-resolution satellite data (>10 m) are accessible without costs (e.g. Sentinel-2, Landsat 7/8), the acquisition of very high-resolution data are still liable to costs (e.g. WV3: around 3,300 €/100 km²). A further challenge, particularly in the Arctic, is the high cloud coverage which requires a broad acquisition timeframe and high revisit time. Similar problems with cloud coverage were described by Acuña-Ruz et al. (2018) for acquisitions in Chiloé, Chile.

We considered satellite data from WV3 to be the most suitable for the application on beach litter surveys due to their spatial resolution and the spectral properties of its sensors that are ideally located to identify the characteristic absorption bands of plastic (Figure 5, Figure 8).

Figure 7:   **Overview of sensors used for detection of plastic litter**

Satellite/Sensor name	Spatial resolution (HR $\geq$ 10 m; VHR < 10 m)	Spectral resolution	Wavelength range [nm]	Platform	Instrument
Sentinel-2/ MSI	HR	Multispectral	450–1,400	Satellite	
TanDEM-X	HR			Satellite	SAR
WorldView3	VHR	Multispectral	400–2,365	Satellite	
PlanetScope	VHR	Multispectral	455–860	Cubesat/Satellite	
SASI	VHR	Hyper	950–2,450	Airborne imager	
APEX	VHR	Hyper	372–2,540	Airborne imager	
AVIRIS-NG	VHR	Hyper	380–2,510	Airborne imager	
ASD FieldSpec Pro	VHR	Hyper	350–2,500	Handheld	
Spectra vista corporation	VHR	Hyper	350–2,500	Handheld	
Spectral evolution	VHR	Hyper	350–2,500	Handheld	

Source: Maximenko et al. (2019) licensed under

2.2.3.2   Review of aircraft-based surveys for plastic litter detection

Airborne-based digital surveys for litter detection over both oceanic and coastal environments were conducted in several studies (Garaba et al. 2018, Garaba et al. 2018b, Moy et al. 2018, Pichel et al. 2012 and Garcia-Garin et al. 2019). For RGB imagery, a spatial resolution of a few centimetres could be reached (Garcin-Garin et al. 2019, Moy et al. 2018). Typical flight altitudes for airborne-based acquisition vary between 250 m and 650 m, resulting in different spatial resolutions (Garaba et al. 2018, Moy et al. 2018, Pichel et al. 2012, Garcia-Garin et al. 2019). Garaba et al. (2018b) combined RGB-acquisitions with NIR and SWIR imagery. In a visual analysis, objects with a size of ten times the spatial resolution could be identified (Garaba et al. 2018b, Moy et al. 2018, Garcia-Garin et al. 2019). Spectral analysis was applied for semi-automatic detections. Garaba et al. (2018b) describe a theoretic detectability on sub-pixel level over the ocean, using hyperspectral imagery. Objects filling 5% of a certain pixel should be detectable. Garaba et al. 2018b applied the hyperspectral ITRES SASI-600 imager with 100 wavebands between 950 nm and 2,450 nm and a spatial resolution of 0.5 m across 1.2 m along the track. Assuming a detectability of plastic litter covering down to 5% of a certain pixel, results in a possible detection of plastic items with a size around 0.025 m across 0.06 m. In a second study, Garaba et al. (2018) acquired Hyperspectral data from AVIRIS (Airborne Visible/ Infrared Imaging Spectrometer) with 224 contiguous spectral bands with a coverage between 400 - 2,500 nm. A spatial resolution of 7.1 m was reached, resulting in a potential detection of objects of around 2.5 m^2, assuming a detectability at sub-pixel scale of 5% of a pixel (Garaba et al. 2018).

The studies cited above highlight a great potential of airborne-based acquisitions in the detection of plastic litter. The higher spatial resolution compared to satellite platforms permit optical detection of macroplastics at RGB-composites with items roughly ten times the spatial resolution (Garaba et al. 2018, Moy et al. 2018, Pichel et al. 2012, Garcia-Garin et al. 2019). Spectral analysis furthermore permits semi-automatic classification approaches that make use of the spectral characteristics of plastics in the NIR and SWIR spectrum. Classification of plastic litter covering down to 5% of a pixel seems to be possible over open water (Garaba et al. 2018, Garaba et al 2018b), although a sub-pixel detection over beaches has not been tested yet. The spatial resolution of hyperspectral images from airborne systems can go down to a sub-metre scale, whereas the planned hyperspectral satellite mission EnMAP only reaches a resolution of tenths of metres (30 m). The main limitation of airborne-based acquisition is its relatively high cost, especially for an Arctic-wide acquisition. A simplified and basic expenditure breakdown will include instrument running costs, aircraft/ fuel overheads and personnel and living

expenses. Coulter et al. (2007) describes the costs of aerial hyperspectral surveys for moderate relief, clear skies, and good infrastructure in many areas in North America to be around $ 100/Line km. In contrast, the costs of surveys with high relief, unpredictable weather conditions and poor infrastructure rise to $ 150/Line km. Coulter et al. (2007) describes a Line km to cover around 1 km^2 for a spatial resolution of 3 m. The given costs assume a large survey area of at least 1,000 km^2; smaller surveys result in a significant increase in Line km costs. In addition to this, the costs for mobilization and demobilization are not included. Operational limitations due to cloud coverage below the flight altitude can often result in a significant amount of standby days. Mobilization and demobilization costs can vary significantly depending on survey area, flight planning, permits, insurance, custom bonds, personnel transportation, and logistics. Moreover, the cost estimation in question was conducted for 2007 and can therefore vary with any change due to variable fuel costs (Coulter et al. 2007). BCSH has several years of experience in aerial surveys in North Sea and Baltic Sea, and assumes a possible coverage of around 800 km of coastline per day at a flight altitude of 550 m. This value is a rough estimation of the possible coverage and is highly dependent on coastal characteristics.

2.2.3.3   Review of UAV based surveys for plastic litter detection

UAVs permit the lowest flight altitudes of the three platforms and therefore the highest spatial resolution, even though airborne-based sensors can come close. The integration of different sensors to an UAV is defined by the sensor size as well as by the UAVs maximum take-off weight (MTOW). The MTOW varies by the size and engine of the UAVs and influences the maximum flight area that can be covered. RGB sensors usually reach a spatial resolution of several centimetres but can go down to sub-centimetre (Martin et al. 2018). Multispectral sensors can reach a GSD of several centimetres (e.g. MicaSense Altum, Parrot Sequia +). Both multispectral and RGB cameras can be mounted on UAVs with a MTOW up to 5 kg. Hyperspectral sensors on UAVs can have a spectral coverage from VNIR to SWIR but are very expensive (e.g. Micro-Hyperspec).

Martin et al. (2018) und Bao et al. (2018) successfully applied semi-automatic classification of plastic litter using UAV RGB imagery. Bao et al. (2018) applied image segmentation thresholding for a successful classification with an overall accuracy of 98.6%. This classification was performed over uniform sandy beaches. However, the authors point out a potential overestimation of plastic litter with their approach; it seems to have easily misclassified natural items such as shells, branches and leaves as plastic litter. Martin et al. (2018) applied a visual census as well as a machine learning approach. For the machine learning approach, a random forest classifier was used, employing a histogram of orientated gradients (HoG) descriptor. Random forest is an ensemble learning method often applied in land use/ cover classification. HoG is a widely utilized feature representation method used for example in human face detection (Dalal & Triggs 2005). Martin et al. (2018) described a significant positive correlation between items abundance found with the visual census and random forest (Spearman correlation, r=0.61, p=0.026, n=13) which showed that images with a higher number of plastic items lead to higher number of detected items by the random forest algorithm. However, the authors also described an overestimation of plastic items using the machine learning algorithms by a factor of five, as well as missed detections. Both studies clearly showed eligibility of UAVs for automatic detection of plastic beach litter using. It is worth noting that the study areas in both cases were uniform sandy beaches, therefore the possibility to transfer these methods to gravel beaches has not been tested yet. Finally, Martin et al. (2018) also compared the time for a UAV survey to the time needed to conduct a traditional beach survey. The authors described the UAV approach as being thirty-nine times faster. UAVs permit high resolution data with multi and hyperspectral coverage. They can monitor beaches with high time efficiency and are eligible for (semi-) automatic classification. Promising studies used RGB imagery for successful automatic

detection on sandy beaches, although a successful transfer to gravel beaches must still be investigated. Moreover, the application of multi or hyperspectral data has to be tested. Some of the main limitations of this method are the accessibility of remote areas as well as the costs arising from applications of hyperspectral sensors. The main climatic limitations for consideration are rain or strong wind speeds (above 30 km/h); due to their low flight altitude UAVs are less likely to be affected by clouds.

2.2.3.4   Ongoing literature research

Literature research was an ongoing process during the project with a high number of studies published within this period. Recent publications focused on drone-based beach litter detection using high resolution RGB cameras (e.g. Andriolo et al., 2020; Gonçalves et al., 2020a and 2020b; Escobar Sánchez et al., 2021; Martin et al., 2021; Merlino et al., 2020; Papakonstantinou et al., 2021). The GSD used varied between 0.18 cm (Merlino et al., 2020) and 1.2 cm (Andriolo et al., 2020) covering areas between <1 ha and 5 ha (Papakonstantinou et al., 2021). Only Martin et al. (2021) applied a large-scale application covering a much larger area by monitoring a total of 44 beaches along the Saudi Arabian coast with a linear distance of 1400 km. To achieve very low GSD low flight altitudes (down to 6 m) were necessary and the acquisition of transects of 100 m beach lengths could take up to half an hour flight time.

Various classification approaches were tested and compared with in-situ OSPAR monitoring. Classification approaches were manual screening of drone footage, pixel- and object-based (semi-) automatic classifications using machine learning algorithms (e.g. Random Forest, Support Vector machine, Maximum Likelihood Classifier) and Convolutional Neural Networks.

Marine litter was identified by manual screening with detection rates between 20% (Merlino et al., 2020 (GSD = 0.18 cm)) and up to almost 100% (Andriolo et al., 2020 (GSD = 1.2 cm); Escobar Sánchez et al., 2021 (GSD = 0.27 cm); Gonçalves et al., 2020b (GSD = 0.55 cm)).

The detection success applying (semi-) automatic classification approaches to detect beach litter varied among different studies between 25% - 74% (Escobar Sánchez et al., 2021), 64% - 73% (Gonçalves et al., 2020), 62% - 82% (Martin et al. 2021) and 77.26% (Papakonstantinou et al., 2021).

For further investigations the application of multispectral sensors was discussed and suggested in several studies (Gonçalves et al., 2020a und 2020c; Escobar Sánchez et al., 2021; Wolf et al., 2020).

2.2.3.5   Classification methods

In addition to the literature review of the three remote sensing platforms, a focus was also placed on the applicability of (semi-) automatic classification methods. Therefore, the classification methods that are applied on the abovementioned platforms were evaluated in terms of an Arctic-wide application. The most common approaches have been image recognition tools and imaging spectroscopy (e.g. Acuña-Ruz et al. (2018), Biermann et al. 2020, Garaba et al. 2018, Martin et al. 2018). Martin et al. (2018) applied image recognition tools on very high resolution drone data. Imaging spectroscopy for beach litter detection was applied by Acuña-Ruz et al. (2018). Imaging spectroscopy describes the measurement, analysis and interpretation of electro-optical spectra acquired over large areas. It permits the classification of remote sensing imagery on a pixel level by using spectral characteristics of the occurring land cover (Shaw and Burke 2003). For a large-scale application as in the Arctic, a pixel-based classification of satellite data is of high interest. Therefore, publications about reflection spectrum of plastic litter were further investigated. Garaba et al. (2020) present a continuous spectrum of plastic litter in the RGB, NIR and SWIR spectra. Various wavelength ranges were identified that are

characteristic for the presence of plastics. Those are around 931 nm, 1215 nm, 1417 nm, and 1732 nm and thus in the range between NIR and SWIR (Figure 8). Sensors that cover these wave ranges are particularly well-suited for (semi-) automatic detection of plastic litter as they promise a potential discrimination from the surrounding materials.

2.2.4 Conclusion

An overview of the spectral and spatial resolution, spectral and area coverage, acquisition limitations and costs of the three platforms is presented in Table 2. As the single characteristics should not be evaluated in the same way and each method has its own advantages and limitations, the application of a ranking is quite subjective and was therefore not applied. The literature study emphasizes that any of the three platforms can give a satisfying solution on its own. Maximenko et al. (2019) suggests therefore an integration of several observing systems for the detection of marine debris.

Table 2: **Comparison of remote sensing platforms for an Arctic-wide beach litter monitoring**

	Satellite (WorldView 3)	Airborne	UAV
Spectral coverage	RGB. NIR, SWIR	RGB, NIR, SWIR	RGB, NIR, (SWIR)
Spatial Resolution	0.31 m/ 1.24 m/ 3.7 m	Centimetre to metre	**Sub-centimetre**
Area Coverage	**Large-scale**	Medium-scale	Small-scale
Flight altitude	617 km	250 m to 650 m	10 m to 100 m
Detectable litter size	optical: >3 m spectral: 1.24 m (pixel-level)	Optical: >10/ 20 cm Spectral: >5% of a pixel over sea	Optical: >10 cm Spectral: pixel-level
Costs for Arctic-wide monitoring	**Medium**	High	Very high
Limitation	Relative Coarse Resolution, cloud coverage	High costs in general and very high costs for small areas, cloud coverage	Small scale, Accessibility of remote areas, wind, high costs for SWIR sensors

Comparison of spectral and spatial resolution, area coverage, acquisition limitations and costs. WorldView 3 was chosen as an example for a potential satellite platform because of its high spatial resolution compared to other satellite systems.

BCSH agrees with the idea of an integrated approach and suggests a combination of two platforms, namely satellite and UAV, for Arctic-wide beach litter detection. As the main limitation of satellite platforms is the relatively coarse spatial resolution, BCSH suggests a combination with very high-resolution UAV imagery. The advantages of both platforms can then be combined to create some very promising survey characteristics. Multispectral satellite imagery permits the investigation of pixel-based classification approaches as conducted by Acuña-Ruz et al. (2018), making use of the characteristic plastic reflectance spectrum (Figure 8). At present, the relatively coarse spatial resolution is limiting the detection of single pieces of beach litter but has great potential for an Arctic-wide identification of hotspots of beach litter accumulation. Identified hotspots can be further investigated using UAV imagery. UAVs permit a time-efficient acquisition on a centimetre scale at low costs and can complement traditional beach litter surveys. Anyhow, an Arctic-wide application of drone imagery is not feasible because of costs and difficult accessibility. UAV imagery furthermore permits the evaluation of different spatial resolutions for plastic litter detection, acquired by different flight altitudes or down sampling. Therefore, UAVs can offer an outlook for future satellite missions, overcoming the restrictions of spatial resolution. The main limitation of UAVs is the low spectral coverage for

the low to middle price sector and the low area coverage, both being compensated by satellite imagery.

Figure 8: **Example reflectance spectrum of plastic litter**
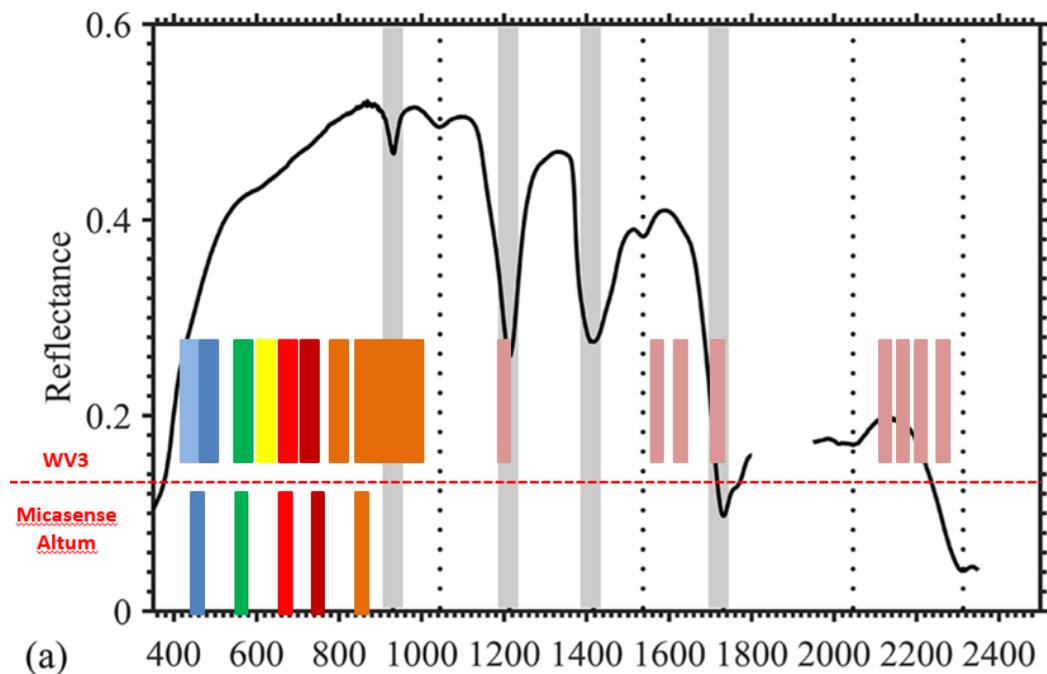
Example of a reflectance spectrum of plastic litter given by Garaba et al. (2020) with the four characteristic absorption areas marked in grey. WV3 and MicaSense Altum spectral bands are added to indicate their overlap with the characteristic plastic features. WV3 bands are above the red line, MicaSense Altum bands below.
Source: Adjusted from Garaba et al. 2020 licensed under <u>CC BY 4.0</u>

For this study, BCSH suggests the application of WorldView3 (WV3) satellite imagery. WV3 combines very high spatial resolution (1.24 m for VIR and 3.7 m for SWIR) with multispectral coverage and has already been applied successfully for beach litter detection (Acuña-Ruz et al. 2018). Its panchromatic band (0.31 m) can be used for pan-sharpening of the RGB, resulting in a colour image of 0.31 m resolution which in turn enables a visual identification of large plastic litter items. The multispectral bands can furthermore be used for (semi-) automatic classification approaches. As a UAV sensor, BCSH recommends the MicaSense Altum. The MicaSense Altum consists of five individual sensors in the VIR (RGB, redEdge and NIR) with a spatial resolution of 3.4 cm at a flight altitude of 80 m. Aside from the promising studies of airborne-based systems with high spatial and spectral resolutions, the large costs could make airborne-based acquisitions unattractive for an Arctic-wide application.

3 Work package 2

3.1 Statistical analysis to identify enrichments of beach litter

3.1.1 Methods

Sources of beach litter can roughly be subdivided into land-based sources and sea-based sources. The latter comprise fishing, shipping, offshore mining activities, and wind power installations, while the former include tourism, uncontrolled emissions from landfills, harbours, and input via tributaries (i.e. rivers and estuaries). Attribution of beach litter types to sources has been attempted by Tudor and Williams (2004), who invented the matrix score technique. However, except for fishing-related items, source attribution is highly uncertain, partly, because many litter types may originate from several sources.

Rivers are assumed to substantially emit litter to marine waters (Gasperi et al., 2014; Rech et al., 2014), but until present, the quantifications of source terms is uncertain (Morritt et al., 2014), because empirical studies partly neglect the high temporal variability of discharge at the rivers' mouths, and model studies mainly rely on generic parameterizations and suffer from lack of information on boundary conditions.

Information on riverine input of litter in the Arctic is subject to speculations, while Bergmann et al. (2017), Kirkfeldt (2016), and Nashoug (2016) mainly assign beach litter on Svalbard and Greenland, respectively, to sea-based sources, and among these mainly to fishing. Kirkfeldt (2016) discusses landfills as potential sources of beach litter on Greenland. Even scientific field studies have been suspected to contribute to pollution with litter (Nashoug, 2016). The lack of information of litter emissions to the Arctic requires acquisition of data on potential drivers of beach litter in the Arctic.

Therefore, AE listed up data requirements, necessary for the identification of enrichments of beach litter on Greenland and Svalbard. These demands include geo-referenced data on settlements, harbours, landfills, currents, and wind, the latter two of which can explain the beaching of litter from sea-based sources. Information on landfills and incineration plants on Greenland was depicted from Eisted and Christensen (2011). Data on settlements, harbours, landfills and mining activities on Greenland were provided by OA and are visualized in B.1 and B.2.

The following data were acquired from the web by AE:

► Wind data of Svalbard (https://www.unis.no/resources/weather-stations/),

► Wind data of Greenland (https://www.promice.org/PromiceDataPortal/),

► Data on currents and Lagrangian transport deriving from a model of computational fluid dynamics (CFD) of the waters around Svalbard (Hattermann et al., 2016; https://data.npolar.no/home/),

► Topographical data (shape files) of Greenland given in 1:1.000.000 (https://eurogeographics.org/),

► Topographical data (shape files) of Svalbard given in 1:100.000 (https://data.npolar.no/home/),

► Coordinates of harbours on Greenland (http://capetotrade.com/getportlist.php?ref1=GL),

► Coordinates of harbours on Svalbard (http://ports.com/browse/europe/svalbard-and-jan-mayen/).

AE geo-referenced the data and created shape files from positions of meteorological stations, harbours, and settlements (Figure 2,Figure 3). The latter data were provided by BCSH. Moreover, wind data were analysed and illustrated using the R routine CLIFRO (https://cran.r-project.org/web/packages/clifro/index.html).

ASTD data on shipping and fishing activities in the Arctic were provided by the PAME International Secretariat. These data were aggregated to total occurrences (= abundances) of fishing vessels, passenger ships, and other ships in the waters of West Greenland (i.e. west of -44.0° longitude) and East Greenland (i.e. east of -44.0° longitude), respectively.

Finally, statistical analyses were carried out to identify drivers of beach litter abundances on Greenland. In detail, a permutational analysis of covariance (PERMANCOVA) was calculated with mean abundances of fishing vessels [counts] as factor, and numbers of inhabitants [counts], distance to next settlement [km], distance to next harbour [km], and overnight stays in the respective district of Greenland [counts] as covariates. Mean total abundances of beach litter [counts] on Greenland, taken from the OSPAR beach litter database, were used as dependent variable. Abundances of passenger ships and other ships were excluded as factors, to avoid over-parameterization and redundancy of information of the statistical model. Information on wind could not be directly linked to surveyed beach locations, because the distances between OSPAR beaches and the closest meteorological stations partly amounted to several hundred kilometres. Therefore, wind data were excluded from the statistical model, as well, and the interpretations of wind data was qualitatively rather than quantitatively. Additional tests for autocorrelation, normality, and homogeneity of variance were calculated to ensure that the results of the PERMANCOVA were not biased. The statistical software Systat 12.0 (Systat Inc.) was used for all statistical analyses.

3.1.2   Results and Discussion

According to Eisted and Christensen (2011), on Greenland, there are six small- to medium-sized incineration plants and 30 straw incineration plants. Nearly every settlement has its own landfill, partly household garbage is directly littered onto beaches. Information on the amount and composition of garbage on Greenland is rudimentary or completely lacking. An overview on the positions of landfills and their respective size is shown in B.2. Apart from inputs from shipping and fishing, wind erosion from landfills and direct littering on the beach substantially contribute to the amounts of beach litter on Greenland (Kirkfeldt, 2016).

However, sea-based sources of beach litter are dominating in the Arctic (Bergmann et al., 2017; PAME, 2019; Tekmann et al., 2017). Accordingly, wind driven transport leads to accumulation of beach litter at location highly exposed to wind (Schulz et al., 2015a). Abundances of beach litter are low when the portion of westerly winds is negligible (Figure 9). According to Figure 9,Figure 10, and Figure 11, at the western coast of Greenland, high portions of wind from the west and southwest might lead to accumulations of beach litter. This agrees well with analyses of beach litter data in the North Sea, where wind drift has been discussed as the major driving force of the beaching of litter (Neumann et al., 2014; Schulz et al., 2015a). Besides, the vicinity to cities, such as Nuuk, can be regarded as additional factor, causing enrichments of beach litter (Kirkfeldt, 2016).

Figure 9:   Map of Greenland with a wind rose
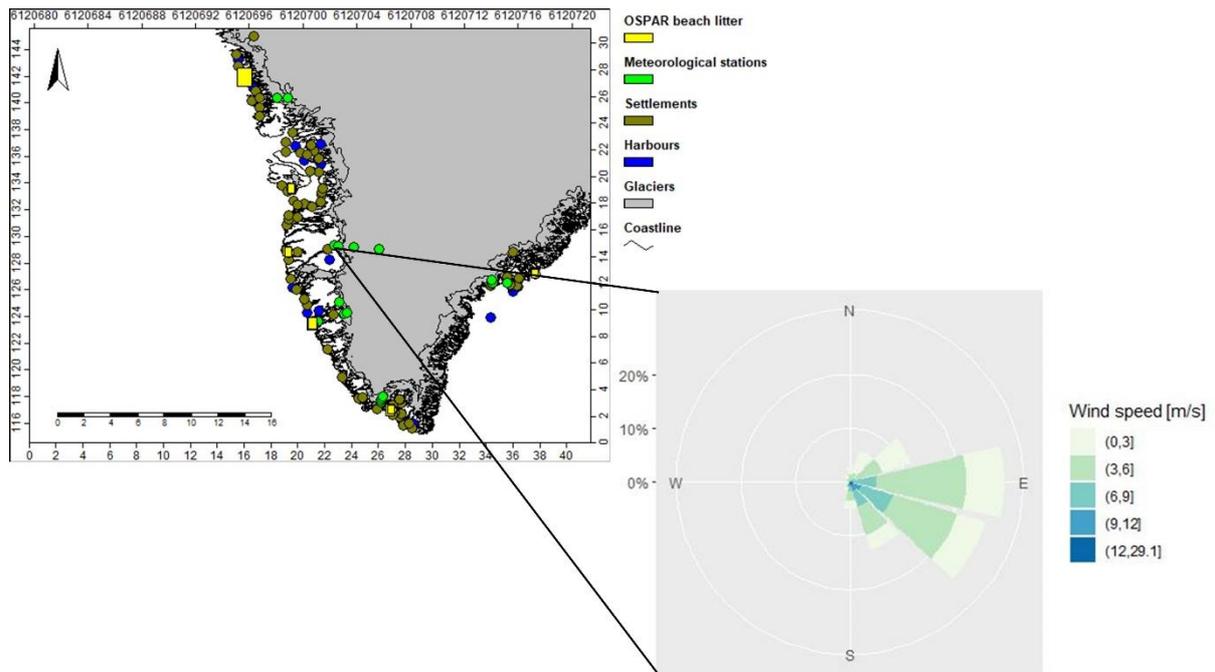
Black lines hint at the position of the respective meteorological station north of Nuuk. Source: Compiled by AE

Figure 10:   Map of Greenland with a wind rose
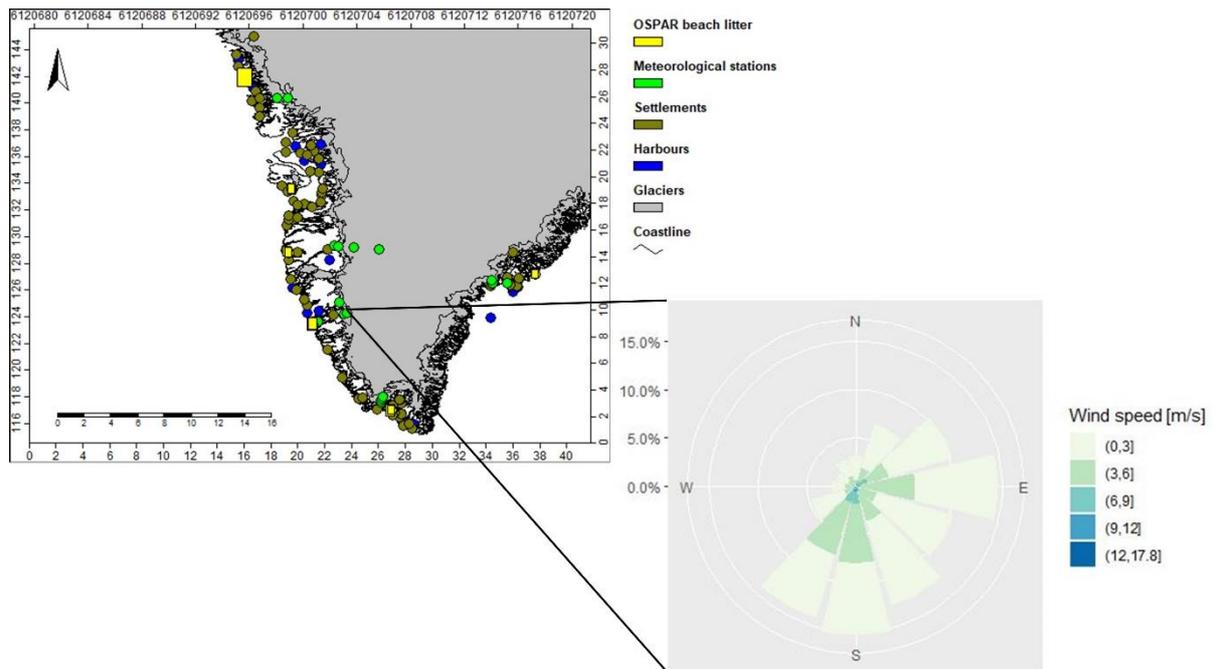
Black lines hint at the position of the respective meteorological station near Nuuk. Source: Compiled by AE

Figure 11:   Map of Greenland with a wind rose
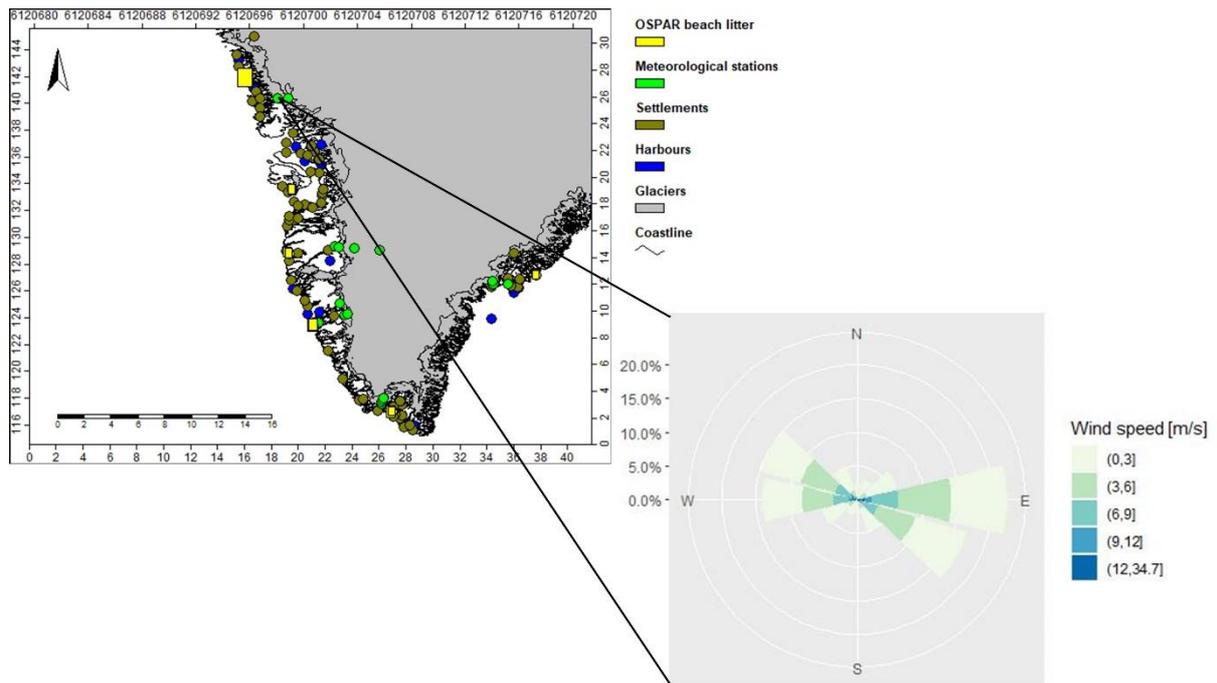
Black lines hint at the position of the respective meteorological station near Upernavik. Source: Compiled by AE

In addition, the prevailing residual currents partly explain increased abundances of beach litter at the western coast of Greenland (Figure 12). Thus, hot spots of beach litter are expected to be present at Nuuk, Ilulissat and Upernavik (B.1), which made these sections of coastline appropriate as beaches to be surveyed in this project. The observed low abundances of beach litter on the western coast of Svalbard (Weslawski and Kotwicki, 2018) agree with the positions sheltered from the predominant easterly winds (Figure 13). In the north of Svalbard, high amounts of beach litter were found on beach sections exposed to wind and currents (Bergmann et al., 2017). Note that for Svalbard, the wind rose given in the figure above is exemplary of a number of wind roses generated (i.e. for five locations on Svalbard), which exhibit similar patterns of prevailing easterly winds. In contrast to Greenland, information on potential driving forces of beach litter was limited on Svalbard. Therefore, here, interpretations of beach litter data have to be cautious and confined.

Figure 12:   Illustration of predominant residual currents in the waters around Greenland
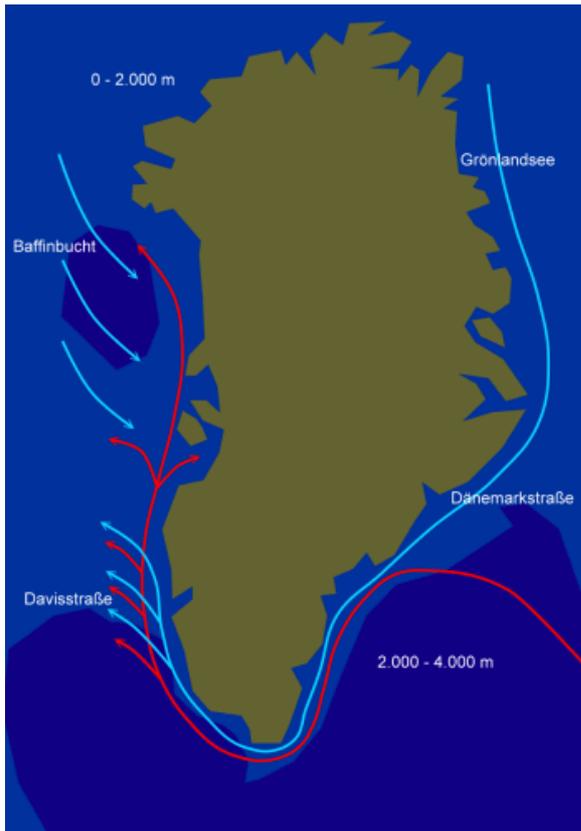
Red arrows reveal warm currents, Blue arrows show cold currents. Source: arktis-reise.de

Figure 13:   Map of Svalbard with a wind rose
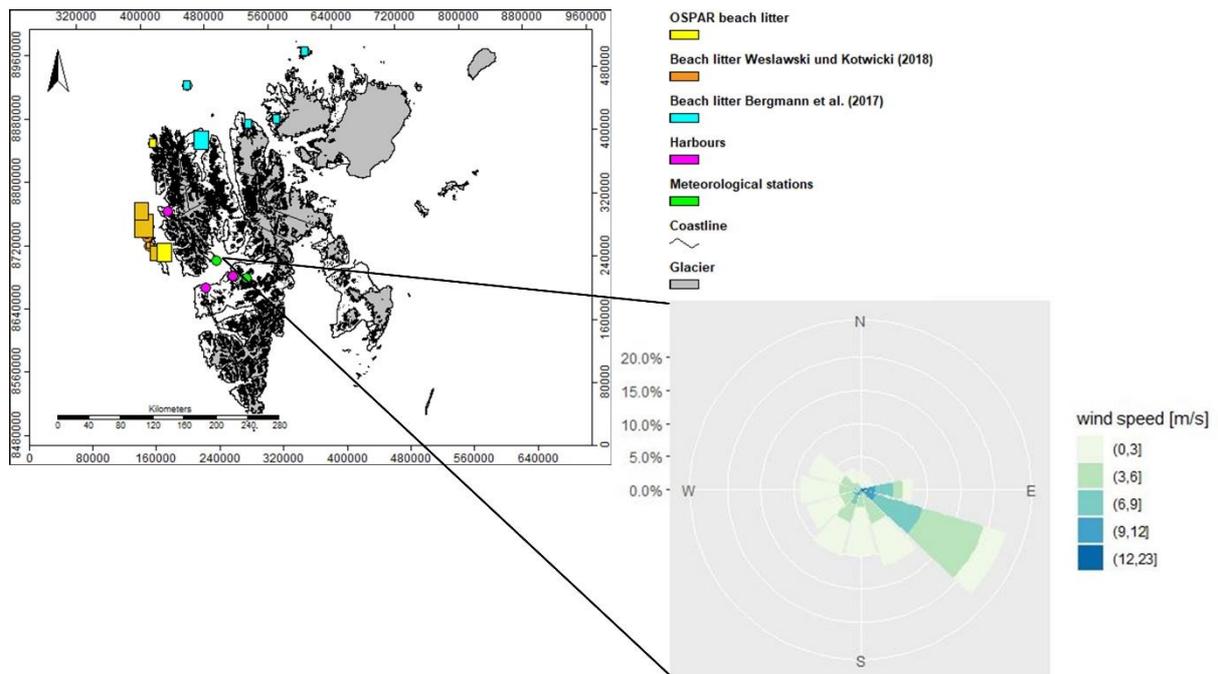
Black lines hint at the position of the respective meteorological station near Barentsburg.
Source: Compiled by AE

Table 3 shows aggregated data on ship traffic in the waters around Greenland. Abundances of shipping differed by one order of magnitude between West and East Greenland.

Table 3: **Aggregated data on shipping in the waters around Greenland, exemplarily given for the year 2016.**

Subcategory	Region	Arithmetic mean [abundance]	Standard deviation [abundance]
All ships	Greenland	19,515	4,237
	East Greenland	744	958
	West Greenland	18,771	3,579
Fishing vessels	Greenland	7,839	2,097
	East Greenland	256	149
	West Greenland	7,583	2,125
Passenger ships	Greenland	941	268
	East Greenland	62	138
	West Greenland	878	184
Other ships	Greenland	10,735	2,566
	East Greenland	425	785
	West Greenland	10,310	1,977

Table 4 summarizes results of the PERMANCOVA, calculated to identify drivers of beach litter on Greenland. While autocorrelation could not be excluded definitely, normality (Kolmogorov-Smirnov test) and homogeneity of variance (Levene test) were given. Therefore, the model is assumed not or barely to be biased. The model fit was good with a measure of determination of 0.83. Fishing and overnight stays were identified as significant factor and covariate, respectively, determining beach litter abundances. This is in line with previous studies on beach litter in the Arctic, which highlighted the dominance of sea-based sources of beach litter (Bergmann et al., 2017; Kirkfeldt, 2016; Tekmann et al., 2017). Overnight stays not necessarily represent the impact of tourism. They are high in areas with sufficient infrastructure, where emissions of domestic waste via landfills and direct littering are high, as well. In this respect, Strietmann et al. (2021) identified domestic waste as dominant portion of beach litter at several locations at the western coast of Greenland, while fisheries contributed another major percentage of beach litter. This is also in line with results from Mallory et al. (2021) who found that the distance to settlements significantly explained beach litter densities at the western coast of Greenland, while long-range transport from sea-based sources was assumed to add to high densities of beach litter even on remote beaches at high latitudes. In contrast, applying a Lagrangian transport model, Strand et al. (2021) supposed that fishery-related litter at beach sites on Svalbard stems from local sources (the Barents and Norwegian Seas). Accordingly, Vesman et al. (2020) attributed sea-based sources as major contributor to beach litter on the Northern Island of the Novaya Zemlya archipelago.

Table 4:   **Results of a permutational analysis of covariance (PERMANCOVA)**

Factor/ covariate	t [-]	p [-]	d [-]
Fishing vessels	15.196	0.002	
Inhabitants	1.551	0.239	
Distance of settlement	0.573	0.465	
Distance to harbour	0.274	0.611	
Overnight stays	36.823	<0.001	
Test			
Levene test		0.217	
Kolmogorov Smirnov test		0.438	
Dublin-Watson test			1.634

Results of a permutational analysis of covariance (PERMANCOVA) on driving forces of beach litter abundances on Greenland (r^2 = 0.827, p<0.05, n = 17).

3.2 Modelling approach to identify enrichments of beach litter

3.2.1 Methods

The main objective of this analysis was to predict locations around Svalbard with high probabilities of beaching of macroplastics. BCSH applied two datasets on currents with distinct resolutions and assessed the outputs from both models. For Greenland no high-resolution dataset of sea current was available and therefor the modelling approach was not applied.

3.2.1.1 Data sources

3.2.1.1.1 Fishing data

Sea-based sources of litter dominate in large parts of the Arctic with fishery being the main source (Bergmann et al., 2017; PAME, 2019; Tekmann et al., 2017). Bergmann et al. (2017) describes 44 to 100% of the mass of litter collected during beaches surveys in Svalbard consisting of fisheries-related items. Therefore, fishing data were used for a first estimation of the monthly sources of litter, derived from fishing activity of the region surrounding the Island of Svalbard. Fishing activity in the area north of 65° latitude and between 5°W and 65°E of longitude were included as sources (Figure 14).

First, fishing activity was obtained from the monthly AIS data for 2019, available at the EMODNET website (https://www.emodnet-humanactivities.eu/view-data.php), and downloaded in raster format. Each monthly raster represents the number of routes of fishing boats by square kilometre (Figure 14). Then, based on the monthly raster density, we randomly deployed plastic particles at sea every 100 km of route (0.1% of the total points). Finally, each monthly particle dataset was saved as an independent text file to be loaded later in the current models.

Figure 14:　　　　**Cumulated traffic density**
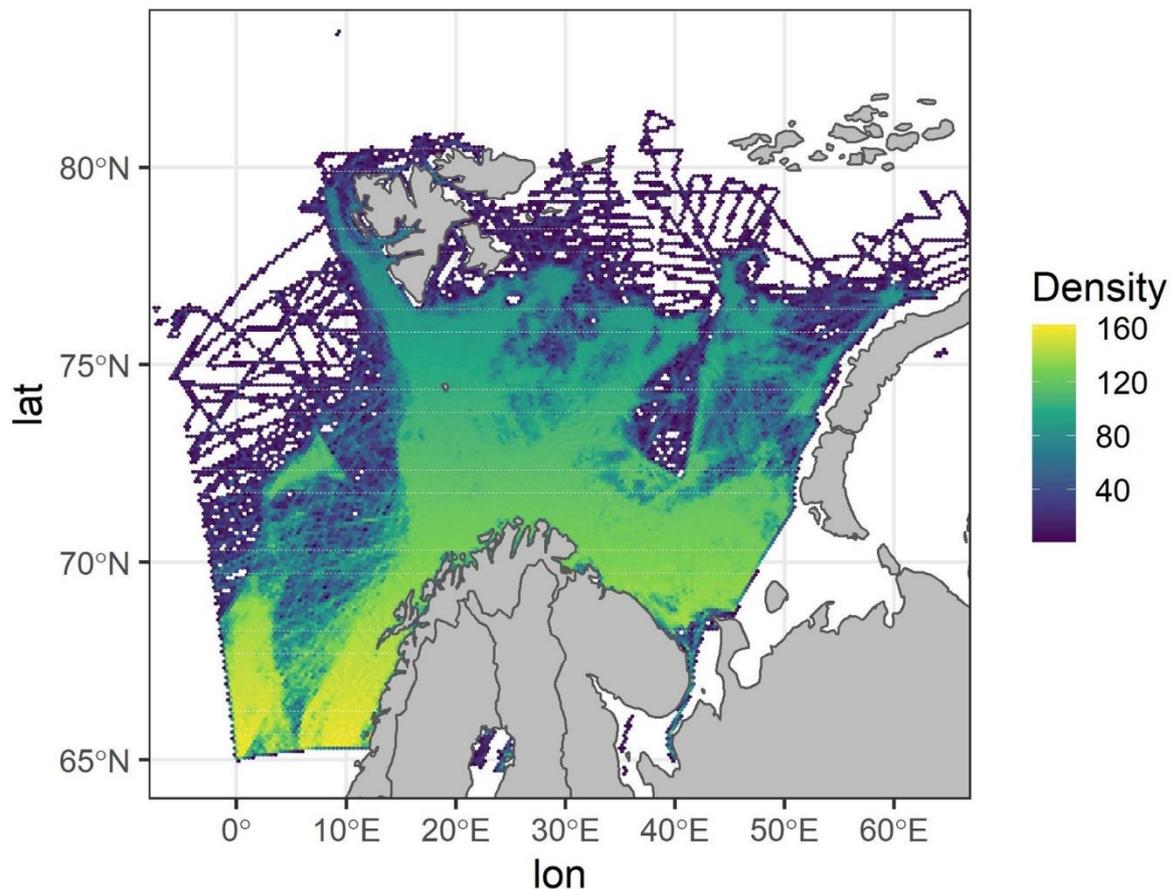
Cumulated traffic density (number of routes by square kilometre) for year 2019 with input data from the EMODNET website. Source: Compiled by BCSH

3.2.1.1.2　　Data on ocean currents

Two different daily currents datasets were assessed, each at a different spatial resolution and covering a different area extension:

A lower resolution and larger spatial coverage dataset was downloaded from the Marine Copernicus website (https://marine.copernicus.eu), with a resolution of 0.025 x 0.025 degrees and a spatial extension ranging from -10°W to 60°E and from 55°N to 90°N and a temporal coverage extended for two years, from 15th August 2008 to 15th August 2010 (C.1). The objective of this layer was to simulate the trajectories of plastic particles released according to the monthly fishing intensity during the full period. It was also used to feed particles to the higher resolution model.

On the other hand, daily data on currents at higher resolution on a more restricted spatial coverage around Svalbard are available on the Norwegian Meteorological Service website (https://api.npolar.no). This dataset is available from 2005 until 2010 at 800 m spatial resolution and covers a narrow area around Svalbard. Therefore, it could not cover the full fishing area considered for this study (C.2), and plastic particles had to be fed from the lower resolution model. Data on currents were downloaded for a temporal range of one year, from 16th August 2009 to 16thAugust 2010.

3.2.1.1.3 Wind data and wave data

Even if wind and wave stoke data are relevant components to predict the trajectories of floating objects at sea, no reliable data were found for the considered area and period. Most of the satellite-based ocean wind and waves products were either not available for the years 2008 - 2010 or did not cover latitudes higher than 80°N. In addition, resolution was never higher than 0.25°. Therefore, wind and wave data could not be considered for the present study.

3.2.1.2 Particle Trajectory Models

Modelling was performed with OpenDrift v1.0, an open-source Python based framework for Lagrangian particle modelling (Dagestad et al. 2018).

Lagrangian trajectory models are of common use to predict the pathways of floating objects in the ocean. Data on currents were fed as NetCDF files (Copernicus) or NetCDF ROMS files (API), while particles were multi-seed randomly inaugurated along the corresponding month from text files.

As mentioned above, modelling was performed in two phases. In the first phase (LR-model) the particle trajectory model was applied on the simulated marine litter released each month since August 2008 until August 2010 for the whole area, using the low-resolution current data from Copernicus. During the second phase (HR-model), particles from the LR-model entering the area covered by the high-resolution current data (API) were fed into the high-resolution current data from API from August 2009 until August 2010 (Figure 15).

The standard configuration of the PlasticDrift module was used for both models. However, due the lack of information regarding wind and wave stoke, a 10% uncertainty factor was included in the currents, thus giving the particles more room to spread over a larger surface area and to simulate more real-life conditions than those originated by the sea current alone.

Figure 15: Schematics of the seeding and modelling time periods
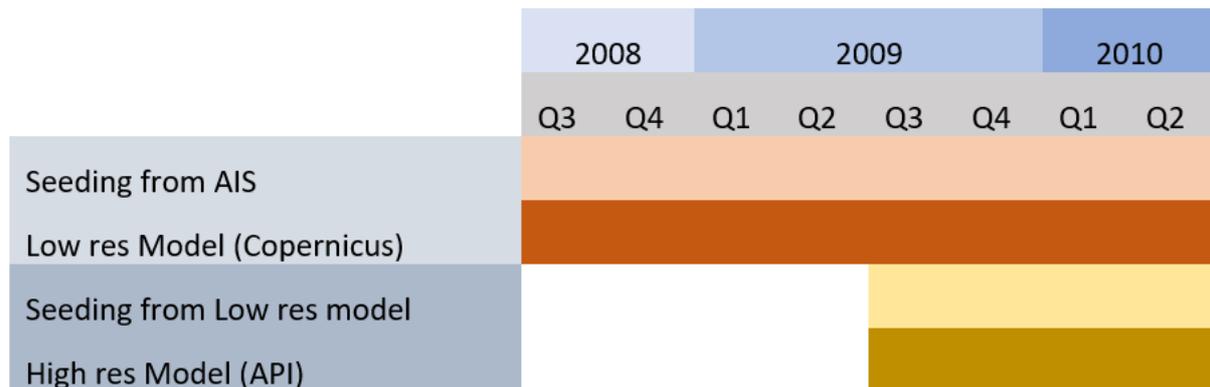
Schematics of the seeding and modelling time periods for the low-resolution and high-resolution plastic trajectory models. Particle seeding from AIS data was done during the entire period for the Low-resolution model, while particle seeding in the high-resolution model using the low-resolution model output as particle source was done between August 2009. Source: Compiled by BCSH

Cumulative stranding of litter particles at the end of the study period (August 2010) was obtained for the LR model and HR model and was stored as NetCDF files for further analysis.

3.2.1.3 Stranding density and comparison between models

Stranded particles in the Svalbard area from both, the LR model and HR model, were imported into R software (RCore, 2019) and visualized using the ggplot2 library (REF). Statistical correlation between both point distributions was assessed by means of a linear model. In

addition, statistical differences between the spatial density pattern of stranded particles along the coast were assessed, computing the significance of the pairwise differences relative to the mean and variance of all differences between the two input datasets. This allowed for identifying regions of significant difference between the two datasets (i.e. Bateman et al 2010; Januchowski et al. 2010). Significant spatial differences were analysed using the SDMTools library (REF) in R software.

3.2.2   Results

3.2.2.1   Litter beaching results from different model resolutions

Results obtained at the end of the period study (August 2010) are shown for the LR model (Figure 16) and HR model (Figure 17). In both cases, green dots represent the initial plastic particles released (42,912 particles for the LR model and 13,948 particles for the HR model).

Figure 16:   Start and end situation of the LR model
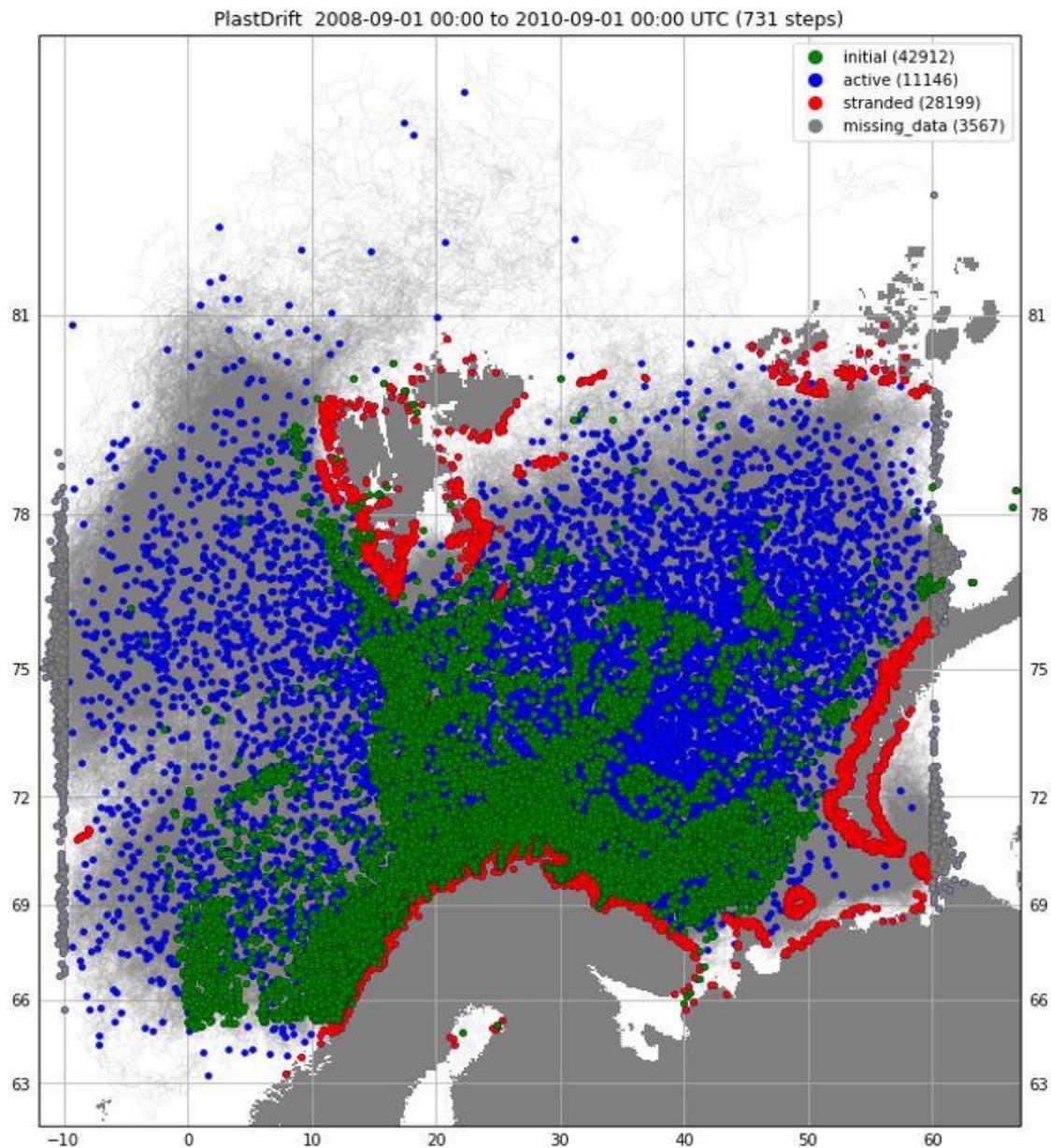
Green dots represent initial particles locations estimated from the AIS data, blue dots represent final (active) particle

locations, red dots represent stranded particles in the last time step and grey dots are missing particles (leave the study area). Grey lines show the trajectory of the particles. Source: Compiled by BCSH

Figure 17: **Start and end situation of the HR model**
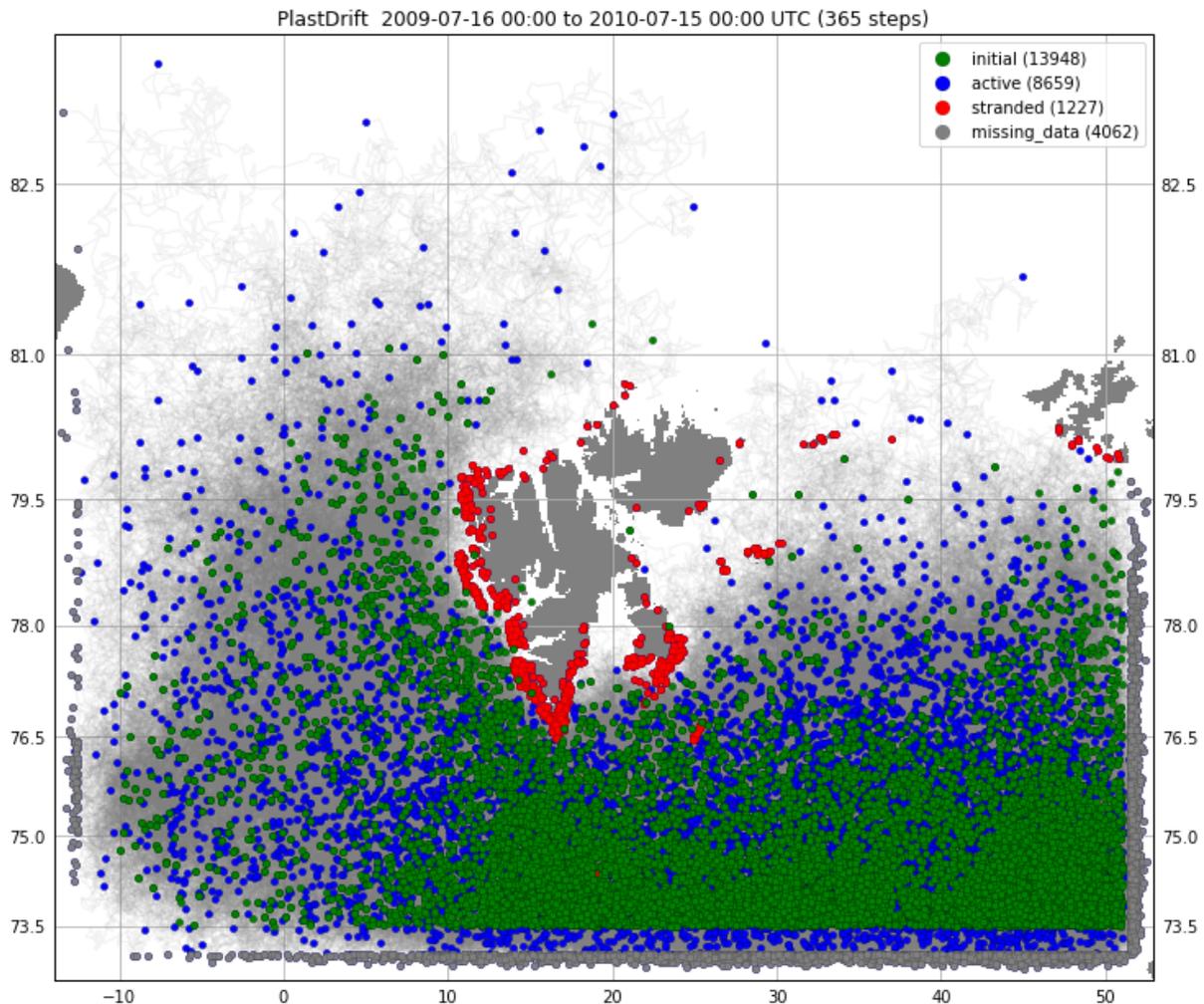
Green dots represent initial particles locations from the LR model, blue dots represent final (active) particle locations, red dots represent stranded particles in the last time step and grey dots are missing particles (leave the study area). Grey lines show the trajectory of the particles. Source: Compiled by BCSH

Detailed stranding points and density of points are shown in Figure 18 for the LR model and Figure 19 for the HR model, respectively. Both particle trajectory models show similar stranding patterns regardless of their spatial resolution. Highest stranding probability occurred at the Southern and Western coast of Svalbard.

3.2.2.2 Comparison of litter beaching results

To make two comparable datasets, points were aggregated in cells of 5 km x 5 km intersecting the coastline (Figure 20). A linear regression model showed a significant relation between both density distributions along the coast of Svalbard with a measure of determination (R^2) of 0.758. The residuals of the linear regression model showed higher variations between predicted standings in the southern coast (Figure 21). A spatial significant difference analysis confirmed the results from the linear regression, showing significant differences in the southern and western area (Figure 22). The HR trajectory model showed more probability of stranded particles at the western coast and the island of Hopen (indicated by orange squares) while the

60

LR trajectory model showed more probability of stranded particles in the Storfjorden and on Edgeøya (indicated by green squares).

Figure 18:   Stranding particles and particle density of the LR model
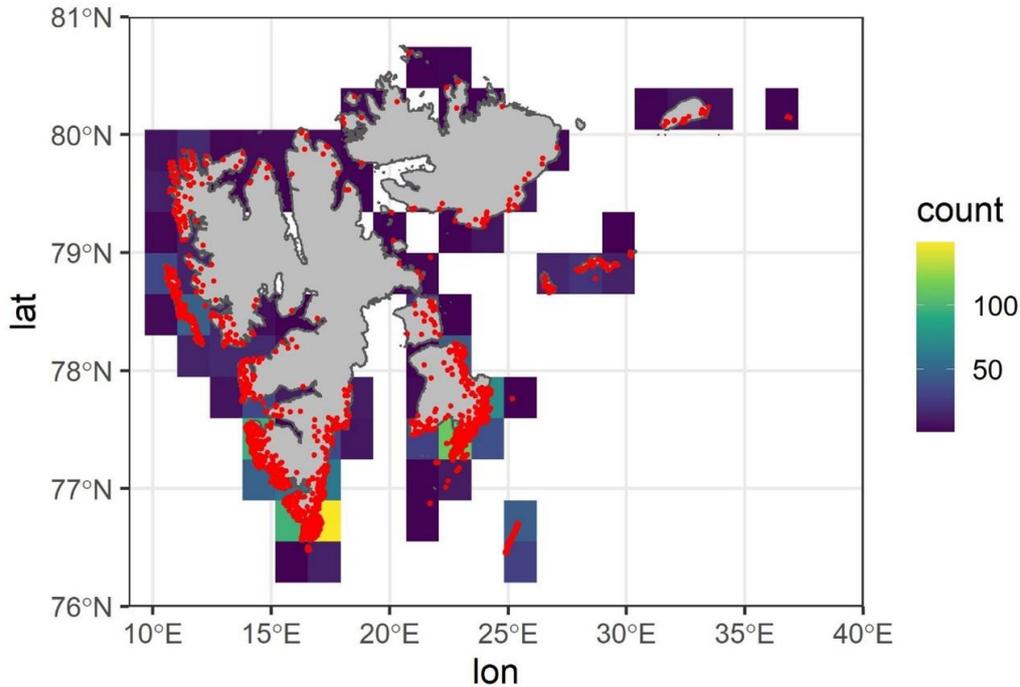
Stranding particles (red dots) and particle density estimates along the coast of Svalbard derived from the Low-Resolution model at the last time step. Source: Compiled by BCSH

Figure 19:   Stranding particles and particle density of the HR model
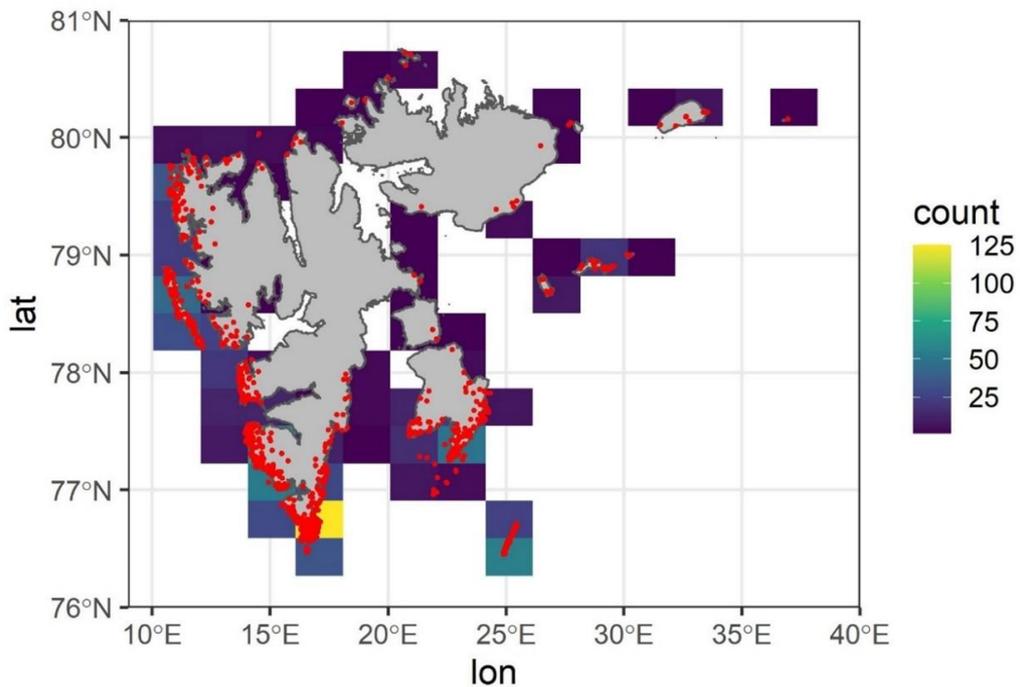
Stranding particles (red dots) and particle density estimates along the coast of Svalbard derived from the High-Resolution model at the last time step. Source: Compiled by BCSH

Figure 20:   **Comparison of the density of stranding particles for the two models**
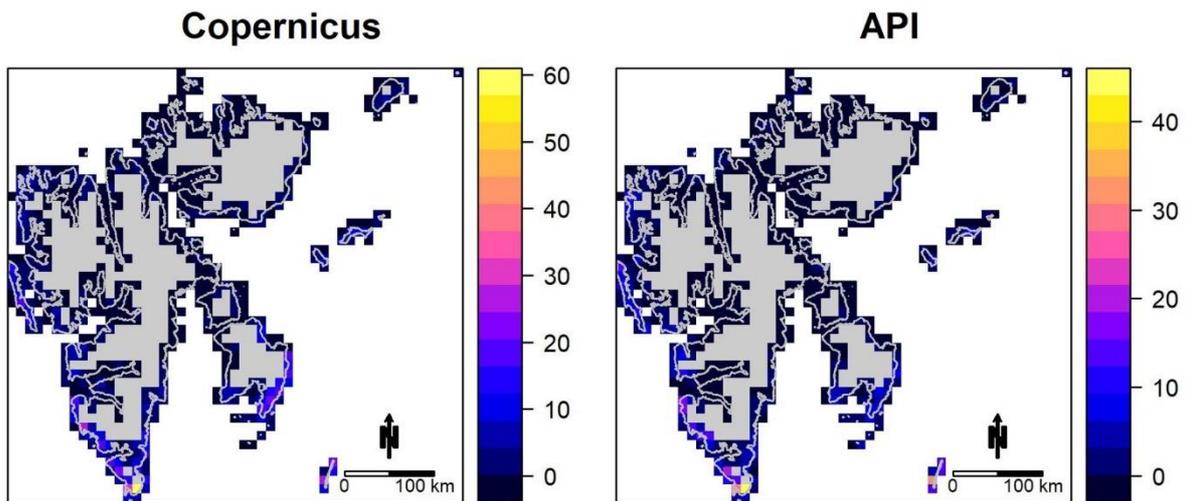
Density of stranding particles by 5 km x 5 km cell used for the comparison between the LR model and HR model stranding output at the last time step. Scale is in number or particles by 5 km x 5 km cell. Source: Compiled by BCSH

Figure 21:   **Residuals of the linear model comparing the stranding density obtained from the HR and LR models**
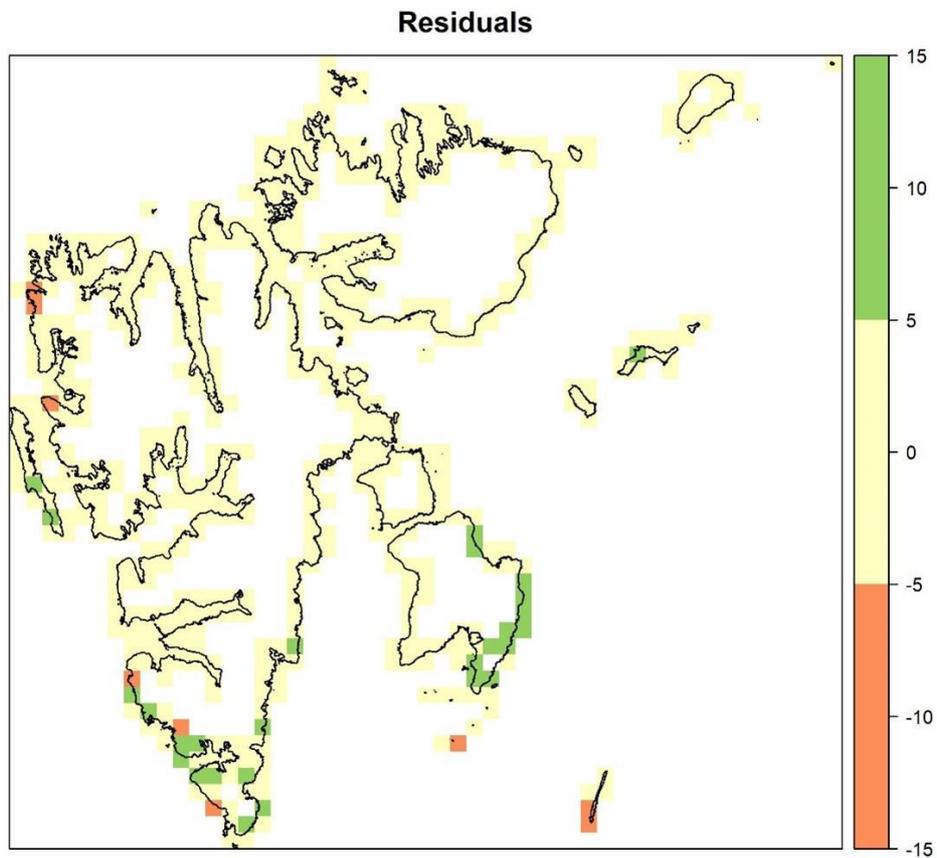
Although both results were well correlated (R^2 = 0.795), high residual values were observed at the southern and west coast. Red squares show areas with much higher probability of stranding shown by the HR model. Green squares show areas with much higher probability of stranding probability shown by the LR model. Cell size has 5 km x 5 km resolution.
Source: Compiled by BCSH

Figure 22:   Pattern Differences of LR and HR
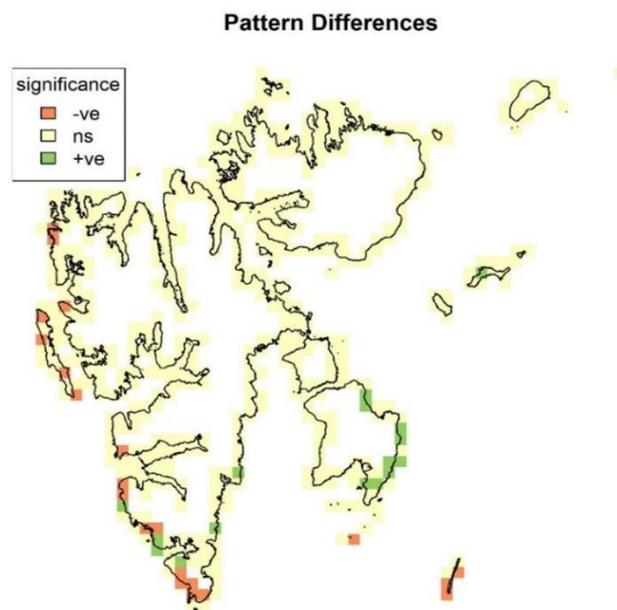
Significant spatial differences between the stranding densities obtained from the HR and LR models. Red squares show areas with higher significant probability of stranding in the HR model. Green squares show areas with higher significant probability of stranding in the LR model. Cell size has 5 km x 5 km resolution. Source: Compiled by BCSH

3.2.3   Discussion

The results of the two model approaches indicated potential beach litter accumulation to be mainly on the West coast of Svalbard (Figure 18, Figure 19). However, studies on beach monitoring as well as OSPAR studies on Svalbard indicated relatively low abundances on the western coast of Svalbard compared to higher abundances in the North. One explanation could be the predominating Easterly winds. As wind and wave data were not available for the study area during the study period, results might change once these are included. Wind and wave data are also a requirement for the Leeway configuration, that is a more accurate model configuration for big size plastics compared to PlasticDrift which focuses on microplastics. Anyhow, interpretations of beach litter data on Svalbard have to be cautious as reference data from conducted beach litter surveys are only available on the northern and western coasts and information on potential driving forces of beach litter is limited. Assuming that the results from both models (Copernicus and API) have enough similarities, it could be possible to run a new low-resolution model for 2019 with wind and wave data. Moreover, data on abundances of beach litter received by this studies' fieldwork should be included for future approaches. Additional input parameter could further improve the model. The results of this study show the numbers of particles arriving to the coast. The elements are deactivated once they touch the coastline. However, they could be returned to the water immediately if it is not a sedimentary environment (i.e. cliffs), or if they arrive on a beach during high tide or a storm. This will depend on various factors such as the type of coast (beach or cliff), the slope of the beach or wave angle incidence. A follow-up study could build on the existing model and include wind and wave data as well as information about the coast type. Single coast type information could be acquired during the fieldwork for this study. Anyhow, because of the relative broad resolution of ocean circulation models and wind and wave data, modelling approaches can so far just give a first indication of the location of accumulation sites of marine litter in the Arctic. Arctic-wide higher resolution input data would be necessary to identify single beaches.

4　Work package 3

4.1　Fieldwork Summary

On April 30, 2021, M. Schulz (AE), A. Trobisch (AE) and M. Schnurawa (BCSH) performed a test beach litter survey on the dog beach of Rantum on the island of Sylt, applying the OSPAR method (2010) and a drone survey. Two 50 m-sections of beach were surveyed between the dune belt and the waterline applying the OSPAR method (OSPAR, 2010). Coordinates of large pieces of beach litter were measured applying a high-resolution Global Positioning System (GPS). Subsequently, the drone survey was carried out by M. Schnurawa on a 600 m-long beach section, which included the two 50 m-long sections of beach conventionally surveyed.

In the first six months of 2021, A. Trobisch (AE) and M. Schnurawa (BCSH) organized the expeditions to Svalbard and Greenland evaluating the different options for logistics. Promising approaches were land-based operations with local partners using zodiacs or small boats, the chartering of a private vessel and the cooperation with cruise operators.

Regarding cruise operators, BCSH was in exchange with the Association of Arctic Expedition Cruise Operators (AECO). The environmental specialist of AECO introduced the project to AECOs members and several cruise operators (e.g. Arctic Sailing Escape, Hurtigruten, Marine Incognita, OceanWide and Secret Atlas) and contacted BCSH discussing potential cooperation. Due to the uncertainties of Covid-19 pandemic BCSH and AE decided against a cruise-based fieldwork. Anyhow, a cooperation with a cruise operator is an attractive option which should be considered for future campaigns as it can combine scientific research, public relation, and education. Furthermore, it permits the application of citizen science approaches as well as an implementation within an existing infrastructure.

Chartering a private vessel was just considered shortly, as prices were too costly and only feasible when sharing the vessel with other stakeholders or scientists which was not possible due to Covid-19 pandemic.

Finally, BCSH and AE decided to apply the land-based approach for all study areas on Svalbard and Greenland collaborating with local partners. From the base station on land, the surveys were planned, and the equipment could be prepared. A boat operator took all participants with a zodiac or small boat to the study area. As the results of working package 2 could only give rough indications of beaches with high litter accumulations, local expertise was essential. On Svalbard Geir Gabrielsen from the NP provided detailed knowledge about potential beaches. For Greenland BCSH and AE exchanged with WSP and WWF for Nuuk and the boat driver Jan Banemann and the municipality of Qeqqata for Sisimiut. The land-based operation allowed a flexible survey planning and short-term decisions required by the weather conditions.

The expedition to Svalbard lasted from June 25, 2021 to July 17, 2021. A quarantine of ten days was necessary in Oslo, before the journey to Ny-Ålesund via Longyearbyen could be continued on July 5, 2021. During their stay in Ny-Ålesund, A. Trobisch and M. Schnurawa participated in scientific meetings and exchange within the research camp. Due to unfavourable weather conditions, apart from a test of the equipment outside the Radio Silent Zone on July 7, 2021, only one one-day excursion for a beach litter survey at Kapp Mitra was possible (Figure 24). The day before excursion, strong winds prevailed, probably influencing the beaching of floating litter. Apart from A. Trobisch and M. Schnurawa, a boat guide, two students and two polar bear observers participated in the excursion. A total of two 50 m sections of beach were completely surveyed applying the OSPAR method. Coordinates were determined by means of a high-resolution GPS. The coordinates of single pieces of litter could not be measured because of lacking access to internet and radio. An additional section of beach could only partly be surveyed

by as much as 20%, because a polar bear was approaching, and therefore, the excursion had to be terminated. Table 5 shows the properties and coordinates of the surveyed beach sections.

Before starting the OSPAR monitoring, drone surveys were conducted (Two RBG and two VIR surveys; Figure 23). The surveys included the two sections monitored after OSPAR and comprised around 2.75 km of coastline and an area of around 25 ha.

WV3 satellite images of Kapp Mitra were already acquired on July 25, 20121 The image covers an area of around 7500 ha with a coastline of 23 km length and includes 17 spectral bands (1x pan-chromatic, 8x RGB/ NIR, 8x SWIR). Due to potential cloud coverage an acquisition time span of up to three weeks before the fieldwork was chosen. The final offset between satellite acquisition and fieldwork was eight days.

Figure 23   Kapp Mitra, Landing of WingtraOne – Picture: Geir Wing Gabrielsen

Table 5   Coordinates and properties of the two sections of beach at Kapp Mitra, surveyed on July 11, 2021. Coordinates are given as Universal Transverse Mercator (UTM) – coordinates with the date WGS84.

Property	Kapp Mitra 1	Kapp Mitra 2
Position Start	550671.3 N, 8783090.3 E	550621.2 N, 8783078.4 E
Position End	550621.2 N, 8783078.4 E	550571.6 N, 8783066.1 E
Beach fact sheet	no	no
Beach substrate	80% rock, 20% sand	80% rock, 20% sand
Beach inclination	10%	10%

65

Property	Kapp Mitra 1	Kapp Mitra 2
Mean width at high water nip tide	15 m	15 m
Total length of the beach	175 m	175 m
Beach surroundings	Cobbles, drift wood, tundra	Cobbles, drift wood, tundra
Exposition to sea	SSE	SSE
Prevailing wind direction	E – SE	E – SE
Prevailing residual currents	No information	No information
Objects influencing currents	No	No
Distance to next settlement/harbour	23 km (Ny-Ålesund)	23 km (Ny-Ålesund)
Position of next settlement/harbour	SE	SE
Number of inhabitants of next settlement	50 (winter) – 150 (summer)	50 (winter) – 150 (summer)
Type of next harbour	Mainly zodiacs and small arctic circle boats, few seal hunting boats and ships for supply	Mainly zodiacs and small arctic circle boats, few seal hunting boats and ships for supply
Size of next harbour	small (10 – 15 vessels)	small (10 – 15 vessels)
Next river mouth	No information	No information
Wastewater discharge	no	no
Tourism	no	no
Infrastructure behind the beach	no	no
Facilities for food consumption at the beach	no	no
Beach cleaning activities	no	no
Litter collection	no	no
Accessibility	By boat with polar bear observers	By boat with polar bear observers
Beach usage	For the first time for scientific purposes, no usage before	For the first time for scientific purposes, no usage before

The expedition to Greenland took place from September 13, 2021, to September 28, 2021. Due to technical problems of the airline, the flight back had to be postponed to September 29, 2021. During the stay on Greenland, beach litter surveys were performed near Nuuk and Sisimiut with up to four replicate sections of beach (Table 6). Travelling to the selected beaches was done by boat with a boat guide. Conventional surveys were carried out applying the OSPAR method (OSPAR, 2010), where the most recent categorization of 2021 was used. Categorization was amended by beach litter types, which are characteristic of the western coast of Greenland. These new litter types were derived from the SUMAG project and are based on the splitting of OSPAR beach litter types, as was similarly done by the Joint Category list of the MSFD. Furthermore, GNSS measurements were made to determine the exact position of the found litter as a reference for the drone surveys. Due to technical problems with the GNSS receiver, this approach was only available for the two survey areas around Sisimiut.

Before starting the OSPAR monitoring, for each of the six survey areas drone surveys were conducted (two RBG and two VIR surveys; Table 7). WV3 satellite images of Nuuk and Sisimiut were acquired on September 13, and September 15, 2021respectively. All study sites were cloud free. An overview of the satellite coverage is shown in Figure 25 and Figure 26. For Nuuk, the time difference between satellite acquisition and fieldwork was only two to four days, for Sisimiut six to seven days.

Table 6 Coordinates and properties of surveyed sections of beach near Nuuk and Sisimiut. Coordinates are given as Universal Transverse Mercator (UTM) – coordinates with the date WGS84.

Property	Nuuk Day 1	Nuuk Day 2-1	Nuuk Day 2-2	Nuuk Day 2-3	Nuuk Day 3	Sisimiut Day 4-1	Sisimiut Day 4-2	Sisimiut Day 4-3	Sisimiut Day 4-4	Sisimiut Day 5-1	Sisimiut Day 5-2
Date	Sep. 15, 21	Sep. 16, 21	Sep. 16, 21	Sep. 16, 21	Sep. 17, 21	Sep. 23, 21	Sep. 23, 21	Sep. 23, 21	Sep. 23, 21	Sep. 24, 21	Sep. 24, 21
Position Start	461901 N, 7109578 E	462106 N, 7109376 E	462850 N, 7105287 E	462855 N, 7105248 E	462037 N, 7094362 E	404196 N, 7424987 E	404180 N, 7425034 E	404138 N, 7425123 E	404114 N, 7425167 E	409475 N, 7424354 E	409590 N, 7424297 E
Position End	461862 N, 7109570 E	462093 N, 7109345 E	462879 N, 7105285 E	462840 N, 7105163 E	462022 N, 7094386 E	404180 N, 7425034 E	404138 N, 7425123 E	404114 N, 7425167 E	404086 N, 7425208 E	409505 N, 7424323 E	409578 N, 7424332 E
Beach substrate	90% gravel	50% gravel, 50% blocks	50% gravel, 50% sand	100% gravel	90% gravel	100% sand	100% sand	100% sand	100% sand	85% sand, 12,5% gravel, 2,5% blocks	85% sand, 15% gravel
Beach cleaning	No	No	No	Seasonally in May/ June and September / October	No	No	No	No	No	No	No
Mean width at high water nip tide	15 m	17 m	8 m	5 m	11 m	9 m	9 m	9 m	9 m	8 m	14 m
Total length of the beach	40 m	30 m	30 m	90 m	30 m	250 m	250 m	250 m	250 m	50 m	50 m
Beach surroundings	Tundra	Tundra	Tundra and grassland	Tundra	Tundra	Tundra	Tundra	Tundra	Tundra	Tundra	Tundra
Exposition to sea	W	N	S/ SE	E	S/ SE	WSW	WSW	WSW	WSW	W	W

Property	Nuuk Day 1	Nuuk Day 2-1	Nuuk Day 2-2	Nuuk Day 2-3	Nuuk Day 3	Sisimiut Day 4-1	Sisimiut Day 4-2	Sisimiut Day 4-3	Sisimiut Day 4-4	Sisimiut Day 5-1	Sisimiut Day 5-2
Prevailing wind direction	NE	E/ NE	S/ SE	S/ SE	S/ SE	W	W	W	W	W	W
Prevailing direction of currents	NE	E/ NE	S/ SE	S/ SE	S/ SE	W	W	W	W	W	W
Objects, which influence near-shore currents	No	No	No	No	No	No	No	No	No	No	No
Distance to next settlement/ harbour	6.5 km	6.5 km	11 km	11 km	21 km	20 km	20 km	20 km	20 km	27 km	27 km
Position of next settlement/ harbour	NE	NE	N	N	N	WNW	WNW	WNW	WNW	NW	NW
Number of inhabitants of next settlement	17,500	17,500	17,500	17,500	17,500	5,500	5,500	5,500	5,500	5,500	5,500
Type of next harbour	Small and big ships, ferries and container ships	Small and big ships, ferries and container ships	Small and big ships, ferries and container ships	Small and big ships, ferries and container ships	Small and big ships, ferries and container ships	Sea port	Sea port	Sea port	Sea port	Sea port	Sea port
Size of next harbour	big	big	big	big	big	small	small	small	small	small	small
Tourisms	no	no	no	no	no	no	no	no	no	no	no

Property	Nuuk Day 1	Nuuk Day 2-1	Nuuk Day 2-2	Nuuk Day 2-3	Nuuk Day 3	Sisimiut Day 4-1	Sisimiut Day 4-2	Sisimiut Day 4-3	Sisimiut Day 4-4	Sisimiut Day 5-1	Sisimiut Day 5-2
Infrastructure behind the beach	none	none	none	none	none	none	none	none	none	none	none
Facilities for food consumption at the beach	none	none	none	none	none	none	none	none	none	none	none
Beach cleaning activities	no	no	no	yes	no	no	no	no	no	2019	2019
Litter collection	no	no	no	yes	no	no	no	no	no	2019	2019
Accessibility	By boat	By boat	By boat	By boat	By boat	By boat	By boat	By boat	By boat	By boat	By boat
Beach usage	Fishing nearby	Fishing nearby	Fishing nearby	Fishing nearby	Fishing nearby	No information	No information	No information	No information	No information	No information

Figure 24   Study Area Svalbard: Kapp Mitra
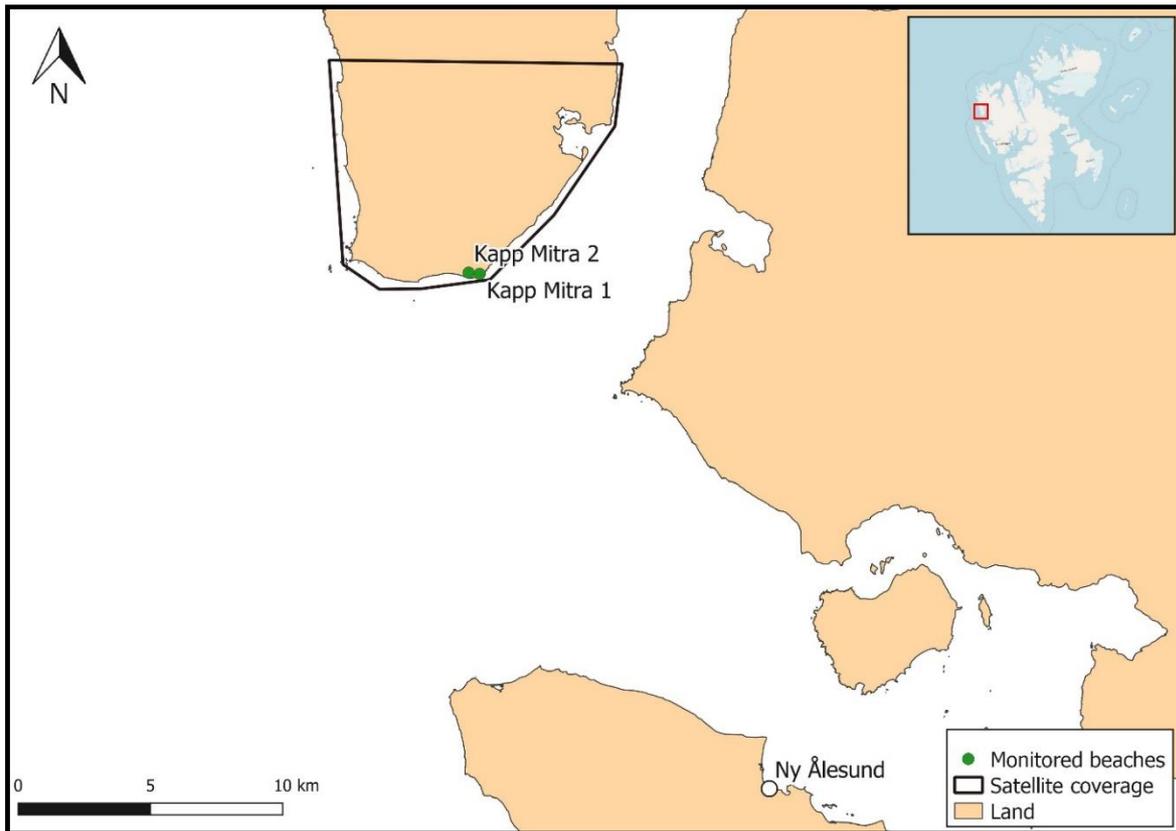
Source: Compiled by BCSH

Figure 25   Study Area Greenland: Nuuk
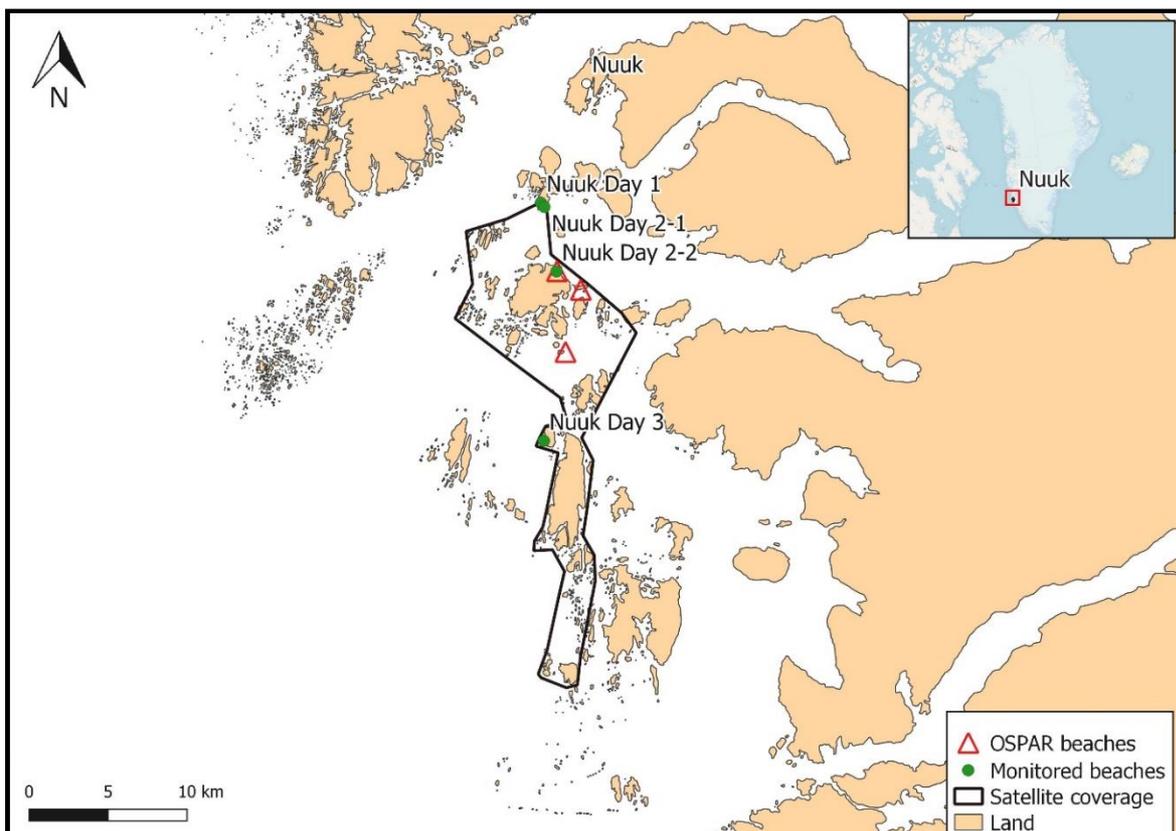
Source: Compiled by BCSH

Figure 26   Study Area Greenland: Sisimiut
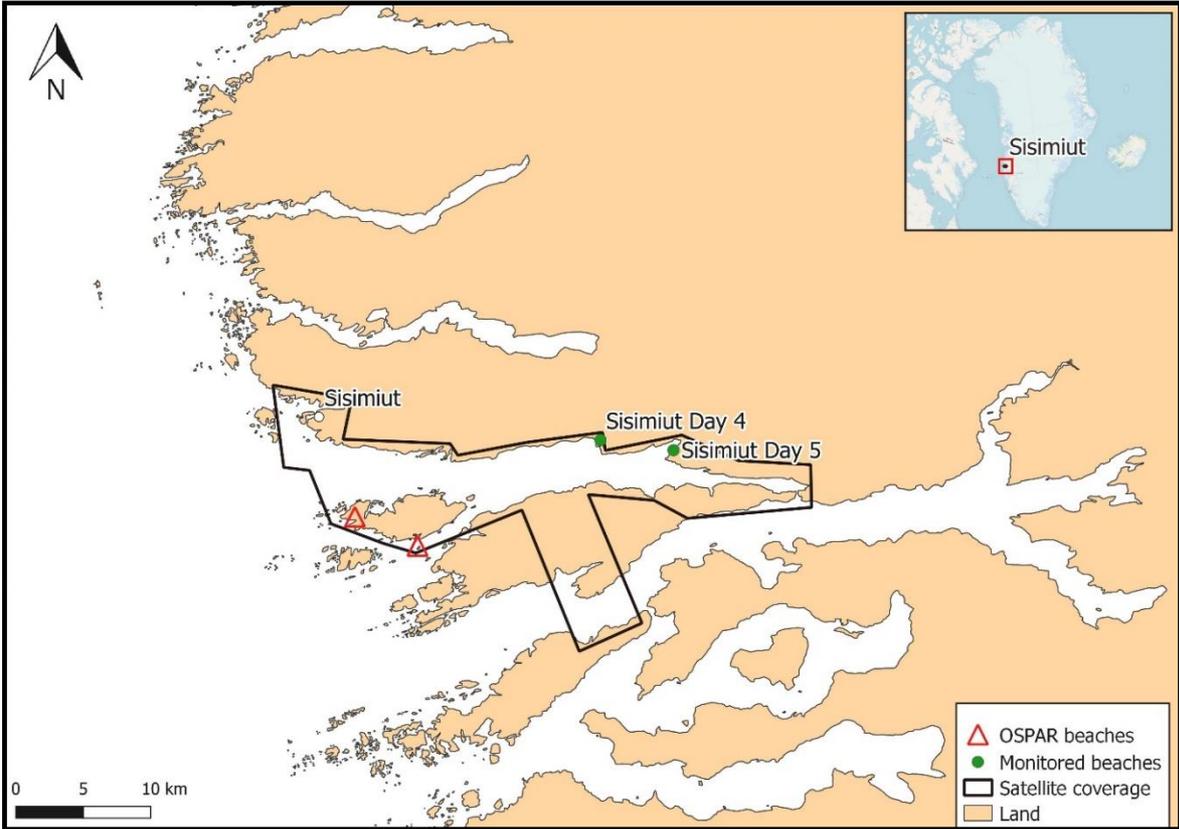
Source: Compiled by BCSH

4.2 Methods

4.2.1 OSPAR

For in situ beach litter monitoring the OSPAR method was applied as described on page 35 section 2.1.2.1.

Based on the survey data of the sections of beach on Svalbard and Greenland, mean abundances were calculated. From these mean abundances, mean material compositions and the most common litter categories were deduced (Table 2). Power analyses on one-sided one-sample Wilcoxon rank sum tests were calculated to estimate how long beach sections have to be for a significant detection of beach litter reduction by as much as 10%, 20%, 30%, 40% and 50%, respectively. For this purpose, a significance level of 0.05 was used. Power analyses were done with the statistical software Gpower (® University of Kiel, Germany).

4.2.2 Drone Survey

To obtain high-resolution aerial images of the study area, drone surveys were applied. Therefore, a WingtraOne drone with two payloads, a Sony QX1 20MP camera and a MicaSense Altum multispectral sensor, were used. Only one payload at the time could be mounted, resulting in at least two surveys per study area. The WingtraOne operates like a fixed wing drone and is therefore specialized for large scale applications as required in the Arctic. The GSD for all surveys were 1.4 cm and 3.4 cm for the Sony QX1 and the MicaSense Altum respectively, resulting in a flight altitude of 68 m and 80 m. Those GSD were the lowest possible for this drone-sensor configuration. Frontal and side overlap were set to 56% and 80 - 90% for the Sony QX1 and 70% and 80 - 90% for the MicaSense Altum. The frontal overlap was limited by the drones' minimum flight speed, the GSD and the trigger time of the sensor and was maximized for both configurations. The flight time of the single surveys varied between 06:56 and 49:26 minutes covering areas between 1.31 ha and 33.85 ha. Between 108 and 1877 single images were recorded and geo-tagged per survey (Table 7). The WingtraOne's PPK module and a Stonex S900T GNSS receiver were used to correct the absolute geo-location with an accuracy of around 2 cm for the RGB images. The geo-tagged images were processed in Pix4D mapper version 4.6.4 to create point clouds and to calculate elevation models, othomosaics and index maps. As the PPK correction was only available for the Sony QX1, Ground Control Points (GCPs) were used for the alignment between multispectral and RGB images.

Table 7   Overview of Drone surveys conducted on Svalbard and Greenland

Study Area	Date	Local time VIR- survey	Nr.of pictures [RGB/ VIR]	Flight duration [min] [RGB/ VIR]	Area covered [ha][RGB/VIR]
Kapp Mitra 1	June 11, 21	10:38 – 11:28	1223/ 1877	45:11/ 50:26	32.1/ 33.85
Kapp Mitra 2	June 11, 21	12:50 – 13:36	1165/ 1790	37:03/ 46:42	33.97
Nuuk Day 1	September 15, 21	10:07 – 10:24	163/ 240	09:00/ 17:19	2.93
Nuuk Day 2-1	September 16, 21	09:31 – 09:46	186/ 285	10:56/ 15:47	2.13
Nuuk Day 2-2	September 16, 21	13:09 – 13:23	142/ 205	09:55/ 14:08	1.31
Nuuk Day 3	September 17, 21	13:53 – 14:12	988/ 454	33:45/ 19:34	21.09/ 5.45
Sisimiut Day 1	September 23, 21	11:22 – 11:25	108/ 168	08:08/ 03:56	1.98
Sisimiut Day 2	September 24, 21	12:35 – 12:51	226/ 345	10:55/ 16:49	4.24

4.2.2.1 Manual Screening

For the manual screening, the orthomosaics (RGB, GSD: 1.4 cm) were loaded into QGIS 3.16.6. A 5 x 5 m grid and a scale of 1:20 were used to visually search the study area for litter objects. All images were screened by the same person, making comparison between surveys more reliable. In-field knowledge of the main litter categories was used to facilitate the categorization.

The boundaries of litter objects were marked and saved into a shapefile format. If possible, information about the material was added. An indicator of certainty described the confidence whether an object could be identified as litter with values between 1 - 3 (uncertain, possible, certain). Objects that could not be assigned to a category were classified as *others*. After finishing the manual screening, the area of the objects was calculated, and the height information was extracted from the DSM. The DSM was further used to create a 0 m (ASL) contour line from which the closest distance to each litter object was calculated.

Photos taken with a digital camera were used to verify the marked objects together with in-situ GNSS measurements. Due to technical problems, GNSS measurements were only available for the two surveys close to Sisimiut (Greenland) and were performed on a subset of the objects, which were found and registered with the OSPAR method. It was not possible to measure all the objects because of limited accessibility and time. For the two surveys close to Sisimiut, the results of the manual screening were compared to the objects measured via GNSS, and a detection rate was calculated.

To visualize the spatial distribution, a density map of 1.2 x 1.2 m, 3.7 x 3.7 m and 10 x 10 m was calculated with both, the number of objects and the area coverage per unit. A grid size of 1.2 x 1.2 m and 3.7 x 3.7 m was selected to fit the cell size of WorldView3 imagery. The density maps were aligned to the raster cells of WorldView3 satellite images and used to indicate satellite pixels covered by plastic.

4.2.2.2 Statistical Analysis

For comparing the abundances of litter found by OSPAR and drone surveys within the same areas, litter objects were counted for each surveyed location. The success of the drone surveys was assessed by calculating the percentage of objects found on the drone footage, considering the OSPAR results to be the total abundance of litter.

For all statistical analyses regarding the properties of the detected litter objects, data from Greenland and Svalbard drone flights were combined. All litter objects found on the aerial images were considered, regardless of their position in- or outside of the OSPAR reference areas. Objects for which material was not classified as plastic were excluded from further analyses, but an overview table showing the number of litter objects found for each material category can be found in the appendix (E.1). Size distribution of the plastic objects is strongly right skewed, with only few objects of large size. Specifically, only five objects are larger than 1 m^2 (ranging up to 4.7 m^2), so these were also excluded from further analyses. Moreover, we excluded all objects of colours that were found less than 20 times.

Each object was additionally allocated to a more specific type of litter. If no specific type could be recognised, the type "other" was allocated. Based on this classification a new variable "*type_ident*" was introduced. This variable is a binary code for whether an object could be classified as a specific type of litter (1) or not (0). Similarly, the binary variable "*certain*" was introduced based on the values for how certain an object was classified as litter (with 1 = possibly litter, 2 = probably litter, 3 = certainly litter). Only litter objects classified with certainty (3) were coded 1, others (1-2) were coded 0.

We investigated the properties of the detected litter objects to explore why some litter objects might be found on drone footage while others might not. We assessed the distributions of size and colour among plastic litter objects. To investigate their effects on whether its type could be classified, we used a binomial generalised linear model (glm). The saturated model included "*type_ident*" as the response variable and the explanatory variables colour and area (cm^2), as well as their interaction term (type_ident ~ colour * area). We applied the same model structure to explore the effects of size and colour on the certainty of an object to be identified as litter on the drone footage (certain ~ colour * area).

We gradually reduced all models and compared them to the saturated models by ranking their AIC (Akaike information criterion) values. See E.3 and 0 in the appendix for an overview of all full and reduced models, including their AIC ranking. Models with an ΔAIC <2 were considered equally fitting. Homogeneity of variance was assessed by plotting the residuals of the best-fitting models (appendix). We excluded any outliers that were influential by standard of the Cook's distance. In order to avoid mulit-collinearity, we computed the variance inflation factor (VIF) for all models (Zuur et al., 2010). The visreg package was used to plot the final models. All data processing and statistical analysis was performed in R version 3.6.3 (RCore Team, Vienna).

4.2.2.3   (Semi-) Automatic classification

For the (semi-) automatic classification of plastics, machine learning (Random Forest (RF), Support Vector Machine (SVM)) and deep learning approaches (neural network) were applied and compared in their performance. Because of the selected GSD of 1.4 cm and 3.4 cm for RGB and VIR respectively, a pixel-based approach was applied as objects down to the size of single pixels were expected.

To evaluate the benefit of multispectral-drone footage, classifiers using RGB and VIR datasets were furthermore compared. Therefore, the RGB dataset was extended by calculating colour spaces of LAB, HSV, YcBcR. The VIR dataset was extended by the same colour spaces as well as by five indices: NDVI, BNDVI, GNDVI, NDRE and rendvi2.

4.2.2.3.1   Indices and colour spaces.

In order to expend the input features for semi-automatic classification approaches, indices and colour spaces were calculated.

The applied indices were derived from the spectral bands of the MicaSense Altum sensor and limited by the sensor's spectral coverage. They are normalized combinations of two single bands which are widely used in drone applications, mostly in vegetation analyses (Beyer & Grenzdörfer, 2018; Pourazar et al., 2019). The applied indices are: *Normalized Difference Vegetation Index (NDVI), Blue NDVI (BNDVI), Green NDVI (GNDVI), Normalized Difference Red Edge (NDRE) and redEdge NDVI (Rendvi2)* (Table 8)*.*

The use of colour spaces for beach litter detection was already successfully applied in several studies (Gonçalves et al., 2020a, 2020b, 2020c and Pinto et al., 2021). The colour spaces used were HSV, CIE-LAB and YCbCr (Table 9). They were calculated based on the RGB images.

Table 8:   Indices derived from MicaSense Altum sensor

NDVI	BNDVI	GNDVI	NDRE	Rendvi2
$\dfrac{NIR-red}{NIR+red}$	$\dfrac{NIR-blue}{NIR+blue}$	$\dfrac{NIR-green}{NIR+green}$	$\dfrac{NIR-redEdge}{NIR+redEdge}$	$\dfrac{redEdge-red}{redEdge+red}$

Table 9:   Colour spaces derived from RGB channels

HSV	CIE-LAB	YCbCr
H: Hue	L: perceptual lightness	Y: luminance intensity
S: Saturation	A: green-red axis	Cb: blue chrominance component
V: Brightness	B: blue-yellow axis	Cr: red chrominance components

4.2.2.3.2   Extraction of training and validation data

In order to train and validate different classification approaches two reference datasets were created. Therefore, land cover classes were defined: For Svalbard five classes were defined including *plastic*, *rock*, *vegetation*, *water,* and *wood* and for Greenland nine classes including a*lgae*, *plastic*, *rock*, *sand*, *snow*, *vegetation*, *water*, *wood, and shadow*. The class shadow was added in a later step after evaluating first classification results and was only applied in scenario (3). The difference in number of classes between Svalbard and Greenland are the result of different environmental conditions.

For all survey areas polygons were created covering the occurring land cover classes, including the screened plastic objects. Those polygons were used to randomly sample reference points per class with a minimum distance of three-times the GSD. Reference points from the surveys on Svalbard and Greenland were separated into two datasets, as the land cover classes differed.

For Svalbard, a total number of 26091 points were sampled with the smallest class plastic with 660 samples and the biggest class rock with 10881 samples. A sub-sampling into 496 points per class for training and 150 for validation was done to assure a balance between the classes.

For Greenland, a total number of 39015 points were sampled with the smallest class shadow with 2000 samples and the biggest class vegetation with 7482 samples. A sub-sampling into 1500 points per class for training and 500 for validation was done to assure a balance between the classes.

The created points were used in R version 3.6.3 to extract the pixel values of different feature spaces to train the single classifiers. For Greenland three scenarios of feature spaces were tested:

1. 1.4 cm RGB data combined with LAB, HSV and YcBcR resulting in a total of 12 features
2. 3.4 cm VIR data combined with five Indices resulting in 10 features
3. 3.4 cm VIR data combined with LAB, HSC, YcBcR and five indices giving 19 features

4.2.2.3.3   Classifier:

For the (semi-) automatic classification of beach litter three classifiers were applied and compared in their performance. The comparison was applied to images from Greenland only, because many reference data were available here. As the classification accuracy of Random Forest, Support Vector Machine and deep learning were comparable, only Random Forest was applied to images from Svalbard, being the least computationally intensive.

4.2.2.3.3.1   Random Forest

Random Forest (RF) is an ensemble classifier widely used in remote sensing (e.g. Gonçalves et al., 2020a, 2020b, 2020c; Martin et al., 2018). RF was introduced by Breiman in 2001 and combines independent decision trees into an ensemble learning algorithm. The assumption is

that independent trees produce individual errors which are not produced by the majority of other trees. The combination of independent classifiers furthermore reduces the variance and the bias of the algorithm (Breiman, 2001; Waske, B., Braun, M., 2009). Through random subsets of training data, input data is generated which is then used to train the classifier. Each subset usually uses two-thirds of the dataset for training. The last third is used for evaluation of the classification as the so-called out-of-bag (oob). Decision trees decide at each node based on a random split rule how to divide the input data. Split rules are defined by a random selected feature subset (here spectral bands, indices, and colour spaces) which reduces the correlation between trees and therefore also the generalization error (Breiman 2001). The size of the feature subset is usually defined by the square root of input features. To determine the best decision at each node, several techniques such as gain-ratio (Quinlan, 1993), Gini Index (Breiman et al., 2017) and Chi-square (Mingers, 1989) can be applied. The most used approach for Random Forest is the Gini index which calculates the impurity before and after a split to quantify the best decision. The highest decrease in impurity reflects the best decision. The Gini index leads to a simple and compact tree structure. Each decision tree returns the most frequent class to the input. Usually, a simple majority vote creates the final classification result. The number of trees and the size of the feature subset is defined by the user.

Random forest was applied within the open-source software R version 3.6.3. The *randomForest* library was implemented by Liaw and Wiener using the Fortran original of Breiman et al., 2011. To obtain the best classification accuracy, the number of trees and input features were optimized for each scenario.

4.2.2.3.3.2   *Support Vector Machine*

Support Vector Machines (SVM) are another machine learning classifier widely used in remote sensing and already applied in beach litter detection (Escobar Sánchez et al., 2021; Gonçalves et al., 2020c).

SVM uses training data to create a decision surface (hyperplane) which is separating the input data into user-defined classes. If the classes are not linearly separable, SVM transforms them into a high-dimensional space using kernel methods. A commonly used kernel function is the radial basis function (Waske et al, 2007) which was also applied in this study. Applying SVM with the radial basis function, two parameters can be adjusted. The cost parameter C which defines the amount of training data that can be misclassified and the parameter Υ which changes the shape of the decision boundary (Gonçalves et al., 2020c). To determine optimal values for C and Υ the grid search method was applied for each scenario. The grid search method applies a defined range of different combinations of C and Υ parameters to determine the combination with the highest classification accuracy (Gonçalves et al., 2020c; Mather & Tso, 2016).

4.2.2.3.3.3   *Deep Learning*

For comparison with the machine learning algorithms a simple three layer fully connected neural network was tested for litter classification using the same reference data set. The neural network is composed of an input-, hidden and output layer. Three scenarios of class weighting for cross entropy loss were tested to improve the classification of plastics. The first scenario did not use any weight, the second scenario used a medium weighting for plastics (a factor of 4 for plastic and 1 for other classes) and the third scenario a strong weighting with a factor of 4 for plastic and 0.1 for other classes. The neural work was trained and inferred in Python v.3.9.7 (in the PyTorch framework).

4.2.2.3.3.4   *FasterRCNN*

Object Detection using Bounding boxes was tested using FasterRCNN. Due to limited objects available for training, the model over fitted to the input data (results in low test accuracy) which

may indicate the possibility of an acceptable model in future with significantly more data. Besides this first evaluation, the FasterRCNN was not further tested.

4.2.2.4   Accuracy Assessment

Validation was applied to determine the quality of the classification results. Validation was performed with the means of a confusion matrix. The confusion or error matrix provides a cross tabulation of the classes predicted by the classification against the validation data. Validation data represented about 25% of the reference data obtained by sampling. The confusion matrix obtains information about overall and per Class Accuracy and User's and Producer's Accuracy. Furthermore, areal estimations can be obtained. The use of a confusion matrix implies the assumption that each pixel can be associated to a single class. The accuracy obtained by the confusion matrix can be understood as a thematic accuracy. The diagonal of the matrix represents the correctly classified data whereas the off-diagonal represents false classification. User's Accuracy (UA) is computed using the number of correctly classified validation data to the total amount of validation data assigned to a particular class. Pixels mistakenly assigned to a particular class represent the error of commission. Producer's Accuracy (PA) reflects the number of validation data of a certain class correctly assigned to the corresponding class. Validation data not assigned to the corresponding class account to the omission error of this particular class. Therefore, an error of omission of one class, represents an error of commission in the other (Strahler, 2006).

4.2.3   Satellite imagery

For a large-scale application of remote sensing techniques for beach litter detection in the Arctic, satellite imagery was tested. Therefore, WV3 satellite imagery was acquired close to the fieldwork dates, recording VIR and SWIR imagery with a resolution of 1.2 m and 3.7 m, respectively. The WV3 imagery was ordered by Maxar Technologies as a level 3 ortho product with atmospheric compensation (for Greenland only) and radiometric correction. The spectral signals of the WV3 imagery were supplied with 16 bits of dynamic range and transformed to surface spectral reflectance. As for Svalbard the acquired WV3 imagery did not include atmospheric compensation, an atmospheric correction approach was applied as described in de Grandpré et al. (2022).

As the litter objects expected were of sub-pixel size, a manual screening was not possible. In addition, a comparison with the OSPAR monitoring was not applied, as the detection and classification of macro litter would require a much higher resolution (see drone imagery). Instead, we focused on whether satellite imagery can help to locate areas of high beach litter accumulations. Studies of floating marine litter showed that even litter in sub-pixel size could be detected by satellites (Biermann et al., 2020).

To evaluate the potential of WV3 imagery for beach litter detection, the density maps of the manual drone screening were used to create a reference data set. Having an information about the litter coverage per pixel, reference data were summarized to pixels covering litter in two 5% classes between 0 - 10% and in nine 10% classes from 10 – 100%. Additionally, pixels covering the main background groups were summarized into the following classes: Vegetation, rocks, wood, and sand. The reference data were then used to extract and compare the spectral signal of the satellite imagery.

Besides, an approach of spectral unmixing was applied to test whether pixels partially covered by plastic can be identified by using a linear unmixing model. Therefore, spectral endmember (pure spectra of each class) of all occurring classes were calculated by averaging the spectral signal of pixels which were covered by 100% of the corresponding class. For plastic, the spectral

signal of the pixel with highest plastic coverage (96.7%) was used. Using the endmember, the linear unmixing model calculates the fraction of the different classes within a pixel.

4.3 Results and discussion

4.3.1 Results of conventional surveys

On Svalbard, mean total abundance was as low as 28.5 pieces of litter per 50 m beach. Beach litter consisted of 57.9% plastic, 40.4% wood and 1.7% rubber. The high percentage of plastics agrees well with the results of previous beach litter studies on Svalbard (Bergmann et al., 2017). Among the top-14 litter categories, six are attributable to fishing and shipping. (Table 10) sources is highlighted, which is in line with the results of previous studies (Bergmann et al., 2017; Weslawski and Kotwicki, 2018).

Table 10 **The most abundant litter categories on the two sections of beach near Kapp Mitra, surveyed according to the OSPAR protocol (OSPAR, 2010). Numbers in rectangular brackets give the OSPAR beach litter IDs.**

Rank	OSPAR beach litter category	Mean abundance [-]
1	Wood: Other <50 cm [74]	8
2	Plastic: Nets and net rests <50 cm [115]	4.5
3	Plastic: Plastic fragments 2.5 cm ><50 cm [461 a]	3.5
4	Wood: Other >50 cm [75]	3
5	Plastic: Styrofoam [45]	2
6	Plastic: Caps and lids[15]	1.5
6	Plastic: Strapping bands [39]	1.5
8	Plastic: Tangled nets [331 a]	1
8	Plastic: Fish boxes [341 a]	1
10	Plastic: Ropes [31]	0.5
10	Plastic: Strings and cords [321 a]	0.5
10	Plastic: Other [48]	0.5
10	Rubber: Balloons [49]	0.5
10	Wood: Cork [68]	0.5

Near Nuuk and Sisimiut, mean total abundances of beach litter amounted to 116.4±111.9 pieces and 87,3±33,5 pieces on a 50 m section of beach, respectively. This was in the same order of magnitude as Strietmann et al. (2021) found on the western coast of Greenland. Material composition was dominated by plastics and processed wood (Figure 27 and Figure 28), which again agrees with the results of Strietmann et al. (2021). Beach litter types, which can definitely be assigned to sea-based sources, were the most common among the top-ten litter types (Table 11 and Table 12). This finding is partly contradictory to the results of Strietmann et al. (2021),

who found that domestic waste was predominant. These differences are likely because the latter authors partly surveyed beaches in closer vicinity to settlements than in the study at hand. However, the precise distances to settlements are not given in the study of Strietmann et al., 2021).

Power analyses gave statistical powers <50% and >80% for beaches near Nuuk and Sisimiut, respectively, for beach litter reductions of 50% (Figure 29). At lower reduction rates, statistical power was even lower and thus not sufficient. Therefore, surveyed beach sections should have a minimum length of 200 m, because near Sisimiut, replicate data of four beach sections of 50 m length were used for power analyses. In turn from a logistic point of view, this supports the application of UAVs for beach litter surveys on Greenland.

Figure 27   Mean material composition of beach litter on the beaches near Nuuk.
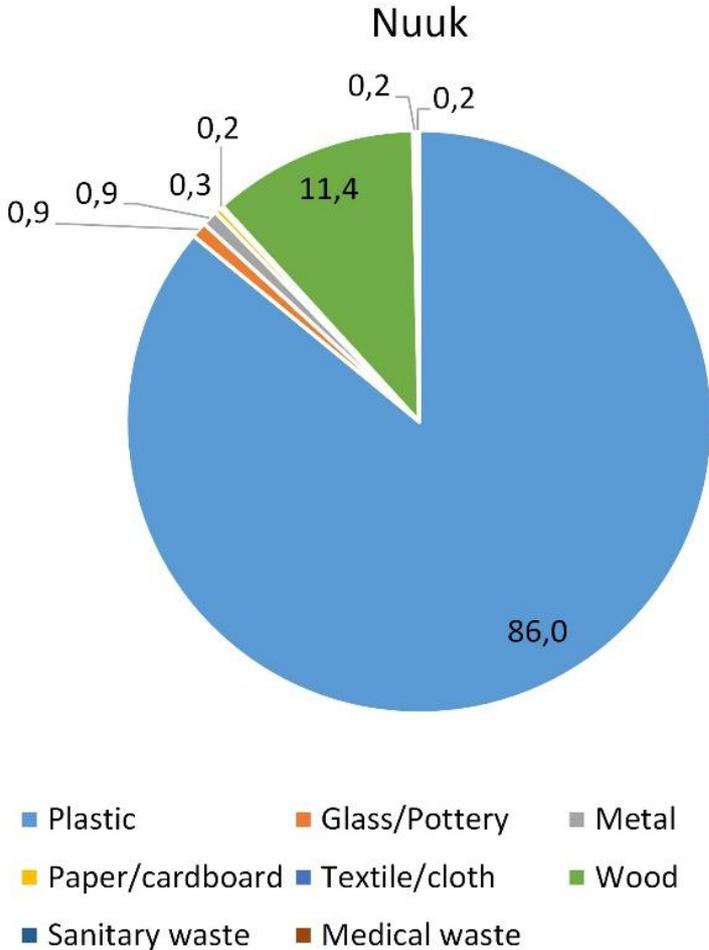
Source: Compiled by AE

Figure 28   Mean material composition of beach litter on the beaches near Sisimiut.

Sisimiut
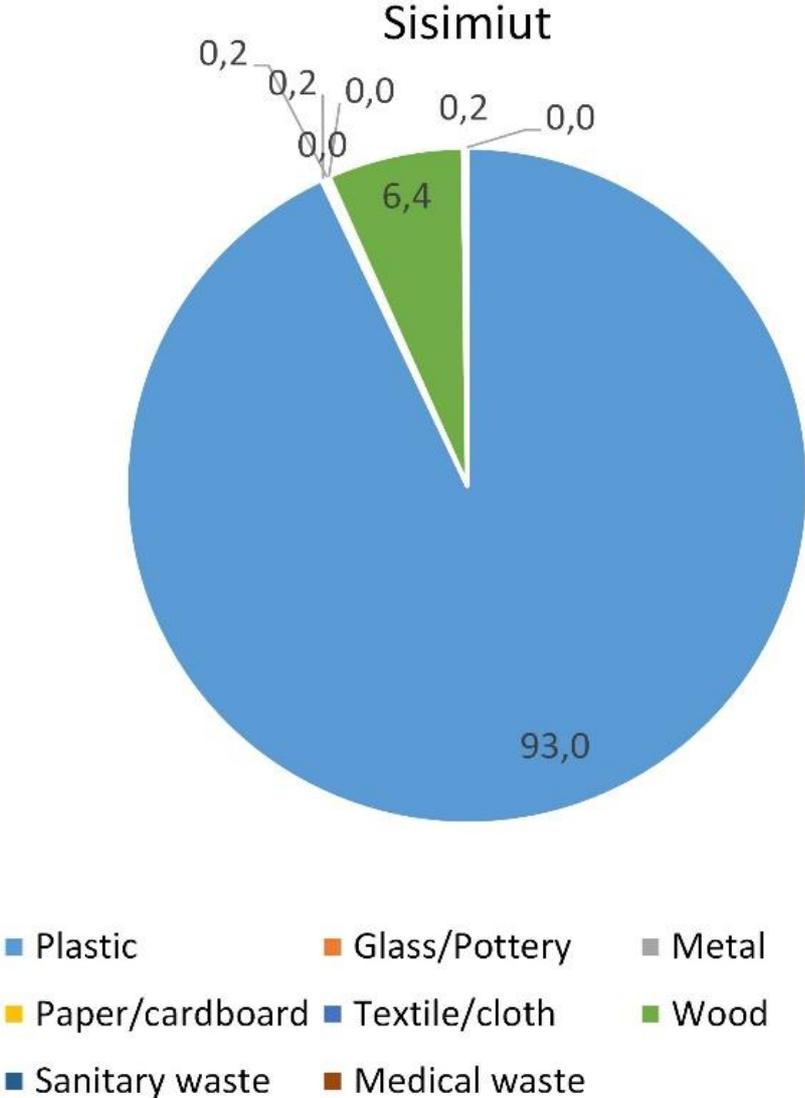
Source: Compiled by AE

Table 11   The most abundant litter types on conventionally surveyed sections of beach near Nuuk on Greenland. Numbers in rectangular brackets give OSPAR beach litter type IDs.

OSPAR beach litter type	Arithmetic mean [-]	Standard deviation [-]	Percentage of total abundance [%]	Coefficient of variation [%]
Plastic fragments 2.5 cm – 50 cm [461]	27.4	23.5	23.5	85.8
Styrofoam [other items]	11.6	20.5	10.0	176.8
Shotgun cartridges [43]	11.4	10.2	9.8	89.6
Rope >1 cm [31]	11.4	11.5	9.8	100.7
Other wood <50 cm [74]	7.8	8.5	6.7	108.6
Other wood >50 cm [75]	5.2	6.2	4.5	119.6
Strapping bands [39]	4.0	3.8	3.4	95.2
Other plastic items [48]	3.4	7.6	2.9	223.6
Drinks [4]	2.8	4.8	2.4	170.2
String and cord <1 cm; not from dolly ropes or unidentified [321]	2.6	4.3	2.2	166.8

Table 12   The most abundant litter types on conventionally surveyed sections of beach near Sisimiut on Greenland. Numbers in rectangular brackets give OSPAR beach litter type IDs.

OSPAR beach litter type	Arithmetic mean [-]	Standard deviation [-]	Percentage of total abundance [%]	Coefficient of variation [%]
Plastic fragments 2.5 cm – 50 cm [461]	23.3	22.1	26.7	94.5
Shotgun cartridges [43]	10.4	5.1	11.9	48.6
String and filaments [322]	5.9	6.9	6.8	117.1
String and cord <1 cm; not	4.8	3.6	5.4	76.4

OSPAR beach litter type	Arithmetic mean [-]	Standard deviation [-]	Percentage of total abundance [%]	Coefficient of variation [%]
from dolly ropes or unidentified [321]				
Strapping bands [39]	4.8	2.8	5.4	59.5
Rope >1 cm [31]	4.0	2.0	4.6	50.0
Styrofoam [other items]	3.9	4.5	4.5	114.3
Other wood <50 cm [74]	3.4	3.6	3.9	104.9
Tangled dolly rope [332]	3.2	1.9	3.6	61.3
Nets and pieces of net <50 cm [115]	2.8	2.5	3.2	87.6

Figure 29   **Statistical Power for various reductions of total abundance of beach litter. Results were calculated by means of one-sided one-sample Wilcoxon rank sum tests using a significance level of 0.05.**

Source: Compiled by AE

83
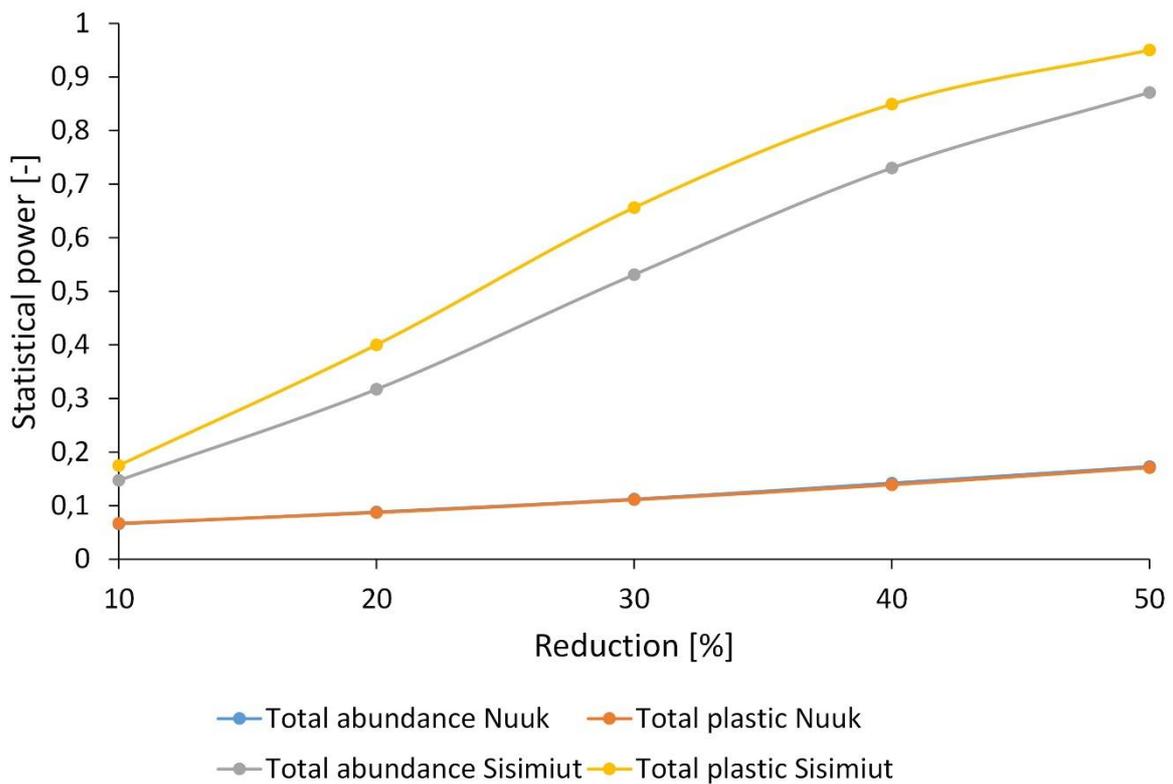
4.3.2   Results and discussion of the drone survey

4.3.2.1   Manual Screening

The manual screening of the drone surveys performed in QGIS 3.16.6 resulted in litter abundances between 0 and 63 pieces for the monitored beach areas (Table 13). The manual screening of the drone footage took between 05:00 minutes and 30:00 minutes.

Images from Kapp Mitra 2 were not compared to results applying the OSPAR method, because the monitoring and drone surveys had to be interrupted due to a polar bear sighting and exact boundaries of the study side necessary for comparison were not given. The area was screened anyhow as input for automatic detection and as reference for the satellite images.

Density maps were calculated showing the spatial distribution of litter covering the monitored beaches and the surroundings. The density maps with grid size 1.2 x 1.2 m are shown in Figure 30, Figure 31, Figure 32 and Figure 33. A further analysis of the manual screening was performed in 4.3.2.2. Additionally, Figure 30 shows the number of litter items found at Kapp Mitra in a 10 x 10 m grid.

Table 13   Number of litter objects found by manual screening and in situ OSPAR monitoring

Name of beach	Time for screening	Number of litter items	Number of plastic objects	Number of litter items OSPAR	Number of plastic litter items OSPAR
Kapp Mitra 1	15 min	15	11	57	33
Kapp Mitra 2					
Nuuk Day 1	20 min	43	35	273	232
Nuuk Day 2-1	10 min	32	24	162	107
Nuuk Day 2-2	5 min	10	9	130	113
Nuuk Day 3	5 min	0	0	11	10
Sisimiut Day 4	30 min	63	55	325	316
Sisimiut Day 5	20 min	10	38	251	251

Figure 30   Density map of area coverage and number of plastic litter objects at Kapp Mitra
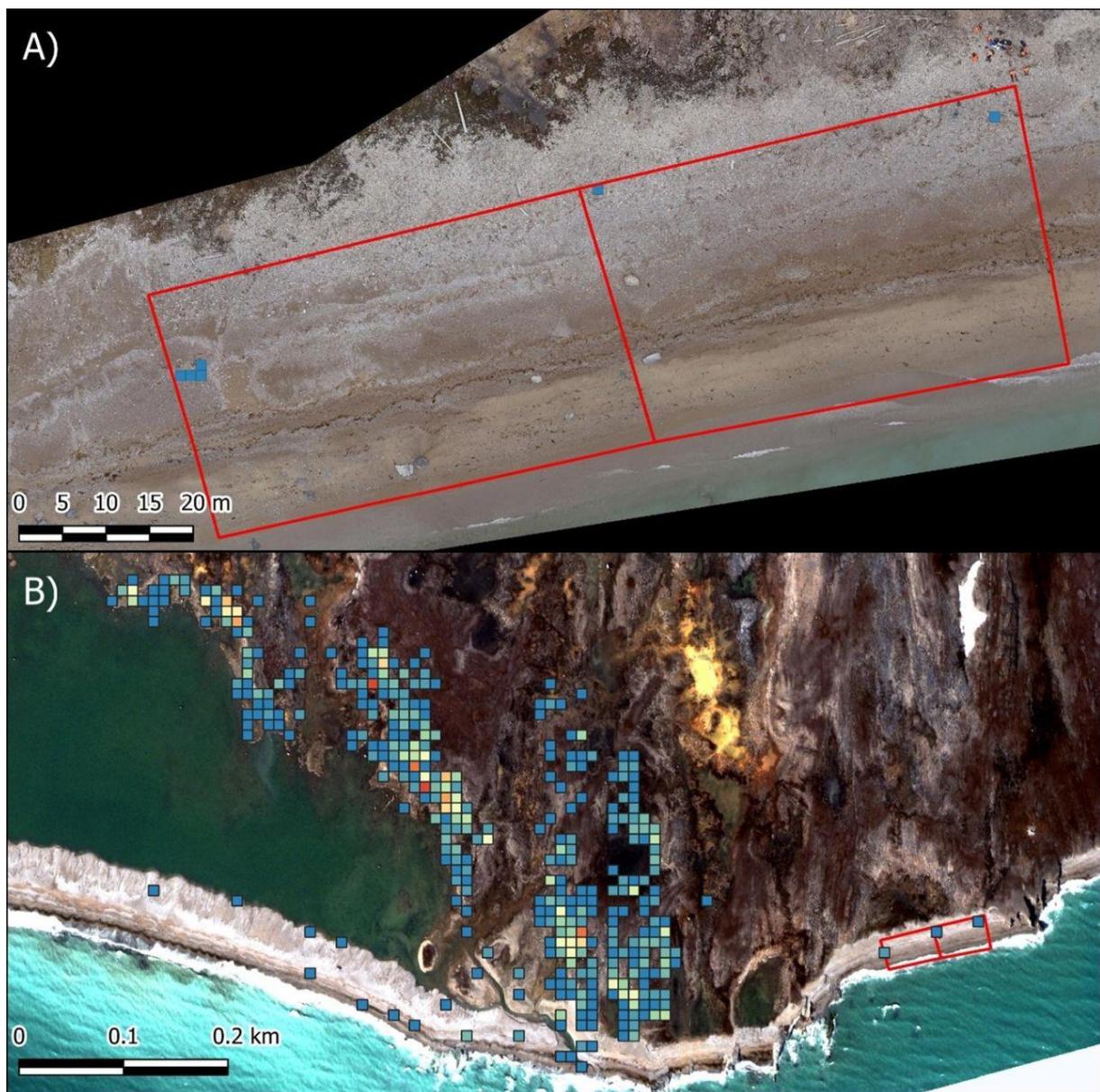
A) Density map of study area Kapp Mitra (1 and 2) with the area covered by plastic items in a 1.2 x 1.2 m grid. B) Density map of number of plastic items found at Kapp Mitra area in a 10 x 10 m grid. Marked in red is the area monitored by the OSPAR method. Source: Compiled by BCSH

Figure 31   Density map of area coverage of plastic litter on the study sites around Nuuk
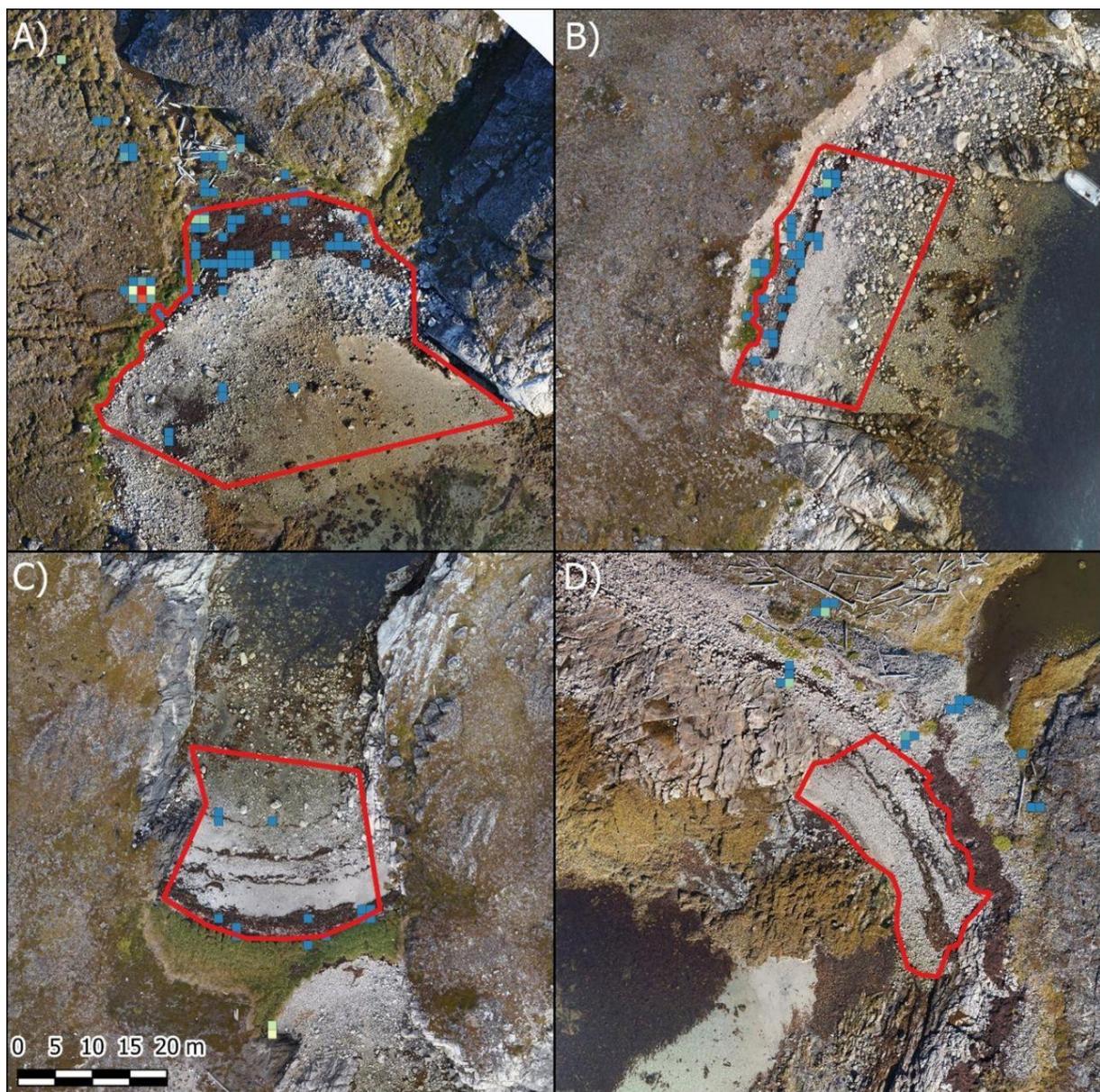
A), B), C) and D) show the beach segments of the study areas Nuuk Day 1, Nuuk Day 2-1, Nuuk Day 2-2, and Nuuk Day 3 respectively. Marked in red is the area monitored by the OSPAR method. The density maps show the area covered by plastic items in a 1.2 x 1.2 m grid. Source: Compiled by BCSH

Figure 32   **Density map of area coverage of plastic litter on the study side Sisimiut Day 4**
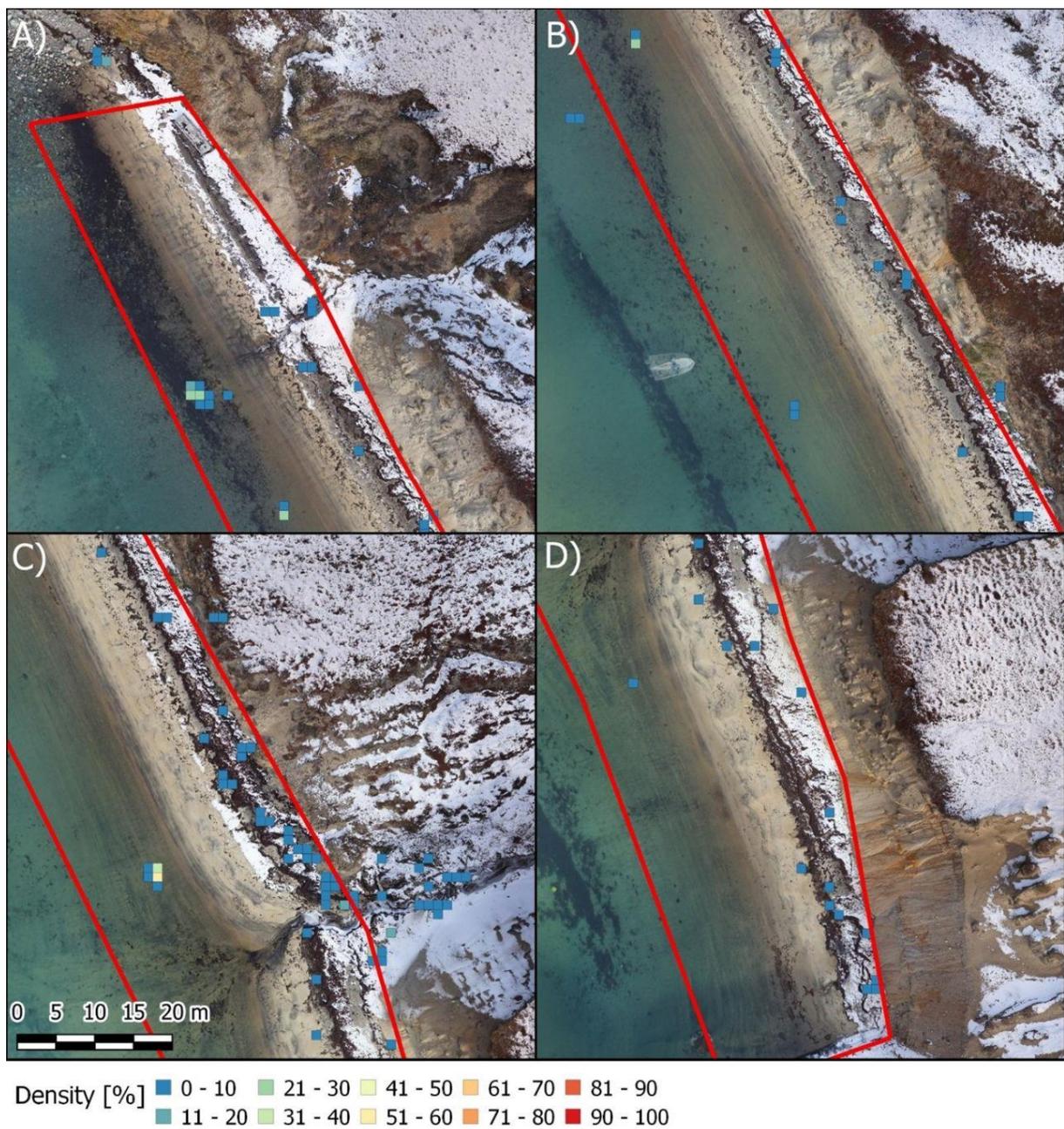
A), B), C) and D) are single beach segments of the same study area. Marked in red is the area monitored by the OSPAR method. The density maps show the area covered by plastic items in a 1.2 x 1.2 m grid. Source: Compiled by BCSH

Figure 33   Density map of area coverage of plastic litter on the study side Sisimiut Day 5
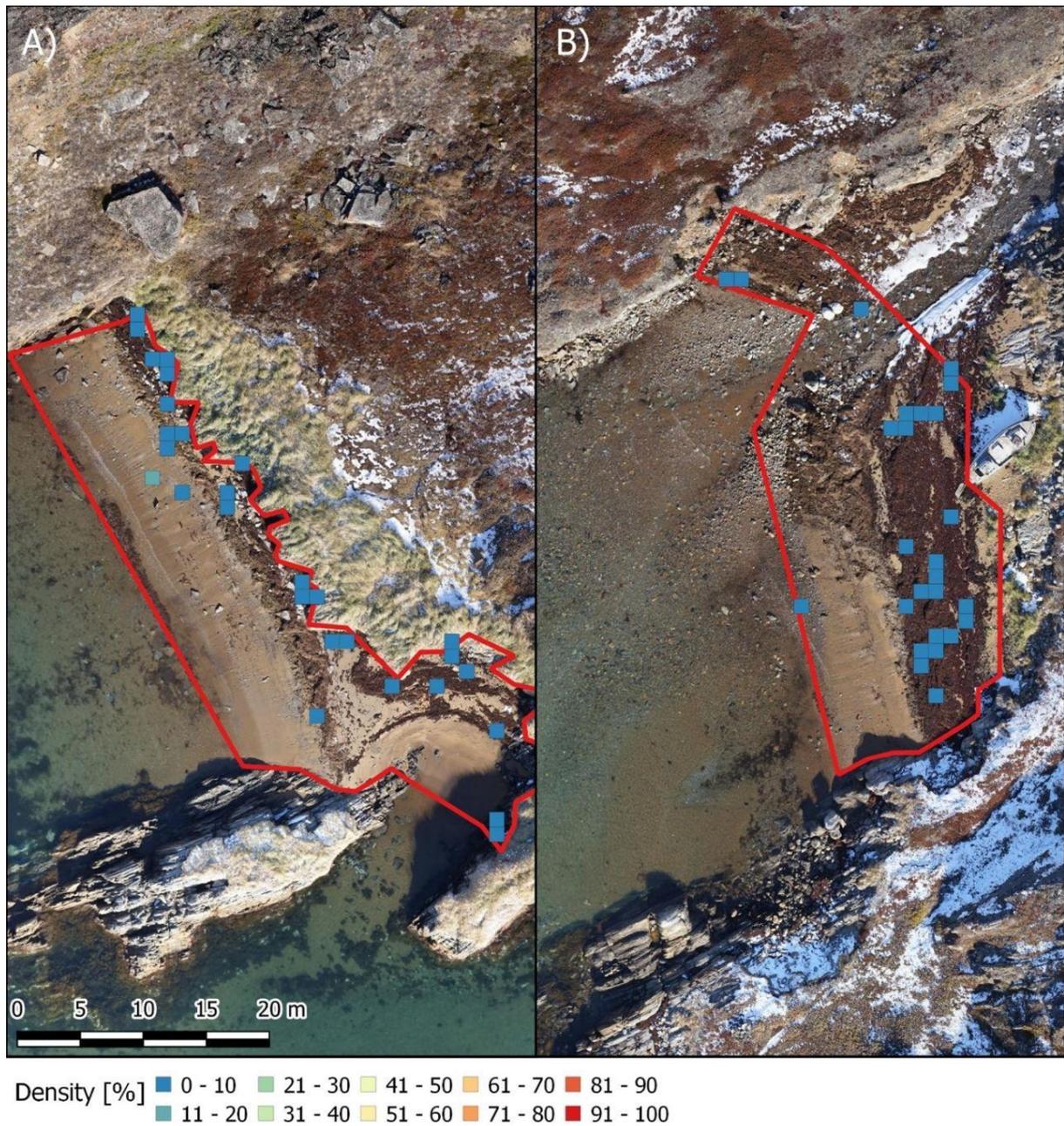
A) and B) are single beach segments of the same study area. Marked in red is the area monitored by the OSPAR method. The density maps show the area covered by plastic items in a 1.2 x 1.2 m grid. Source: Compiled by BCSH

For Sisimiut, a comparison between the objects found during the manual screening and the objects which were measured in the field via GNSS receiver was performed. A total of 112 and 55 objects were measured at Sisimiut Day 4 and Sisimiut Day 5, respectively. During the manual screening 44.64% (Sisimiut Day 4) and 54.55% (Sisimiut Day 5) of those objects could be detected (Table 14). Figure 34 and Figure 35 show the spatial distribution of the objects that were found with both methods and the litter objects missed during the manual screening. Photos of all objects measured via GNSS receiver can be found in the appendix D.1 and D.2. with an indication whether they were also detected during the manual screening of the drone imagery.

Figure 34 **Spatial distribution of the objects marked via GNSS receiver, and the objects found during the manual screening on the study site Sisimiut Day 4**

Bag: 104; buoy: 30; bottle: 83,105,112,120; brush: 76; container: 20, 44, 80, 84, 88 ,103; cup: 98, foam: 62, 102; foil: 41, 71, 75, 110; food container: 4, 58, 85, 121; lid: 31; net: 25; plastic fragment: 6, 21, 26, 28, 50, 82, 93; plastic other: 42, 101; rope: 12, 16, 36, 43, 49; shotgun: 5, 11, 14, 22, 24, 57, 68, 70, 81 ,89, 107; strapping band: 32, 118; string & filament: 18, 37, 69, 90, 119; 89; styrofoam: 23, wood: 7, 8, 79. Source: Compiled by BCSH

Figure 35 **Spatial distribution of the objects marked via GNSS receiver, and the objects found during the manual screening on the study site Sisimiut Day 5**
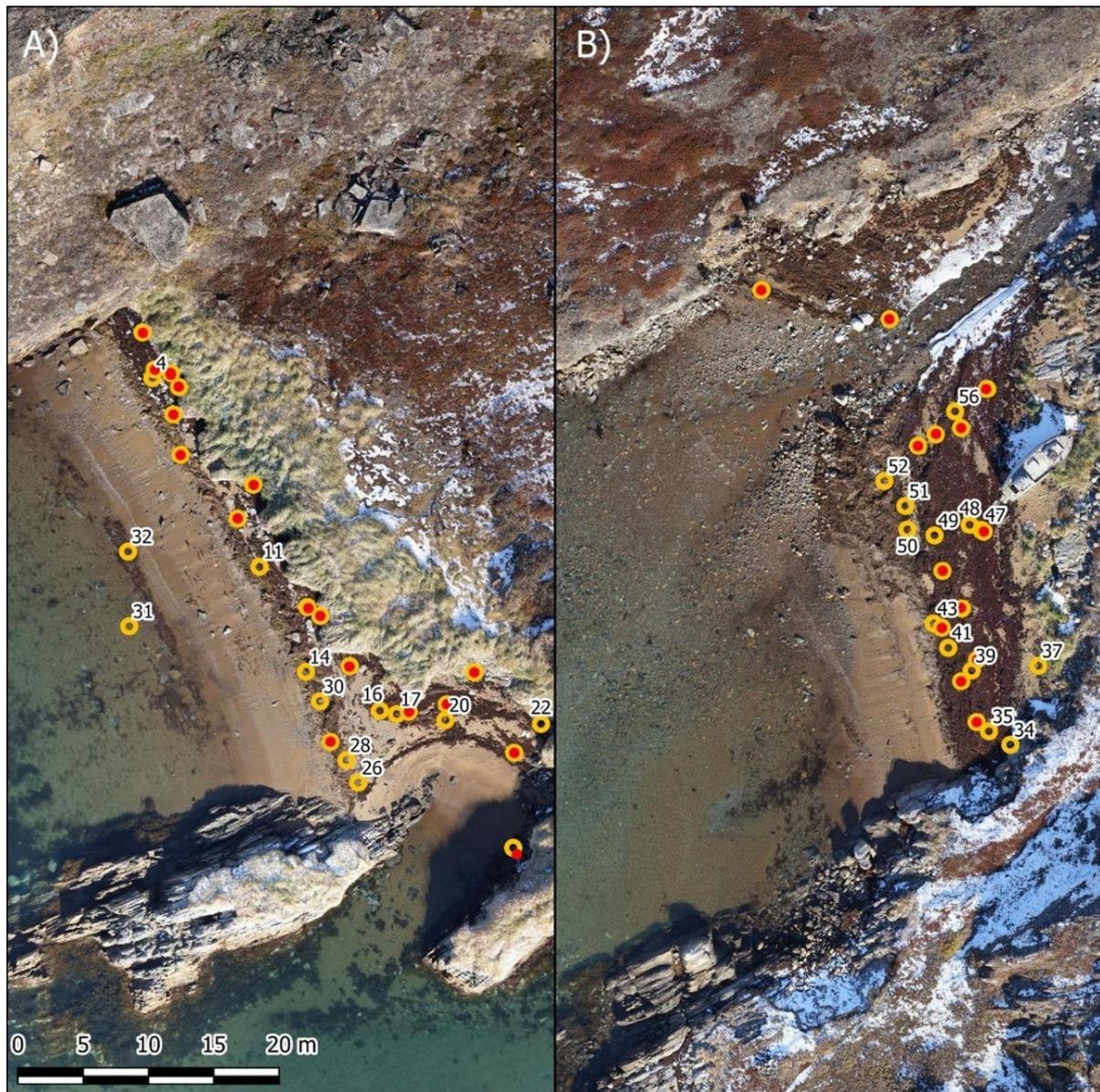
Foil: 11, 14, 22, 30; food container: 39; lid: 31; rope: 4, 32, 37; shotgun: 16, 28, 34, 35 ,41, 43, 48, 49, 52, 56; strapping band: 20; string & filament: 26, 50; tube: 17, 47. Source: Compiled by BCSH

Table 14 **Comparison between GNSS measurement of objects and the manual screening**

Study Area	Objects marked with exact geolocation via GNSS receiver	Objects marked during the manual screening of the drone footage	Percentage of the objects found during the manual screening compared to those marked via GNSS receiver
Sisimiut Day 4	112	50	44.64%
Sisimiut Day 5	55	30	55.55%

4.3.2.2 Statistical Analysis

We compared the numbers of litter objects found within the same defined areas during the OSPAR and drone surveys and calculated the percentage of objects found on drone footage per area. Including all found objects, also of wooden material, a maximum of 24%, mean of 14.8%, median of 15.8% and minimum of 0% of objects was found on aerial images (Figure 36 A, n = 8 different locations). When wooden objects were excluded a maximum of 17.5% were found (Figure 36 B).

Figure 36 **Percentage of litter objects found within distinct areas during drone surveys (n = 8 different, independent locations). A: all objects, including wood. B: only litter objects that were not classified as wood.**
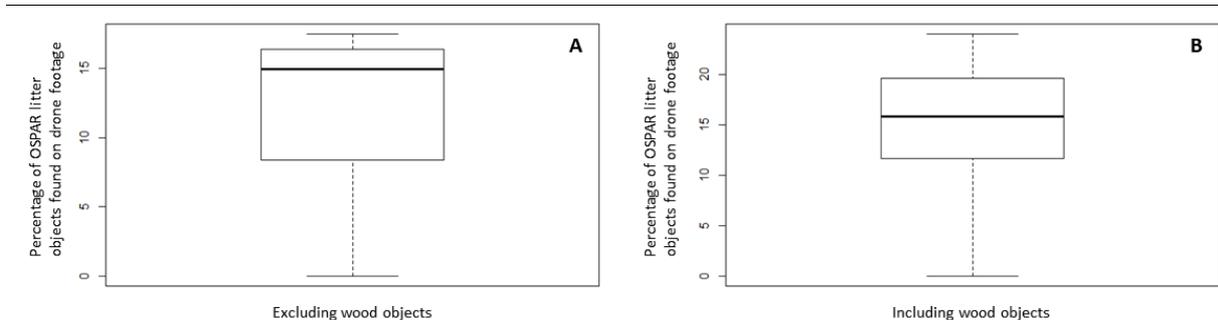
Source: Compiled by BCSH

The size-distribution of litter objects was strongly right-skewed with few objects of large size (>1 m^2), which is reflected in the deviation of median and mean values (Table 15). Excluding objects that were not classified as plastic, as well as objects of a size >1 m^2 reduced this margin.

Table 15 **Size distribution of the litter objects found on the drone footage. The table lists the minimum, median, mean, and maximum size in cm² of all objects, only plastic objects and plastic objects smaller than 1 m².**

Data	Minimum size [cm²]	Median size [cm²]	Mean size [cm²]	Maximum size [cm²]
All litter	19.9	392.3	4563.3	1,733,736.8
Only plastic litter	19.9	262.7	924.7	47,541.6
Only plastic <1 m²	19.9	223.6	616.2	9,192

Plotting the size distribution for each colour of the plastic litter that was found at least 20 times showed a very similar right-skewedness for all four colours (Figure 37). Only white plastic

objects showed a different size distribution with a higher median than the other colours (blue = 161.10 cm^2; red = 127.29 cm^2; white = 372.74 cm^2; yellow = 161.57 cm^2.

Figure 37 **Boxplots showing the distribution of object size (as area [cm²]) for each colour category of plastic litter objects found on drone footage.**
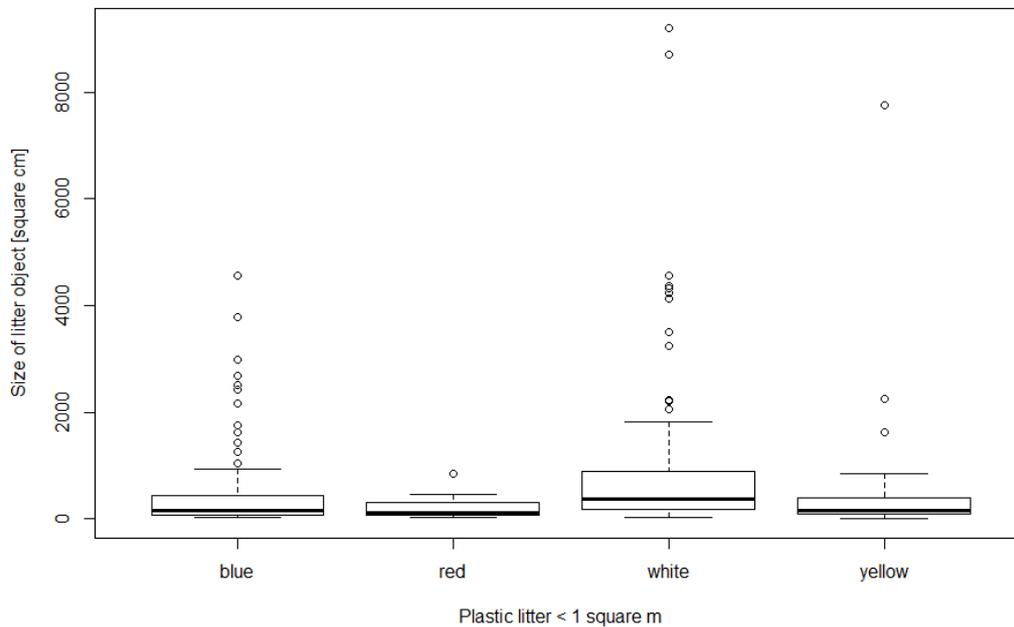
Source: Compiled by BCSH

When testing for effects of the objects' size and colour on the likelihood for the type to be identified, the initial saturated model showed very high VIFs (Area = 21.72, Colour = 77.59, 602.79 = Area*Colour). According to Zuur et al. (2010) we thus removed the parameter with the highest VIF value. Reduced models with area and colour as well as area alone as explanatory variables had very similar AICs with ΔAIC <2, so they were considered equally fitting (see E.30). The model including area (highly significant) and colour (non-significant) was chosen for presentation, plotting the effect of the size of the object [cm^2] on whether the type of litter could be identified for each colour found (Figure 38). Large object size favours type-identification significantly, but this effect is similar for the different colours.

Table 16 **Model results estimating the likelihood for a plastic litter object's type to be identified from drone footage.**

Parameter	Chi-squared	df	*p*	VIF
Area [cm]	28.10	1	1.15e-07***	1.34
Colour	4.31	3	0.23	1.34

Results of an analysis of variance (ANOVA) of a binomial generalised linear regression with a logit link function are shown. The variance inflation factor (VIF) was computed to avoid multi-collinearity. Asterisks indicate significance (* p<0.5; ** p<0.01; *** p<0.001).

Figure 38   Effects of size (area [cm²]) and colour on the likelihood of a plastic litter object's type to be identified from drone footage.
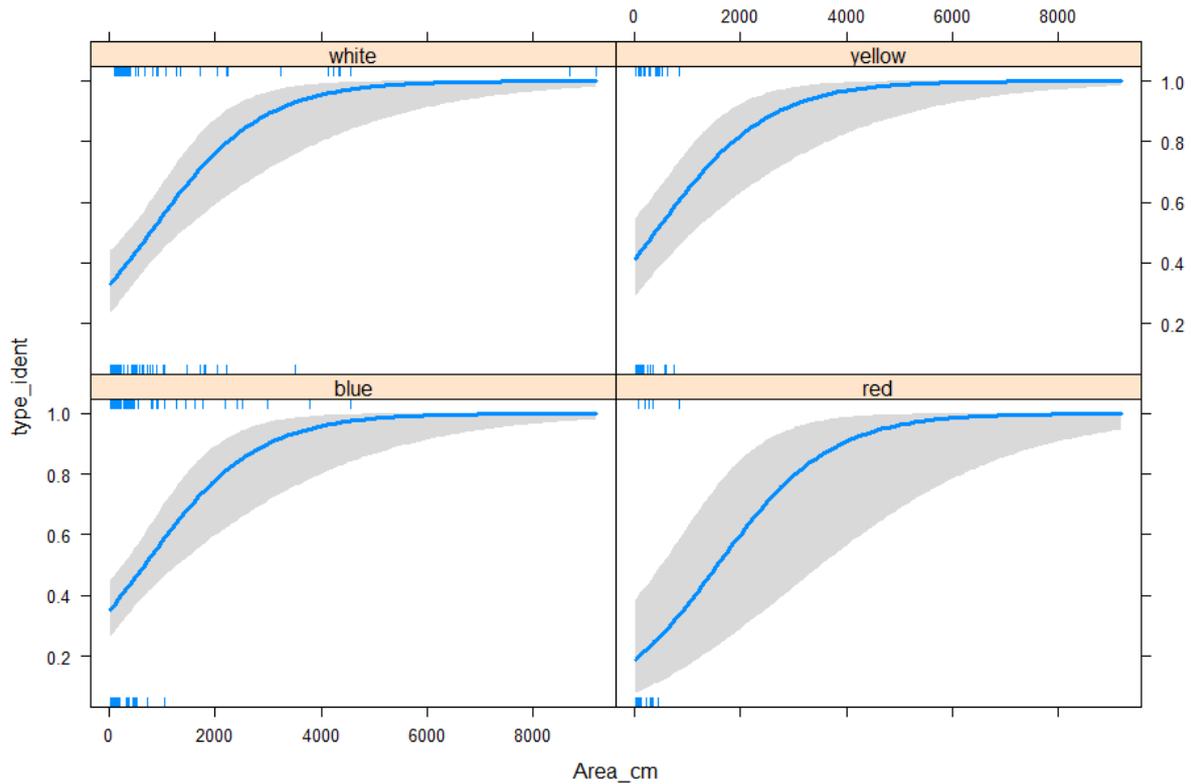
The blue line represents the regression line of the binomial model. Confidence bands are shown as shaded grey area.
Source: Compiled by BCSH

Regarding the effects of the objects' size and colour on the likelihood for a plastic object to be identified as litter with certainty, the initial full model showed very high VIFs (VIF Area = 8.81; VIF Colour = 4.66, VIF area*colour = 24.55). Thus, the model was reduced by the term with the highest VIF value. Reduced models with area and colour as well as area alone as explanatory variables had very similar AICs with ΔAIC <2, so they were considered equally fitting (see E.3). The model including both parameters, area, and colour, but no interaction term resulted to be the best fitting one by AIC (see 0). Therefore, it is displayed here in Figure 39. Similar to type-identification, large object sizes increase the likelihood of certain classification significantly. However, also the colour of the object has a significant effect with white objects generally less likely classified as litter with certainty, while blue and red objects showing the highest likelihoods.

Table 17   Model results estimating the likelihood for a plastic litter object to be identified as such with certainty.

Parameter	Chi-squared	df	p	VIF
Area [cm]	48.98	1	2.58e-12***	1.34
Colour	24.69	3	1.793e-05***	1.34

ANOVA results of a binomial generalised linear regression with a logit link function are shown. The variance inflation factor (VIF) was computed to avoid multi-collinearity. Asterisks indicate significance (* p<0.5; ** p<0.01; *** p<0.001).

Figure 39   Effects of size (area [cm²]) and colour on the likelihood of a plastic litter object to be identified as such with certainty.
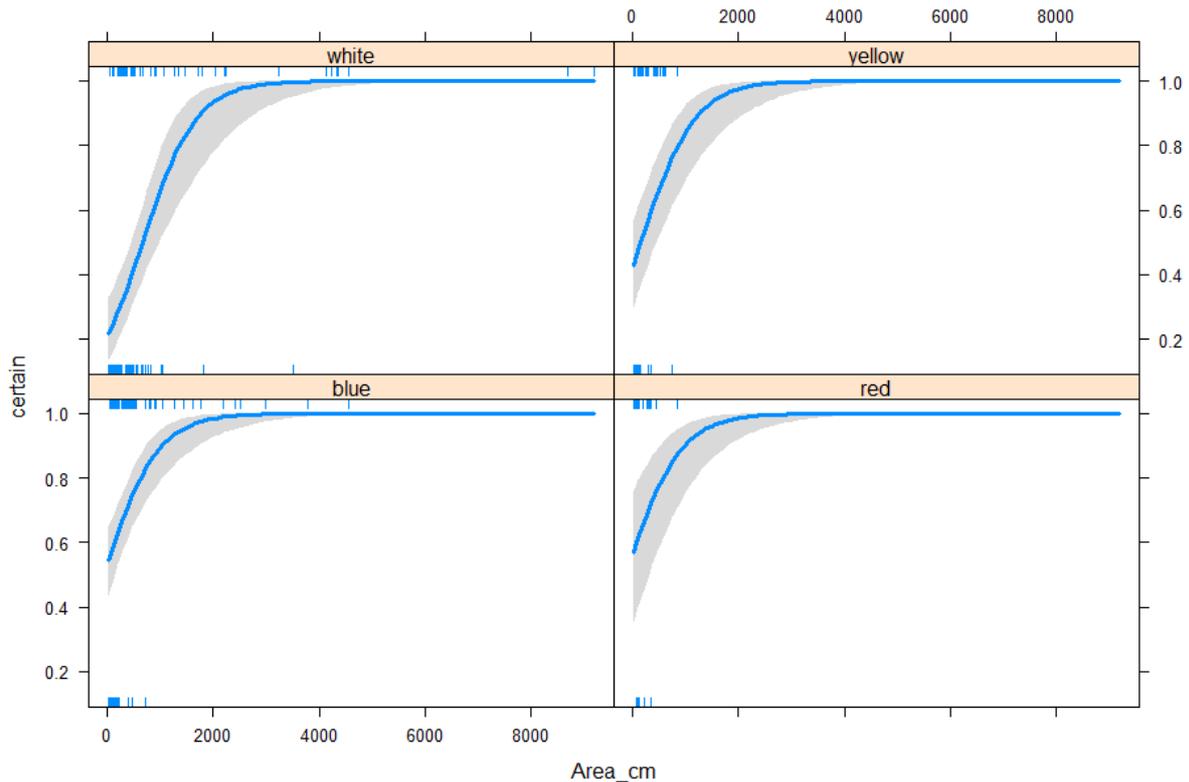
The blue line represents the regression line of the binomial model. Confidence bands are shown as shaded grey area.
Source: Compiled by BCSH

4.3.2.3   (Semi-) Automatic classification

(Semi-) Automatic classification was applied for the drone surveys on Greenland and Svalbard and the results were compared in terms of their respective Overall-, User- and Producer-Accuracy. For Greenland the best OA was 90.6% and 85.8% on Svalbard, both applying RF v3. The highest User- and Producer- Accuracy were reached with 95.2% and 82.9% for Greenland and 76.2% and 65.3% for Svalbard respectively (Table 18and Table 19). To obtain the best classification results, different classifier and feature spaces were tested for Greenland.

The results vary between 62.5% and 90.6% for Overall- 52.6% and 95.2% for Producer's- and 25.4% and 82.9% for User's Accuracy (Table 18). The PA of 95.2% was reached applying the heigh weighted NN indicating a possible overfitting of the plastic class which is confirmed by its corresponding UA of only 40.6%. In general, the different classifier showed comparable results for the single scenarios. Comparing the different scenarios, v3 achieved the highest OAs for all classifiers. The classification results were furthermore compared to the manual screening of the drone footage in terms of the area covered by litter resulting in an area classified as plastic up to 10 to 2000- times higher for the RGB input and up to 40 to 425- times for VIR input (Table 18). The highest overestimation of plastics occurred for NN with height class weighting, confirming a strong overfitting for this scenario.

Table 18   Classification Accuracy Greenland

		Accuracy of plastic detection			Proportion of area covered by Plastic [%]					
		OA [%]	PA Plastic [%]	UA Plastic [%]	Nuuk Day 1	Nuuk Day 2-1	Nuuk Day 2-2	Nuuk Day 3	Sisimiut Day 4	Sisimiut Day 5
Manual Screening of drone images					0.1	0.1	0.1	0.06	0,04	0.04
RF RGB	v1	75.6	54.0	63.0	6.3	2.7	7.1	2.5	18.3	6.4
SVM RGB	v1	77.3	52.6	70.3	5.1	1.0	5.5	1.6	20.3	5.6
NN RGB	v1a	78.0	52.8	67.2	3.0	2.9	1.8	2.5	46.1	4.7
	v1b	77.3	52.8	67.7	8.3	10.7	10.1	10.0	62.3	10.7
	v1c	62.5	92.6	25.4	33.3	46.0	50.9	45.6	86.6	32.8
RF VIR	v2	89.6	67.8	**82.9**	8.6	7.0	5.6	7.5	4.2	10.1
	v3	**90.6**	72.4	81.5	4.3	5.4	5.4	6.7	4.3	7.5
SVM VIR	v2	89.5	64.0	81.6	9.7	8.2	5.5	9.5	5.6	11.2
	v3	90.4	72.0	79.8	5.5	6.8	6.0	5.8	7.7	8.3
NN VIR	v2a	90.1	72.0	79.5	9.0	7.9	8.1	11.4	4.2	10.9
	v2b	88.1	83.6	62.5	17.6	22.6	17.4	19.4	9.6	19.2
	v2c	79.8	**95.2**	40.6	27.5	42.5	33.5	32.7	15.4	27.0
	v3a	89.9	71.6	76.3	7.6	8.0	6.8	6.6	4.4	9.0
	v3b	84.5	89.6	48.6	10.5	19.5	16.1	18.4	8.6	15.7
	v3c	87.9	84.0	62.4	17.2	29.4	20.1	17.6	9.6	17.5

OA: Overall Accuracy, UA: User's Accuracy and PA: Producer's Accuracy.

v1: RGB with colour space, v2: VIR with spectral indices, v3: VIR with spectral indices and colour spaces

NN: a: no weight on single classes, b: medium weight on plastic class, c: height weight on plastic class

In red: highest values for OA, UA, and PA

Table 19   Classification Accuracy Svalbard

		Accuracy of plastic detection		
		OA [%]	PA Plastic [%]	UA Plastic [%]
RF RGB	v1	78.6	64.0	**76.2**
RF VIR	v2	83.8	63.3	69.3
	v3	**85.8**	**65.3**	74.8

OA: Overall Accuracy, UA: User's Accuracy and PA: Producer's Accuracy.

v1: RGB with colour space, v2: VIR with spectral indices, v3: VIR with spectral indices and colour spaces

In red: highest values for OA, UA, and PA

For a better understanding of potential misclassifications, confusion matrices were created. The results showed that the highest misclassification of plastics occurred with rocks and wood (F.1-F.18). Misclassifications occurred in either way, plastic items being falsely assigned to the class *rocks* or *wood* and *rocks* or *wood* falsely assigned to the class *plastic*. Depending on the beach environment, the misassignments of those classes had a greater impact on the overall performance.

To understand potential sources of misassignments of the three critical classes, we evaluated the reference data sets and extracted reflectance values, indices, and colour spaces. In Figure 40 the average reflectance of the different classes for Greenland are shown. The average value of plastic does not overlap with any other class. To see the distribution of reference data in more detail, a density function of the reflectance of *plastic*, *rocks* and *wood* is shown as most misclassification occurred in those classes (Figure 41, Figure 42). In contrast to Figure 40, a high overlap of all the spectral bands can be observed. Figure 43 to Figure 45 show a similar pattern for the applied indices and colour spaces with high overlaps between the classes Those overlaps may have led to the difficulties of the machine- and deep learning algorithms to distinguish between those three classes. A similar pattern of overlap between *plastic, rock and wood* was observed for the reference data of Svalbard (G.1 to G.6).

A comparison between the average reflectance values from the reference data set with the reflectance of the geo-referenced in situ plastic objects was done in Figure 46 and Figure 47. The reflectance of the in-situ points confirmed the observed wide range of reflectance values for plastic objects.

Figure 40   Average reflectance of input classes from the reference data of Greenland
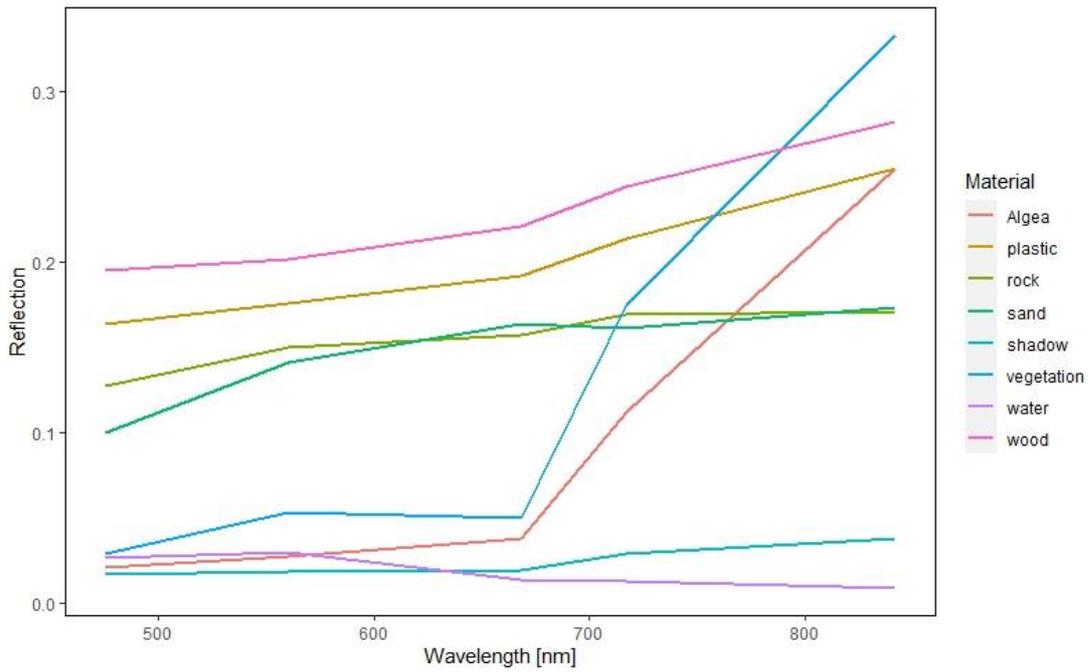
Class snow was excluded for visualization as its reflection value for all wavelength is outside the y-axis range used in this visualization. It was set to better display the differences in pattern of the other classes. Source: Compiled by BCSH

Figure 41   Density functions of the reflectance extracted of the reference data on Greenland for wavelengths of 475, 560, 668, 717 and 842 nm.
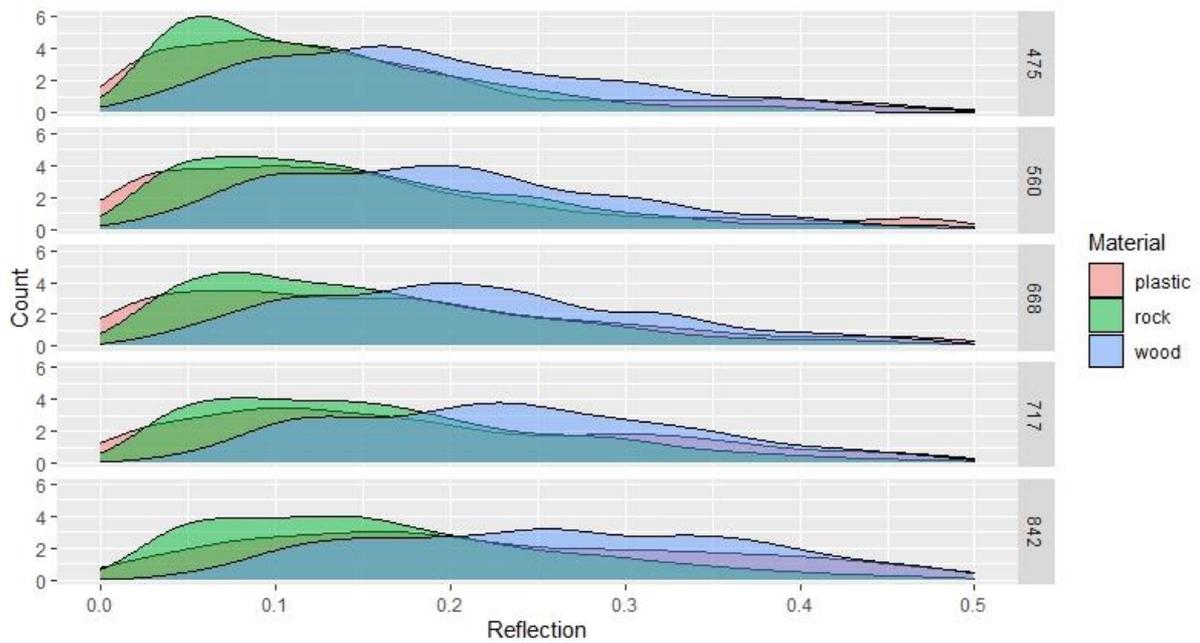
Source: Compiled by BCSH

Figure 42 **Density functions of the applied indices extracted of the reference data on Greenland**
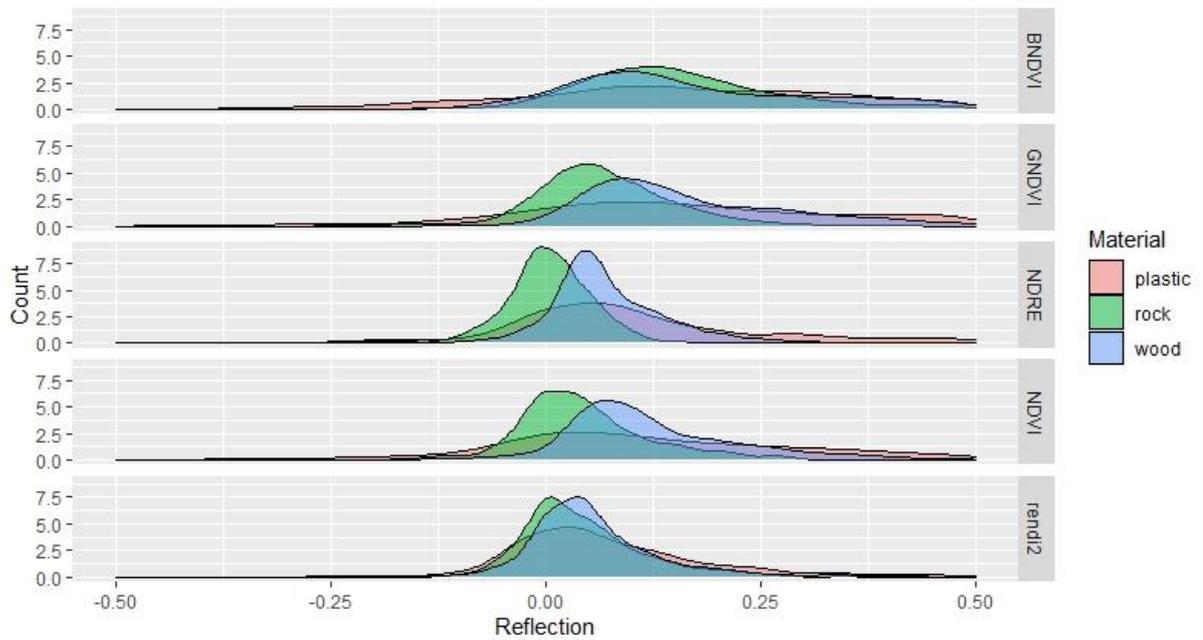
Source: Compiled by BCSH

Figure 43 **Density functions of HSV colour space extracted of the reference data on Greenland**
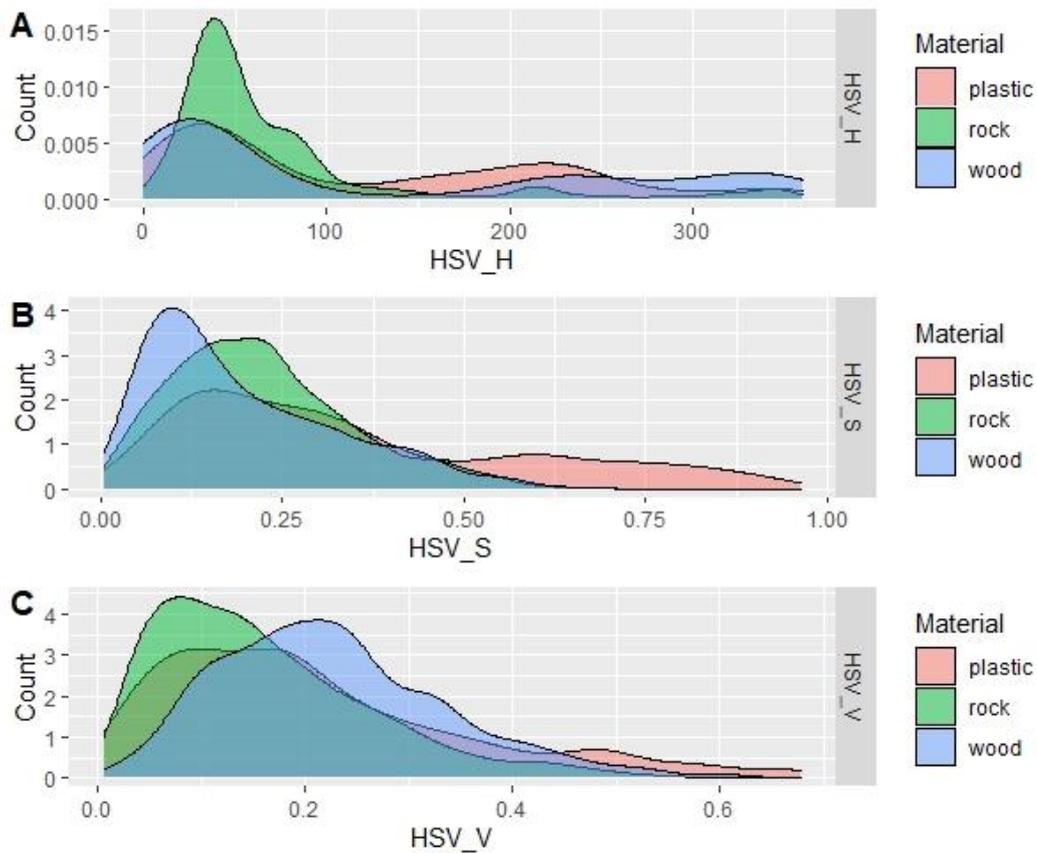
Source: Compiled by BCSH

Figure 44   Density functions of LAB colour space extracted of the reference data on Greenland
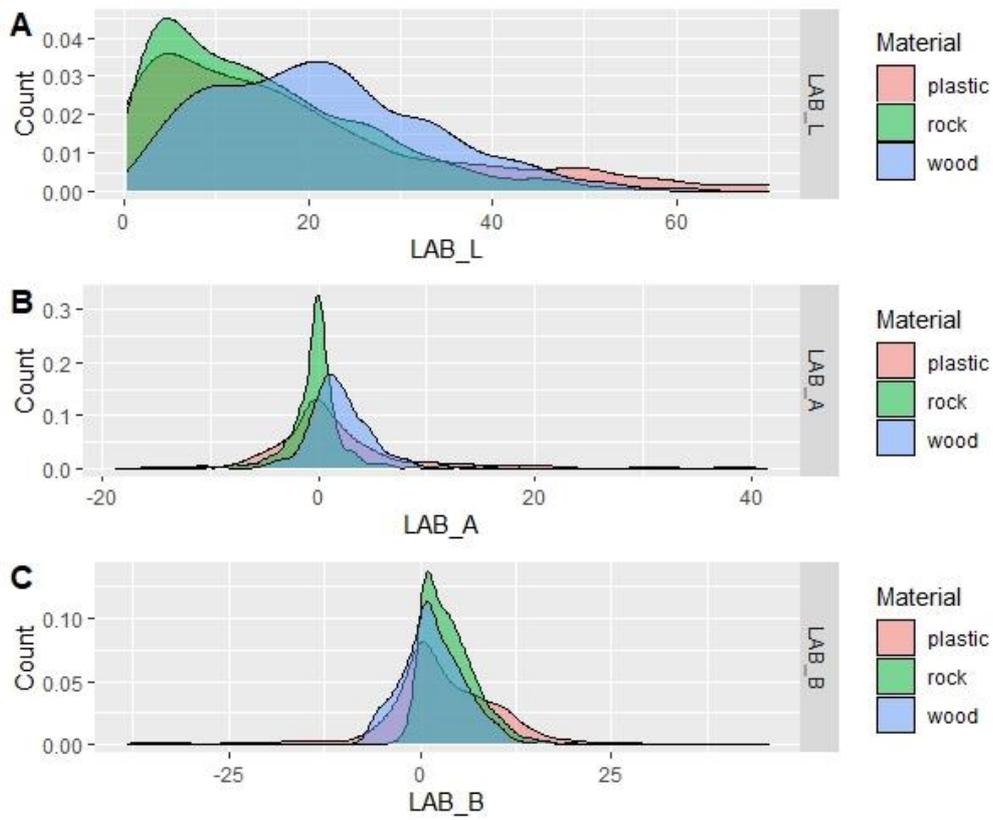
Source: Compiled by BCSH

Figure 45   Density functions of YcBcR extracted of the reference data on Greenland
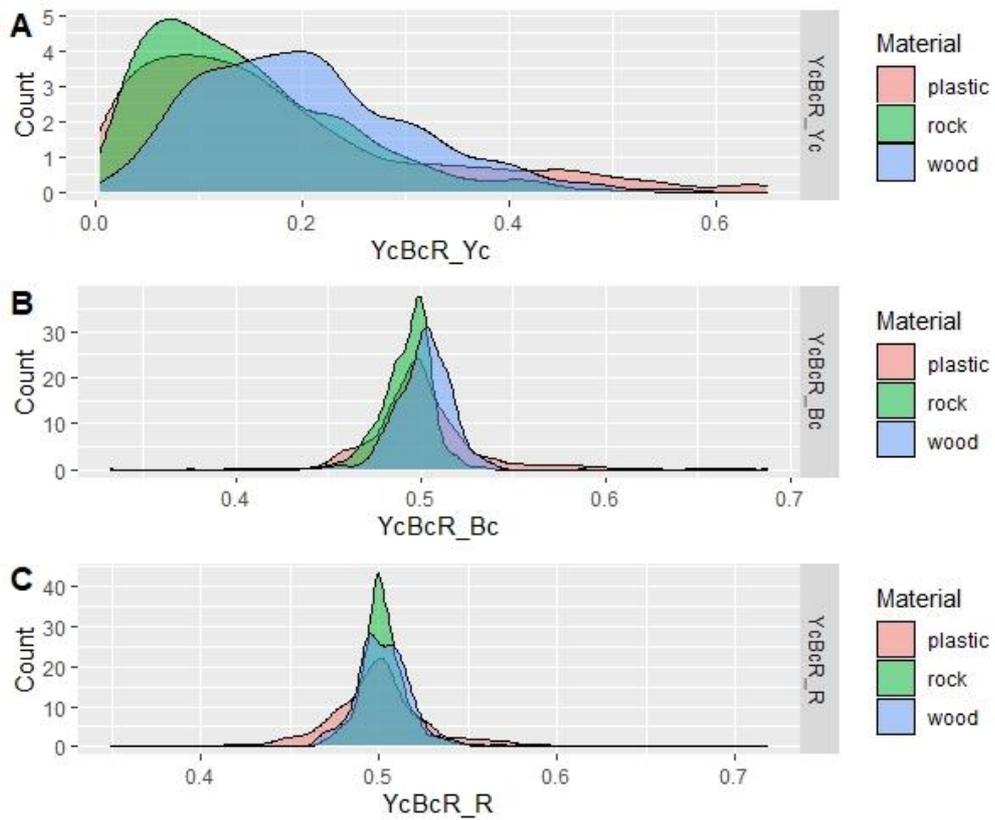
Source: Compiled by BCSH

Figure 46   Comparison of the reference data set with the in situ geo-referenced plastic objects
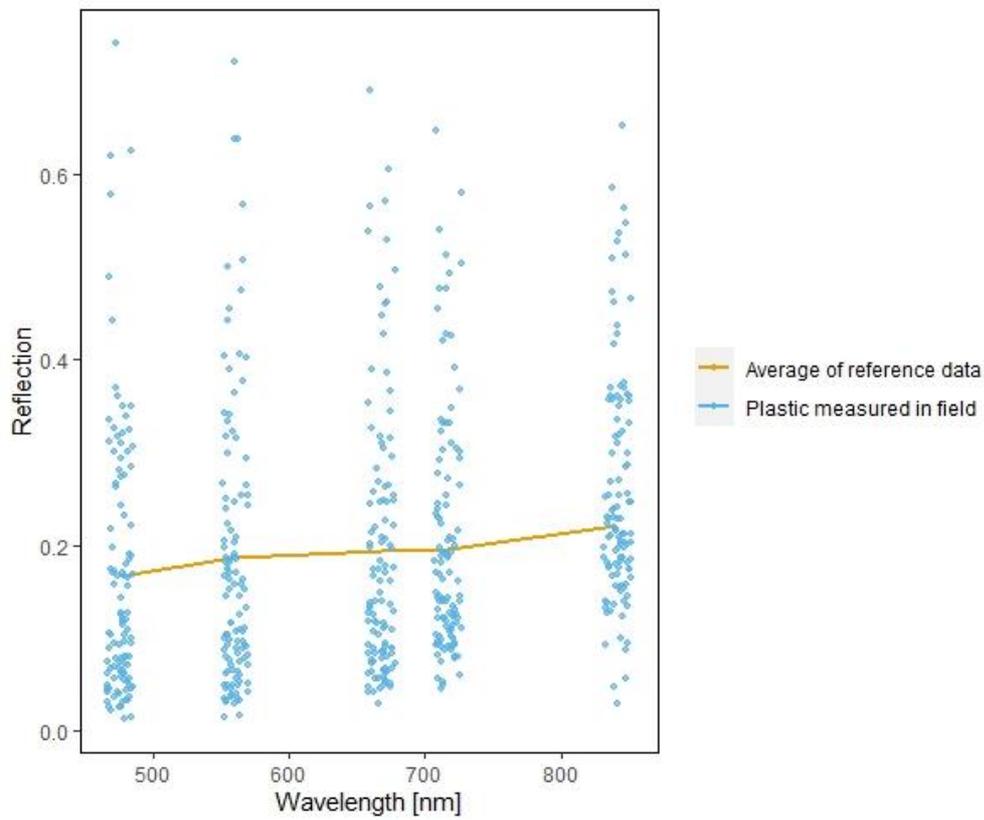
Comparison of reflection values of plastic objects for Sisimiut Day4. Source: Compiled by BCSH

Figure 47   Comparison of the reference data set with the in situ geo-referenced plastic objects
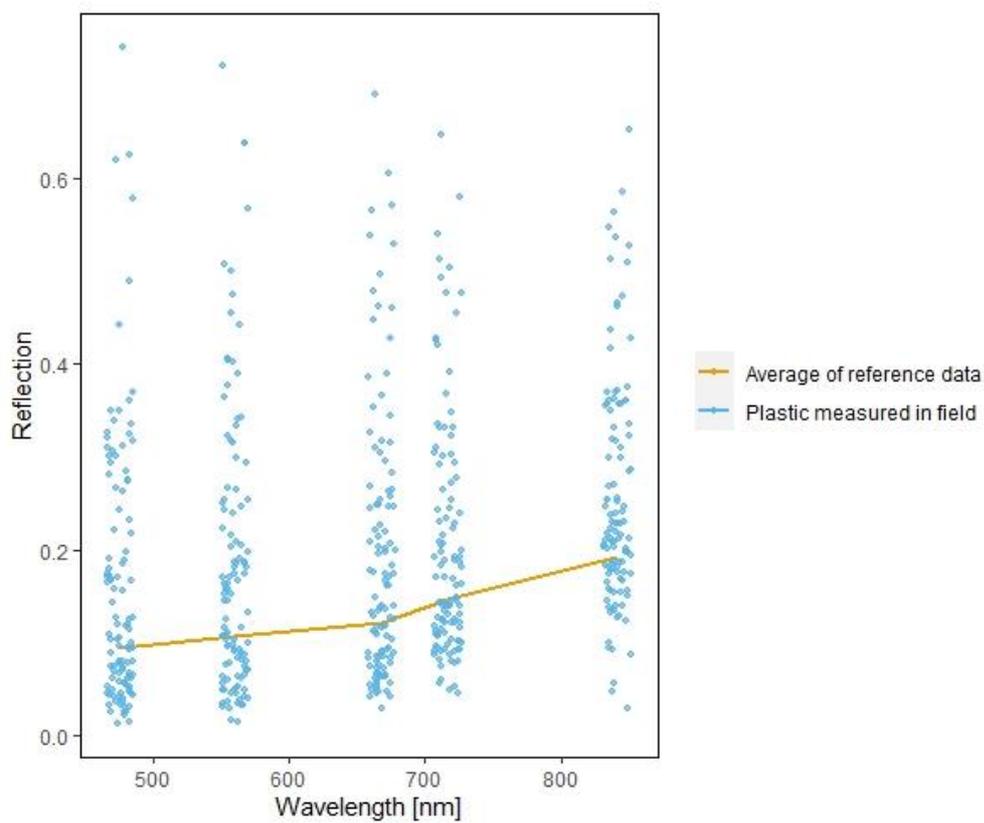
Comparison of reflection values of plastic objects for Sisimiut Day5. Source: Compiled by BCSH

To visually compare the distribution of plastics between the (semi-) automatic classification and the manual screening, density maps were created for all surveys on Greenland (Figure 48 to Figure 51). For the (semi-) automatic classification the results of RF v3 were used as they reached the highest OA. The created density maps confirm the high overestimation of plastic classifications observed in Table 18. Especially areas covered by rocks and wood show high densities of plastics for RF classifications while areas of vegetation, water, and sand show more accurate results.

Figure 48   Density map of area coverage of plastic litter from manual screening and RF v3
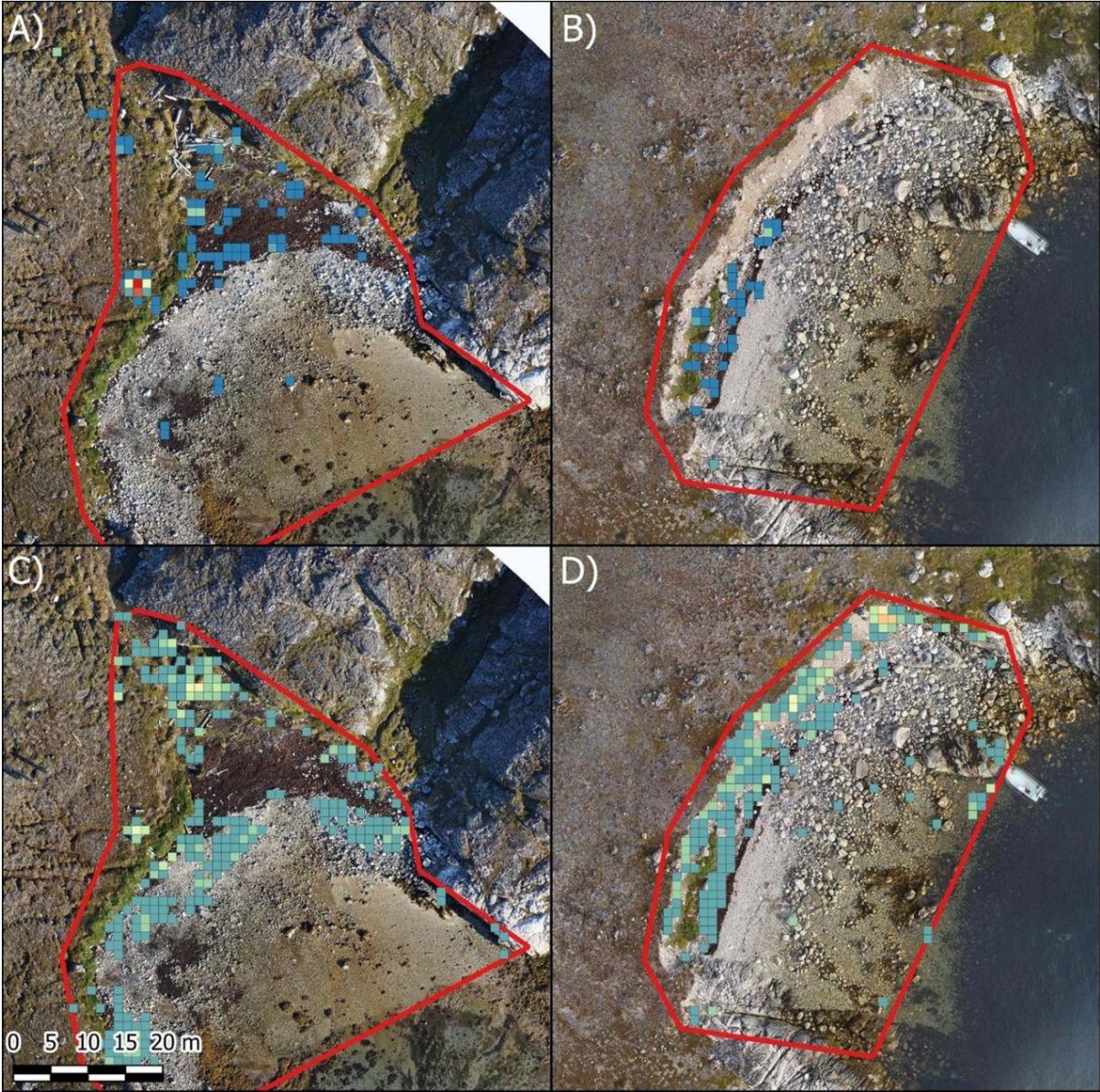
A) and C) Nuuk Day 1 manual screening of drone images and RF v3, respectively

B) and D) Nuuk Day 2_1 manual screening of drone images, D and RF v3, respectively.

Raster cells with density 0 – 0.1 were excluded for C) and D) for better visual comparison. Source: Compiled by BCSH

Figure 49   Density map of area coverage of plastic litter from manual screening and RF v3
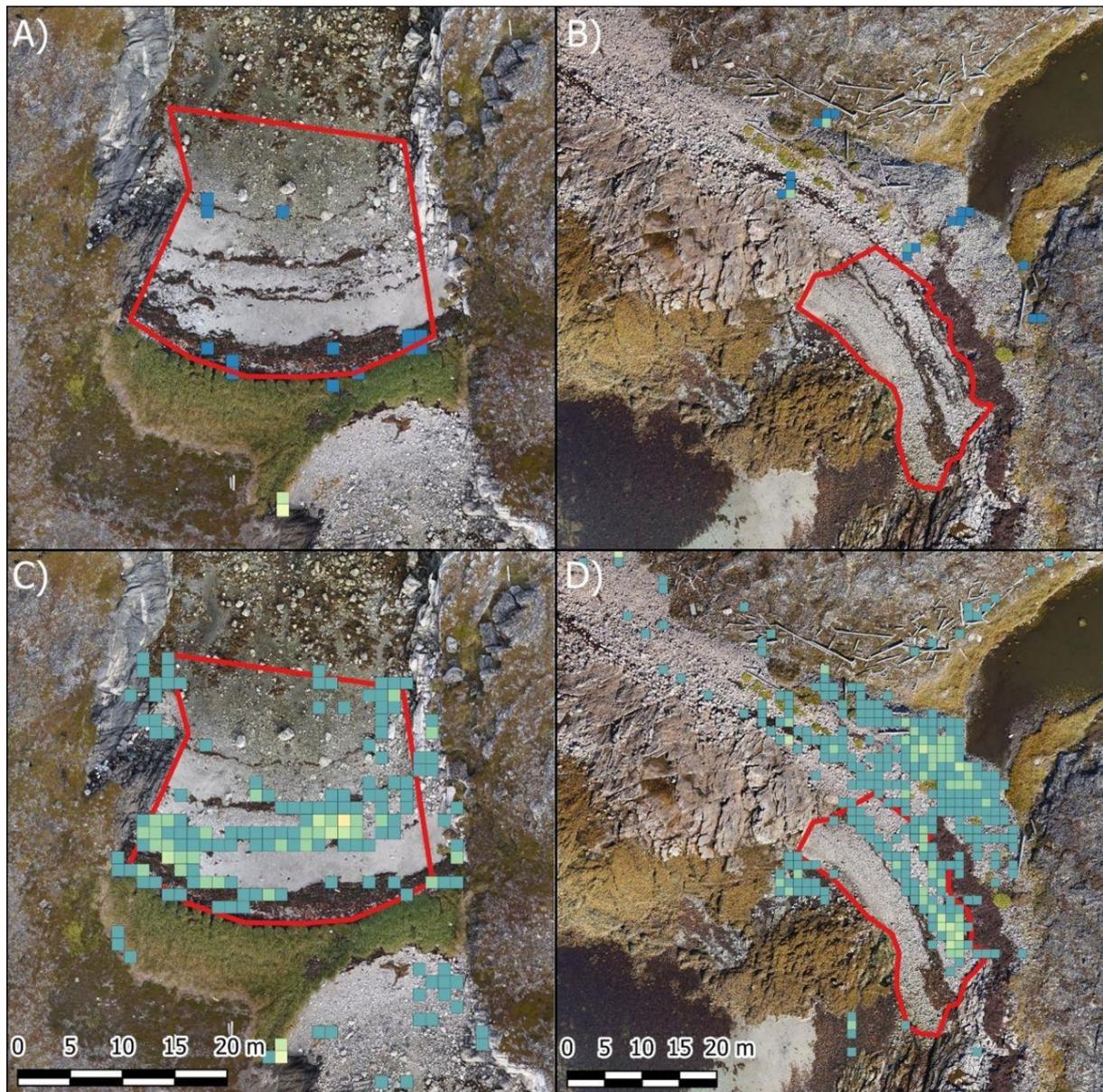
A) and C) Nuuk Day 2_2 manual screening of drone images and RF v3, respectively

B) and D) Nuuk Day 3 manual screening of drone images, D and RF v3, respectively.

Raster cells with density 0 – 0.1 were excluded for C) and D) for better visual comparison. Source: Compiled by BCSH

Figure 50   **Density map of area coverage of plastic litter from manual screening and RF v3**
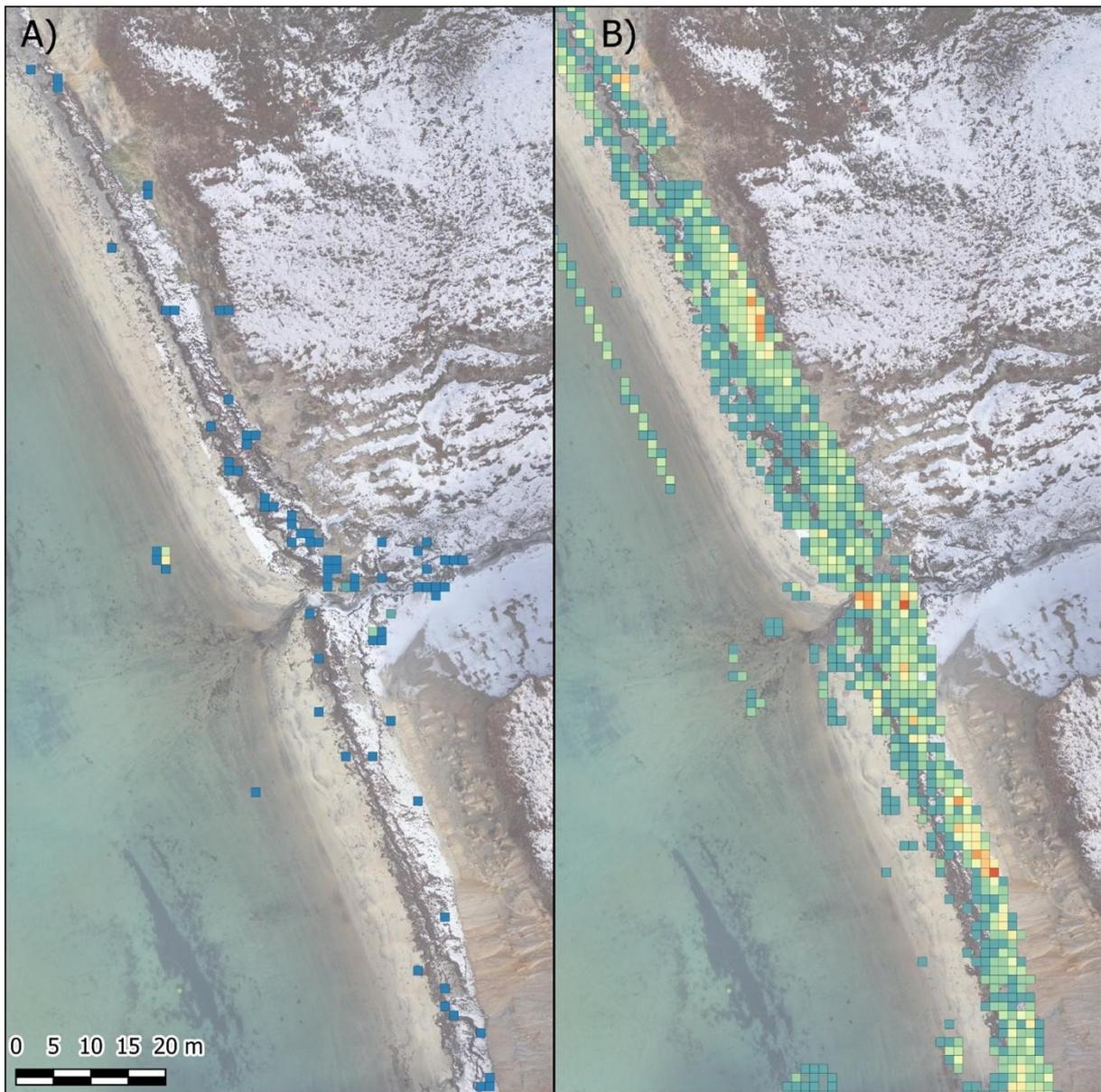
A) and B) Sisimiut Day 4 manual screening of drone images and RF v3, respectively

Raster cells with density 0 – 0.1 were excluded for B) for better visual comparison. Source: Compiled by BCSH

Figure 51   Density map of area coverage of plastic litter from manual screening and RF v3
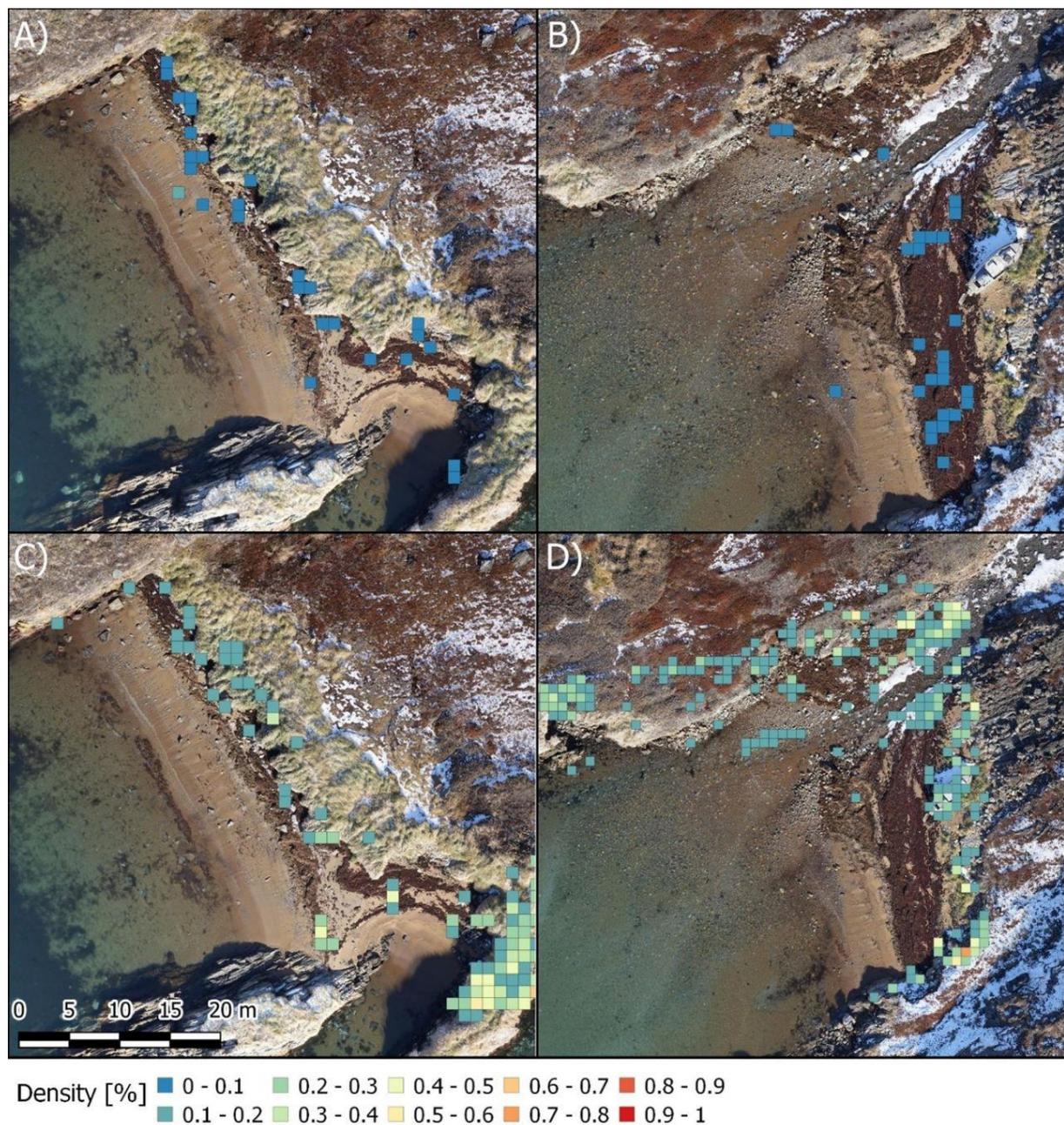
A) and C) Sisimiut Day 5-1 manual screening of drone images and RF v3, respectively

B) and D) Sisimiut Day 5-2 manual screening of drone images, D and RF v3, respectively.

Raster cells with density 0 – 0.1 were excluded for C) and D) for better visual comparison. Source: Compiled by BCSH

4.3.2.4   Discussion drone survey

The aim of this study was the evaluation of remote sensing techniques for beach litter monitoring in the Arctic with a focus on macro litter.

Drone surveys were conducted to acquire high resolution RGB and VIR imagery to perform a manual screening and to test machine learning approaches for (semi-) automatic classifications. The results of the manual screening were furthermore used as reference data for the satellite imageries. For a ground truthing of the litter abundance, OSPAR beach litter monitoring was applied on all survey areas.

Manual screening of drone footage

During the manual screening between 0 and 17.5% of plastic items could be found compared to those found during the OSPAR monitoring. This value is lower compared to most of the previous studies using drone footage for litter detection which achieved a detectability between 18% (Merlino et al., 2020) and up to almost 100% (Andriolo et al., 2020; Escobar Sánchez et al., 2021; Gonçalves et al., 2020b) (Table 20). Only Lo et al. (2020) described a lower recovery rate with 3.7%, but these authors separated litter objects into different size classes. The detectability of 3.7% was achieved for objects between 2.5 - 5 cm.

For the low recovery rates in this study, three potential reasons were identified and will be discussed in the following: The litter size-distribution, the beach environment, and the ground sample distance (GSD) of the drone footage.

The litter size distribution has a great impact for litter detectability with larger objects being easier to spot. Martin et al. (2018) and Merlino et al. (2020) described that "smaller items" (<4 cm) had a higher probability to not be found in their respective studies. This was confirmed by Lo et al. (2020) who described increasing detection rates for an increase in object size. Therefore, depending on the litter size distribution at the study site, the total proportion of detected litter objects can vary significantly between different sites. This might be one reason why Andriolo et al. (2020) found almost 100% of litter objects whereas Merlino et al. (2020) only detected around 18 – 20% even though the GSD used by Andriolo et al. (2020) was coarser and the beach environment was more complex. Looking into the objects found by Andriolo et al. (2020), over 50% were assigned to classes which have typically a size >10 cm and are therefore easier to spot. For this study, only an approximate estimation of litter size distribution can be given as the OSPAR monitoring does not include a size measurement of objects but only gives size ranges for some categories. At the study sites at Kapp Mitra (Svalbard), Nuuk (Greenland) and Sisimiut (Greenland) *plastic fragments (2.5 cm - 50 cm)* were the most or second most frequent plastic litter category with most objects being confirmed to be <10 cm by the infield surveyors. Other abundant litter types were *Styrofoam*, *strapping bands*, *strings and cords*, *string and filaments*, *nets and net rests <50 cm*, *caps and lids* and *Shotgun cartridges* (Table 10, Table 11 and Table 12) which could also be confirmed to be mainly of a size <10 cm. The dominance of small objects might have led to the relative low total detectability compared to other studies. This can be confirmed by looking into the size distribution of the manual screening where small objects seem to be underrepresented. Large objects (>10 cm) make up between 70% and 80% of the plastic objects found during the manual screening.

A second parameter impacting the detectability of beach litter is the beach environment. This includes the type of the beach (substrate: sand or gravel), its homogeneity (e.g., occurrence of vegetation, wood, or rocks) and the resulting potential of litter to be buried or hidden. The more heterogenic a beach is, the more difficult it is to distinguish litter objects from the background. The bright colours which make it easy to identify litter objects might not be dominant over vegetation or coloured pebbles. White, black, and transparent objects might be confused by rocks, pebbles, or wood. Vegetation, rocks, or wood might furthermore cover or hide litter objects on the drone image (Merlino et al., 2020). Moreover, snow coverage can burry objects and make an identification of white or transparent litter difficult (Figure 53). Most of the previous studies were performed on sandy beaches (Escobar Sánchez et al., 2021; Falatti et al., 2019; Gonçalves et al., 2020b; Martin et al., 2018 and Merlino et al., 2020) (Table 20). As an exception, Lo et al. (2020) studied both, sandy and gravel beaches. Information about the beach homogeneity and the potential of objects being hidden or buried are difficult to retrieve from previous studies if not mentioned explicitly. Anyhow, if available the orthomosaics and pictures of the study area give a first impression of the homogeneity of the beaches (Figure 52, Andriolo et al., 2020; Falatti et al., 2019, Gonçalves et al., 2020b; Martin et al., 2018).

Figure 52   Overview of the beach environment from previous studies applying manual screening of UAV imagery

1) Escobar Sánchez et al., 2021; 2) Merlino et al., 2020, both licensed under CC BY 4.0

Figure 53   Examples of the beach environment on Svalbard and Greenland
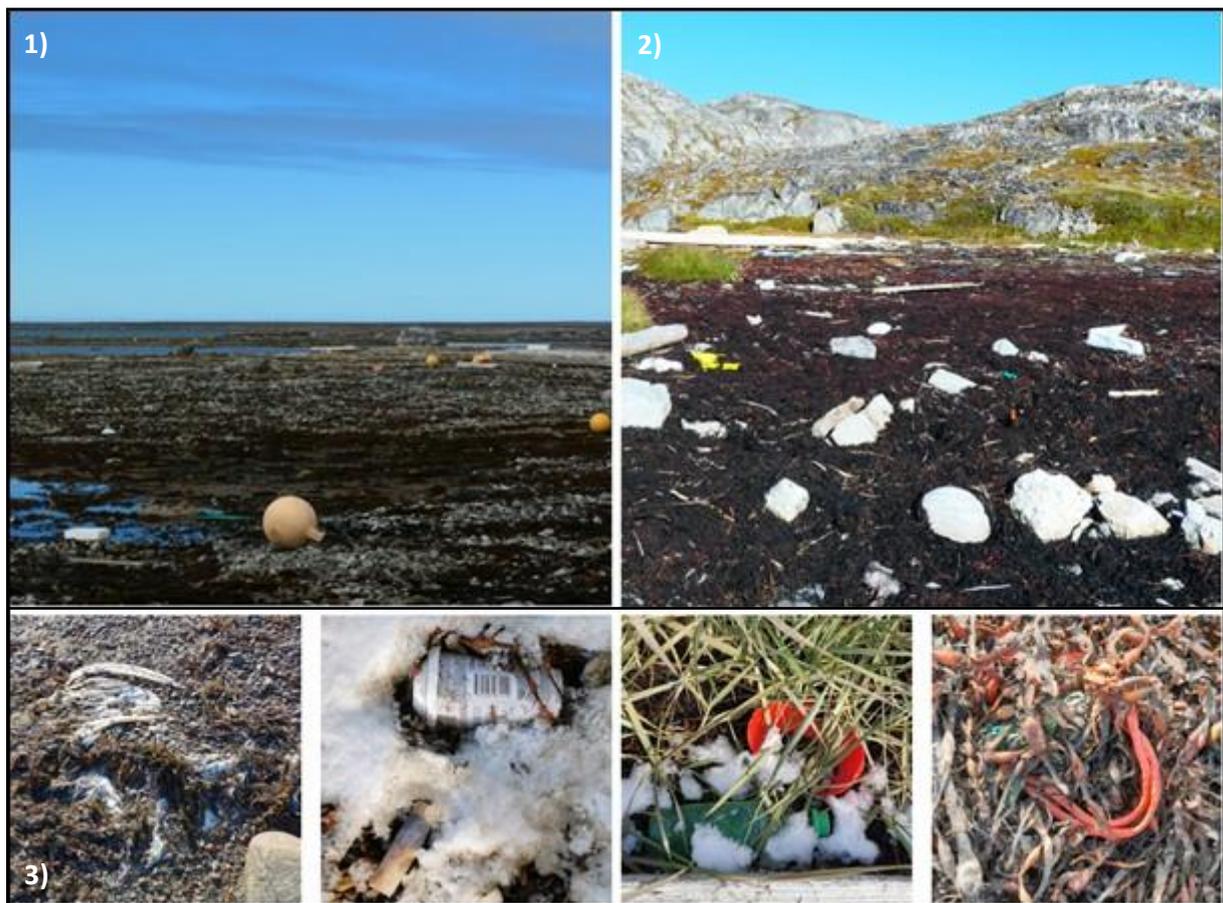
1) Area attached to the beach of Kapp Mitra, Svalbard 2) Beach at survey area of Nuuk Day 1, Greenland
3) Buried or hidden litter items at Sisimiut Day 4 and Sisimiut Day 5, Greenland. Source: Compiled by BCSH

The beaches monitored in the studies of Escobar Sánchez et al., (2021), Falatti et al., (2019) and Gonçalves et al., (2020b) appear to be mostly homogeneous with little patches of vegetation. The beaches monitored by Martin et al. (2018) show a higher abundance of vegetation, whereas Andriolo et al. (2020) and Merlino et al. (2020) studied the most heterogeneous environment with high amounts of vegetation and driftwood (Figure 52). Lo et al. (2020) does not give an overview of the beaches but described one beach with gravel and pebbles and the second beach

as homogeneous with mostly sand as beach substrate. Comparing the litter detectability, the studies with gravel beaches and more heterogeneous beach environments had low detection rates (Martin et al., 2018; Merlino et al., 2020 and Lo et al., 2020). Covering both, sandy and gravel beaches, Lo et al. (2020) showed that sandy beaches had a higher detection rate than gravel beaches. An exception was given by Andriolo et al. (2020) who could detect almost all objects, even though the beach environment was complex. A possible reason of the high detectability in Andriolo et al. (2020) is the impact of the litter size distribution and the high number of large items. In this study, a total of seven Arctic beaches (including various beach segments) were monitored, including five gravel and two sandy beaches. The beach environment of all study sites was defined as heterogeneous with high abundances of driftwood, rocks, and vegetation and with some snow coverage on the two sandy beaches close to Sisimiut (Figure 53). Comparing the beach types, the two sandy beaches were among those beaches with the highest detection rates (15% and 17.4%). The low detectability of the two sandy beaches compared to studies with similar beach substrates, might be influenced by the snow coverage, the litter size distribution, and the GSD. Figure 37 shows that from the plastic items found on the drone image, white items had a different size distribution compared to coloured items showing a two to three times larger median object size than white objects. Assuming no colour-based difference in the actual size distribution, white or transparent objects were difficult to identify, especially over bright backgrounds such as rocks and snow. This could be confirmed at the study site of Sisimiut Day 4, where 50% of the objects which were measured via GNSS, and which could not be found during the manual screening, were white or transparent. The complex background with vegetation and coloured pebbles might furthermore have decreased the detectability of coloured objects and influenced the overall detection rate.

Beside the litter size distribution and the beach environment, the GSD of the drone image has the biggest impact on the detectability of beach litter as it defines the number of pixels per object. A higher GSD (lower GSD value) decreases the minimum size of an object sufficient to be identified. Items with roughly ten times the GSD can be visually detected on RGB imagery (Moy et al. 2018, Pichel et al. 2012, Garcia-Garin et al. 2019). The GSD is the only of the above-mentioned parameter that can be directly defined by the operator as it depends on the chosen sensor and the flight altitude. Previous studies used a GSD between 0.18 cm (Merlino et al., 2020) and 1.2 cm (Andriolo et al., 2020) (Figure 52). Comparing the detectability of the abovementioned studies, the detection rate did not always increase with a decrease of GSD values. This confirms the impact of additional parameters, such as the litter size distribution and the beach environment. To evaluate the impact of the GSD, Lo et al. (2020) compared different GSD settings for the same beaches and confirmed that lower GSD values result in an increase in detectability. Anyhow, the lowest GSD value may not always be the best choice, as a decrease in the GSD value requires a lower flight altitude and results in less area coverage. In addition, fixed-wing drones, which are typically used for large area monitoring, are limited in their minimum flight altitude. As fixed-wing drones require a minimum flight speed to be stable and as the shutter speed is limited by the sensor, only the flight altitude can be adjusted to ensure that the footprint of the images is sufficient to not produce any data gaps. Multicopter can adjust the flight speed and therefore also operated at very low altitudes but have typically lower area coverage. Therefore, they are favourable for experimental purposes but not for large-scale applications. The previous studies of manual screening of beach litter all applied a Multicopter focusing on small test sites. Only Martin et al. (2018) describes a large-scale application along the Saudi-Arabian coastline. As for litter detection in the Arctic, time and area coverage are crucial, for this study it was decided to use a WingtraOne fixed-wing drone. The minimum GSD of the applied drone-sensor configuration was 1.4 cm for the RGB sensor. Assuming a possible detectability of objects with a minimum size of ten times the GSD, it was expected to detect mainly litter items >14 cm. Nevertheless, due to the shape and colour of plastic items, also

smaller items were expected to be found. Therefore, one goal of this study was to test whether a GSD of 1.4 cm is still sufficient to reflect the litter distribution on Arctic beaches. The results of the manual screening showed that this is not the case, and a decreased GSD will be necessary for future applications. As described above, the environmental background made the colour-based detection of small items difficult. Moreover, comparing the detection rate of this study with the results of Andriolo et al. (2020) who used a similar GSD, shows huge differences which can be addressed to differences in the litter size distribution.

Table 20   Overview of the litter found by manual screening of drone footage and the corresponding litter size distribution, the beach environment, and the GSD

	Plastic litter found by manual screening compared to ground truthing	Size distribution	Beach environment	GSD [cm]
This study	0 - 17.5%	Majority of objects <10 cm	gravel beach and sandy beach	1.4
Andriolo et al., 2020	"non-worthy number missed"	~ 50% of plastic items were bottles, bags, and octopus pots	coastal dune area	1.2
Escobar Sánchez et al., 2021	87.5 - 99%	NA	sandy beach	0.27
Falatti et al., 2019	87.8 - 95.9%	Between 2 – 35% of items were <5 cm	sandy beach	0.44
Gonçalves et al., 2020b	98%	NA	sandy beach and partly dune	0.55
Lo et al. 2020	3.7 - 100%	Five size classes from 2.5 – 5 cm to 30 – 50 cm	gravel/ pebbles and sandy beach	0.2 – 0.6
Martin et al., 2018	61.8%	NA	sandy beach	0.5 – 0.7
Merlino et al., 2020	18 - 20%	NA	sandy beach	0.18

NA: No information regarding litter size distribution

Further minor factors that can impact the detectability are the daytime of the acquisition, weather conditions, and the user performing the manual screening. Lo et al. (2020) and Martin

et al. (2018) suggested to conduct surveys around noon to minimize shadow effects which can have a negative effect on the detectability. In the Arctic, the avoidance of shadow effects is challenging as the sun angle remains low throughout the entire day. Additionally, weather conditions can have an impact on the image's film speed (ISO) and therefore the image quality. In remote areas like the Arctic, it is not feasible to expect perfect conditions in terms of daytime and weather condition as the time for monitoring is normally limited. For the manual screening, User's, being familiar with the study site and the expected litter items, are favourable as they will be more likely able to identify the litter items on the drone images.

Beside the detectability, also the time which was needed for the drone surveys and the manual screening was evaluated. The duration of the drone surveys varied between 20 and 60 minutes, including approximately 15 minutes for the drone setup. The data processing and preparation was mainly automatic and took around 01:30 hour of manual working time per survey. Large surveys, like those performed at Kapp Mitra, could cover up to fifty 50 m transects per hour, resulting in a working time (drone survey and data processing) of around 3 minutes per 50 m transect, if all transects are monitored. Smaller areas, covering a single 50 m transect still need a minimum of 20 minutes for the drone survey and around 1,5 hours for the data processing. The manual screening of the 50 m transects took between 5 to 20 minutes, depending on the beach background and the number of litter items. Including the drone survey, the data processing and the manual screening, the total time for a drone-based monitoring of a 50 m transect could vary between 8 and 130 minutes (Table 21).

Table 21   Time required for a drone-based monitoring of beach litter

Monitored beach	Drone survey	Data processing	Manual Screening	Total	OSPAR
50 m transect	01:10 – 20:00 min	01:50 – 90:00 min	05:00 – 20:00 min	08:00 – 130:00 min	90:00 – 180:00 min

In the Arctic, the OSPAR monitoring was performed by two to three people who took between 30 minutes and one hour per 50 m transect, resulting in a total working time between 90 and 180 minutes. Compared to the calculated time for a drone-based monitoring, the OSPAR method requires up to 22-times longer for large areas but could be faster for small areas. The total time for a drone-based litter monitoring can be reduced by a successful integration of (semi-) automatic classification and detection approaches, saving the time for the manual screening. Martin et al. (2018) described that their drone surveys were 39-times faster compared to a standard in situ approach, but without including the time required for the manual screening. As the beaches in the Arctic are often remote, the time at the beach can be limited. Therefore, a faster beach coverage can already be of high value.

After all, the manual screening of drone images is a promising method for beach litter detection as already shown in several studies (e.g. Escobar Sánchez et al., 2021; Falatti et al., 2019; Gonçalves et al., 2020b and Lo et al. 2020). Drone surveys allow a fast, non-intrusive acquisition of large and remote areas. Furthermore, they can give information about the spatial distribution and the impact of beach morphology as well as about the litter size distribution. However, the environmental conditions and the requirements of a large area coverage and a short available acquisition time, make an implementation in the Arctic challenging. The results of this study show that a GSD of 1.4 cm is not sufficient for the detection of small objects. To meet up with the complex environmental conditions and the litter size distribution, a sub-centimetre GSD is

recommended. Yet even applying a GSD of 0.18 cm to 0.4 cm (Merlino et al., 2020 and Lo et al., 2020 respectively), the authors described difficulties in litter detection for objects <10 cm. Besides, further limitations must be considered. To conduct a drone survey in the Arctic, a drone pilot is required in remote areas and in case the litter should not only be monitored but also collected, the beaches must be accessed in person neglecting one of the major advantages of drone surveys. Additionally, the detectability of litter will always be limited for buried or covered litter objects and an information about the litter weight cannot be given straight forward. To match the requirement of fast acquisitions of large areas, fixed-wing drones are favourable. Professional fixed wing monitoring drones like the WingtraOne and the senseFly eBee achieve up to date only GSDs of 0.7 - 1 cm with their sensors, which might not be sufficient to record small litter items. This can change in the future with the integration of new sensor technology, which is already available in the very high price sector (e.g. PhaseOne sensors). Alternatively, the implementation of citizen science with an application of low-cost multicopter could be an option for a smaller-scale integration of Arctic beach litter monitoring. Especially an integration in existing cruise tours or research programs could be of interest. The area around Nuuk showed that beside the long coastline, only a small part of it was suitable for beach litter detection and could therefore also be covered by a low-cost multicopter. A well-defined operational framework could be developed and shared with the different parties. Drone surveys could be applied during existing cruise tours, university excursions or research programs and be collected and interpreted by experts afterwards. To match the Arctic requirements, such a framework is still needed to be developed.

Besides, for evaluating the feasibility of the application of fixed wing drones for manual litter detection in the Arctic, the results of the manual screening were used as reference for the (semi-) automatic classification approaches of the multispectral UAV sensor as well as for the satellite imagery.

(Semi-) automatic classification

The advantages of drone applications for beach litter monitoring were already mentioned and discussed for a manual screening. One of the biggest advantages of drone surveys compared to traditional in situ methods, the large area coverage, also produces the challenge of screening huge amounts of data and images. On Greenland and Svalbard, a total number of 4201 RGB- and 5364 VIR- images were collected covering a total area of around 100 ha per sensor. Multicopter flying on lower altitudes to achieve a higher GSD, would acquire even more images for the same area as the image footprint is smaller. A manual screening of large areas can be very time-consuming resulting to be neither time- nor cost-efficient. Therefore, for a large-scale monitoring, an application of (semi-) automatic classification approaches is necessary. Several approaches have already been tested for beach litter detection like threshold methods (Bao et al., 2018), Maximum-Likelihood classifier (ML) (Bao et al., 2018; Escobar Sánchez et al., 2021), Random Forest (RF) (Escobar Sánchez et al., 2021; Gonçalves et al., 2020a; Gonçalves et al., 2020b; Gonçalves et al., 2020c; Martin et al., 2018; Wolf et al. 2020), Support Vector Machine (SVM) (Escobar Sánchez et al., 2021; Wolf et al. 2020), K-nearest neighbors (KNN) (Gonçalves et al., 2020c), and Convolutional Neural Networks (Bak et al., 2019; Falatti et al., 2019; Gonçalves et al., 2020b; Papakonstantinou et al., 2021; Wolf et al. 2020). The above-mentioned studies used high resolution RGB drone imagery with a GSD between 0.2 cm (Wolf et al., 2020) and 3 cm (Bao et al., 2018). The classification approaches further varied between pixel-based, object-based, and tile-wise classifications. The performances of the classification approaches are summarized in Table 22. As mentioned in Goncalves et al. (2020a), a comparison between different classifiers in different studies has always to be taken with care, as factors like the beach environment, the number of the testing areas and the litter size distribution may have a bigger impact than the classification method itself. Bao et al. (2018) was amongst the studies with highest overall

accuracy (OA) even using the lowest GSD of all studies mentioned and a relatively simple classification approach. Here, the very homogeneous beach environment may have had a bigger impact than the GSD or the classifier. Beside the OA, which was just available in some of the mentioned studies, the UA (User's Accuracy; also mentioned in some studies as sensitivity or Recall) and the PA (Producer's Accuracy; also mentioned as precision or positive predicted value (PPV)) are used for accuracy assessment as well as the F-score. The F-score is a measure of the overall quality of a classifier and combines UA and PA:

$$F_{score} = 2 * \frac{UA * PA}{UA + PA}$$

It was used by the majority of the studies on automatic beach litter detection (Falatti et al., 2019; Gonçalves et al., 2020a, b, c; Martin et al., 2018; Papakonstantinou et al., 2021; Wolf et al. 2020).

In this study, three classifiers were applied (RF, SVM and NN) using RGB and VIR imagery. The multispectral imagery was tested to determine whether the additional spectral information can compensate the low applied GSD values. The GSD of the two sensors was 1.4 cm and 3.4 cm for RGB and VIR respectively. The OA achieved with the RGB footage was between 62.5% - 78.6% with highest results for RF. The reached OA is comparable to the results of Escobar Sánchez et al. (2021) and Papakonstantinou et al. (2021) even though those studies used a GSD of 0.27 cm and 0.5 cm respectively. The similar OA compared to Escobar Sánchez et al. (2021) who only used the RGB colour space, might be the result of an extension of the feature space by the HSV, CIE-LAB and YcbCr colour spaces as previous applied in Gonçalves et al. (2020a, b, c). Bao et al. (2020) and Wolf et al. (2020) using a GSD of 1 – 3 cm and 0.5 cm respectively reached higher OA with up to 98% and 83% respectively. The achieved F-score in this study varied between 40% and 70% for the RGB imagery and is also comparable to the studies summarized in Table 22, being only slightly lower than the studies of Falatti et al. (2019), Gonçalves et al. (2020a, c), Papakonstantinou et al. (2021) and Wolf et al. (2020) who reached F-scores between 73% and 81%.

Comparing the performance on RGB and VIR imagery, in this study, the classifier using VIR outperformed those using RGB images with the highest OA of 90.6% against 78.6% and F-score of 77% against 70%. For both, the best results of OA and F-score were achieved using RF, but SVM showed similar results. Using a Neuronal Network (NN) with weighting, an overfitting of the model could be observed with very high PA and low UA values. The results on RGB and VIR imagery indicate that the feature space used for classification had a greater impact on the classification performance than the GSD, making the application of VIR sensors promising for the future. Anyhow, when looking at the actual classification maps, a huge overestimation of plastics could be observed in this study for both sensors. Despite the good statistical results, it was not possible to separate the different classes by their spectral information properly (Figure 40 to Figure 45). Especially the classes wood and rock were often misclassified as plastics. As those classes were much more abundant at the study sites than plastics, already a small percentage of misclassified rocks or wood could increase the number of pixels classified as plastic significantly. This emphasises again the strong impact of the beach environment on the classification performance. For VIR acquisitions, the weather conditions have also a major impact. Before and after each survey an image of a reference reflection panel is taken, to transform the recorded digital values into reflectance values. Therefore, the weather conditions (especially cloud coverage) should not change within a single survey as this could lead to corrupted reflection values and impact the classification results. Yet, if this can be assured, the reflectance values of VIR acquisitions allow for a transferability of the reference data and the models for spectral-based classification approaches. In contrast, RGB images for colour-based classifications are highly weather and light-dependent and difficult to transfer.

This study confirms the great potential of (semi-) automatic classification approaches offering a method for a time- and cost- efficient large-scale detection of beach litter using drone imagery. The statistical approaches permit an evaluation of the performance of the classifier, and as the classification is user-independent, no previous expert knowledge of the litter types is necessary. However, the promising statistical results have always to be confirmed by the actual classification maps. Here, the results show that a strong overestimation of plastic occurs applying classification approaches that focus on the spectral characteristics alone. The feature spaces of the applied sensors were not able to distinguish between the occurring land cover classes. Therefore, the spectral coverage of the sensors must be extended, or other structural parameters have to be included. The application of an object-based classification could reduce the misclassification of the wood class but would require a better GSD. Image recognition methods are also promising, requiring a better GSD and a huge amount of training data. Furthermore, most studies so far only focused on litter detection and not litter type identification as performed with the OSPAR surveys. Among the above-mentioned studies, only Wolf et al. (2020) presented an approach for plastic type identification using image recognition techniques. For a future implementation into ongoing monitoring programmes, this is an essential step, which must be further investigated and developed, taking into account the required area coverage (limiting the GSD), the beach environment and the litter size distribution.

The applied methods in this study were not sufficient for a litter type detection or identification mainly due to the limitations of the GSD and the complex beach environment. Nevertheless, the results give a first impression on VIR sensors for beach litter detection and recommendations for further drone-based litter monitoring in the Arctic. Due to the complex beach environments and the litter size distribution in the Arctic, a sub-centimetre GSD or an extension of the spectral coverage is recommended. The method with highest OA in this study was RF with comparable results to SVM.

Table 22 **Overview of the performance of different machine learning approaches for beach litter monitoring using drone imagery.**

	Litter found by (semi-) automatic detection	Classifier used	Beach environment	Sensor	GSD [cm]
This study RGB	OA: 62.5% - 78.6% PA: 52.6% - 92.6% UA: 25.4% - 70.3% F-score: 40% - 70%	RF, SVM, NN	gravel beach and sandy beach	RGB	1.4
This study VIR	OA: 79.8% - 90.6% PA: 63.3% - 95.2% UA: 40.6% - 82.9% F-score: 57% - 77%	RF, SVM, NN	gravel beach and sandy beach	VIR	3.4
Bak et al., 2019	OA: 98%	CNN	Gravel and sandy beach	RGB	0.4

	Litter found by (semi-) automatic detection	Classifier used	Beach environment	Sensor	GSD [cm]
Bao et al., 2018	OA: 98.6% - 98.9%	ML and Threshold	Sandy beach	RGB	1-3
Escobar Sánchez et al., 2021	OA: 25% - 74% PA: 0 - 100% UA: 0 - 11%	Object-based ML, RF, SVM	sandy beach	RGB	0.27
Falatti et al., 2019	PA: 12% - 69% UA: 25% - 100% F-score: 19% - 81%	CNN	sandy beach	RGB	0.44
Gonçalves et al., 2020a	PA: 53% - 77% UA: 62% - 74% F-score: 57% - 76%	RF	sandy beach and dune	RGB	0.55
Gonçalves et al., 2020b	PA: 65% - 71% UA: 55% - 70% F-score: 60% - 70%	RF, CNN	sandy beach and dune	RGB	0.55
Gonçalves et al., 2020c	PA: 61% - 70% UA: 67% - 78% F-score: 64% - 73%	Object-based RF, SVM, KNN	Sandy beach	RGB	0.55
Martin et al., 2018	PA: 40% UA: 8% F-score: 13%	RF	sandy beach	RGB	0.3 - 0.5
Papakonstantinou et al., 2021	OA: 61% - 78% PA: 63% - 98% UA: 15% - 84% F-score: 28% - 77%	Five different CNNs	sandy beach	RGB	0.5
Wolf et al. 2020	OA: 83%** PA: 77%** UA:77%** F-score: 77%**	CNN, SVM, RF	Beach and riverine system	RGB	0.2

The terms PA/ sensitivity/ recall and UA/ PPV/ precision were used as in the original publication

OA: Number of correctly identified reference objects divided by total number of reference objects

PA/ sensitivity/ recall of plastic: Number of correctly identified plastic objects divided by total number of reference plastic objects

UA/ Positive predicted value (PPV)/ precision of plastic: Number of correctly identified plastic objects divided by the number of reference objects classified as plastic

F–score: Two-times the product of UA and PA divided by the sum of UA and PA

** Results for CNN classifier.

4.3.3 Results and discussion of WV3 satellite imagery

The spectral information of WV3 imagery was extracted from pixels covering plastics and the most common land cover classes (rock, sand, snow, vegetation water and wood). *Plastic-Pixels* were further divided by the area by which the pixels were covered with plastic. For all other classes, pixels with a class coverage of 100% were used. An overview of the number of pixels per class are shown in Table 23 and Table 24 for the WV3 multispectral imagery and the WV3 SWIR imagery, respectively. For the higher resolution multispectral imagery (1.2 x 1.2 m), 95% and 74% of all identified *Plastic-Pixels* were covered by less than 5% of plastic for Svalbard and Greenland, respectively. Zero and five pixels were covered by more than 50% of plastic for Svalbard and Greenland, respectively (Table 23). For the SWIR imagery, only three pixels were covered by more than 10% of plastics, all identified on Greenland.

Table 23   Numbers of pixels used to extract spectral information of WV3 multispectral imagery

Material	Pixel Coverage	Svalbard (n pixel)	Greenland (n pixel)
Plastic	0 – 1%	**604**	**120**
	1 – 5%	**336**	**127**
	5– 10%	**33**	**47**
	10 – 20%	**10**	**18**
	20 – 30%	0	**11**
	30 – 40%	**1**	**3**
	40 – 50%	0	**4**
	50 – 60%	0	**2**
	60 – 70%	0	**1**
	70 – 80%	0	**0**
	80 – 90%	0	**1**
	90 – 100%	0	**1**
Rock	100%	>2000	>250
Sand	100%	-	>1000

Material	Pixel Coverage	Svalbard (n pixel)	Greenland (n pixel)
Snow	100%	>250	0
Vegetation	100%	>1000	>750
Water	100%	>8000	>8000
Wood	100%	>100	<50

Table 24   **Numbers of pixels used to extract spectral information of WV3 SWIR imagery**

Material	Pixel Coverage	Svalbard (n pixel)	Greenland (n pixel)
Plastic	0 – 1%	642	122
	1 – 5%	5	32
	5– 10%	0	4
	10 – 20%	0	2
	20 – 30%	0	1
	30 – 100%	0	0
Rock	100%	250	100
Sand	100%	0	150
Snow	100%	50	0
Vegetation	100%	150	100
Water	100%	1000	1000
Wood	Not defined	30	15

The extracted spectral information for Svalbard and Greenland is displayed as average values per extracted class in Figure 54 and Figure 55, respectively. As plastic pixels were mixed pixels, the spectral response was also displayed in terms of the plastic coverage (Figure 56 and Figure 57). The reflection values and shape differed between the two study sites, because distinct methods for atmospheric correction were applied. Comparing the average spectral shape of all plastic pixels with the other classes, a strong similarity of plastic pixels to wood and vegetation pixels can be observed. As the plastic pixels had a relatively small plastic coverage, the other classes, mainly vegetation and wood, dominated the plastic pixels signal. This can be confirmed by Figure 56 and Figure 57, showing the different plastic coverage per pixels. The lower the plastic coverage was, the more similar the signal was to other classes but even within the highest plastic coverage the signal seemed to have a strong influence of vegetation. Anyhow, looking into the high plastic coverage, SWIR values are not always available, because the SWIR imagery was coarser than the multispectral WV3 product, and the plastic proportion was therefore too small.

Moreover, the variation within the classes is important to consider, to determine whether different classes can be separated spectrally. An overview is given in the appendix in H.1 and H.2. Both figures show a huge variance of reflectance values for the classes *sand*, *rock,* and *plastic* and *plastic, rock and wood* for Greenland and Svalbard, respectively. As the plastic coverage of pixels was very low, a comparison to the spectral pattern of plastic with its typical spectral features as described in Garaba et al. (2020) and shown in Figure 8 was not possible. A detection of beach litter at a sub-pixel scale was tested by spectral unmixing of WV3 imagery. The average spectral signal for each class was calculated to perform a spectral unmixing using the Rstoolbox library within R version 3.6.3. Previous studies showed the ability to detect floating marine litter on a sub pixel scale using Sentinel 2 satellite imagery with a plastic coverage of 30 – 55% (Biermann et al., 2020; Topouzelis et al., 2019). Sentinel 2 has a similar spectral coverage as WV3 imagery with 12 bands from visible to SWIR, but a resolution of 10 - 60 m. A sub-pixel detection of plastics over open water for objects even smaller, covering down to 5% of a pixel, was described by Garaba et al. (2018b) using hyperspectral imagery. In this study, the performed spectral unmixing of the reference data from manual drone screening was not able to detect the plastic coverage of the reference data. In contrary to floating marine litter, the background of beach environments is more complex (e.g. vegetation, sand, rocks, wood) and the litter size distribution on the study area was very low. When including SWIR imagery, the sub-pixel size of 30 – 55% as described in Biermann et al. (2020) and Topouzelis et al. (2019) was not achieved for any pixel. Compared to Garaba et al. (2018b), the spectral coverage seemed to be insufficient for a sub-pixel detection down to 5%, also partly influenced by the complex background signals.

The results of the WV3 imagery showed that beach litter detection with satellite imagery is still limited by its spatial resolution. Only large accumulations of litter seem to be detectable (Acuña-Ruz et al., 2018). The litter size and the spatial distribution of litter on Arctic beaches were too low to be identified on WV3 imagery. A higher spatial resolution or a larger spectral coverage would be required to also detect lower litter accumulations on beaches from satellites. In future studies, spectral unmixing approaches for beach litter detection should be further tested for survey areas with very high litter abundances to determine the coverage of plastics that is required for a successful detection. The results of such an evaluation can then further be used to determine the resolution of satellite imagery needed to detect litter accumulations as found on Greenland and Svalbard. Another promising approach would be the evaluation of accumulations of driftwood as a proxy for high abundances of marine litter, as driftwood was detectable at WV3 imagery. Beside the spatial resolution, the costs of commercial satellites are still limiting a large-scale application (WV3: around $ 3500 for 100 km^2). To benefit from the large area coverage of satellite imagery, the detectability of smaller accumulations of beach litter still needs to be made feasible and the costs of the data acquisition must be reduced.

Figure 54   Average spectral signal extracted from WV3 imagery on Svalbard
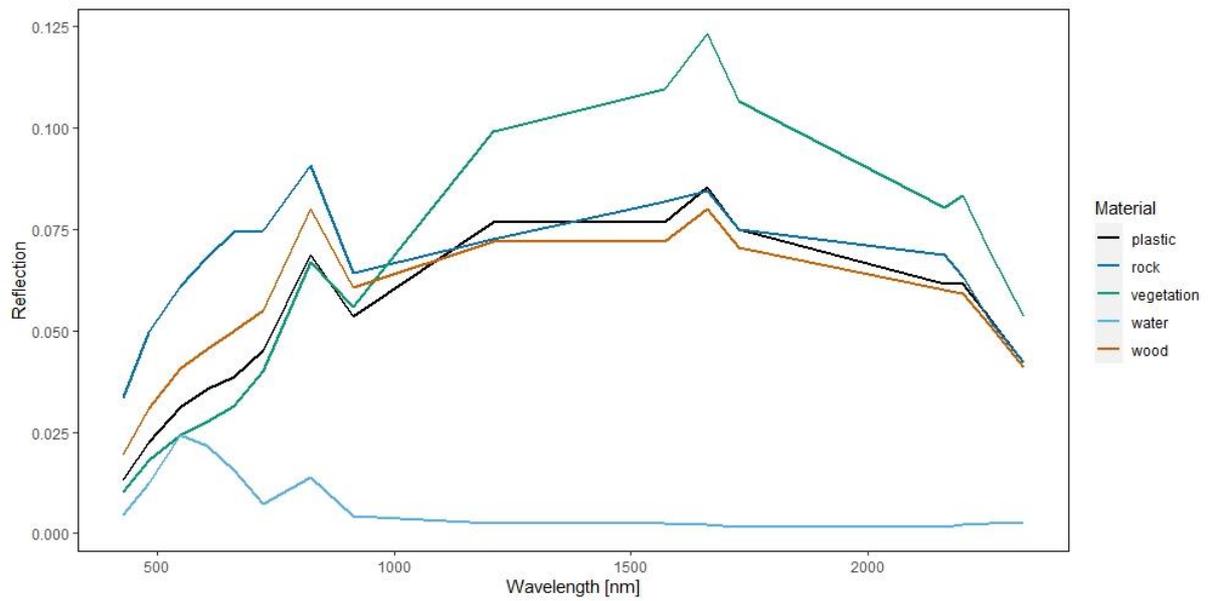
Source: Compiled by BCSH

Figure 55   Average spectral signal extracted from WV3 imagery on Greenland
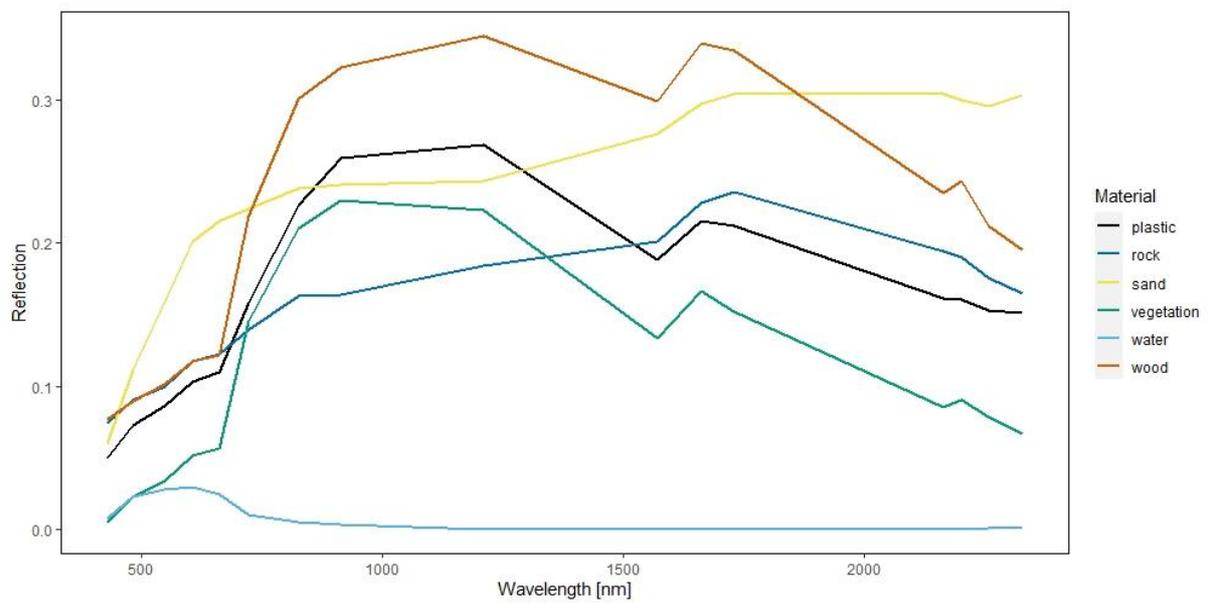
Source: Compiled by BCSH

Figure 56   Average spectral signal extracted from WV3 imagery on Svalbard for different plastic coverages
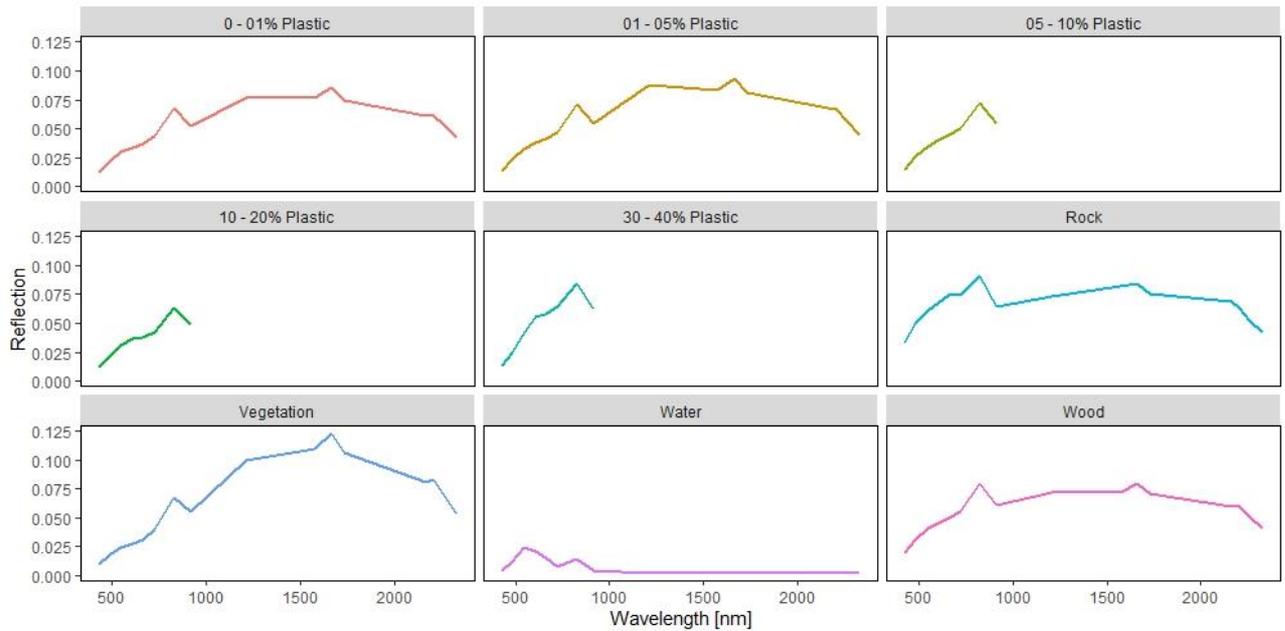
SWIR reflectance values were not always available due to the coarser resolution. Source: Compiled by BCSH

Figure 57   Average spectral signal extracted from WV3 imagery on Greenland for different plastic coverages
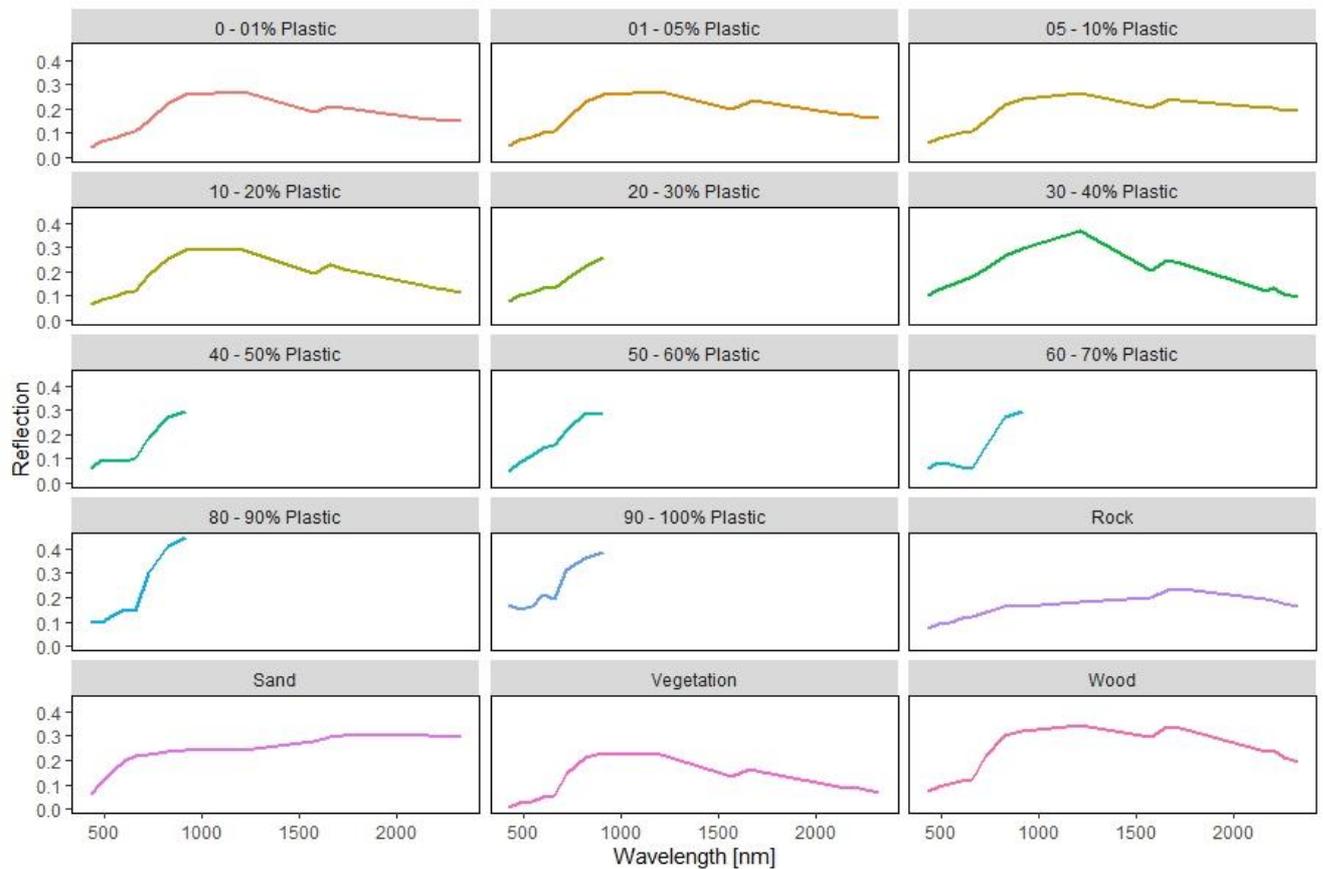
SWIR reflectance values were not always available due to the coarser resolution. Source: Compiled by BCSH

5   Work package 4

M. Schnurawa (BCSH) has participated in the Arctic Plastic Symposium in Reykjavik, Iceland (02 to 04 March2021). He presented a scientific talk, which has been made permanently available for download by the organizer of the symposium.

M. Schulz (AE) has participated in the Arctic Plastic Symposium in Reykjavik, Iceland (02 to 04 March 2021).

M. Schnurawa (BCSH) participated in the Nationales Forum für Fernerkundung und Copernicus 2022, Berlin (21 to 23 June 2022) He presented a scientific talk about the results of AP3.

6   List of references

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

Alkalay R., Pasternak G., Zask A., (2007): Clean-coast index—A new approach for beach cleanliness assessment. Ocean & Coastal Management 50, 352–362.

Andriolo, U., Gonçalves, G., Bessa, F., & Sobral, P. (2020). Mapping marine litter on coastal dunes with unmanned aerial systems: A showcase on the Atlantic Coast. Science of the Total Environment, 736, 139632.

Bao, Z., Sha, J., Li, X., Hanchiso, T., & Shifaw, E. (2018): Monitoring of beach litter by automatic interpretation of unmanned aerial vehicle images using the segmentation threshold method. Marine pollution bulletin, 137, 388-398.

Barnes, D. K. A., Fraser, K. P. P., (2003): Rafting by five phyla on man-made flotsam in the Southern Ocean. Marine Ecology Progress Series 262, 289–291.

Barnes, D. K. A., Milner, P., (2005): Drifting plastic and its consequences for sessile organism dispersal in the Atlantic Ocean. Marine Biology 146, 815–825.

Bateman, B.L., VanDerWal, J., Williams, S.E. & Johnson, C.N. (2010): Inclusion of biotic interactions in species distribution models improves predictions under climate change: the northern bettong Bettongia tropica, its food resources and a competitor. Journal of Biogeography, In Review.

Bergmann M., Lutz B., Tekman M. B., Gutow L. (2017): Citizen scientists reveal: Marine litter pollutes Arctic beaches and affects wild life. Marine Pollution Bulletin 125, 535–540.

Beyer, F., & Grenzdörfer, G. (2018): Klassifikation von Vegetationstypen auf Moorstandorten unter Verwendung von multisensoralen Drohnendaten.

Biermann, L., Clewley, D., Martinez-Vicente, V., & Topouzelis, K. (2020). finding plastic patches in coastal Waters using optical Satellite Data. Scientific reports, 10(1), 1-10.

Bravo M., de los Ángeles Gallardo M., Luna-Jorquera G., Núñez P., Vásquez N., Thiel M., (2009): Anthropogenic debris on beaches in the SE Pacific (Chile): Results from a national survey supported by volunteers. Marine Pollution Bulletin 58, 1718–1726.

B Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Breiman, Leo, et al. Classification and regression trees. Routledge, 2017.

Browne, M. A., Ayake, A., Galloway, T. S., Lowe, D. M., Thompson, R. C., (2008): Ingested Microscopic Plastic Translocates to the Circulatory System of the Mussel, Mytilus edulis (L.). Environmental Science & Technology 42, 5026–5031.

Browne, M. A., Galloway, T. S., Thompson, R. C., (2010): Spatial Patterns of Plastic Debris along Estuarine Shorelines. Environmental Science & Technology 44, 3404–3409.

Bullimore, B. A., Newman, P. B., Kaiser, M. J., Gilbert, S. E., Lock, K. M., (2001): A study of catches in a fleet of "ghost-fishing" pots. Fishery Bulletin 99, 247–253.

Buhl-Mortensen L., Buhl-Mortensen P., (2017): Marine litter in the Nordic Seas: Distribution composition and abundance. Marine Pollution Bulletin 125, 260–270.

Campbell, J. B., & Wynne, R. H. (2011): Introduction to remote sensing. Guilford Press, 31-37.

Cheshire A., Adler E., Barbière J., Cohen Y., Evans S., Jarayabhand S., Jeftic L., Jung R.-T., Kinsey S., Kusui, E. T., Lavine I., Manyara P., Oosterbaan L., Pereira M. A., Sheavly S., Tkalin A., Varadarajan S., Wenneker B.,

Westphalen G., (2009): UNEP/IOC Guidelines on Survey and Monitoring of Marine Litter. Regional Seas Reports and Studies No. 186. IOC Technical Series No. 83: 96 pp.

Coulter, D., Hauff, P. L., & Kerby, W. L. (2007): Airborne hyperspectral remote sensing. In Proceedings of the 5th Decennial International Conference on Mineral Exploration (pp. 375-378). Toronto, Ontario, Canada.

Dagestad, K. F., Röhrs, J., Breivik, Ø., & Ådlandsvik, B. (2018): OpenDrift v1. 0: a generic framework for trajectory modelling.

Dalal, N., & Triggs, B. (2005): Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

de Grandpré, A., Kinnard, C., & Bertolo, A. (2022): Open-Source Analysis of Submerged Aquatic Vegetation Cover in Complex Waters Using High-Resolution Satellite Remote Sensing: An Adaptable Framework. Remote Sensing, 14(2), 267.Erdle, K., Mistele, B., & Schmidhalter, U. (2011): Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars. Field Crops Research, 124(1), 74-84.

Eisted R., Christensen T. H. (2011): Waste management in Greenland: current situation and challenges. Waste Management & Research, 29(10), 1064–1070.

Escobar-Sánchez, G., Haseler, M., Oppelt, N., & Schernewski, G. (2021). Efficiency of aerial drones for macrolitter monitoring on Baltic Sea beaches. Frontiers in Environmental Science 8, 283.

Fallati, L., Polidori, A., Salvatore, C., Saponari, L., Savini, A., & Galli, P. 5.1 Anthropogenic Marine Debris assessment with Unmanned Aerial Vehicle imagery and Deep Learning: a case study along the beaches of the Republic of Maldives. New insights into coral reef threats and restoration perception: a case study in the Republic of Maldives, 124.

Fowler, C. W., (1987): Marine Debris and Northern Fur Seals: a Case Study. Marine Pollution Bulletin 18, 326-335.

Galgani, F., Leaute, J. P., Moguedet, P., Souplet, A., Verin, Y., Carpentier, A., Goraguer, H., Latrouite, D., Andral, B., Cadiou, Y., Mahe, J. C., Poulard, J. C., Nerisson, P., (2000): Litter on the Sea Floor along European Coasts. Marine Pollution Bulletin 40, 516-527.

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.

Garaba, S. P., Aitken, J., Slat, B., Dierssen, H. M., Lebreton, L., Zielinski, O., & Reisser, J. (2018): Sensing ocean plastics with an airborne hyperspectral shortwave infrared imager. Environmental science & technology, 52(20), 11699-11707.

Garaba, S. P., & Dierssen, H. M. (2020): Hyperspectral ultraviolet to shortwave infrared characteristics of marine-harvested, washed-ashore and virgin plastics. Earth System Science Data, 12(1), 77-86.

Garcia-Garin, O., Aguilar, A., Borrell, A., Gozalbes, P., Lobo, A., Penadés-Suay, J., Rage, J. A., Revuelta, O., Serrano, M., Vighi, M. (2020): Who's better at spotting? A comparison between aerial photography and observer-based methods to monitor floating marine litter and marine mega-fauna. Environmental Pollution, 258, 113680.

Gasperi J., Dris R., Bonin T., Rocher V., Tassin B. (2014): Assessment of floating plastic debris in surface water along the Seine River. Environmental Pollution 195, 163-166.

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

Gonçalves, G., Andriolo, U., Pinto, L., & Duarte, D. (2020b). Mapping marine litter with Unmanned Aerial Systems: A showcase comparison among manual image screening and machine learning techniques. Marine pollution bulletin 155, 111158.

Gonçalves, G., Andriolo, U., Gonçalves, L., Sobral, P., & Bessa, F. (2020c). Quantifying marine macro litter abundance on a sandy beach using unmanned aerial systems and object-oriented machine learning methods. Remote Sensing, 12(16), 2599

Gregory, M. R., (2009): Environmental implications of plastic debris in marine settings—entanglement, ingestion, smothering, hangers-on, hitch-hiking and alien invasions. Philosophical Transactions of the Royal Society B 364, 2013–2025.

Howell, E. A., Bograd, S. J., Morishige, C., Seki, M. P., Polovina, J. J., (2012): On North Pacific circulation and associated marine debris concentration. Marine Pollution Bulletin 65, 16–22.

Januchowski, S. R., Pressey, R. L., VanDerWal, J., & Edwards, A. (2010): Characterizing errors in digital elevation models and estimating the financial costs of accuracy. International Journal of Geographical Information Science, 24(9), 1327-1347.

Kirkfeldt T. S., (2016): Marine Litter in Greenland. Aalborg University, Denmark, 60 pp.

Majer, A.P., Vedolin, M.C., Turra, A., (2012) Plastic pellets as oviposition site and means of dispersal for the ocean-skater insect Halobates. Marine Pollution Bulletin 64, 1143–1147.

Mallory M. L., Baak J., Gjerdrum K., Mallory O. E., Manley B., Swan C., Provencher J. F. (2021): Anthropogenic litter in marine waters and coastlines of Arctic Canada and West Greenland. Science of the Total Environment 783, 146971.

Martin, C., Parkes, S., Zhang, Q., Zhang, X., McCabe, M. F., & Duarte, C. M. (2018): Use of unmanned aerial vehicles for efficient beach litter monitoring. Marine pollution bulletin, 131, 662-673.

Martin, C., Zhang, Q., Zhai, D., Zhang, X., & Duarte, C. M. (2021). Enabling a large-scale assessment of litter along Saudi Arabian red sea shores by combining drones and machine learning. Environmental Pollution, 277, 116730.

Mather, P., & Tso, B. (2016). Classification methods for remotely sensed data. CRC press.

Matsuoka, T., Nakashima, T., Nagasawa, N., (2005): A review of ghost fishing: scientific approaches to evaluation and solutions. Fisheries Science 71, 691–702.

Maximenko, N., et al., (2019): Towards the integrated marine debris observing system. Frontiers in marine science, 6, 447.

Merlino, S., Paterni, M., Berton, A., & Massetti, L. (2020). Unmanned aerial vehicles for debris survey in coastal areas: Long-term monitoring programme to study spatial and temporal accumulation of the dynamics of beached marine litter. Remote Sensing 12(8), 1260.

Micasense, Inc (2020): Altum Specifications. https://micasense.com/wp-content/uploads/2019/11/Altum-Specifications-1.pdf

Mingers and J. An empirical comparison of pruning methods for decision tree induction. Machine learning, 4:227–243, 1989

Morritt D., Stefanoudis P. V., Pearce D., Crimmen O. A., Clark P. F. (2014): Plastic in the Thames: A river runs through it. Marine Pollution Bulletin 78, 196–200.

Moy, K., Neilson, B., Chung, A., Meadows, A., Castrence, M., Ambagis, S., Davidson, K. (2018): Mapping coastal marine debris using aerial imagery and spatial analysis. Marine pollution bulletin, 132, 52-59.

Nashoug B. F. (2016): Sources of Marine Litter" –Workshop Report, Svalbard 4th -6th September 2016: 23 pp.

Neumann D., Callies U., Matthies M., (2014): Marine litter ensemble transport simulations in the southern North Sea. Marine Pollution Bulletin 86 (1-2), 219-228.

Opfer S., Arthur C., Lippiatt S., (2012): NOAA Marine Debris Shoreline Survey Field Guide, 14 pp.

OSPAR, (2010): Guideline for Monitoring Marine Litter on the Beaches in the OSPAR Maritime Area. OSPAR Commission, ISBN 90-3631-973.

PAME, (2019): Desktop Study on Marine Litter including Micro-plastics in the Arctic https://www.pame.is/images/03_Projects/Arctic_Marine_Pollution/Litter/Desktop_study/Desktop_Study_on_marine_litter.pdf

Papakonstantinou, A., Batsaris, M., Spondylidis, S., & Topouzelis, K. (2021). A citizen science unmanned aerial system data acquisition protocol and deep learning techniques for the automatic detection and mapping of marine litter concentrations in the coastal zone. Drones, 5(1), 6.

Pettorelli, N. (2019). Satellite remote sensing and the management of natural resources. Oxford University Press, 4-9.

Pichel, W. G., Veenstra, T. S., Churnside, J. H., Arabini, E., Friedman, K. S., Foley, D. G., Brainard, R. E, Kiefer, D., Ogle, S., Clemente-Colón, P., Li, X. (2012): GhostNet marine debris survey in the Gulf of Alaska–Satellite guidance and aircraft observations. Marine pollution bulletin, 65(1-3), 28-41.

Pinto, L., Andriolo, U., & Gonçalves, G. (2021). Detecting stranded macro-litter categories on drone orthophoto by a multi-class Neural Network. Marine Pollution Bulletin, 169, 112594.

Pourazar, H., Samadzadegan, F., & Dadrass Javan, F. (2019). Aerial multispectral imagery for plant disease detection: Radiometric c Rech S., Macaya-Caquilpán V., Pantoja J. F., Rivadeneira M. M., Jofre Madariaga D., Thiel M. (2014): Rivers as a source of marine litter – A study from the SE Pacific. Marine Pollution Bulletin 82, 66–75.

Quinlan, R. C. (1993). 4.5: Programs for machine learning morgan kaufmann publishers inc. San Francisco, USA.

Ribic, C. A., Sheavly, S. B., Rugg, D. J., Erdmann, E. S., (2010): Trends and drivers of marine debris on the Atlantic coast of the United States 1997–2007. Marine Pollution Bulletin 60, 1231–1242.

Richards, J. A., & Richards, J. A. (1999). Remote sensing digital image analysis (Vol. 3, pp. 10-38). Berlin et al.: Springer., 1-20.

Schulz M., Clemens T., Förster H., Harder T., Fleet D. M., Gaus S., Grave C., Flegel I., Schrey E., Hartwig E., (2015): Statistical analyses of the results of 25 years of beach litter surveys on the south-eastern North Sea coast. Marine Environmental Research 109: 21-27.

Schulz M., Krone R., Dederer G., Wätjen K., Matthies M., (2015): Comparative analysis of time series of marine litter surveyed on beaches and the seafloor in the southeastern North Sea. Marine Environmental Research 106, 61-67.

Schulz M., Neumann D., Fleet D. M., Matthies M., (2013): A multi-criteria evaluation system for marine litter pollution based on statistical analyses of OSPAR beach litter monitoring time series. Marine Environmental Research, doi.10.1016/j.marenvres.2013.08.013.

Schulz M., van Loon W., Fleet D., Baggelaar P., van der Meulen E., (2017): OSPAR standard method and software for statistical analysis of beach litter data. Marine Pollution Bulletin 122, 166-175.

Schulz M., Walvoort D. J. J., Barry J., Fleet D. M., van Loon W. M. G. M., (2019): Baseline and power analyses for the assessment of beach litter reductions in the European OSPAR region. Environmental Pollution 248, 555-564.

Shaw, G. A., & Burke, H. K. (2003): Spectral imaging for remote sensing. Lincoln laboratory journal, 14(1), 3-28.

Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., ... & Woodcock, C. E. (2006). Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. European Communities, Luxembourg, 51(4), 1-60.

Strand K. O., Huserbråten M., Dagestad K.-F., Mauritzen C., Grøsvik B. E., Nogueira L. A., Melsom A., Röhrs J. (2021): Potential sources of marine plastic from survey beaches in the Arctic and Northeast Atlantic. Science of the Total Environment 790, 148009.

Strietman W.J., van den Heuvel-Greve M.J., van den Brink A.M., Leemans E., Strand J., Bach L. (2021): Beach litter in West Greenland: a source analysis. Report, 64 pp.

Tekman M. B., Krumpen T., Bergmann M. (2017): Marine litter on deep Arctic seafloor continues to increase and spreads to the North at the HAUSGARTEN observatory. Deep-Sea Research Part I 120, 88–99.

Tekman, M.B., Gutow, L., Macario, A., Haas, A., Walter, A., Bergmann, M.: Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research. https://litterbase.awi.de/litter

Topouzelis, K., Papakonstantinou, A., & Garaba, S. P. (2019): Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018). International Journal of Applied Earth Observation and Geoinformation, 79, 175-183.

Tudor D. T., Williams A. T. (2004): Development of a 'Matrix Scoring Technique' to determine litter sources at a Bristol Channel beach. Journal of Coastal Conservation 9, 119-127.

van Franeker, J. A., Blaize, C., Danielsen, J., Fairclough, K., Gollan, J., Guse, N., Hansen, P.-L., Heubeck, M., Jensen, J.-K., Le Guillou, G., Olsen, B., Olsen, K.-O., Pedersen, J., Stienen, E. W. M., Turner, D. M., (2011): Monitoring plastic ingestion by the northern fulmar Fulmarus glacialis in the North Sea. Environmental Pollution 159, 2609-2615.

Vesman A., Moulin E., Egorova A., Zaikov K. (2020): Marine litter pollution on the Northern Island of the Novaya Zemlya archipelago. Marine Pollution Bulletin 150, 110671.

Votier, S. C., Archibald, K., Morgan, G., Morgan, L., (2011): The use of plastic debris as nesting material by a colonial seabird and associated entanglement mortality. Marine Pollution Bulletin 62, 168–172.

Waske, B., & Benediktsson, J. A. (2007). Fusion of support vector machines for classification of multisensor data. IEEE Transactions on geoscience and remote sensing, 45(12), 3858-3866.

Waske, B., & Braun, M. (2009). Classifier ensembles for land cover mapping using multitemporal SAR imagery. ISPRS journal of photogrammetry and remote sensing, 64(5), 450-457.

Weslawski J. M., Kotwicki L., (2018): Macro-plastic, a new vector for boreal species dispersal on Svalbard. Polish Polar Research 39, 165–174.

Wolf, M., van den Berg, K., Garaba, S. P., Gnann, N., Sattler, K., Stahl, F., & Zielinski, O. (2020). Machine learning for aquatic plastic litter detection, classification and quantification (APLASTIC-Q). Environmental Research Letters, 15(11), 114042.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. Methods in ecology and evolution, 1(1), 3-14.

A   Appendix

A.1   Coordinates of OSPAR beach litter survey sites on Greenland, Svalbard, and Iceland.

Beach name	ID	Longitude	Latitude
Dagmar Island North St1	GRL001	-17.5637	81.6864
Henryland - East Greenland	GRL002	-23.6400	69.6079
Sand Island, Young Sund	GRL003	-20.1526	74.2653
Kap Mary, Young Sund	GRL004	-20.1884	74.1656
Clavering Island, Young Sund	GRL005	-20.1714	74.2181
Stroem Island	GRL006	-30.6166	68.0859
Moraene Island, Taasiilaq	GRL007	-35.7887	66.0208
Nuuk Fjord St1	GRL008	-51.7491	64.0257
Nuuk Fjord St2	GRL009	-51.7303	64.0614
Nuuk Fjord St3	GRL010	-51.7614	64.0719
Disko Brededal	GRL011	-53.1853	69.3172
Disko Ippik	GRL012	-53.2417	69.3019
Sisimiut St1	GRL013	-53.6022	66.8711
Sisimiut St2	GRL014	-53.4946	66.8533
Qaqortoq St1 (Akia)	GRL015	-46.1317	60.6753
Qaqortoq St2	GRL016	-46.0990	60.7215
Upernavik	GRL017	-56.5772	73.6703
Brucebukta	NO002	78.4448	11.8602
Luftskipodden	NO003	79.6812	10.7670
Raudasandur	IS001	65.4628	-23.9555
Budavik	IS002	64.8276	-23.3572

Bakkavik	IS003	64.1539	-22.0072
Surtsey island East	IS004	63.3058	-20.5962
Surtsey island West	IS005	63.3084	-20.6003
Rekavik bak Hofn	IS006	66.4312	-22.5162
Vikur	IS007	66.1012	-20.2333

B Apppendix

B.1 Cities and settlements on Greenland
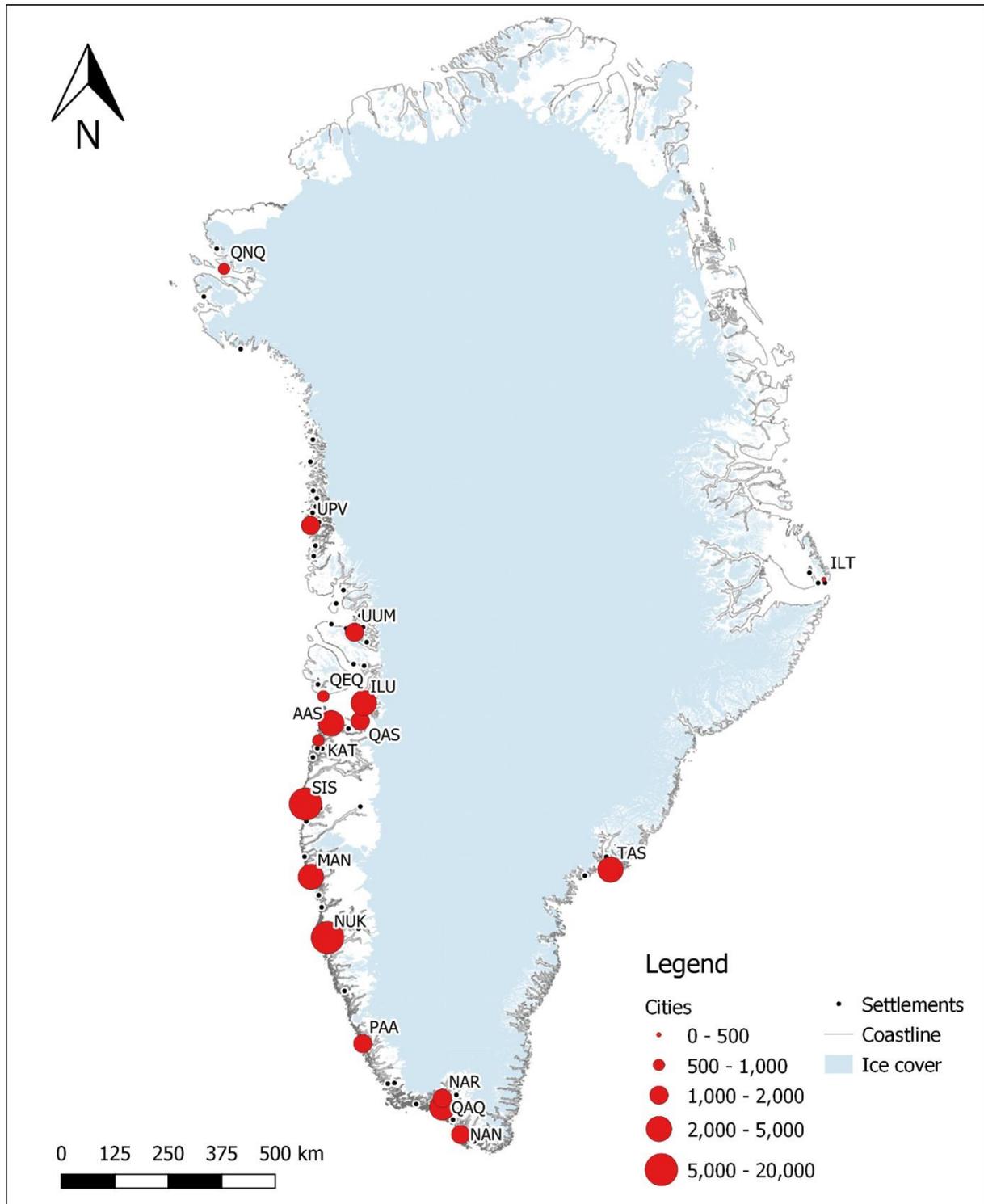
Number of inhabitants of cities on Greenland (https://stat.gl/):

AAS: Aasiaat, ILU: Ilulissat, ILT Ittoqqortoormiit, KAT: Kangaatsiaq, MAN: Maniitsoq, NAN: Nanortalik, NAR: Narsq,

NUK: Nuuk, PAA: Paamiut, QAQ: Qaqortoq, QAS: Qasigiannguit, QEG: Qeqertarsuaq, QNQ: Qaanaaq, SIS: Sisimiut,

TAS: Tasiilaq, UPV: Upernavikm, UUM: Uummannaq

Source: Compiled by BCSH

B.2 Landfills and mining activities on Greenland
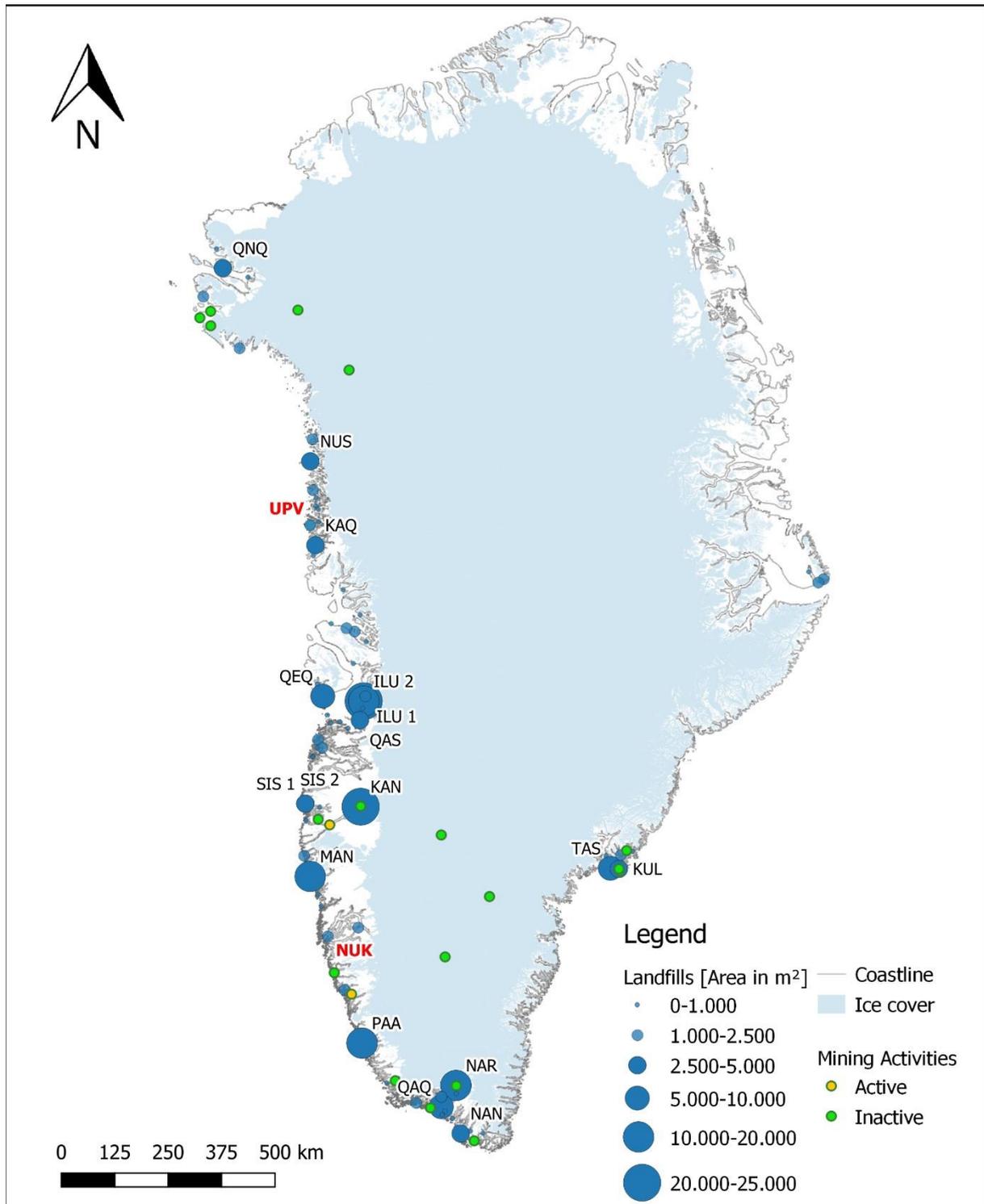
Landfills with an area coverage above 2.500 m² are indicated by name (Correspondence WSP Arctic):

ILU: Ilulissat, KAN: Kangerlussuaq, KAQ: Kangersuatsiaq, KUL: Kulusuk,, MAN: Maniitsoq, NAN: Nanortalik, NAR: Narsq, NUS: Nuussuaq, PAA: Paamiut, QAQ: Qaqortoq, QAS: Qasigiannguit, QEG: Qeqertarsuaq, QNQ: Qaanaaq, SIS: Sisimiut, TAS: Tasiilaq;

Landfills in NUK: Nuuk, and UPV: Upernavik have smaller area coverage; Active mining activities are open pit mining.

Source: Compiled by BCSH

C Appendix

C.1 Daily current Intensity from the Copernicus data portal

Daily current intensity from the Copernicus data portal used for the low-resolution model at 4 different time steps.
Source: Compiled by BCSH

C.2 Daily current Intensity from the Norwegian Polar Institute
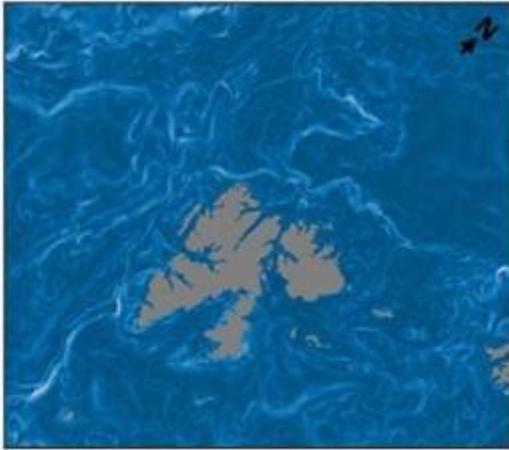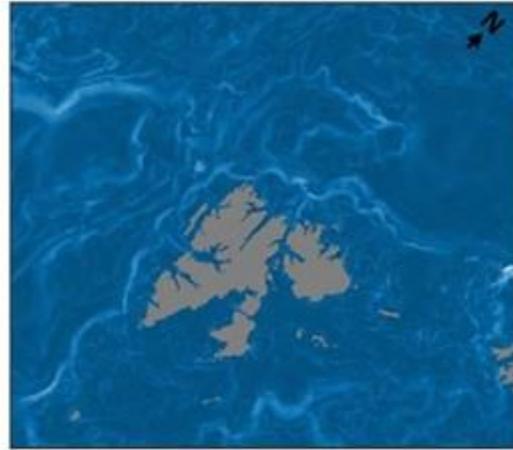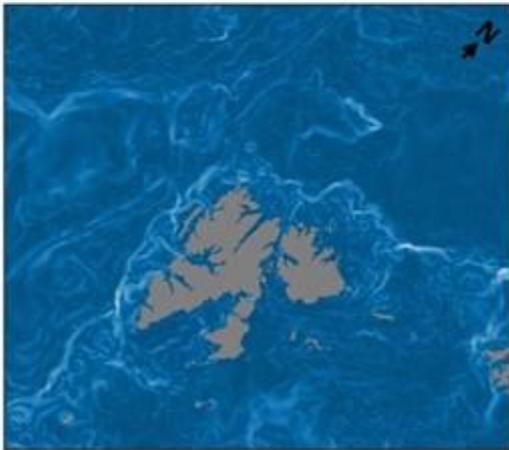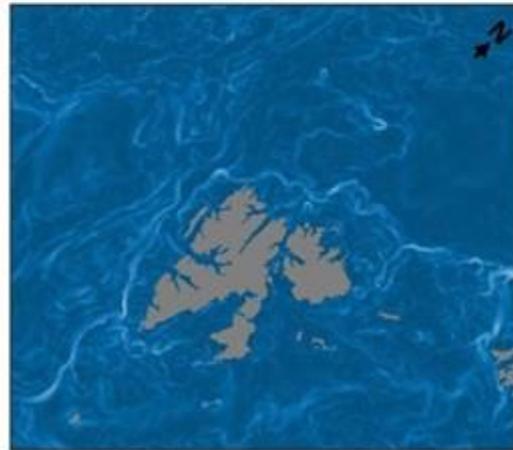
Daily current Intensity from the Norwegian Polar Institute data portal used for the High-resolution model (API) at 4 different time steps. Please note different Spatial projection (south pointing to the lower-left corner). Source: Compiled by BCSH

D.1 Photos of the litter objects measured with a GNSS receiver in the survey area Sisimiut Day 4 taken with a digital camera

In red objects which were not found during the manual screening of the drone imagery. Source: Compiled by BCSH

In red objects which were not found during the manual screening of the drone imagery. Source: Compiled by BCSH

In red objects which were not found during the manual screening of the drone imagery. Source: Compiled by BCSH

In red objects which were not found during the manual screening of the drone imagery. Source: Compiled by BCSH

D.2 Photos of the litter objects measured with a GNSS receiver in the survey area Sisimiut Day 5 taken with a digital camera

In red objects which were not found during the manual screening of the drone imagery. Source: Compiled by BCSH

In red objects which were not found during the manual screening of the drone imagery. Source: Compiled by BCSH

E   APPENDIX

E.1 Number of objects found for each detected category of materials

Material	glas	metal	plastic	wood
Number	1	4	357	165

E.2 Size distribution of all plastic litter objects that were smaller than 1 m²
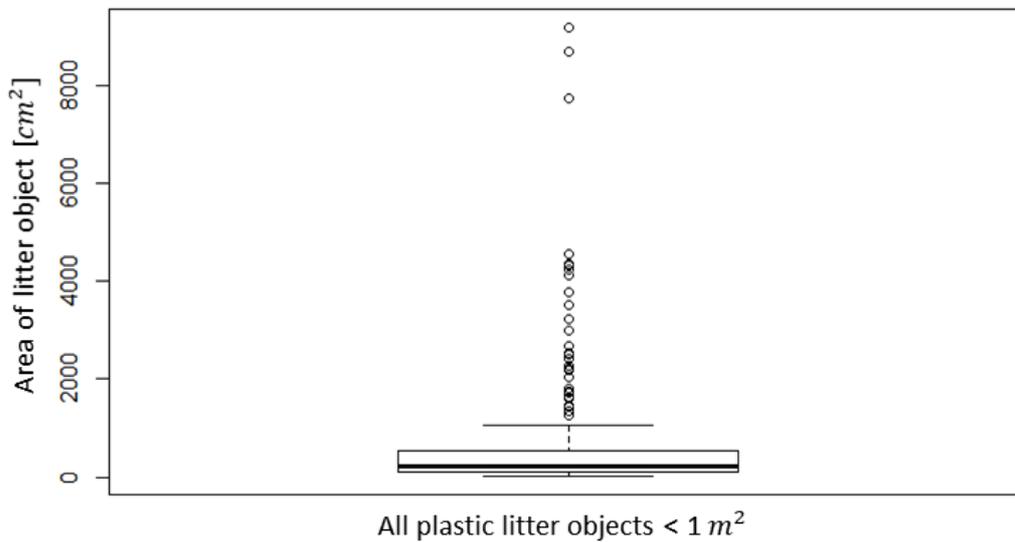
Source: Compiled by BCSH

E.3 Summary table of the model ranking by AIC for the effects of several parameters on the probability to identify the type of a plastic litter object. The model chosen for presentation is highlighted in bold letters.

Model	Variables	df	AIC
Id_01	Area + Colour + Area * Colour	8	383.47
Id_02	Area + Colour	5	393.72
Id_03	Area	2	392.02

E.4 Summary table of the model ranking by AIC for the effects of several parameters on the likelihood of a plastic litter object to be identified as such with certainty. The model chosen for presentation is highlighted in bold letters.

Model	Variables	df	AIC
Certain_01	Area + Colour + Area * Colour	8	348.51
Certain_02	Area + Colour	5	361.07
Certain_03	Area	2	379.76

E.5 Residual scatterplot of the final model (id_02) describing the probability to identify the type of a plastic litter object.
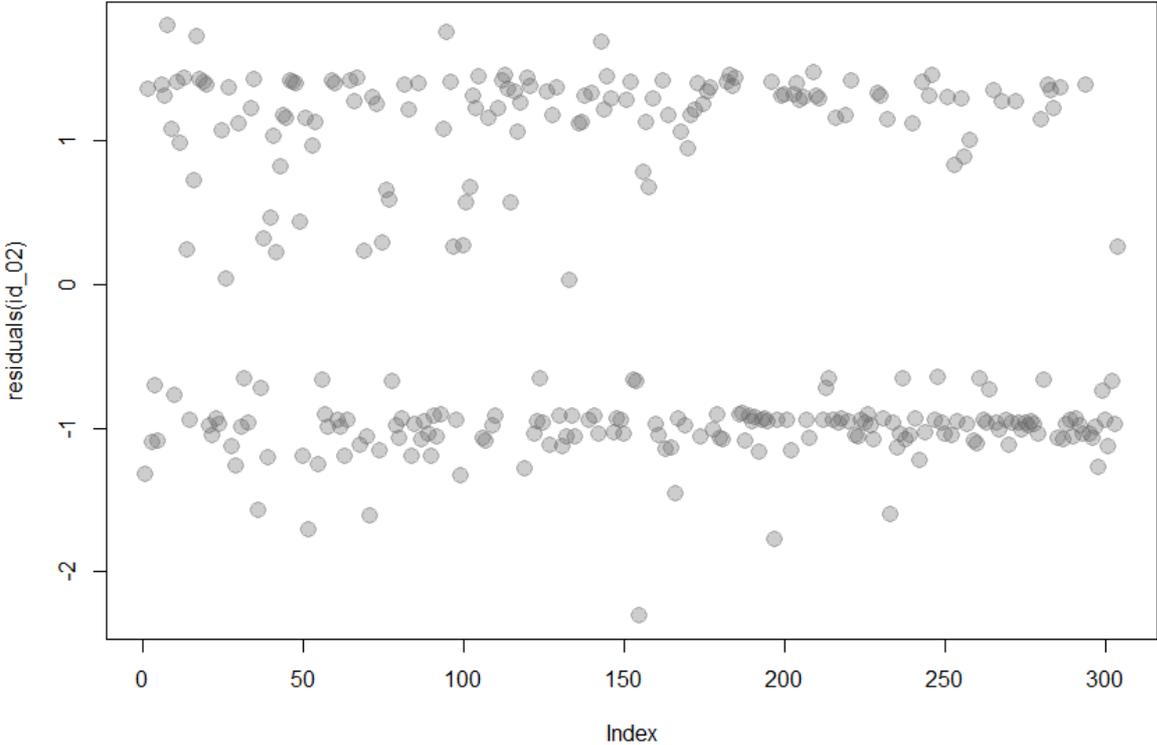
Source: Compiled by BCSH

E.6 Residual scatterplot of the final model (caertain_02) describing the likelihood of a plastic litter object to be identified as such with certainty
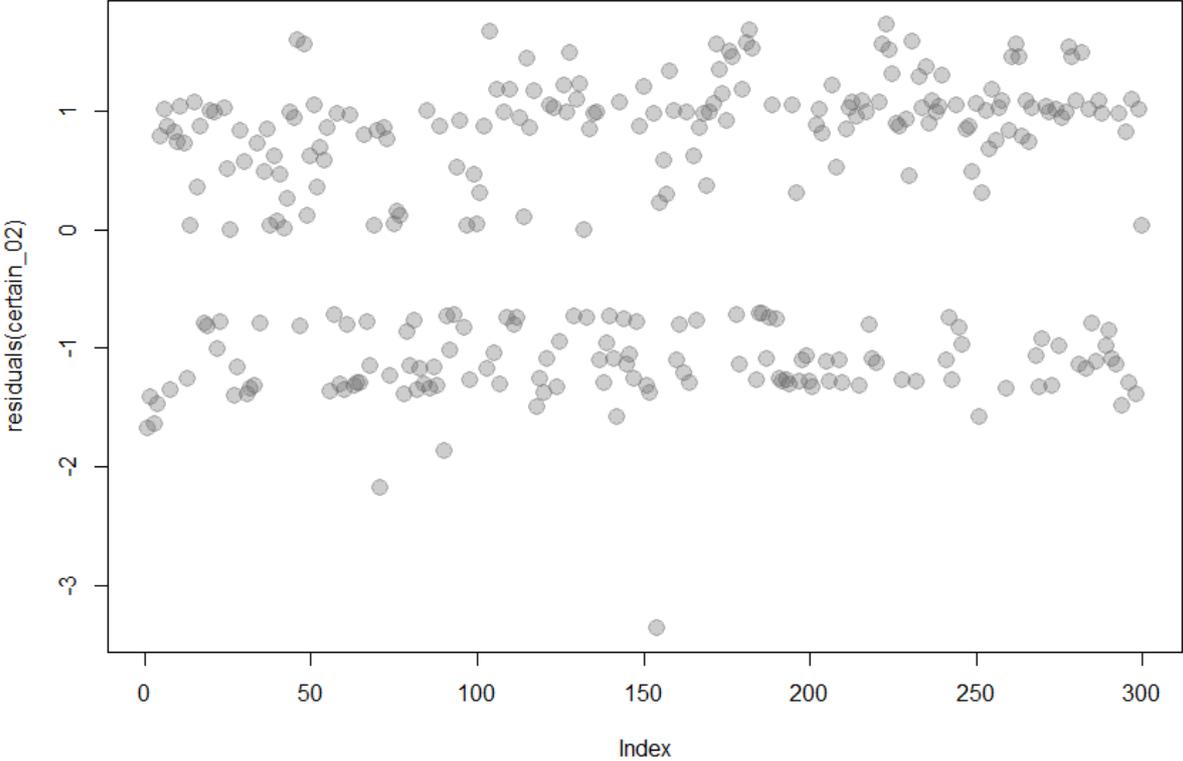
Source: Compiled by BCSH

F   APPENDIX - Confusion Matrices: (Semi-) Automatic Classification

F.1   Greenland Random Forest v1

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**416**	3	11	5	34	30	0	1	0	500	**83.2**
plastic	9	**252**	73	8	28	15	13	87	15	500	**50.4**
rock	19	46	**259**	7	42	24	2	95	6	500	**51.8**
sand	7	4	5	**470**	0	6	0	8	0	500	**94.0**
shadow	22	15	18	1	**365**	23	3	5	0	452	**80.8**
snow	38	6	22	26	14	**386**	0	12	0	504	**76.6**
vegetation	0	9	0	0	24	0	**488**	0	0	521	**93.7**
water	4	58	104	22	16	17	1	**267**	11	500	**53.4**
wood	0	6	0	0	0	0	0	4	**451**	461	**97.8**
sum	515	399	492	539	523	501	507	479	483		**OA: 75.6%**
UA [%]	**80.8**	**63.2**	**52.6**	**87.2**	**69.8**	**77.0**	**96.3**	**55.7**	**93.4**		

F.2 Greenland Support Vector Machine v1

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**432**	2	7	10	27	22	0	0	0	500	**86.4**
plastic	6	**263**	74	14	28	12	10	80	13	500	**52.6**
rock	18	33	**288**	13	40	20	2	78	8	500	**57.6**
sand	8	1	3	**477**	0	7	0	4	0	500	**95.4**
shadow	22	11	13	2	**378**	20	4	2	0	452	**83.6**
snow	40	4	33	25	15	**380**	2	5	0	504	**75.4**
vegetation	0	4	1	0	22	0	**494**	0	0	521	**94.8**
water	2	51	108	25	17	14	0	**270**	13	500	**54.0**
wood	0	5	0	0	0	0	0	6	**450**	461	**97.6**
sum	528	374	527	566	527	475	512	445	484		OA: 77.3%
UA [%]	**81.8**	**70.0**	**54.6**	**84.3**	**71.7**	**80.0**	**96.5**	**60.7**	**93.0**		

F.3 Greenland Neuronal Network v1a

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**433**	8	5	11	21	0	20	0	2	500	**86.6**
plastic	7	**264**	72	10	27	17	14	14	75	500	**52.8**
rock	15	42	**291**	7	35	5	27	2	76	500	**58.2**
sand	7	2	5	**474**	0	0	8	0	4	500	**94.8**
shadow	22	13	15	1	**370**	0	22	5	4	452	**81.9**
snow	0	8	0	0	0	**448**	0	0	5	461	**97.2**
vegetation	33	6	21	19	16	0	**399**	3	7	504	**79.2**
water	0	4	1	0	22	0	0	**494**	0	521	**94.8**
wood	3	46	101	18	12	14	17	0	**289**	500	**57.8**
sum	520	393	511	540	503	484	507	518	462		OA:78.0%
UA [%]	83.3	67.2	56.9	87.8	73.6	92.6	78.7	95.4	62.6		

F.4 Greenland Neuronal Network v1b

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**433**	8	7	9	21	0	21	0	1	500	**86.6**
plastic	7	**264**	73	10	26	18	12	13	77	500	**52.8**
rock	16	37	**284**	9	38	7	26	3	80	500	**56.8**
sand	7	1	7	**471**	0	0	10	0	4	500	**94.2**
shadow	23	16	21	1	**362**	0	19	6	4	452	**80.1**
snow	0	7	0	0	0	**448**	0	0	6	461	**97.2**
vegetation	36	3	22	17	17	0	**398**	3	8	504	**79.0**
water	0	6	1	0	22	0	0	**491**	1	521	**94.2**
wood	3	48	103	20	10	13	19	1	**283**	500	**56.6**
sum	525	390	518	537	496	486	505	517	464	OA: 77.3%	
UA [%]	**82.5**	**67.7**	**54.8**	**87.7**	**73.0**	**92.2**	**78.8**	**95.0**	**61.0**		

F.5 Greenland Neuronal Network v1c

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**371**	81	5	6	22	0	15	0	0	500	**74.2**
plastic	3	**463**	3	4	8	2	10	3	4	500	**92.6**
rock	14	390	**61**	5	11	0	12	2	5	500	**12.2**
sand	9	76	0	**408**	0	0	6	0	1	500	**81.6**
shadow	11	152	1	2	**275**	0	8	3	0	452	**60.8**
snow	0	89	0	0	0	**372**	0	0	0	461	**80.7**
vegetation	39	107	12	11	8	0	**324**	3	0	504	**0.0**
water	0	42	0	0	22	0	1	**456**	0	521	**0.2**
wood	1	422	8	11	3	2	8	0	**45**	500	**0.0**
sum	448	1822	90	447	349	376	384	467	55		OA: 62.5%
UA [%]	82.8	25.4	67.8	91.3	78.8	98.9	84.4	97.6	81.8		

F.6 Greenland Random Forest v2

	algae	plastic	rock	sand	snow	vegetation	water	wood	sum	PA [%]
algae	**479**	2	0	0	0	19	0	0	500	**95.8**
plastic	6	**339**	64	6	1	7	1	76	500	**67.8**
rock	0	27	**414**	18	1	3	3	34	500	**82.8**
sand	0	0	8	**491**	0	0	0	1	500	**98.2**
snow	0	1	6	0	**486**	0	0	7	500	**97.2**
vegetation	12	3	2	0	0	**482**	0	2	501	**96.2**
water	0	2	3	0	0	0	**495**	0	500	**99.0**
wood	3	35	45	5	1	12	0	**399**	500	**79.8**
sum	500	409	542	520	489	523	499	519		
UA [%]	95.8	82.9	76.4	94.4	99.4	92.2	99.2	76.9		**OA: 89.6%**

F.7 Greenland Support Vector Machine v2

	algae	plastic	rock	sand	snow	vegetation	water	wood	sum	PA [%]
algae	**489**	0	1	0	0	9	0	1	500	**97.8**
plastic	7	**320**	76	5	1	11	1	79	500	**64.0**
rock	0	29	**414**	26	1	3	3	24	500	**82.8**
sand	0	1	3	**496**	0	0	0	0	500	**99.2**
snow	0	1	8	0	**485**	0	0	6	500	**97.0**
vegetation	16	3	3	0	0	**475**	0	3	500	**95.0**
water	0	4	3	0	0	0	**493**	0	500	**98.6**
wood	3	34	38	3	1	11	1	**409**	500	**81.8**
sum	515	392	546	530	488	509	498	522		
UA [%]	**95.0**	**81.6**	**75.8**	**93.6**	**99.4**	**93.3**	**99.0**	**78.4**	**OA: 89.5%**	

F.8 Greenland Neuronal Network v2a

	algae	plastic	rock	sand	snow	vegetation	water	wood	sum	PA [%]
algae	**485**	1	0	0	0	13	0	1	500	**97.0**
plastic	6	**360**	56	2	2	8	1	65	500	**72.0**
rock	0	37	**413**	21	2	2	0	25	500	**82.6**
sand	0	1	6	**493**	0	0	0	0	500	**98.6**
snow	0	0	6	0	**486**	0	0	8	500	**97.2**
vegetation	10	3	1	0	0	**483**	0	4	501	**96.4**
water	0	2	3	0	0	0	**495**	0	500	**99.0**
wood	4	49	45	1	0	11	1	**389**	500	**77.8**
sum	505	453	530	517	490	517	497	492		
UA [%]	**96.0**	**79.5**	**77.9**	**95.4**	**99.2**	**93.4**	**99.6**	**79.1**	**OA: 90.1%**	

F.9 Greenland Neuronal Network v2b

	algae	plastic	rock	sand	snow	vegetation	water	wood	sum	PA [%]
algae	**483**	3	0	0	0	13	0	1	500	**96.6**
plastic	4	**418**	31	4	2	5	1	35	500	**83.6**
rock	0	105	**349**	22	0	1	1	22	500	**69.8**
sand	0	3	10	**487**	0	0	0	0	500	**97.4**
snow	0	7	10	0	**477**	0	0	6	500	**95.4**
vegetation	13	6	2	0	0	**480**	0	0	501	**95.8**
water	0	3	3	0	0	0	**494**	0	500	**98.8**
wood	3	124	25	1	1	9	1	**336**	500	**67.2**
sum	503	669	430	514	480	508	497	400		
UA [%]	96.0	62.5	81.2	94.7	99.4	94.5	99.4	84.0	**OA: 88.1%**	

F.10 Greenland Neuronal Network v2c

	algae	plastic	rock	sand	snow	vegetation	water	wood	sum	PA [%]
algae	**476**	12	0	0	0	12	0	0	500	**95.2**
plastic	4	**476**	5	4	0	3	0	8	500	**95.2**
rock	0	297	**172**	23	0	0	1	7	500	**34.4**
sand	0	17	8	**474**	0	0	0	1	500	**94.8**
snow	0	19	5	0	**474**	0	0	2	500	**94.8**
vegetation	13	36	0	0	0	**452**	0	0	501	**90.2**
water	0	8	0	0	0	0	**492**	0	500	**98.4**
wood	2	306	3	4	0	9	1	**175**	500	**35.0**
sum	495	1171	193	505	474	476	494	193		
UA [%]	**96.2**	**40.6**	**89.1**	**93.9**	**100.0**	**95.0**	**99.6**	**90.7**	**OA: 79.8%**	

F.11 Greenland Random Forest v3

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**486**	2	0	0	3	0	9	0	0	500	**97.2**
plastic	6	**362**	56	2	10	0	6	0	58	500	**72.4**
rock	0	27	**411**	13	22	2	2	1	22	500	**82.2**
sand	0	0	8	**491**	0	0	0	0	1	500	**98.2**
shadow	8	15	15	1	**459**	0	8	4	2	512	**89.6**
snow	0	2	0	0	0	**497**	0	0	1	500	**99.4**
vegetation	12	1	4	0	7	0	**475**	0	4	503	**99.4**
water	0	1	0	0	4	0	0	**495**	0	500	**99.0**
wood	1	34	34	1	5	0	12	0	**413**	500	**82.6**
sum	513	444	528	508	510	499	512	500	501		**OA:90.6%**
UA [%]	**94.7**	**81.5**	**77.8**	**96.7**	**90.0**	**99.6**	**92.8**	**99.0**	**82.4**		

F.12 Greenland Support Vector Machine v3

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**485**	2	0	0	1	0	11	0	1	500	**97.0**
plastic	6	**360**	64	3	9	0	5	0	53	500	**72.0**
rock	0	33	**403**	24	19	0	1	0	20	500	**80.6**
sand	0	0	6	**494**	0	0	0	0	0	500	**98.8**
shadow	7	12	8	1	**460**	0	9	3	0	500	**92.0**
snow	0	2	1	0	0	**495**	0	0	2	500	**99.0**
vegetation	11	5	4	0	13	0	**463**	0	4	500	**92.6**
water	0	1	0	0	2	0	0	**497**	0	500	**99.4**
wood	1	36	44	0	2	0	6	0	**411**	500	**82.2**
sum	510	451	530	522	506	495	495	500	491		
UA [%]	**95.1**	**79.8**	**76.0**	**94.6**	**90.9**	**100.0**	**93.5**	**99.4**	**83.7**	**OA: 90.4%**	

F.13 Greenland Neuronal Network v3a

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**484**	2	0	0	1	0	13	0	0	500	**96.8**
plastic	4	**358**	57	4	10	0	6	1	60	500	**71.6**
rock	0	40	**392**	20	22	1	1	0	24	500	**78.4**
sand	0	1	5	**494**	0	0	0	0	0	500	**98.8**
shadow	5	12	12	1	**465**	0	12	3	2	512	**90.8**
snow	0	0	0	0	0	**499**	0	0	1	500	**99.8**
vegetation	10	3	2	0	11	0	**470**	0	7	503	**93.4**
water	0	1	0	0	3	0	0	**496**	0	500	**99.2**
wood	1	52	35	0	2	0	6	0	**404**	500	**80.8**
sum	504	469	503	519	514	500	508	500	498		OA: 89.9%
UA [%]	**96.0**	**76.3**	**77.9**	**95.2**	**90.5**	**99.8**	**92.5**	**99.2**	**81.1**		

F.14 Greenland Neuronal Network v3b

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**483**	4	0	0	1	0	12	0	0	500	**96.6**
plastic	3	**448**	19	3	5	1	5	0	16	500	**89.6**
rock	0	195	**249**	17	24	0	2	0	13	500	**49.8**
sand	0	11	6	**482**	0	0	0	0	1	500	**96.4**
shadow	6	35	7	0	**450**	0	9	5	0	512	**87.9**
snow	0	2	0	0	0	**497**	0	0	1	500	**99.4**
vegetation	13	21	0	0	14	0	**451**	0	4	503	**89.7**
water	0	0	0	0	6	0	0	**494**	0	500	**98.8**
wood	2	205	21	0	3	0	6	0	**263**	500	**52.6**
sum	507	921	302	502	503	498	485	499	298		OA: 84.5%
UA [%]	**95.3**	**48.6**	**82.5**	**96.0**	**89.5**	**99.8**	**93.0**	**99.0**	**88.3**		

F.15 Greenland Neuronal Network v3c

	algae	plastic	rock	sand	shadow	snow	vegetation	water	wood	sum	PA [%]
algae	**485**	3	0	0	1	0	11	0	0	500	**97.0**
plastic	4	**420**	33	4	6	0	6	0	27	500	**84.0**
rock	0	101	**326**	20	32	1	1	1	18	500	**65.2**
sand	0	0	7	**493**	0	0	0	0	0	500	**98.6**
shadow	10	23	10	1	**455**	0	7	5	1	512	**88.9**
snow	0	2	0	0	0	**497**	0	0	1	500	**99.4**
vegetation	14	7	1	0	15	0	**459**	0	7	503	**91.3**
water	0	1	0	0	3	0	0	**496**	0	500	**99.2**
wood	1	116	30	0	3	0	12	0	**338**	500	**67.6**
sum	514	673	407	518	515	498	496	502	392		
UA [%]	**94.4**	**62.4**	**80.1**	**95.2**	**88.3**	**99.8**	**92.5**	**98.8**	**86.2**		**OA: 87.9%**

F.16 Svalbard Random Forest v1

	plastic	rock	vegetation	water	wood	sum	PA [%]
plastic	**96**	13	9	7	25	150	**64.0**
rock	8	**102**	5	0	35	150	**68.0**
vegetation	1	7	**139**	3	0	150	**92.7**
water	2	0	5	**143**	0	150	**95.3**
wood	19	14	7	0	**110**	150	**73.3**
sum	126	136	165	153	170		**OA: 78.6%**
UA [%]	76.2	75.0	84.2	93.5	64.7		

F.17 Svalbard Random Forest v2

	plastic	rock	vegetation	water	wood	sum	PA [%]
plastic	**95**	13	15	0	27	150	**63.3**
rock	15	**121**	0	0	14	150	**80.7**
vegetation	6	0	**142**	0	2	150	**94.7**
water	2	0	0	**148**	0	150	**98.7**
wood	19	7	1	0	**123**	150	**82.0**
sum	137	141	158	148	166		**OA: 85.9%**
UA [%]	69.3	85.8	89.9	100.0	74.1		

F.18 Svalbard Random Forest v3

|	plastic	rock	vegetation	water	wood	sum	PA [%]
plastic	**98**	7	15	0	30	150	**65.3**
rock	5	**132**	0	1	12	150	**88.0**
vegetation	7	0	**141**	0	2	150	**94.0**
water	0	0	0	**150**	0	150	**100.0**
wood	21	3	3	0	**123**	150	**82.0**
sum	131	142	159	151	167		**OA: 85.9%**
UA [%]	74.8	93.0	88.7	99.3	73.7		

G APPENDIX - Spectral signal of drone reference data on Svalbard

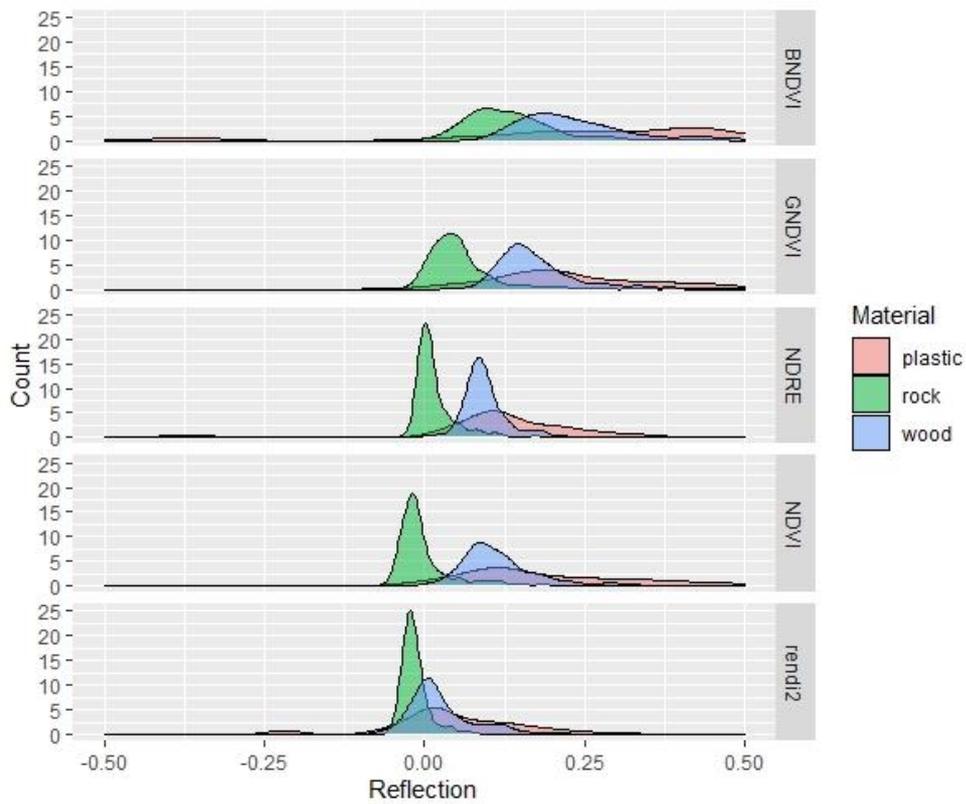## G.1 Average reflectance of input classes from the reference data of Svalbard
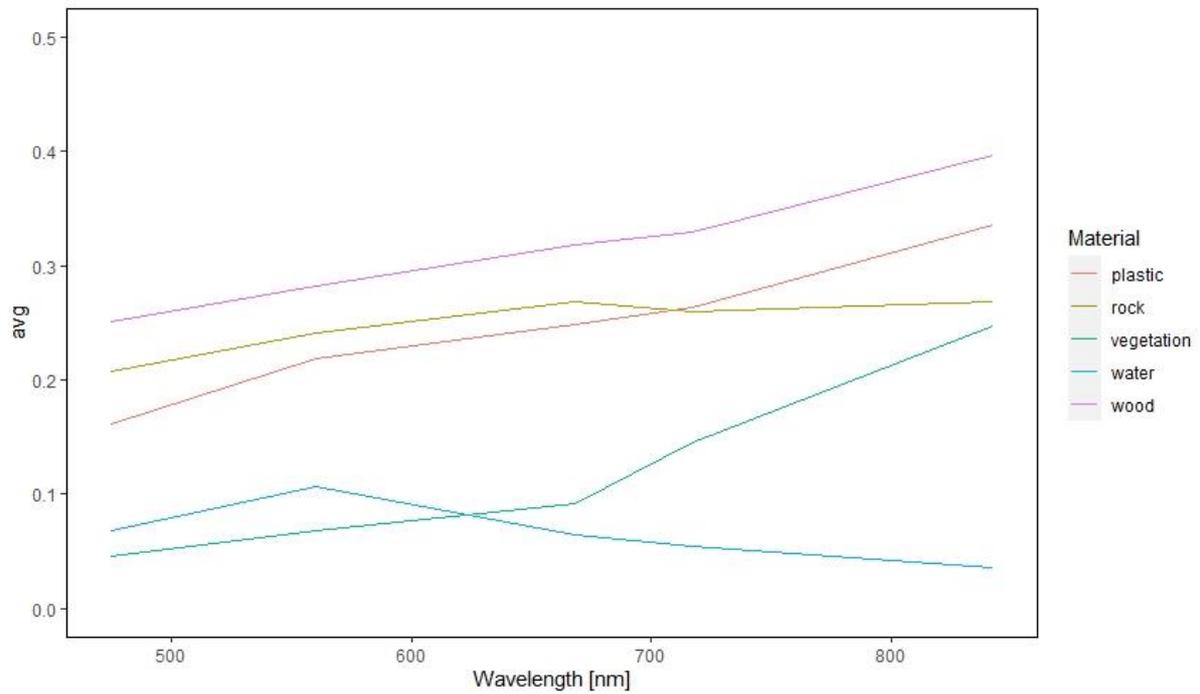
Source: Compiled by BCSH

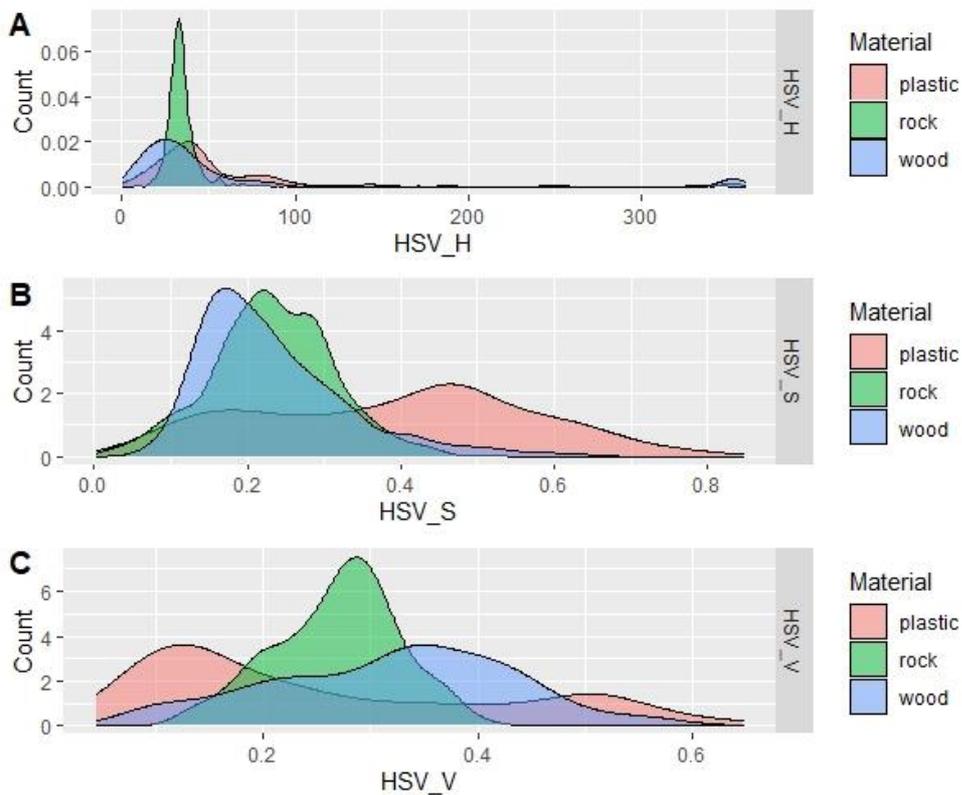## G.2 Density functions of the reflectance extracted of the reference data on Svalbard
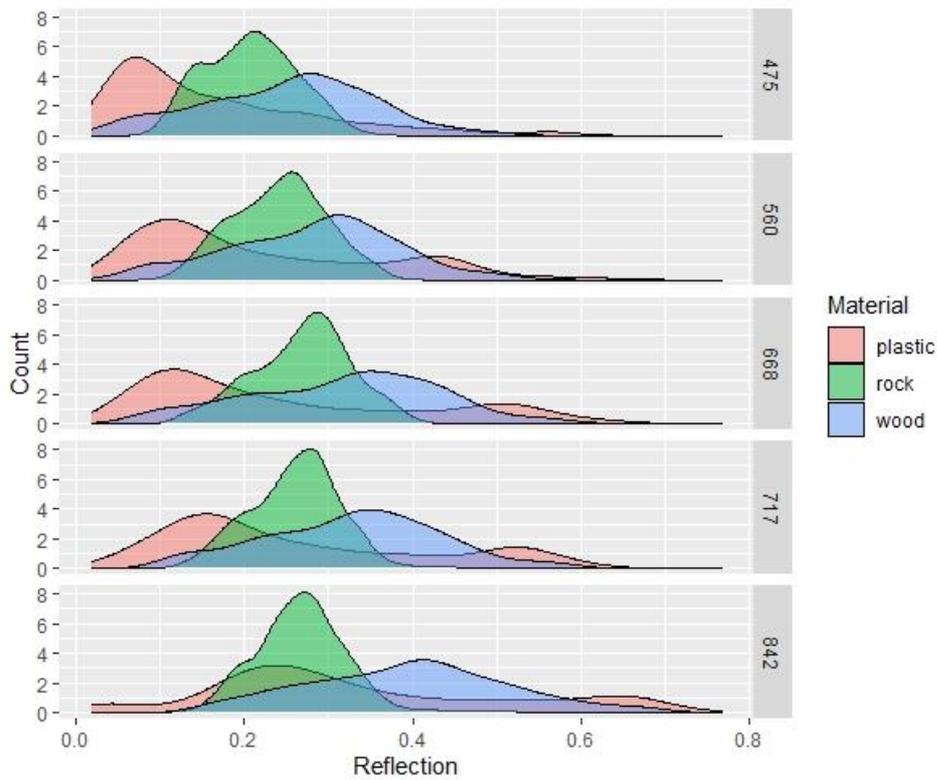
Source: Compiled by BCSH

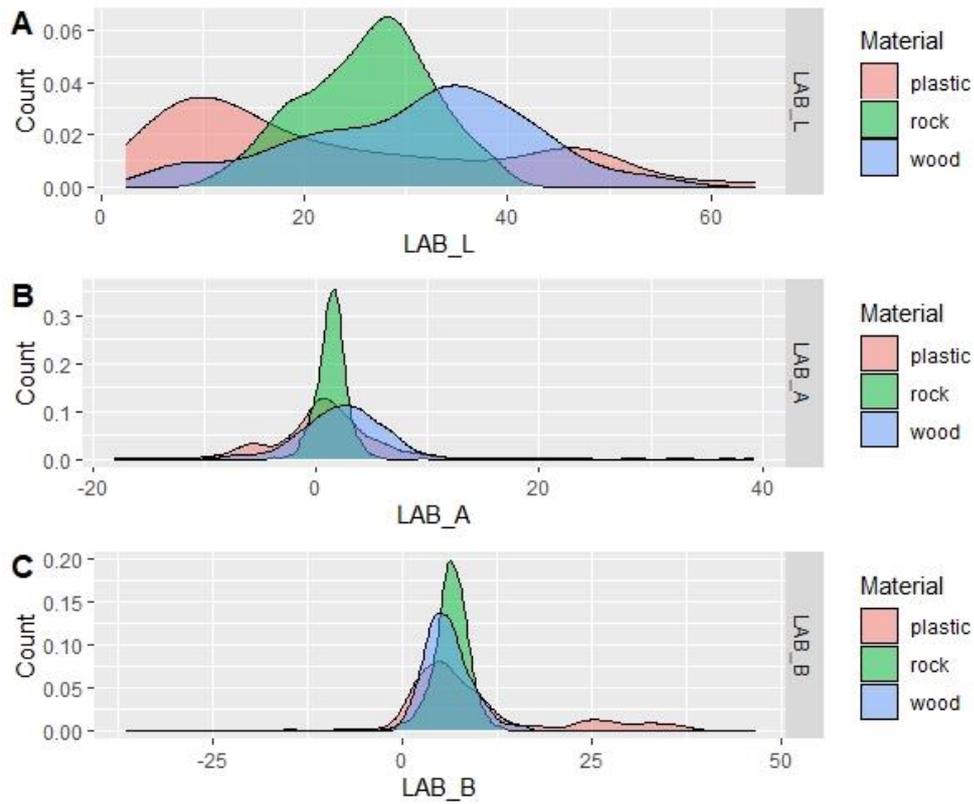## G.3 Density functions of the applied indices extracted of the reference data on Svalbard

Source: Compiled by BCSH

G.4 Density functions of HSV colour space extracted of the reference data on Svalbard
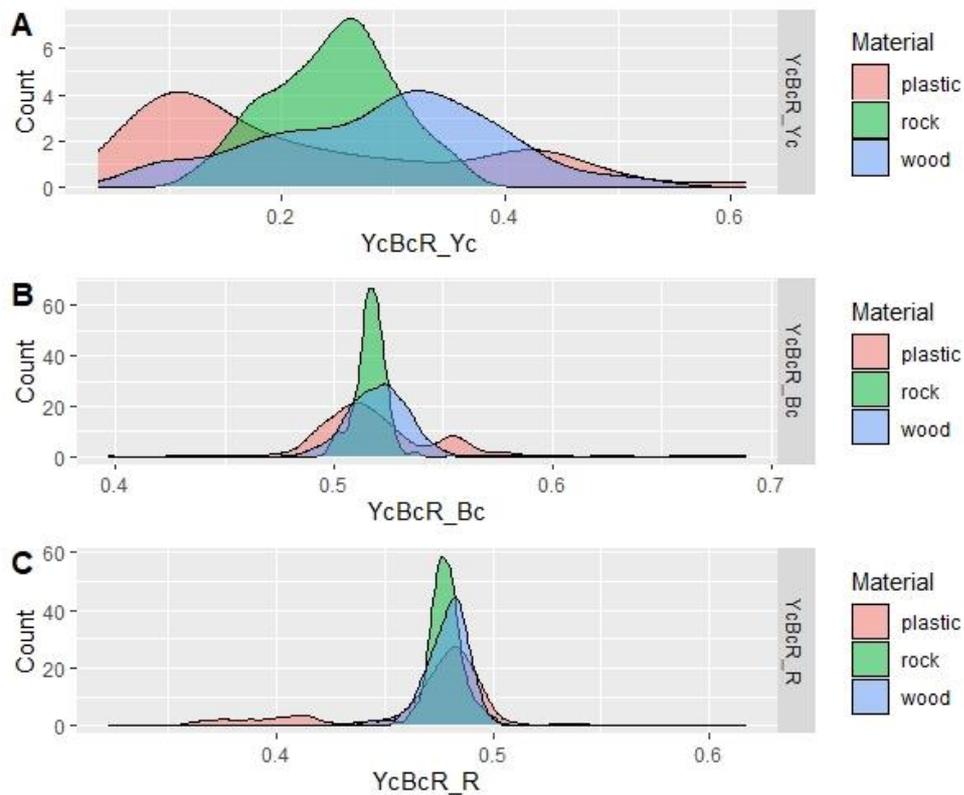

Source: Compiled by BCSH

G.5 Density functions of LAB colour space extracted of the reference data on Svalbard
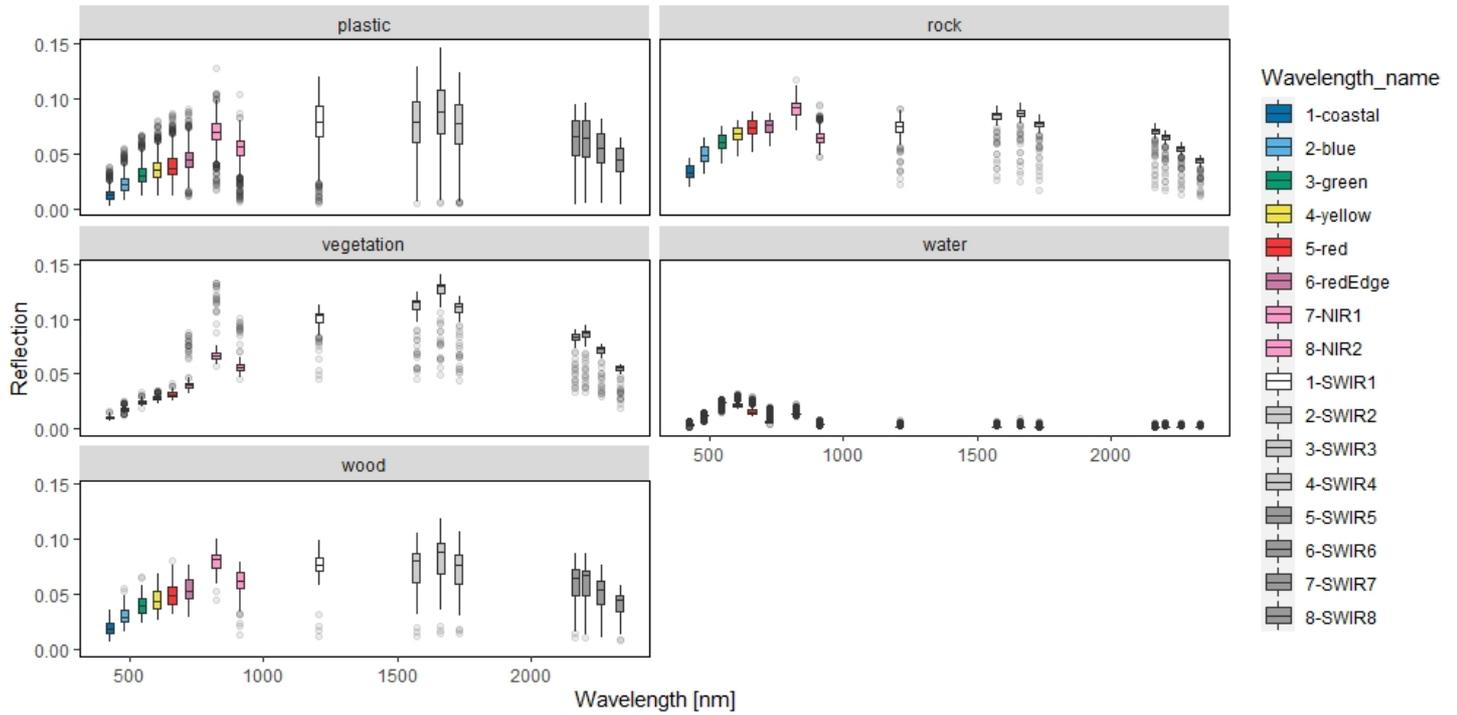

Source: Compiled by BCSH

G.6 Density functions of YcBcR colour space extracted of the reference data on Svalbard
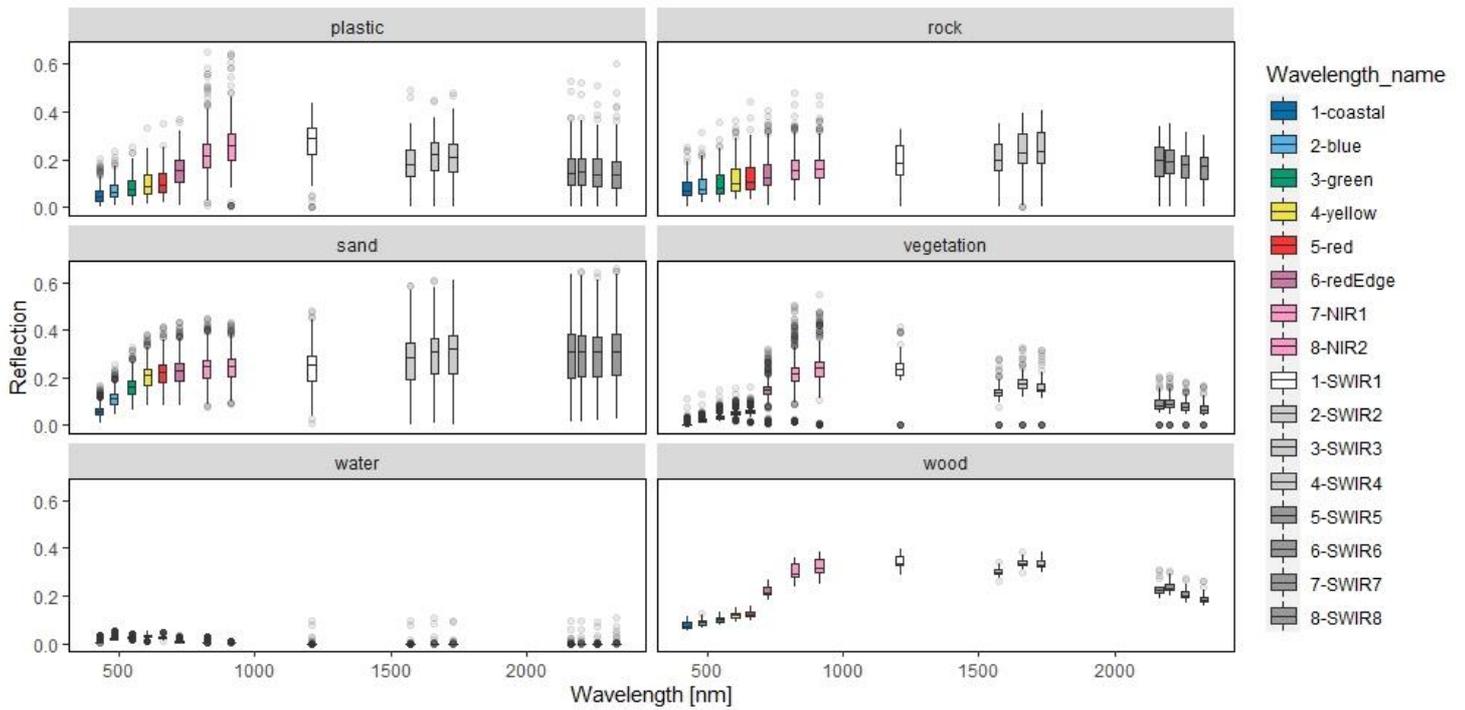

Source: Compiled by BCSH

H   APPENDIX

H.1 Spectral reflectance of WV3 imagery on Svalbard showing variance within the single classes

Source: Compiled by BCSH

H.2 Spectral reflectance of WV3 imagery on Greenland showing variance within the single classes

Source: Compiled by BCSH