Modelling the seasonal distribution of macrozoobenthos in intertidal areas of the Dutch Wadden Sea using novel remote sensing and deep learning methods

Amy Breebaart (6254225)
MSc Thesis
January 22 2023

Utrecht University

Department of Physical Geography

Supervised by:
dr. Wiebe Nijland
dr. Elisabeth Addink

Abstract

The Wadden Sea is the largest intertidal area in the world. As part of the East Atlantic Flyway, this UNESCO World Heritage site provides habitats for 10-12 million birds annually. Macrozoobenthos are an essential constituent of intertidal systems and their food webs. Changes in their distributions affect higher trophic levels of the food web, such as fish and coastal birds, and can additionally serve as indicators for shifts in boundary conditions of intertidal systems. Therefore, it is essential to regularly monitor macrozoobenthic distributions. Field sampling is a costly, labor-intensive and time-consuming undertaking and is therefore not suitable method for frequent monitoring. Remote-sensing methods are also challenging due to the homogeneous spectral signatures of intertidal areas. The present study proposes a novel method to predict the seasonal distribution of macrozoobenthic species, based on feature extraction from a variational autoencoder model (VAE). Seasonal field samples were gathered in September 2021, April 2022 and July 2022. The data was used to test the random forest prediction models performance. The VAE model is trained by Planet Dove multispectral satellite imagery, and delivers additional important information in its layer structure by encoding and decoding the input data. This information is extracted and used in combination with the input multispectral satellite data as features for random forest prediction models, to predict field observations of species presence, biomass and abundance. The multispectral trained presence-absence prediction model achieves similar test data prediction accuracies as the model using VAE model features at 0.72 and 0.69 respectively. The highest accuracy is achieved by a run including VAE model features for Scoloplos armiger at 0.93. The highest abundance prediction test data R-squared was achieved by the VAE-included model for Scoloplos armiger at 0.64. Overall, the biomass and abundance predictions remain challenging. The proposed method requires additional field observations to improve the biomass and abundance predictions and to reduce zero-inflation, but provides a promising option for more affordable and time-efficient macrozoobenthos monitoring on a seasonal scale.

Acknowledgements

I would like to express my profound gratitude to my supervisor, Wiebe Nijland, for guiding me throughout the process and having his door open at all times to answer my every question. I would also like to thank Logambal Madhuanand for her coding advice, arranging the sampling campaigns and providing me with the processed field data and outputs of the variational autoencoder model. Finally, I thank my family and partner for their positivity, patience and support during the field campaigns, modelling and writing phases of this thesis.

Contents

1. Introduction

The Wadden Sea is a shallow coastal region bordering the North Sea and stretches from the Netherlands to northern Denmark. It is the largest intertidal area on Earth and has received multiple environmental statuses which recognize its value. For example: Ramsar status, Natura 2000 area and UNESCO World Heritage site. The rich benthic macrofauna communities of the Wadden Sea sustain an important habitat for fish as well as for birds travelling the East Atlantic flyway, as foraging, breeding and resting grounds (Horn et al., 2017). Food supplies are 10-20 times higher for tidal flat fauna compared to adjacent deeper waters (Reise et al., 2010). Distribution and biomass of macrobenthic fauna in the Wadden Sea display seasonal patterns, such as winter mortalities and spring abundance of species (Beukema & Dekker, 2020). Changes in these seasonal patterns can (1) directly affect coastal bird populations in the Wadden Sea and (2) can be used as indicators for changing conditions within the system such as soil subsidence, eutrophication, change of temperature, exposure time, dredging and human alteration of hydrodynamics (Eriksson et al., 2010; Horn et al., 2020).
The macrobenthic fauna, or macrozoobenthos, across the Dutch Wadden Sea have been monitored on an annual scale since 2008 by the Dutch Royal institute of Sea Research (NIOZ) as part of the Synoptic Intertidal Benthic Survey (SIBES). However, we must increase our understanding of the Wadden Sea system and its drivers on a seasonal scale in order to sustain this UNESCO World Heritage site, including the higher trophic levels such as birds and fish. Within the existing monitoring program SIBES it is impossible to increase the sampling frequency to a seasonal scale. Therefore, there is a need for new macrozoobenthos monitoring methods which can be applied on a seasonal scale. Remote sensing is a widely used method for change detection and image classification, especially with present and increasing availability of multispectral images with high spatial and temporal resolutions. Remote sensing methods are cost and time efficient compared to physical sampling of the entire Wadden Sea. Nevertheless, remote sensing methods are difficult to apply to intertidal areas, as they have poor spectral signatures due to the homogeneous pigmentation, absence of sharp boundaries, continuous moisture and vegetation (Sørensen et al., 2006). Variational autoencoder (VAE), as a deep learning model, is a promising addition to remote sensing methods due to its distribution-free assumptions and unrivalled non-linear approximation for modelling and feature extraction (Zerrouki et al., 2021). Therefore, the present study proposes a novel method to model and predict the seasonal distribution of multiple macrozoobenthic species using random forest models trained by remotely sensed multispectral satellite data in combination with extracted features from a variational autoencoder model.

1.1 The Wadden Sea

The Wadden Sea ranges from the north-west of the Netherlands up to Denmark making it the largest intertidal zone worldwide (Compton et al., 2013). It offers food to a an extensive aquatic fish nursery and millions of coastal birds travelling the East Atlantic flyway. Annually, 10-12 million birds use the intertidal flats of the Wadden Sea (Horn et al., 2017; Horn et al., 2020). Benthic macrofauna are a central constituent of the ecosystem services the Wadden Sea provides. Benthic macrofauna are defined as the invertebrate species larger than 1mm, which live in or on the sediment or hard substrate. Changing boundary conditions in the ecosystem can affect the distribution patterns of macrozoobenthos and therefore influence the functioning of the Wadden Sea food web. Therefore, changes in macrozoobenthos distribution are used as an important indicator of shifting boundary conditions in intertidal systems, as these fauna are sedentary and therefore completely dependent of their

surroundings. Benthic macrofauna can also alter their surroundings by deposit feeding, suspension feeding and bioturbation; many species are therefore recognized as ecosystem engineers (Drent et al., 2017). This characteristic is often used to detect specific communities of benthic macrofauna. Furthermore, macrozoobenthos are at the bottom of the food chain; changes in macrobenthic fauna distribution will affect higher trophic levels such as fish and birds (Compton et al., 2013). In turn, this may lead to unwanted changes in the Wadden Sea food web and ecosystem, which endangers the unique and valuable qualities of the area. The exceptional ecological value of the Wadden Sea was recognized by the UNESCO World Heritage Committee in 2008, when the Dutch and German parts of the Wadden Sea were officially recognized as a World Heritage site. In 2014 the Wadden Sea in its entirety received the World Heritage status (convention and nomination, n.d.). Furthermore, the Wadden Sea has acquired several other environmental statuses such as: Ramsar site, Important Bird Area, Natura 2000 and Man and Biosphere Reserve. Besides the ecological value, the Wadden Sea is also a popular tourist location. A study by Sijtsma et al. (2019) estimated that the community of Wadden Sea fans is approximately fourteen times larger than its number of inhabitants, with only 2,5% of the survey to never have visited the Wadden Sea. Extensive research on the Wadden Sea is funded and performed in order to understand and preserve this unique system with consideration of challenges such as soil subsidence due to fossil fuel extraction, sea level rise, rising temperatures due to climate change, tourism and overfishing (Vermeersen et al., 2018; Wang et al., 2018).

1.2 SIBES

The Synoptic Intertidal Benthic Survey (SIBES) is a sampling program designed by NIOZ in cooperation with the Dutch Oil Company (NAM) to monitor the Dutch Wadden Sea. The survey monitors 28 macrobenthic species and is executed annually since 2008, with approximately which started with approximately 4500 sampling locations and is now extended to approximately 4700 sampling locations (Figure 1) (Compton et al., 2012). The distance between sampling locations is approximately 500 meters with 20% additional random points for estimating autocorrelation parameters (Bijleveld et al., 2012). The survey is designed and executed by NIOZ and funded by the NAM and Rijkswaterstaat (RWS). The sampling program started in 2008 to investigate the potential effects of sediment subsidence, caused by the East Frisian natural gas production area, on macrozoobenthos (Compton et al., 2013). The data is also used to research the changes in macrozoobenthos distributions and the effect it has on migratory bird populations and fish nursing grounds. The sampling locations were selected using gridded sampling with additional randomly interspersed points (Bijleveld et al., 2012). NIOZ staff and volunteers perform the sampling in multiple sampling teams on motorized dinghies and NIOZ research vessel Navicula. The sampling takes approximately 40 days and lasts from June until September. Sample sorting and taxonomy then takes circa one year. SIBES provides only a momentary snapshot of the macrozoobenthos during a specific time in the summer. It uses an intrusive core sampling method and is a time consuming and costly undertaking. SIBES delivers valuable data about the ecosystem of the Dutch Wadden Sea and can be used for a multitude of applications. However its temporal resolution is limited due its high cost and labor intensity.

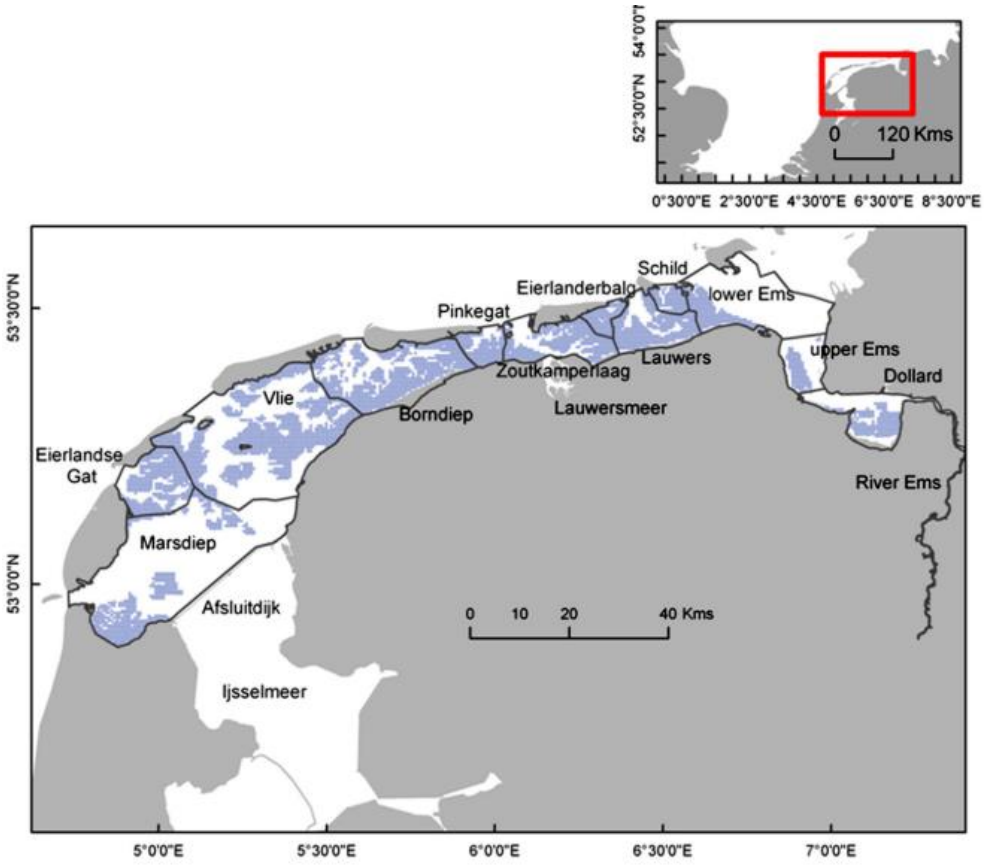


Figure 1. *Sampling locations (blue) from SIBES monitoring program in the Dutch Wadden Sea (Compton et al., 2013, p.105).*

1.3 Seasonality

A study of the Balgzand tidal flats by Beukema & Dekker (2020) found that benthic macrofauna species richness and abundance is affected by seasonality. This study showed that winter-sensitive species such as Lanice conchilega (Pallas, 1766) can halve in biomass during severe winters. Winter sensitive species which were located at relatively higher sites on the intertidal flats showed decreased survival compared to lower lying sites. A study by Beukema et al. (2017) showed increased bivalve growth and biomass with increasing summer temperatures, whereas increasing winter temperatures lead to increased winter weight losses in bivalve biomass (Honkoop & Beukema, 1997). Growth rates of bivalve Macoma balthica (Linnaeus, 1758) reduced during high summer temperatures and survival rates were reduced at both high winter and summer temperatures (Beukema & Dekker, 2020). Hence, different macrozoobenthic species respond differently to seasonal changes in their habitats.

SIBES can only provide annual summer data of macrozoobenthos. Due to the seasonality of macrobenthic fauna distribution, their importance to higher trophic levels in the system and their sensitivity to shifting boundary conditions in the Wadden Sea, it is important to find different methods to monitor and predict seasonal changes of macrozoobenthos distributions.

1.4 Remote Sensing and VAE models

Remote sensing is a non-intrusive method of data acquisition, less labor intensive and available at high spatiotemporal resolutions. Using remotely sensed satellite data for image classification is a widely accepted and adopted concept with many applications. Multispectral satellite imagery provides

possibilities for accurate mapping of among others vegetation cover, vegetation structure but also soil characteristics such as soil moisture. Furthermore, it can be used for among others change detection, image classification and vegetation biomass estimation. These methods use the variability of spectral information from the remotely sensed data. However, intertidal areas of the Wadden Sea contain little spectral variability as they consist of a fairly homogeneous surface (Sørensen et al., 2006). It is therefore difficult to extract sufficient information from multispectral satellite imagery to make seasonal predictions of macrozoobenthos distributions.

Nevertheless, mapping intertidal habitats using remote sensing is also not a new concept. It originated as visual photo interpretation of small intertidal areas (Godet et al., 2009). More recently, remotely sensed multispectral satellite imagery provides the opportunity to map intertidal habitats on a larger spatiotemporal scale. Satellites such as IKONOS, SPOT and Landsat (TM and ETM+) are commonly used for intertidal habitat mapping. Landsat TM and ETM+ is often used due to its large archive and open data (Ryu et al., 2014). However, the spatial and temporal resolution of Landsat data is limited.

Intertidal areas in the Dutch Wadden Sea are largely unvegetated and have a homogeneous surface and spectral signature. These weak spectral characteristics make it difficult to use traditional remote sensing methods to detect macrobenthic species distributions and predict their biomass and abundance.
Benthic habitat mapping using remote sensing methods often concentrates on sediment structures, bathymetry and vegetation. It is focused on habitat mapping by detecting physical indicators of species such as a specific set of substrate characteristics, rather than detecting the actual species. The results of these indicator studies are then tested using biological data (Godet et al., 2009). Remote sensing methods are therefore a promising option for predicting and mapping seasonal distributions of macrozoobenthos, but by itself is not sufficient.
Feature extraction by variational autoencoders, as deep learning convolutional models, have shown promising results in multiple research fields such as anomaly detection, photovoltaic solar power production forecasting and desertification detection (Dairi et al., 2020; Yao et al., 2019; Zerrouki et al., 2021) and provide more accurate prediction results than other commonly used prediction methods. A study by Madhuanand et al. (2023) uses a similar method as the present study with VAE model features. The study predicts physical and ecological variables on intertidal flats in the Dutch Wadden Sea, and the predictions using VAE model features showed improved results compared to multispectral features.

1.5 Aim and research questions

The aim of the present study is to apply and assess a novel method for predicting benthic macrofauna distribution using remotely sensed multispectral satellite imagery in combination with features from a VAE model. We will examine the capabilities and limitations of the proposed method by using Planet 4-band multispectral satellite imagery and extract 64 additional feature layers per image by the variational autoencoder model. We will construct seasonal (1) presence-absence, (2) biomass and (3) abundance prediction models for different macrobenthic species in the Dutch Wadden Sea.

The main questions we try to answer in the present study are:

- Do features from a Variational Autoencoder improve the predicted seasonal distribution of macrozoobenthos compared to multispectral satellite data?
- Is the model applicable over different seasons without distinguishing seasonal data within runs?
    - Do the predictions improve or decline when multi-seasonal data is used?

- What are the capabilities and limitations of the VAE features in terms of modelling presence and absence, biomass and abundance?
    - Is this similar across the modeled species?

We will assess the performance of VAE model features for the different models by comparing the prediction results of (1) multispectral features and (2) multispectral and VAE model extracted features. We then use this data as input for random forest (RF) models in the Python programming language to predict presence-absence, biomass per m^2abundance per m^2 for different macrobenthic species. The proposed models are tested on the following six macrozoobenthic species independently, to detect potential differences in performance between species: 1. Arenicola marina (Linnaeus, 1758), 2. Heteromastus filiformis (Claparède, 1864), 3. Lanice conchilega (Pallas, 1766), 4. Limecola balthica (Linnaeus, 1758), 5. Pygospio elegans (Claparède, 1863) and 6. Scoloplos armiger (Müller, 1776).
We expect that features from the VAE model can be used to train the three distribution models and will provide improved predictions results compared to the multispectral data. We also expect the performance of the models to increase for multi-seasonal runs as they will contain more data to train the random forest models. Per species, we expect the models to have better performance for Lanice conchilega, also known as sand mason worm, as they produce sand cylinders which protrude large areas of intertidal flats. Also, prediction models for larger soil disturbing, ecosystem engineering species such as Arenicola marina are expected to perform well.

2. Methodologies

The study is performed on an intertidal area adjacent to the northeast of barrier island Texel (Figure 2). To construct the proposed prediction models we use core sample data of macrozoobenthos from the intertidal flats of the study area in combination with multispectral satellite imagery, features from a VAE model and random forest models. The seasonal sample data is processed in the laboratory of NIOZ on Texel and delivers data of ash free dry biomass and abundance of the identified species, or other operational taxonomic units, per sample. Multispectral satellite images of the study area, taken at approximately the dates of the sampling campaigns, are selected and used to train a variational autoencoder model. The multispectral data in combination with the features extracted from the VAE model are used to train random forest prediction models. The field data is used for the random forest model training and testing. We will asses the applicability of this novel method and evaluate the seasonal portability of these predictions for six benthic macrofauna species using accuracy scores for classification models and the coefficient of determination, or R-squared values, in combination with observed/predicted plots for the regression models.

2.1 Study Area

The study area is located on intertidal flats to the northeast of Texel. It has a surface area of approximately 56.5 km^2 and contains 55 sampling locations for the April 2022 sampling campaign, 51 locations for July 2022 and 45 locations for the SIBES 2021 campaign (Figure 2). The study area has approximately semi-diurnal tides and falls almost completely dry during low tide, with the exception of several larger tidal channels. The sampling locations of the April and July 2022 campaigns are the same in principle, apart from some alterations after the first campaign in April. The April and July sampling locations are located closer to the shore compared to the SIBES sampling locations, as the April and July

campaigns were executed on foot. The southern portion of the research area is located in a protected bird reserve, for which we received permission to enter for sampling during the campaigns.

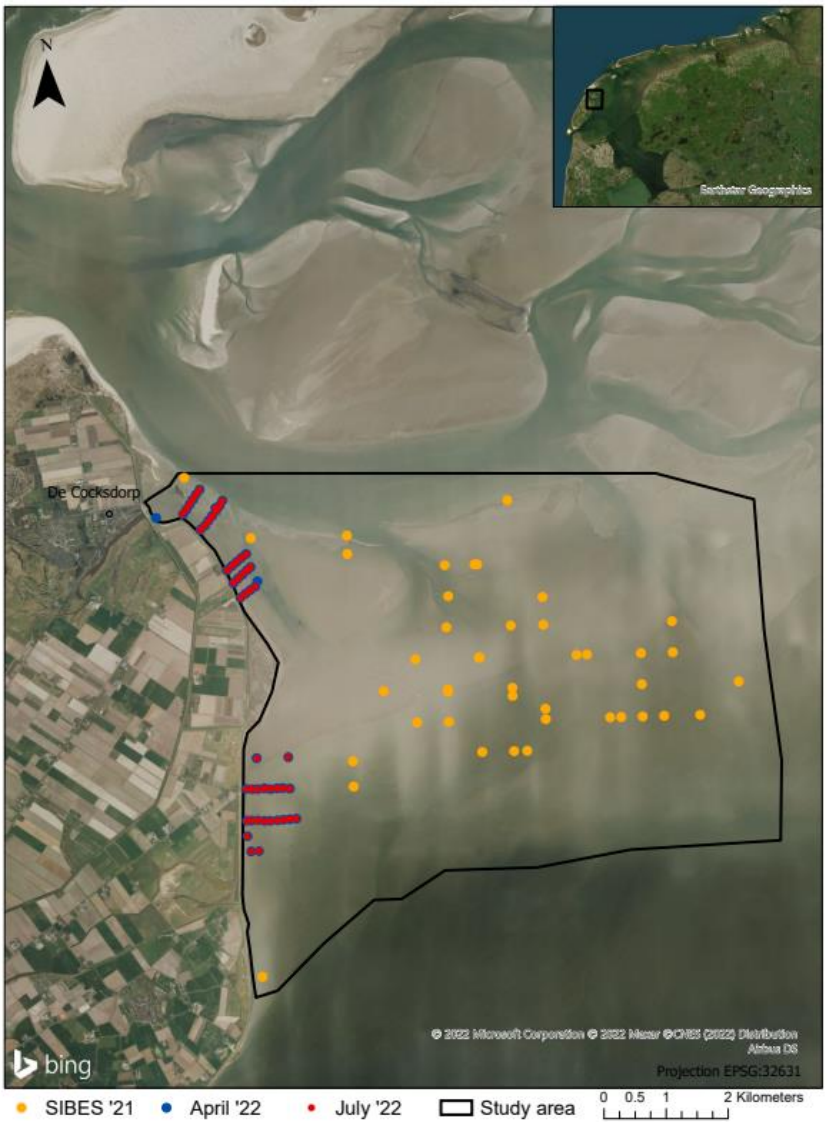


Figure 2. *The study area and sampling locations. April sample locations in blue, July sample locations in red and SIBES 2021 sample locations in yellow.*

2.2 Species

Six macrobenthic species were selected based on their constitution in shorebird diets and variety of ecosystem engineering traits (Table 1)(Compton et al., 2013; Folmer et al., 2017). Additionally, their prevalence throughout the field samples of the April, July and SIBES 2021 campaigns is also essential, as the quantity of input data is important for the random forest model training**.**

Table 1. *Characteristics of the selected species subset: taxonomic classification, approximate longevity in years, habitat type, reproductive mode, depth position in the sediment in cm, feeding mode and approximate size in cm (Adapted from Compton et al., 2013).*

Species	Class	Family	Long-evity	Habitat	Reproductive mode	Position in sediment	Feeding mode	App. Size
Arenicola marina	Polychaeta	Arenicolidae	>5	Burrow	Brooder	>10	Grazers	>5
Heteromastus filiformis	Polychaeta	Capitellidae	<2	Burrow	Broadcast	>10	Deposit	>5
Lanice conchilega	Polychaeta	Terebellidae	>5	Tube	Broadcast	>10	Filter	>5
Limecola balthica	Bivalvia	Tellinidae	>5	Burrow	Broadcast	2 to 5	Deposit	1 to 5
Pygospio elegans	Polychaeta	Spionidae	2 to 5	Tube	Brooder	2 to 10	Deposit	1 to 5
Scoloplos armiger	Polychaeta	Orbiniidae	2 to 5	Burrow	Brooder	>10	Grazers	>5

Arenicola marina, also known as lugworm, is part of the diet of among others the bar-tailed godwit (Folmer et al., 2017). Arenicola marina lives in 20 to 40 cm deep L-shaped burrows in sandy intertidal areas. Their burrowing activities rework the smooth surface into a landscape of mountains and depressions (Flach, 1992). A study of a Danish sandflat by Horn et al. (2020) found Arenicola marina to dominate the total biomass on the sand flat over all seasons. In the Dutch Wadden Sea, the biomass of Arenicola marina accounts for the second largest part of the total macrozoobenthic biomass behind Cerastoderma edule, according to a 2017 study by Folmer et al. (2017). The same study also found that Arenicola marina is sensitive to changes in exposure time and shear stress.

Heteromastus filiformis has been reported in a wide variety of sediment types but primarily in muddy sands (Van der Wal et al., 2008). It is a tube forming species which live in vertical burrows and become 8 to 15 cm long and 1 mm wide. Heteromastus filiformis derive their food from sediment in the anaerobic layer at 10 to 30 cm depth and excretes the remaining sediment as small black pellets at the surface (Cadée, 1979; Hartman & Fauchald, 1971). The 2017 study by Folmer et al. found that Heteromastus filiformis is sensitive to changes in wave intensity.

Lanice conchilega, also known as sand mason, is a tube-building worm and is known as a habitat structuring species. Their tubes consist of sand and shell fragments and protrude 1 to 4 cm above the sediment surface (Van Hoey et al., 2008). They are part of the diet of among others the bar-tailed godwit (Folmer et al., 2017). Communities can reach densities of multiple thousands of individuals per m^2. Due to their protruding tubes, high densities of Lanice conchilega can affect the hydrodynamics which can locally change sedimentation of fine particles (Van Hoey et al., 2008).

Limecola balthica, henceforth Macoma balthica, is a bivalve and is most abundant in areas of relatively fine sediments and long exposure times according to Kraan et al. (2010). They are part of the diet of among others bar-tailed godwits, red knots and oystercatchers (Folmer et al., 2017). The study by Folmer et al. (2017) found a positive relationship between Macoma balthica biomass and exposure time. Relative sea level rise is therefore expected to cause decreased exposure time. Furthermore, they found a negative relationship between biomass and wave forcing.

Pygospio elegans, also known as tubeworm, is a small and sedentary tube-building worm and can reach 15 mm in length (Bolam & Fernandes, 2003). They are associated with sediment of high organic content and are tolerant to a broad variety of environmental conditions (Giere & Pfannkuche, 1982). According to the 2017 study by Folmer et al., numerically they are the most dominant species in the Dutch Wadden Sea behind mud snails (Peringia ulvae (Pennant, 1777)). A study by Compton et al. (2013) also found Pygospio elegans in more than 35% of the sampled points from SIBES 2008 to 2010.

Scoloplos armiger is part of the diet of among the bar-tailed godwit diet and was observed in more than 35% of the SIBES sampled locations from 2008 to 2010 (Compton et al., 2013; Folmer et al., 2017). Folmer et al. (2017) found a slightly negative relationship with exposure time, while increasing waves and grain size have a positive relationship to Scoloplos armiger abundance.

2.3 Samples

The random forest prediction models constructed for the present study are trained and tested by field data. The data is obtained by sampling different locations over three sampling campaigns during different seasons: SIBES data from September 2021 and two campaigns in 2022; April and July. The relative distance between sampling locations of the April and July campaigns is smaller compared to SIBES: approximately 100 meters compared to 500 meters. The macrozoobenthos samples are processed, identified to the highest taxonomic unit and then incinerated and weighed at the NIOZ laboratory to acquire (1) All operational taxonomic units (OTUs) per sample, (2) OTU ash free dry biomass per sample and (3) OTU abundance per sample. In the case of the present study the selected OTUs for the modelling are identified to species level.

2.3.1 Sampling locations

The study area and sampling locations were selected in consultation with NIOZ. It is imperative to preserve the samples at a cool temperature until they reach the NIOZ laboratory. The study area is accessible on foot with the field equipment, and is attached to the same barrier island as the NIOZ laboratory, Texel. The study area consists of multiple gridlines with five to seven sampling points, located approximately 100 meters apart. Apart from the gridlines there are also some additional sampling locations which are also included in SIBES (Figure 2). SIBES sampling locations are located further apart on a 500 meter grid and are mostly inaccessible on foot as they lie further away from the shoreline (Bijleveld et al., 2012). Therefore, only SIBES sampling locations in the study area close to the shore were included in the April and July 2022 sampling campaigns. All sampling points are strictly located in intertidal areas which fall completely dry during low tide. The area must fall dry because the model is used for macrozoobenthos prediction in intertidal areas and does not respond well to the spectral signature of water.

2.3.2 Sample collection

Field samples were collected to identify the present benthic macrofauna, count the individuals and measure their ash free dry mass (AFDM). The sampling was performed using the same methodology as SIBES. However, the samples for the present study were solely collected on foot from intertidal areas during low tide. The sampling campaign of April 2022 occurred from the 26th to the 28th with 55 sample locations. The campaign of July 2022 lasted from the 12th to the 14th with 51 sample locations. Additionally, data of 45 SIBES sample locations of September 2021 are considered in this study, which were taken on September 9th of 2021. During the April and July campaigns we used a Trimble receiver using network RTK and the Dutch 06-GPS network to determine the exact locations of the samples with

accuracies up to 1 cm. A core of approximately 0.018 m^2 to a depth of circa 25 cm is taken at each sampling location. The core is then placed on a 1 mm mesh sieve in the field. Large shellfish are separated from the rest of the sample as they are processed differently in the laboratory. The sieved samples are then placed in a cooler for preservation during the fieldwork.
The bivalves are frozen at the NIOZ laboratory and remainder of the samples are preserved using a 4% formaldehyde solution and stained with Rose Bengal dye (C.A.S. no. 632-68-8) to aid the sorting process.

2.3.3 Taxonomy and abundance

In the laboratory the samples are subjected to a final flushing and cleaning with fresh water over a 0.5 mm sieve. They are then presorted onto petri dishes for taxonomy and counting. Taxonomy is also executed in the same manner as SIBES samples. The identification and counting was executed under a binocular microscope (8-40 x magnification) according to the ISO guidelines (ISO 9001:2008 nr.K57663/01) and then classified according to Hartmann-Schröder (1996), and Hayward and Ryland (1995) (Compton et al., 2012). The macrozoobenthos are identified to the highest possible taxonomic level and all shellfish are identified to species level. During the identification process, the individuals per OTU are also counted. This provides abundance data per sample for all identified OTUs. In this study we consider six OTUs identified to a species level.

2.3.4 Ash free dry mass

Upon completion of taxonomy and counting, the AFDM, or biomass, per OTU per sample is weighed. AFDM of all organisms except shellfish is determined of either single or multiple individuals, their sizes are not considered in the data. This procedure is slightly different for shellfish compared to other OTUs. AFDM for shellfish and other organisms is determined for each individual larger than 8 mm, smaller individuals are measured together. AFDM is determined by placing wet meat of an organism into a crucible and dry it for 48 hours at 60˚C. Then, the sample is cooled in a desiccator and weighed to an accuracy of four decimal places. Finally, the dried sample is incinerated at 560˚C for 5 hours and the ashed sample is cooled and weighed as well (Compton et al., 2013). This procedure delivers the ash free dry biomass per OTU per sample.

2.3.5 Sample data preparation

The processing of all samples results in biomass and abundance data of all OTUs per sample. Some organisms are too small to weigh. For the present study, Pygospio elegans individuals were not weighed as they were too small. Instead, we used a standard biomass of 0.00016 grams per individual, the other five species were weighed as is described in the previous section. If the biomass was lacking from the data, or resembled an unrealistic number we replaced it using the average biomass of the species in our samples. Biomass and abundance of each OTU is then calculated from per sample to per m^2. If species were absent in samples they were assigned a value of 0 for both biomass and abundance.
As the prediction models tend to perform better with normally distributed training data we proceed to check the data we will use to train the prediction models for skewedness. The biomass and abundance data are not normally distributed and zero-inflated. Therefore, we execute a log transformation of both the biomass and abundance data of the six species. The results of the log transformed data are displayed and in chapter 3.

2.4 Remote sensing

The best available 4-band multispectral Planet scene was selected per sampling campaign: September 2021, April 2022 and July 2022. The Planet image data is taken by their Doves constellation. It uses a fleet

of over 150 shoebox-sized satellites which line-scan the Earth daily with a 3x3 meter pixel size. DOVE-C and DOVE-R sensors measure four bands: (1) Blue (455 to 515nm), (2) Green (500 to 590 nm), (3) Red (590 to 670 nm), (4) Near Infrared (780 to 860 nm). The newest generation satellites Super-DOVE measure 8 bands (Planet Labs, 2017). For the present study we only use 4-band data. We used the unharmonized Planet satellite imagery. The Planet images were selected based on several criteria: (1) the date is preferably within a week before the field work and not during or afterwards due to the disturbances caused by the fieldwork itself, (2) the images preferably need to be taken during maximum low tide so there is minimal water present in the study area which can disturb the spectral signature, (3) there should be no cloud cover over the study area. For the SIBES (September) 2021 data we selected the Planet image of the study area of the 5th of September 2021. For the April 2022 we campaign we selected the Planet image of the 23rd of April 2022. Lastly, for July 2022 we selected the Planet image of the 12th of July 2022. The pixel values of the four different bands are extracted at the sampling locations and then used as features for the random forest models**.**

The remotely sensed multispectral satellite images are also used as input for feature extraction by the VAE model. From this VAE model we then extract 64 feature layers per input image. These features are then used as features for the random forest prediction models in combination with the multispectral data.

2.5 The variational autoencoder model

The variational autoencoder model, as a deep learning convolutional neural network, is a regularized variety of the autoencoder model (Kingma & Welling, 2014). The VAE model consists of an encoder, bottle neck and decoder (Figure 3) (Madhuanand et al., 2023). It takes high dimensional input data and encodes the input data into a latent space as distribution (Zerrouki et al., 2021). The purpose of the decoder is to reconstruct the encoded data from the latent space. This is achieved by training the VAE model with the satellite input data to learn its essential features. Then, the VAE model decodes the latent space to reconstruct the original data. Errors of the reconstruction are backpropagated through the network structure to optimize the decoding process (Ma et al., 2020; Zerrouki et al., 2021).

In the present study, the Planet satellite imagery is used as input data for the VAE model, which uses the same modelling structure as is presented by Madhuanand et al. (2023). 64 features per input image are extracted from the VAE latent space in the encoding and decoding process. These features consist of raster data layers are used as training features in the random forest models, additional to the homogeneous multispectral satellite data (Madhuanand et al., 2023).

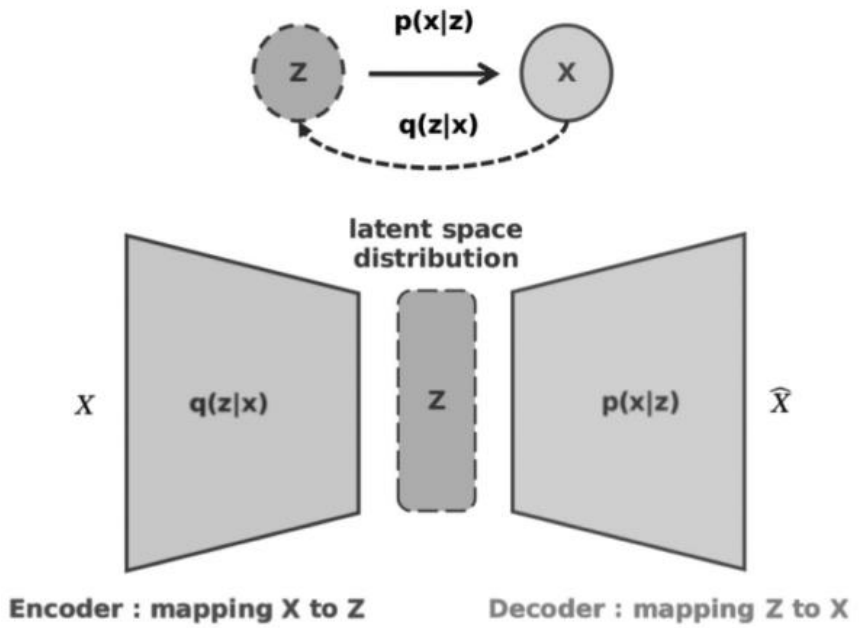


Figure 3. *Diagram of VAE structure: Input data X is compressed to Z by the encoder. Z is then reconstructed by the decoder to* $\hat{X}$*. Errors of reconstruction are backpropagated through the network (top) (Zerrouki et al., 2021, p.205).*

2.6 Random forest modelling structure

A random forest classifier is used to predict the presence and absence of the six selected macrozoobenthic species. A random forest regression is applied for the biomass and abundance modelling. The model is created in Python programming language with use of the scikit-learn package (Pedregosa et al., 2011). Random forest models are considered robust to noise, and applicable for classification and regression (Breiman, 2001). The random forest algorithm is considered one of the most efficient algorithms in terms of run time and prediction accuracy, and has proven to yield good prediction results in ecological modelling compared to other statistical classifiers (Crisci et al., 2012; Cutler et al., 2007). The extracted features from the VAE model and multispectral satellite data are used as features in the random forest models, the field sample data is used as training and testing input data. We model three different predictions per species: presence-absence, biomass and abundance. We execute these predictions for different seasons and combinations of seasons with (1) multispectral features and (2) multispectral and VAE features (Figure 4).

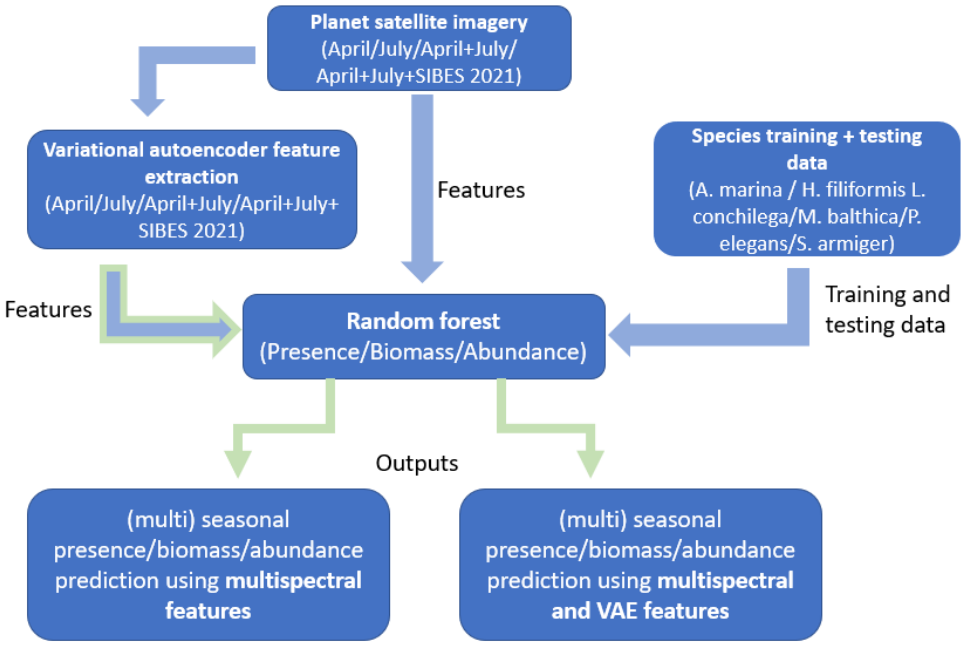


Figure 4. *Modelling structure of the study: blue arrows indicate input, green arrows indicate output. The extracted features from the variational autoencoder are considered to be output of the VAE model, and input for the random forest model. Variables are displayed between brackets.*

2.6.1 Input

As is displayed by Figure 4, the random forest model has two types of feature input; multispectral features and variational autoencoder features. The training and testing data consists of the field data from the three sampling campaigns. The multispectral satellite images are used as features for the random forest models and as input for the VAE model. The output features of the VAE model are then also used as input features for the random forest models. The random forest features consist of the pixel values of the input raster data at the sampling locations. The pixel values of the multispectral satellite data and VAE features are extracted at the coordinates of all sampling locations using the Geopandas and Rasterio Python packages (Gillies, 2013; Jordahl, 2014). This results in geopoints linked to all samples and the corresponding pixel values for all multispectral and VAE model data. Then, the pixel values linked to the location of the sample are used as features for the random forest models.

The input testing and training data consists of the presence-absence, biomass and abundance data of the field samples. The field sample abundance data is used to make a binary presence-absence dataset. We make prediction models of presence-absence, biomass and abundance for all species separately. For the presence-absence prediction we use the RandomForestClassifier function from the scikit-learn package. For the biomass and abundance predictions we used the RandomForestRegressor function. Normally, random forest uses randomly selected training and testing data. However, for the present study the amount of samples is limited. To ensure there is sufficient variance in the training data to make accurate predictions we predetermined training and testing sampling locations (Figure 5).

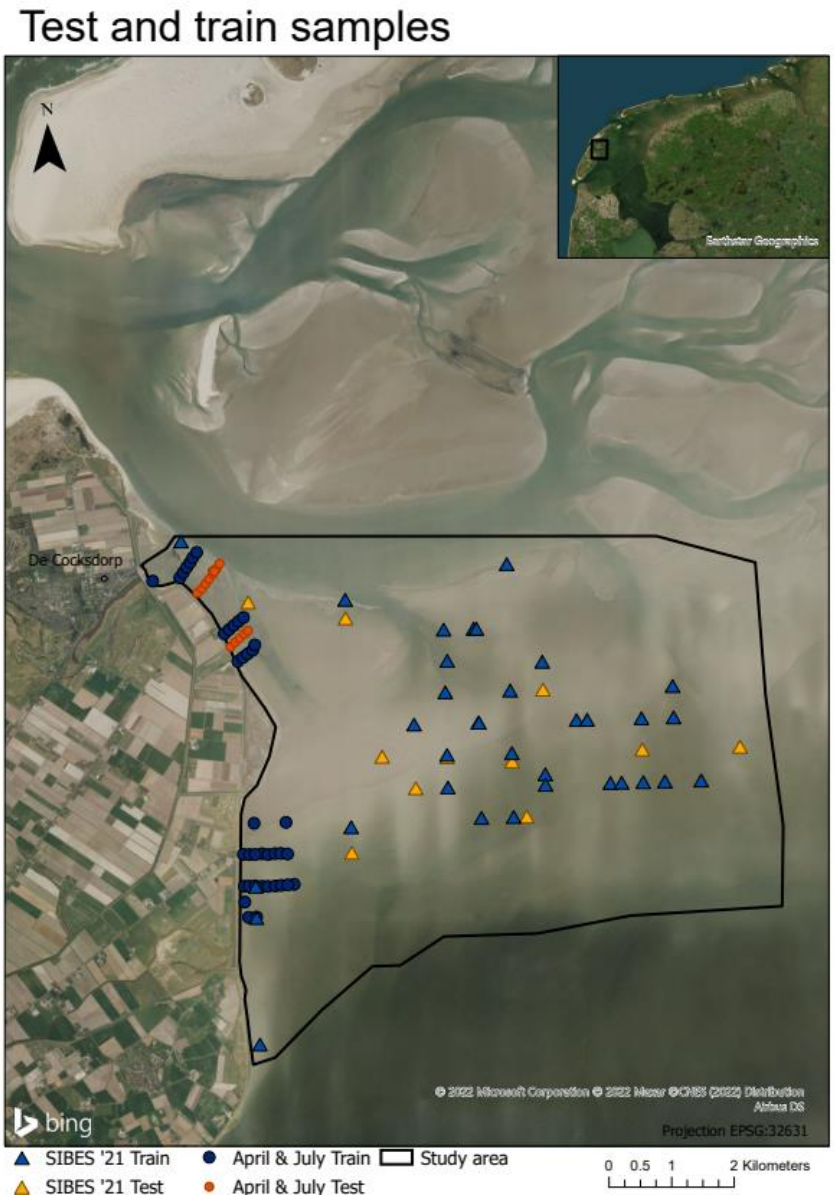


Figure 5. *Training and testing sampling locations for all sampled season. SIBES training and testing data locations are represented by triangles: training data in blue, testing data in yellow., April and July training and testing data locations are represented by circles: training data in dark blue, testing data in orange.*

2.6.2 Parameters

The parameter settings of the model are empirically determined during the presence-absence modelling. The species that produced the best results during the modelling, in our case Arenicola marina, Pygospio elegans and Scoloplos armiger, were used to optimize the random forest parameters. After the empirical optimization, we proceed to run all models using the same settings. We chose this approach as the aim is to produce a more universally applicable model; therefore we also used mostly default settings. The number of trees is set at 1000. The maximum depth of the trees is set at two, minimum samples per split at three and finally minimum number of samples at two. The maximum tree depth is set fairly low at 2 due to relatively low quantity of input data; a larger tree depth causes more overfitting with this data, especially the for single-season runs. All other possible settings from scikit-learn random forest function

such as minimum samples per split, maximum number of features to consider and minimum weighted fraction of the total sum of weights required at a leaf node are set to default.

2.6.3 Outputs

The random forest prediction models deliver among others predictions for both the training and testing data as output. The three prediction models are run for the six selected macrobenthic species separately and with 2 feature variations; solely multispectral features, and VAE and multispectral features. Furthermore, we run the models with four different combinations of seasonal feature data: April, July, April and July and lastly April and July and SIBES 2021 data. This results in eight outputs per prediction type per species, or a total of 24 outputs per species (Figure 6).

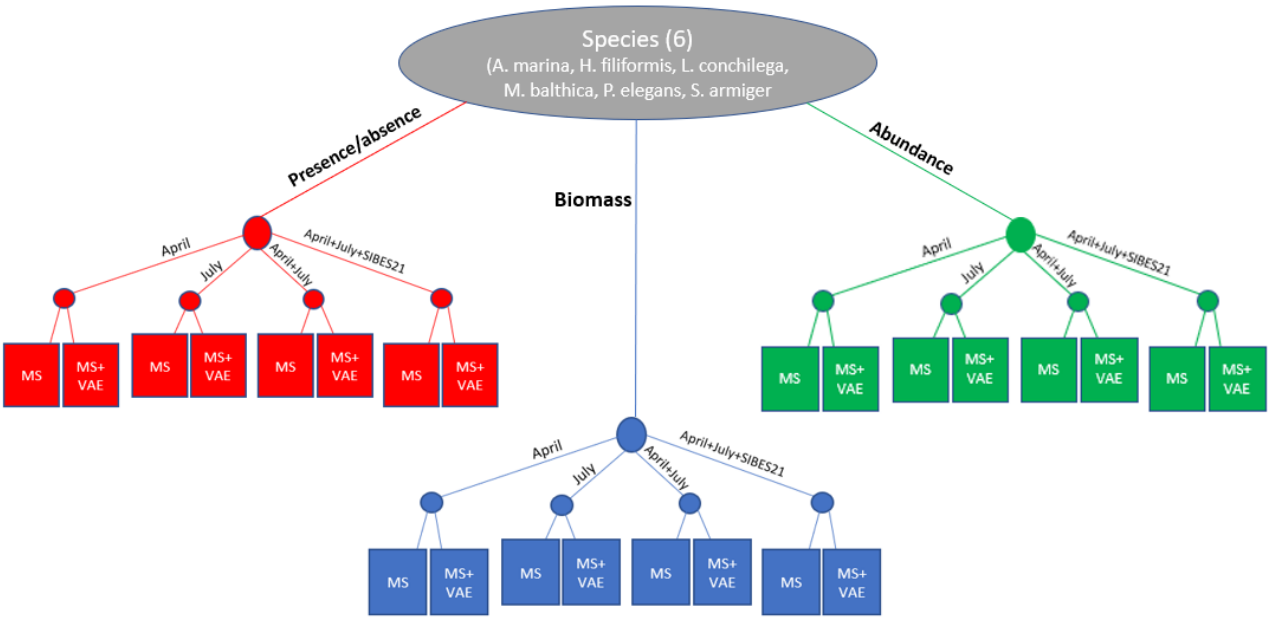


Figure 6. *Outputs for each tested species. There are three prediction types per species; presence-absence in red, biomass (AFDM) in blue and abundance in green. There are outputs for four seasonal variables and two feature variables inputs. The square boxes represent the modeled prediction outputs with the used feature settings. MS: output from multispectral features, MS+VAE: output from multispectral and variational autoencoder features.*

2.6.4 Accuracy assessment

We predetermined the testing and training data points rather than using random data to ensure both sets contain sufficient data variance to train and test the models. This approach was also applied in a study by Puls et al. (2012). The April and July predictions use the same training and testing sample locations, this is different for the SIBES 2021 data as it was sampled at different locations than the 2022 seasonal data (Figure 5). April and July both have 14 test sample locations, April has 41 training samples and July has 37 training samples. SIBES 2021 has 11 test samples and 33 training samples. Thus, the training to testing ratios of the separate sampled seasons are: April 0.75 to 0.25, July 0.73 to 0.27 and SIBES 2021 0.75 to 0.25.

For the presence-absence random forest classifier we use the accuracy score function from scikit-learn metrics and for the random forest regressor we use the R-squared function to assess the prediction accuracies of both the training and testing data. Additionally, we plot the testing and training predictions per species per type of prediction in observed/predicted plots for the different seasonal runs (Mayer &

Butler, 1993). The prediction model for presence-absence is also applied to the raster data of the entire study area to create seasonal distribution maps per species.

3. Results

A subset of six species is selected from the collected field data. These six species are determined to be sufficiently occurrent and abundant throughout the samples of all seasons, to train the random forest models. The modeled predictions per species for presence-absence, biomass and abundance show different patterns in terms of test data prediction accuracies over the different seasonal runs and feature inputs. The accuracy scores of the presence-absence predictions, and R-squared values of the biomass and abundance predictions are displayed by bar charts, and as tabular data in Appendix A. The presence-absence model is used to construct distribution maps of all tested species for April and July 2022 (Appendix B). We also plotted the observed values against the predicted values for the biomass and abundance models to further inspect these predictions (Appendix C and D).

3.1 Field sample data

The three seasonal sampling campaigns result in an extensive amount of data about the presence of different species or other taxonomic units, and their biomass and abundance per sampling location. A total of 52 different OTU's were found in April, 58 in July and 37 in the SIBES samples from September 2021. April samples have an average abundance of 104 individuals per sample over 55 samples, regardless of OTU. July samples have an average of 129 individuals per sample over 51 samples, and the SIBES samples have an average of 63 individuals per sample over 45 samples **(Table 2).** The present study uses a subset of six species from the collected data. The April and July subsets have fewer average individuals per sample than the complete campaign dataset, whereas the SIBES 2021 subset has more than double the average amount of individuals of the entire dataset. The same trend is applicable to the average biomass per sample. Nevertheless, the selected species for the subset were more suitable than other observed species and OTUs, as they are present in most samples throughout the different samples seasons (Figure 7).

Table 2. *Statistics of all gathered sample data in the study area during three different campaigns: number of sampling locations, average individuals per sample, average AFDM per sample in grams and the total identified OTU's per sampling campaign or subset. Statistics are displayed for complete sample campaigns and for the selected species subsets.*

Samples type	Sample locations	Av. indivs. per sample	Av. AFDM per sample (g)	Identified OTUs
April 2022	55	104	0.58	52
April subset	55	73	0.34	6
July 2022	51	129	0.86	58
July subset	51	72	0.69	6
SIBES 2021 (September)	45	63	0.13	37
SIBES subset	45	141	0.67	6

3.1.1 Species occurrence

The occurrence percentages of the selected subset do show variance in occurrence among species and tested seasonal combinations (Figure 7). Pygospio elegans and Scoloplos armiger have the highest occurrence percentages throughout the seasonal combinations with average occurrences of 81% and

80% over all seasonal runs. The average occurrence of Heteromastus filiformis is 69%, for Arenicola marina it is 61%, then Lanice conchilega has an average occurrence of 52% and finally Macoma balthica has an average occurrence of 47%. There is also variance in occurrence between seasons, which is expected due to the seasonality of macrozoobenthic development. The largest difference between April and July occurrence is observed with Macoma balthica at occurrence percentages of 42% in April and 59% in July. The smallest difference between April and July is observed for Heteromastus filiformis at 75% and 71% respectively. Furthermore, the observed occurrence percentages of the seasonal variation which includes SIBES data (A+J+SIB) is lower than the April and July occurrence for Arenicola marina, Heteromastus filiformis and Macoma balthica.

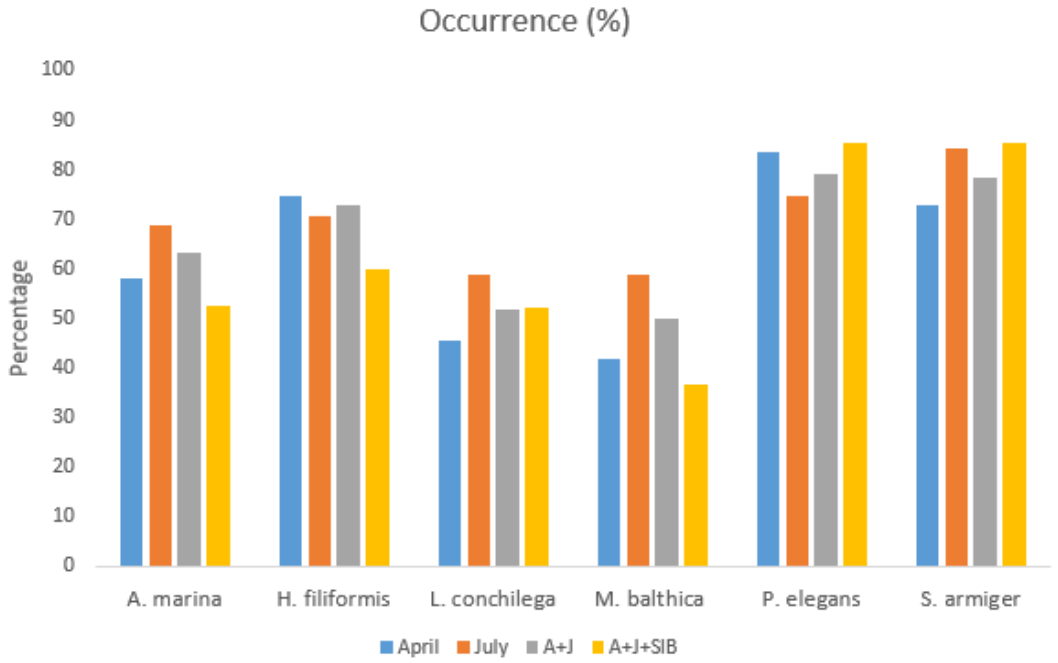


Figure 7. *Species subset occurrence percentages of the field samples for the tested seasonal inputs: April, July, April and July (A+J) and April, July and SIBES (September) 2021 (A+J+SIB).*

3.1.2 Ash free dry mass

The ash free dry mass data distribution the species subset shows strong skewness and 0-inflation (Figure 8). Skewed data distributions affect the performance of the random forest regression model, as the output of the model tends to represent a normal distribution. Therefore, we applied a logarithmic transformation to the biomass data to improve the random forest regression prediction results (Figure 9) **(Madhuanand et al., 2023)**. During this transformation we kept the measured 0-values as 0 in the used data, to ensure enough input data remains to train the model. It also provides the opportunity to see if the model can also predict 0 or low biomass values well. The biomass data is more normally distributed after the logarithmic transformation. We see that there is still 0-inflation, but it appears less strong especially for Pygospio elegans. This is because different scale on the x-axis, as fewer values are rounded into the 0-included values bar of the histogram. It strongly affects the histogram of Pygospio elegans, as the AFDM of Pygospio elegans is especially low per individual. However, the transformation does not affect the true amount of 0-values.

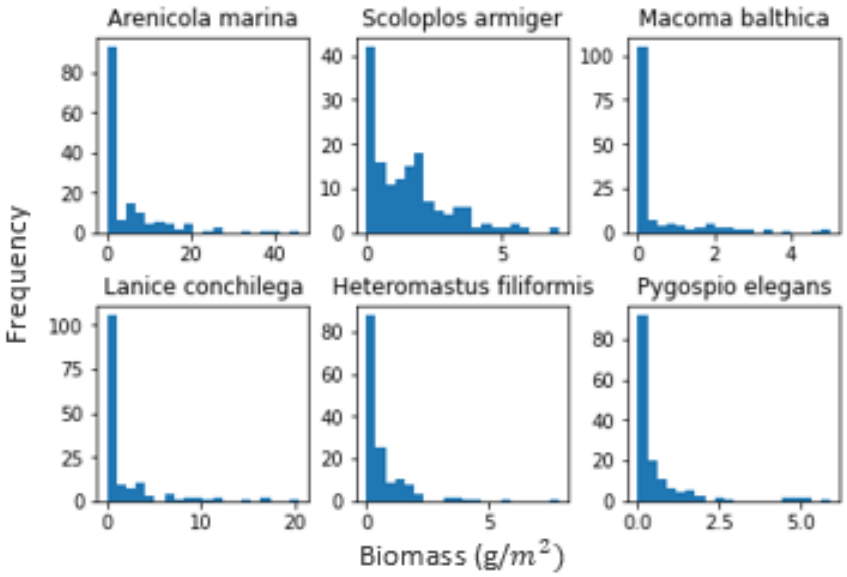


Figure 8. *Ash free dry mass (biomass) distributions per species per sample in frequencies and g/m^2.*

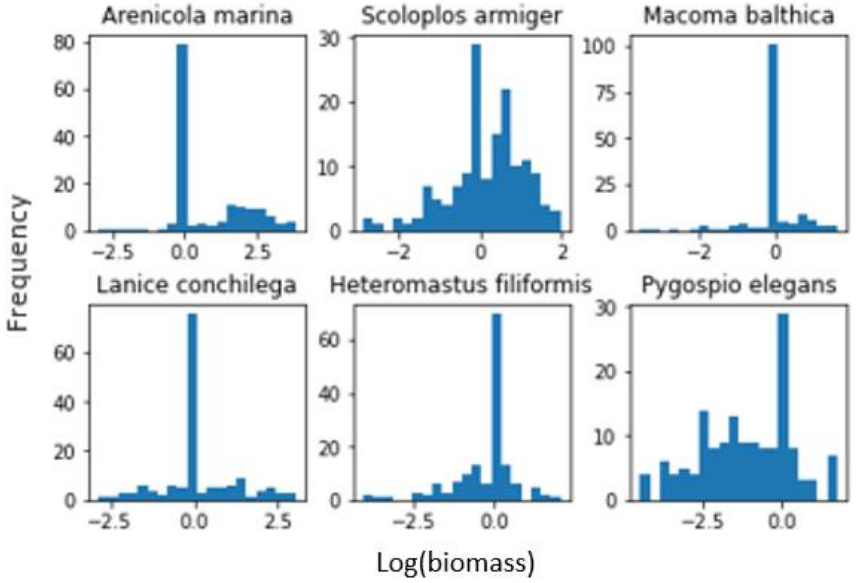


Figure 9*. Ash free dry biomass distribution per species after logarithmic transformation as log(biomass (g/m^2)).*

3.1.3 Abundance

The distribution of the measured abundance data is also skewed and 0-inflated for all selected species (Figure 10). Therefore, we executed the same logarithmic transformation for the abundance data (Figure 11). The logarithmically transformed data is more normally distributed than the original data. However, the 0-values are located left of this distribution. After the transformation the zero-inflation appears to decrease for all species, which is again due to the different x-axes since we kept the 0 abundance values constant throughout the transformation.

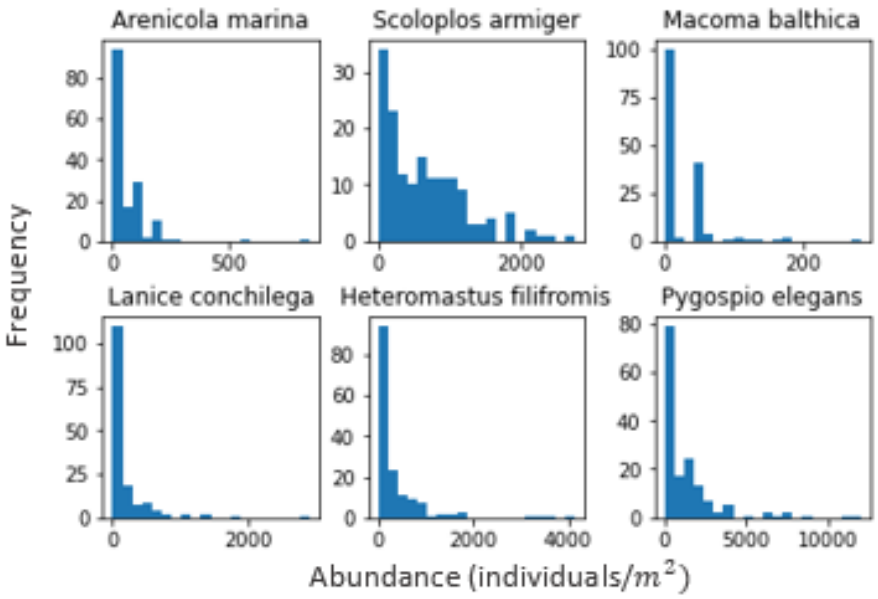


Figure 10. *Abundance data distributions per species per sample in frequencies and individuals per m^2.*

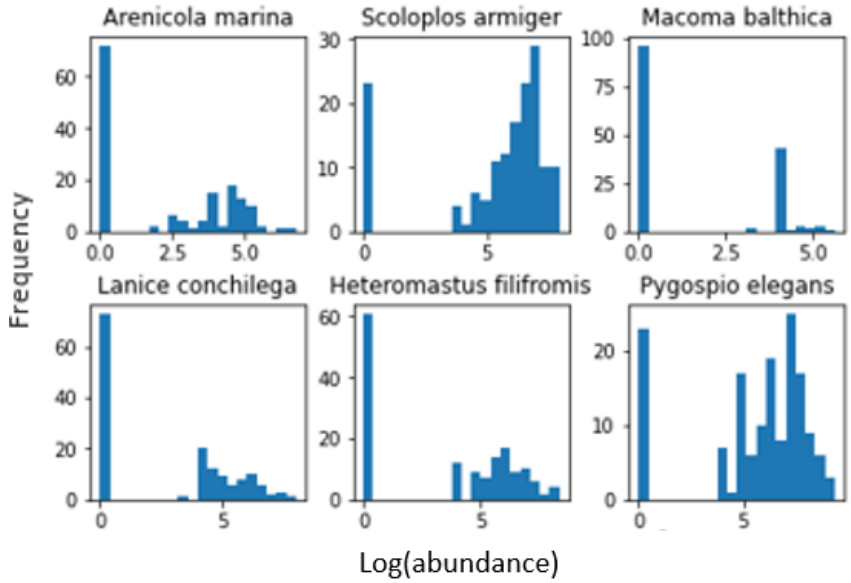


Figure 11. *Abundance data distributions per species per sample after logarithmic transformation.*

3.2 Raster data

Planet multispectral 4-band satellite images are used for the remote sensing and VAE modelling part of the study. The images must be taken at low tide and be free of clouds, especially the sampling locations as these are used to train and test the models. When the predictions are used to create distribution maps the images of the study area must contain as little noise as possible.

3.2.1 Planet satellite images

Three satellite images are selected to apply to the VAE model and use as features in the random forest models. The selected satellite images of the study area show some differences among them (Figure 12).

The image selected to match the SIBES samples from September 2021 was taken on September 5th 2021. It shows some morphological features such as dark channels in the upper part, but also a dark and slightly noisy bottom part (Figure 12 A). However, there are hardly any sampling locations in these areas. The image to match the July campaign is from July 12th 2022. It is also fairly dark and contains a lot of noise and haze on the right side, which does not contain any sampling locations. There is some noise due to clouds on the left side; some of which is located at the sampling locations, and may therefore affect the pixel values we use for the random forest model (Figure 12 B). The image for the April 2022 sampling campaign appears more hazy than that of September 2021, but it has fewer dark sports than the September and July images (Figure 12.C). This is an indication that the tide was likely lower when the April 2022 image was taken compared to the September 2021 image. For each sampling campaign the best available images close to the dates of sampling are selected to represent the study area. For example, the image from the 24th of April is closer in date to the sampling campaign of April from the 26th to the 28th. However, the overall quality of this image is lower, as it contains more noise and the tide was higher at the time the image was taken compared to the April 23rd image (Figure 12.D).

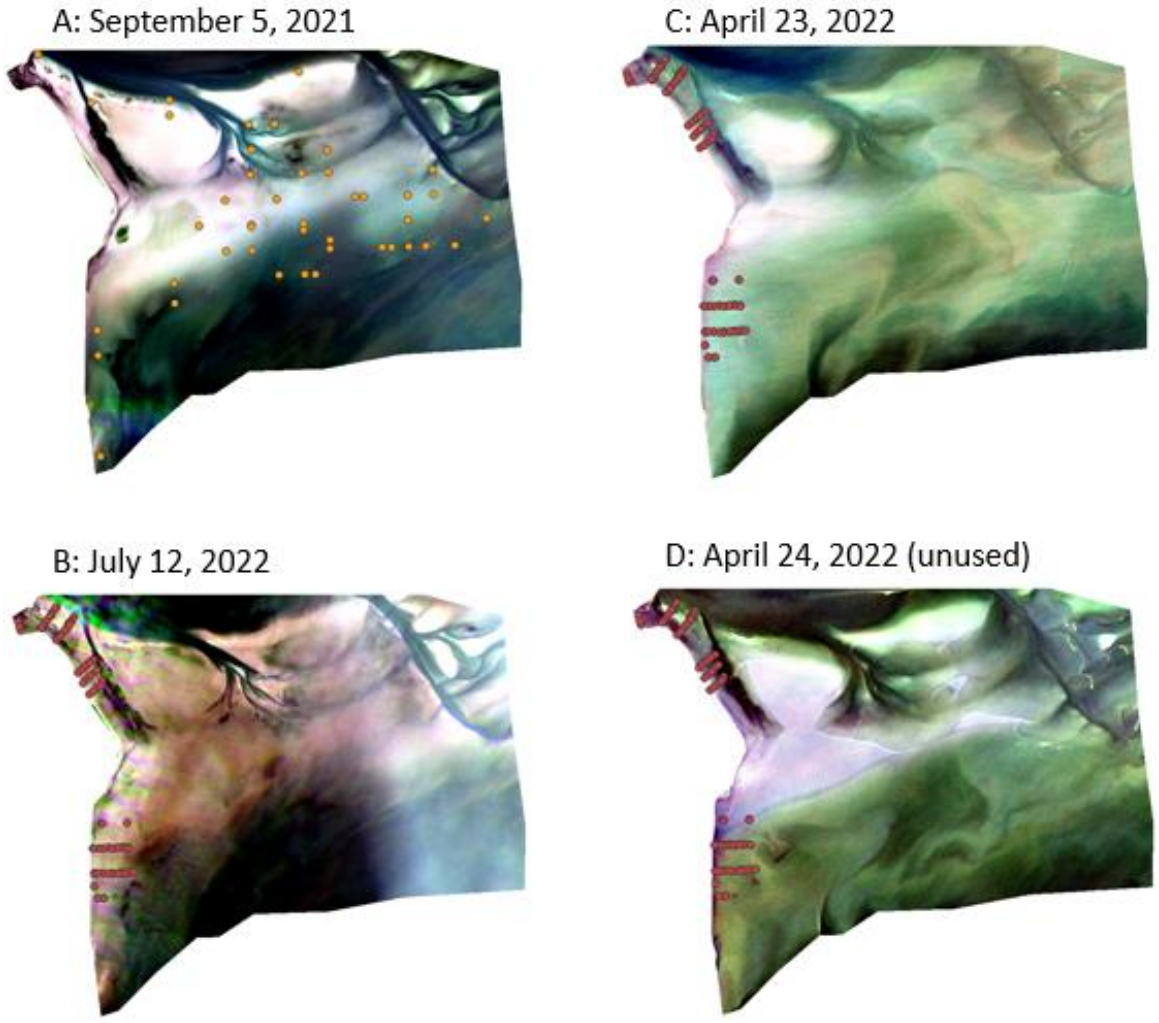


Figure 12. *Planet unharmonized satellite images (RGB) from the study area and their corresponding dates. A is used for the September 2021 data, the SIBES sampling locations are displayed in orange. B is used for the July 2022 data and C for the April 2022 data, the sampling locations are displayed in red. D is an example of an unused image.*

3.2.2 VAE model features

The variational autoencoder model delivers 64 new feature layers per satellite image. This output consists of raster data from the features made by the variational autoencoder model to recreate the original input image. As is displayed by Figure 13, some morphological features can be distinguished in different layers, such as layers (bands) 13, 17 and 26. However, not all layers appear to contain information about the area, such as layers 7 and 10. The meaning of the individual features from the VAE model are difficult to visually interpret, due to the large quantity of layers and their abstractiveness. However, the layers can easily be applied as input features to random forest models. The random forest models also determine feature importance, therefore the layers which contain less information do not affect the final prediction results.

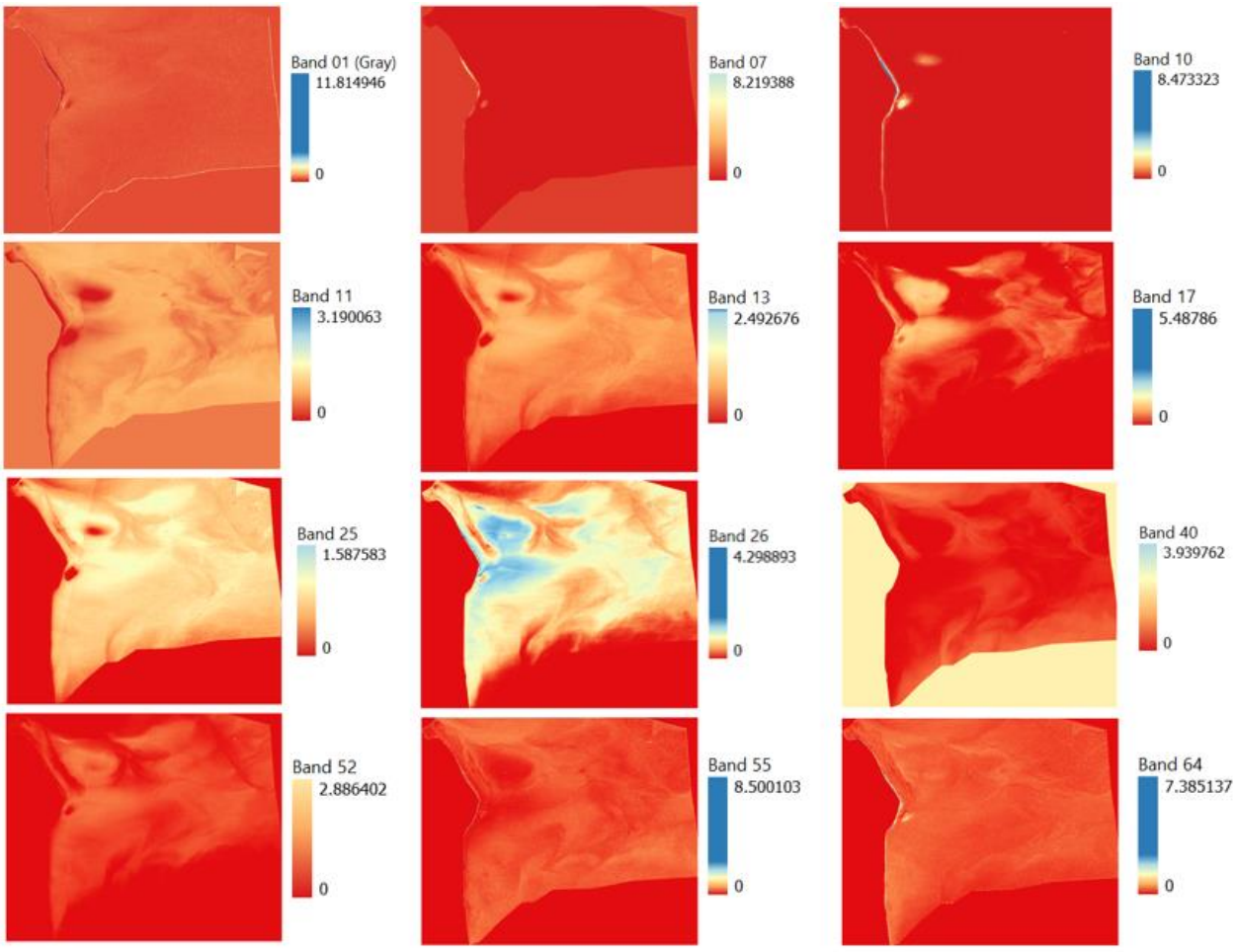


Figure 13. *Nine example layers of the VAE model features extracted from the satellite image of April 23rd 2022. Different layers show different recognizable structures from the original satellite image (Figure 12 C).*

3.3 Presence-absence

The first random forest model is a classifier model which predicts the presence and absence of the species subset at the sampling locations. The test data accuracy scores are calculated for the presence-absence predictions (Figure 14). We differentiated the input features into four seasonal categories: The April and July runs use strictly feature data from the sampling campaign in question. The April+July and April+July+SIBES21 runs use a combination of field sample and raster feature data from all seasons

included in the runs. Furthermore, we performed all runs with two different feature inputs: (1) strictly 4-band multispectral features and (2) both 4-band multispectral and 64 VAE features.

3.3.1 Species performance

The single-season runs, April and July, have high variability in accuracy scores between the different species. Over all seasonal runs, the April Macoma balthica prediction achieves the lowest accuracy score for both tested feature inputs. The April Scoloplos armiger prediction model including the VAE model features achieves the highest accuracy score. These scores of Macoma balthica and Scoloplos armiger represent the largest observed difference between accuracy scores across all presence-absence model variations. The accuracy variability between species is smaller for the multi-seasonal runs than the single-season runs, with April+July+SIBES21 having the least variability.

Overall the presence-absence accuracy scores are reasonable to good (>0.5), with the exception of Macoma balthica and Lanice conchilega. These are also the species with the overall lowest occurrence percentage in the samples (Figure 7). The accuracy score of Macoma balthica for the April runs is 0.43 for both tested feature inputs. Furthermore, it is only slightly above 0.5 for the April+July runs at 0.57 and 0.54; solely multispectral and VAE model features included respectively. Macoma balthica is the lowest performing species for the April+July+SIBES21 run.

The model also achieves relatively low accuracy scores for Lanice conchilega in April at 0.64 and 0.5 for multispectral and VAE-included input features respectively. The April+July accuracies for Lanice conchilega are also low at 0.64 and 0.57. These weaker performing species and runs also show larger differences between the accuracy scores of the training and testing data predictions (Appendix A.1). Arenicola marina has low test accuracy scores for July at 0.57 for both feature inputs, but performs better in the other seasonal runs. The model performs well for the remaining species. The highest accuracy scores throughout the different seasonal runs are achieved by Pygospio elegans and Scoloplos armiger (> 0.7). Heteromastus filiformis also achieves accuracy scores above 0.7, except for the July runs which deliver scores of 0.64 for both feature inputs.

The training accuracy of the model runs is higher than the test accuracy in all cases except Heteromastus filiformis April 4-band and all features, and Macoma balthica July all features: 0.86 test compared to 0.84 train (Appendix A.1). Overall, the difference between test and train accuracy is within reasonable range, as the training data accuracy score mostly stays well within a 0.2 range higher than the testing data accuracy scores.

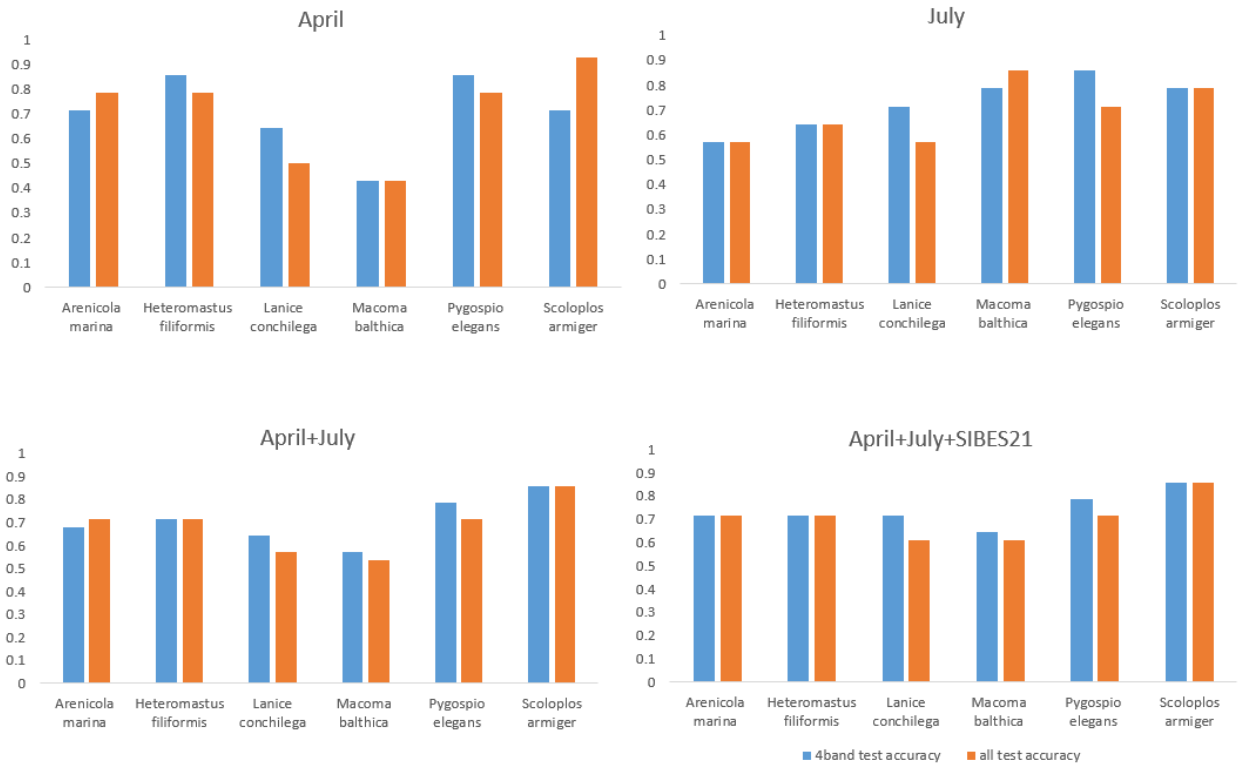


Figure 14. *Presence-absence model test data accuracy scores per species for the different seasonal and feature input data. 4-band multispectral data in blue and multispectral data + 64 VAE model features in orange.*

3.3.2 Distribution maps

The presence-absence prediction models of April and July are used to create distribution maps of the study area. This results in distribution maps of all tested species for April and July (Appendix B). The distribution maps of April show overall plausible distribution patterns for all species which coincide with the observed occurrences in the field samples. Contrastingly, the predicted distributions displayed by the July maps are unlikely. Arenicola marina, Heteromastus filiformis and Macoma balthica are hardly present. Furthermore, the locations where they are predicted to be present all coincide to the same area. Lanice conchilega, Pygospio elegans and Scoloplos armiger show an opposite distribution pattern, as they are predicted to be present in nearly the entire study area. This is understandable for Scoloplos armiger, as the observed occurrence in the July samples is 84%. However, the observed occurrence for Lanice conchilega is lower at 59%. The April distribution map of Pygospio elegans also shows it is predicted to be present across nearly the entire study area (Figure 15). This predicted distribution is understandable, as Pygospio elegans is observed in 84% of the April samples. The July map of Pygospio elegans shows less presence in the study area, which matches the lower observed occurrence of 75%.

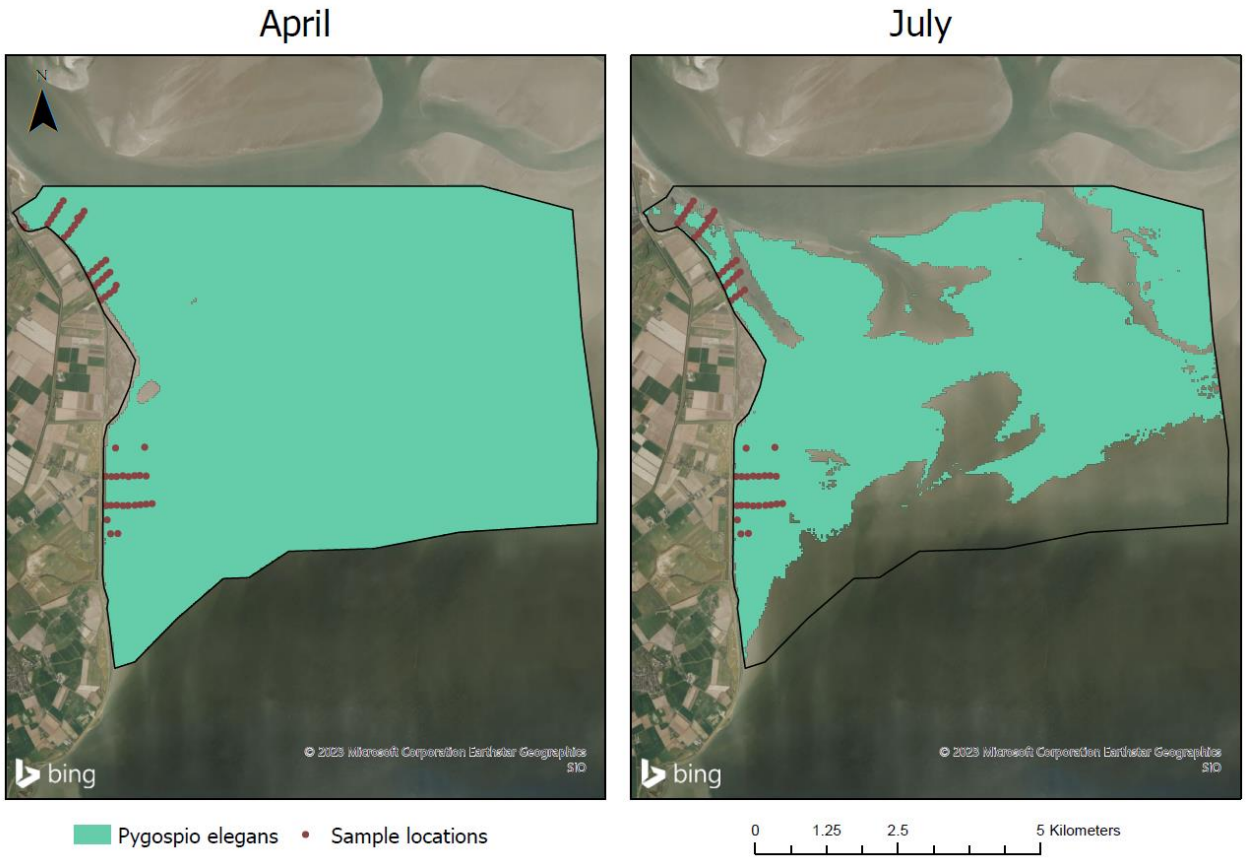


Figure 15. *Predicted presence distribution of Pygospio elegans for April 2022 (left) and July 2002 (right).*

3.4 Biomass

We created RF regression models with the same variables as the presence-absence runs. We altered the RF classifier with presence-absence data to a RF regressor with the logarithmically transformed measured biomass data per species. We use the R-squared values of the different predictions to assess the performance of the executed model runs with the different variables (Figure 16). Upon first glance, the performance patterns are rather different compared to the presence-absence prediction accuracies. The biomass test data R-squared are low across al species. In many cases the testing R-squared value is negative and shows little relation to the training data R-squared values (Appendix A.2)

3.4.1 Species performance

The R-squared values for the biomass test data show that the model as is, is inadequate for biomass predictions (Figure 16). Arenicola marina is somewhat of an exception as it produces (slightly) positive R-squared values for all runs except July. Arenicola marina achieves the best results for the April run with values of 0.29 for only multispectral features and 0.21 with VAE included features. Macoma balthica has an R-squared outlier for the April multispectral run of 0.43. All other R-squared values for Macoma balthica however are negative.

Based on their performance of both the biomass and abundance predictions, the biomass prediction results of Arenicola marina, Lanice conchilega and Pygospio elegans are plot in observed/predicted plots (Appendix C). The scatter plots of Arenicola marina in Appendix C.1 show that the biomass predictions

are within a smaller range of values than the observed data. This effect becomes more evident for the multi-seasonal runs for both training and testing data. We can see compared to April, the multi-seasonal test and train R-squared values decrease as well as the spread of the predicted data. The observed/predicted plots of Lanice conchilega and Scoloplos armiger show little correlation between the observed and predicted test data (Appendix C.2 and C.3). However, the training data has an R-squared of nearly 0.5 or higher for most runs. This indicates that the model likely overfits the training data, while it is not capable of actually learning to predict the test data. It is therefore difficult to interpret patterns and differences between the seasonal and feature input variables.

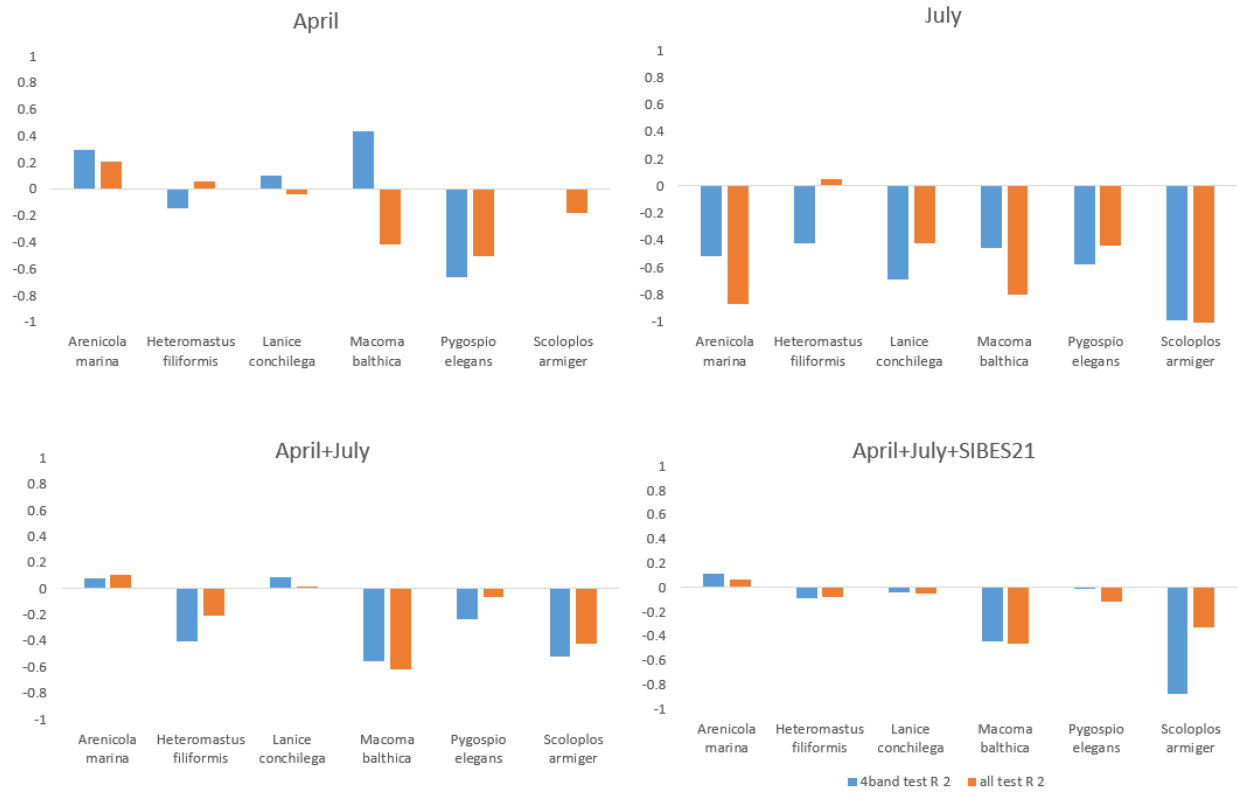


Figure 16. *Biomass test data R-squared values per species for the different seasonal and feature input data. 4-band multispectral data in blue and multispectral data + 64 VAE model features in orange.*

3.5 Abundance

The abundance predictions are also assessed by plotting the R-squared values of the different runs in bar charts (Figure 17). Upon first glance the predictions are improved compared to the biomass model. However, overall there is a high variance between the different species and seasonal runs. There are more positive R-squared values compared to the biomass regression, but there are still multiple negative and near-zero R-squared values.

3.5.1 Species performance

The Arenicola marina prediction performances have a similar pattern compared to the biomass runs; all R-squared values are positive except for July. Heteromastus filiformis, Macoma balthica and Pygospio elegans are predicted poorly; most R-squared values are near 0 or negative. Lanice conchilega achieves positive R-squared with the April predictions using multispectral features, and for both feature inputs

with the remaining seasonal variables. Scoloplos armiger has positive R-squared values for all runs except the July run using multispectral features. Compared to the biomass models, the abundance predictions provide more species that achieve positive test data R-squared values**.** The Scoloplos armiger test data R-squared values for the predictions including VEA model features are noteworthy at 0.57 and 0.62 for the April and April+July+SIBES21 runs respectively. The R-squared values of the abundance predictions of Scoloplos armiger are overall relatively high with the exception of the July runs. The average test data R-squared value of the Scoloplos armiger predictions including VAE model features is 0.44. The average R-squared value of the Scoloplos armiger training data predictions is however considerably higher at 0.80. The predictions based on multispectral features achieve lower R-squared values, with averages of 0.14 for the testing and 0.71 for the training data (Appendix A.3). Although the results of the abundance models are improved compared to the biomass models, there is still visible bias throughout the observed/predicted scatter plots **(Appendix D).** The bias does not particularly increase of decrease with the multi-seasonal predictions compared to the single-season predictions, but the patterns become more evident. Also, 0-values are poorly predicted and the scatter plots show a high variance of predicted values for observed 0-values. The observed/predicted plots of the different runs of Scoloplos armiger, as the best performing species of the abundance prediction model, also display bias for the higher observed values and poorly predicted observed 0-values (Figure 18).

Figure 17. *Abundance test data R-squared values per species for the different seasonal and feature input data. 4-band multispectral data in blue and multispectral data + 64 VAE model features in orange.*

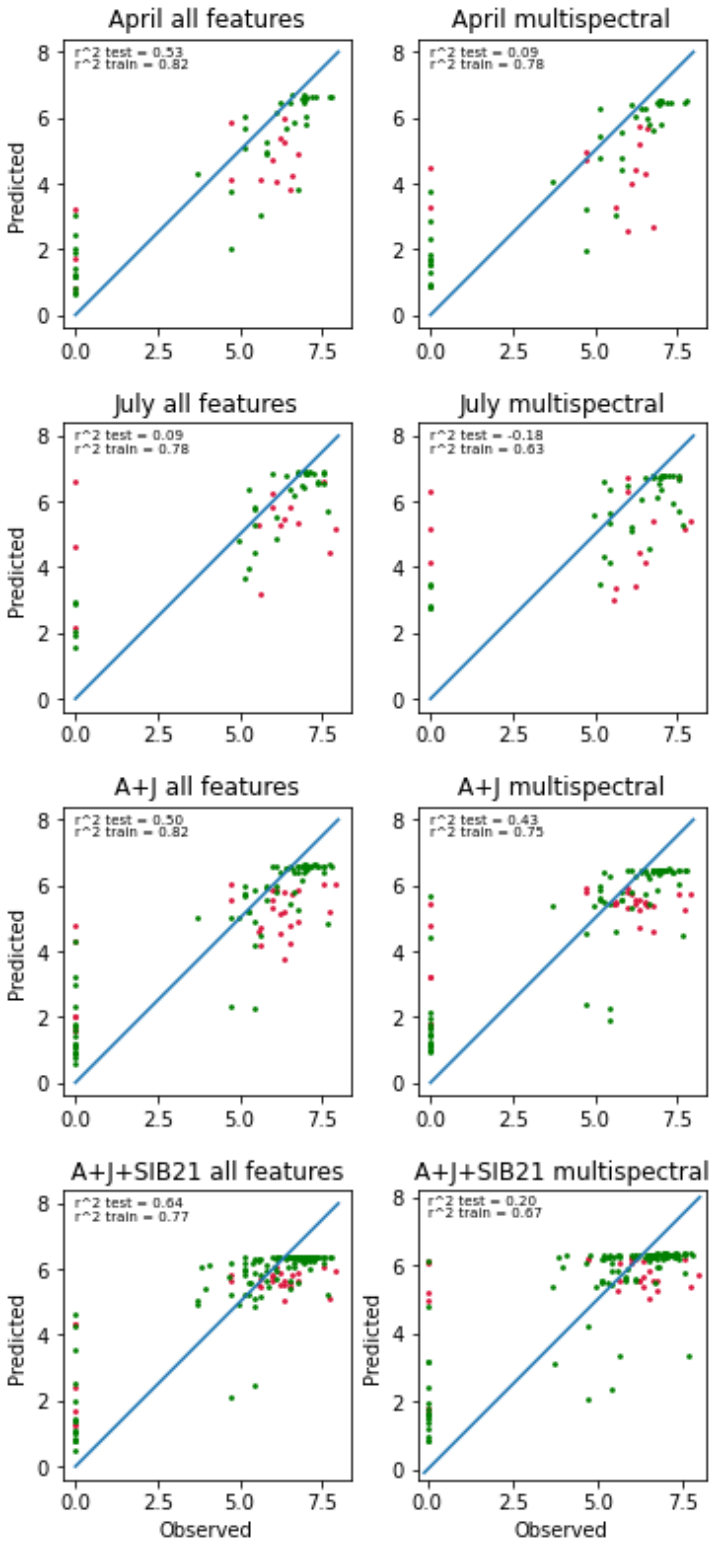


***Figure 18**. Observed / Predicted plots of the different predictions for Scoloplos armiger. Test data is displayed in pink and train data is displayed in green.*

3.6 Seasonal prediction patterns

Although the overall performance of the three prediction models appears quite different, there are similar patterns across these models between the single- and multi-seasonal prediction performances.

Overall, the multi-seasonal runs of the presence-absence predictions perform similarly compared to the single-season runs. However, the multi-seasonal runs show less test accuracy variability between different species (Figure 14). The difference between the training and testing data prediction accuracies also remains similar (Appendix A.1). It is noteworthy that the April+July+SIBES21 produces accuracy scores higher than 0.7 for four out of six species: Arenicola marina, Heteromastus filiformis, Pygospio elegans and Scoloplos armiger. The remaining two species, Lanice conchilega and Macoma balthica, still achieve test data accuracies of above 0.6. The April predictions also produce accuracy scores above 0.7 for the four abovementioned species, but here Macoma balthica has an accuracy score of 0.43 for both feature inputs and Lanice conchilega accuracy scores of 0.64 and 0.5 for respectively multispectral and VAE-included features. The overall performance of the April+July+SIBES21 runs is therefore better than of other seasonal variables.

The biomass test R-squared values are slightly positive only for Arenicola marina (Figure 16). The observed/predicted plots of Lanice conchilega and Scoloplos armiger show little correlation between the observed and predicted test data (Appendix C). Contrastingly, the training data has an R-squared of nearly 0.5 or higher for most runs. This indicates that the model likely overfits the training data, while it is not capable of learning to predict the test data. It is therefore difficult to derive patterns and differences between the seasonal variables.

The test data R-squared values of the abundance predictions show no clear pattern or trend in performance of the different seasonal variables across the different species (Figure 17). We see that the multi-seasonal runs of Arenicola marina and Lanice conchilega provide improved results compared to the single-season runs. The Scoloplos armiger test data R-squared values of the multi-seasonal predictions using VAE model features are slightly lower than the April, but higher than the July predictions. Also, the predictions of species which achieve a positive R-squared value for the April run, achieve a negative or near to 0 R-squared value in July. Lastly, the multi-seasonal runs display similar R-squared patterns across all species compared to the April predictions in terms of positive and negative R-squared values.

3.7 Multispectral and VAE model features

We run each species and seasonal combination with the two different feature inputs. The difference between the predictions using only multispectral features and VAE-included features are relevant to assess the proposed modelling method using VAE model extracted features.

The presence-absence model runs including VAE model features perform well but so do the runs with solely multispectral data (Figure 14). The presence-absence runs including VAE model features perform better than solely multispectral features in the cases of: Arenicola marina for April and April+July, Macoma balthica for July and Scoloplos armiger for April. Contrastingly, the runs using solely multispectral feature input achieve higher accuracy scores than the runs including VAE model features in the cases of April+July and April+July+SIBES21 for Lanice conchilega, Macoma balthica and Pygospio elegans. The performance differences between strictly multispectral and VAE model included features decreases for the multi-seasonal predictions. Overall, the runs including VAE mdoel features perform similar in predicting the presence and absence of the tested species compared to the multispectral data.

There are 13 out of 24 cases where the VAE model additional features produce higher or equal prediction accuracies compared to the run with solely multispectral features.

The effect of the VAE model features on the biomass predictions is difficult to interpret due to the low R-squared values of the test data (Figure 16). Compared to the test data, the training data achieves relatively high R-squared values of around 0.5 (Appendix A.3). Due to the large difference between the testing and training data prediction result we cannot derive trends or patterns within the predictions using multispectral or VAE-included features.

The abundance prediction results show different performance patterns among the predicted species regarding the input feature types (Figure 17). For the April predictions of Arenicola marina the multispectral run performs best with test data R-squared values of 0.16 with multispectral features, compared to 0.05 with additional VAE model features. However, for the multi seasonal runs the VAE included features perform better; 0.11 compared to 0.24 for April+July and 0.03 compared to 0.26 for April+July+SIBES21. There is no distinct difference in performance between the feature inputs of Lanice conchilega; April and July show opposing results. The test data R-squared value of April+July is higher for the multispectral run: 0.32 compared to 0.23 for the run including VAE model features. For April+July+SIBES21 the performances are similar: 0.19 and 0.21. Scoloplos armiger has higher R-squared values for the VAE-included features in every seasonal variation. The April+July predictions multispectral and VAE-included R-squared values are closer together compared to the other seasonal variables: 0.44 compared to 0.5. The best results for Scoloplos armiger are achieved by the April+July+SIBES21 run including VAE model features, with a test data R-squared value of 0.64. The difference between the test data R-squared and training data R-squared of the two variable features remain similar, with the exception of Scoloplos armiger for the April+July+SIBES21 run. Here, the difference of test and train R-squared values the multispectral features is 0.47, whereas the for the predictions including VAE model features it is 0.13 (Appendix A.3).

4. Discussion

The continuous developments and improvements of deep learning structures and their applications for remote sensing provide interesting opportunities for ecological modelling in areas with poor spectral signatures such as intertidal areas. Increasing amounts of available remote sensing data with high spatial and temporal resolutions and the addition of variational autoencoder and random forest models make for an attractive, time and cost efficient method of monitoring macrozoobenthic species on a seasonal scale. The aim of the present study is to model and predict seasonal variations of macrozoobenthos with extracted features from a variational autoencoder model and to assess the capabilities and limitations of this method in terms of predicting presence-absence, biomass and abundance on a seasonal scale. Additionally, we are interested in the difference of prediction accuracies of models with multispectral features and additional VAE model features. Lastly, the portability of the prediction models over multiple seasons is also evaluated by executing predictions with input data of multiple seasons at once.

4.1 Data and modelling results

The sampling locations are measured using a high accuracy Trimble receiver and the Dutch 06GPS network. The field sample data is processed in the NIOZ laboratory using the same preservation, taxonomy and weighing methods as the annual Synoptic Intertidal Benthic Survey. The ash free dry mass is weighed using high precision scales. This delivers high precision and accuracy ash free dry mass and

abundance data of all observed OTUs per sampling location. The observed field data is used as training and testing data in combination with the best available Planet multispectral satellite imagery and VAE model features to train random forest prediction models. The VAE model uses the multispectral satellite data as input and produced 64 new features per satellite image as output (Madhuanand et al., 2023). The satellite and VAE model feature data is used as in Random Forest models to predict species presence-absence, biomass and species abundance. The satellite images used for the present study were visually evaluated but show some variability among them in terms of clouds, noise and tide. The image for July contains some cloud cover at sampling locations, which can affect the modelled predictions. Upon first glance it has no observable effect on the presence-absence July prediction accuracy. However, the species distribution maps of the study area show unlikely distribution patterns for all modelled species, except Pygospio elegans and Scoloplos armiger. Furthermore, for the biomass and abundance predictions the July runs are the poorest performing seasonal run. The results of the July predictions compared to the April and multi-seasonal predictions indicate that the satellite image was less suitable to apply to the models.

This study has shown that it is possible to model the seasonal distribution of presence and absence of the six tested macrozoobenthic species. However, additional VAE model features do not provide much improved prediction results compared to models trained exclusively by multispectral features.
The single-season presence-absence predictions have larger test accuracy variance between species. The multi-seasonal runs are more stable across the different species as they provide less test accuracy variance. Both the single and multi-seasonal runs have an average test accuracy of approximately 0.7. Additionally, the predictions using only multispectral data features perform approximately equally well as the predictions including the VAE model features with an average test data predictions accuracy of 0.72 for the single season and 0.69 for the multi-seasonal predictions. We can therefore conclude that seasonal presence-absence of the six researched species can be modeled using a random forest model with remote sensing and VAE model data. However, multispectral data suffices as feature input for the model as the VAE model features do not provide improved performance. This goes against our initial expectations, since we believed the features would provide additional data which would improve the RF model training.
Furthermore, for the presence-absence predictions we found the accuracy variance between species decreases when the model is used to predict multiple seasons simultaneously. Species presence-absence predictions that performed poorly in at least one of the two single-season predictions, such as Macoma balthica in April, still achieved relatively high accuracy for the multi-seasonal runs. We can therefore conclude that the model can predict seasonal presence and absence of the species concerned by performing a single model run which contains the data of multiple seasons.

The seasonal distribution modelling of biomass and abundance proves to remain a challenge, with overall weak prediction results. The biomass models showed an overall weak performance for all species throughout the different seasons, regardless of the input features. This indicates that the random forest model could not find a relation between the input features and the biomass per species at the sampling locations.
The results of the abundance models are slightly improved compared to the biomass models for Arenicola marina and Lanice conchilega, but also to remain a challenge. The Abundance predictions of Scoloplos armiger are noteworthy, with the highest test data R-squared value being 0.64 for the VAE model included features April+July+SIBES21 run, compared to 0.2 for the multispectral features run. In

contrast with the presence-absence models, the highest test data R-squared values were only achieved by the runs including the VAE model features. Therefore, we can conclude that the random forest model found a relation between the VAE features and the abundance of Scoloplos armiger, with the multi-seasonal runs performing best. For Arenicola marina and Lanice conchilega the best results were also achieved by the April+July+SIBES21 runs including VAE model features. However, the test data R-squared values for Arenicola marina and Lanice conchilega are considerably lower compared to those of Scoloplos armiger. The test data R-squared value is 0.27 for both Lanice conchilega and Arenicola marina, whereas the training data R-squared values are 0.67 for both Arenicola marina and Lanice conchilega; a large gap between the test and train data prediction results remains. This indicates that the model requires more training to improve the predictions.

Arenicola marina is known as a large bioturbator (Compton et al., 2013). Due to its size and bioturbating potential the VAE model may be able to find features that help recognizing this species. Although the abundance prediction results of Arenicola marina are unsatisfactory, they do show improvement with increasing input data. It is therefore likely that the model can improve the detection of Arenicola marina abundance with additional training and testing data. Lanice conchilega, or sand mason, is known as an ecosystem-engineering species which can form dense aggregations (Van Hoey et al., 2008). These tubes protrude the intertidal flats, which we also observed during the sampling campaigns. We therefore expected VAE model to be able to extract features which enhance detection of Lanice conchilega, and therefore perform well for the presence-absence as well as abundance modelling. However, the performance of the Lanice conchilega presence-absence prediction model was among the lowest, and also the abundance models struggled to predict for Lanice conchilega.
Scoloplos armiger is a widely distributed species across the Wadden Sea and is known to inhabit areas with coarse sediment (Folmer et al., 2017; Kraan, 2010). The abundance predictions based on the model including the VAE model features consistently outperform the predictions based on multispectral features. Furthermore, the difference between the test and training data R-squared values is smaller for the predictions using VAE model features. Scoloplos armiger also showed the most normal-like sample data distributions after the logarithmic transformation, together with Pygospio elegans. However, Pygospio elegans is a considerably smaller species than Scoloplos armiger. For this reason we believe the model achieves better predictions for Scoloplos armiger and weak predictions for Pygospio elegans, even though they were the two most abundant species in the sample data.

4.2 Study limitations

The satellite image that was used for the July campaign contains some noise and clouds. Especially the results of the biomass and abundance models show that the July predictions are weaker than the April and multi-seasonal predictions. Furthermore, the presence-absence distribution maps in appendix B show unlikely distribution patterns on the July maps. We therefore believe that the quality of the image for July was insufficient. However, the multi-seasonal prediction models, including the July data, still show relatively high accuracy and R-squared values compared to the July predictions. The study is limited by the available satellite imagery to match the sampling campaigns, as they need to be taken at low tide under favorable weather conditions.
The effect of bias and 0-inflation becomes evident by the biomass and abundance observed-predicted plots in appendices C and D. Scoloplos armiger has the highest occurrence in the field samples and therefore had the least amount of 0 values in the training and testing data for both biomass and abundance (Figure 8-11). This can be a reason to explain that Scoloplos armiger is the best performing

species of the abundance regression model. Pygospio elegans was also present in many samples, however this species is relatively small in size. We therefore suspect that the models cannot learn to detect the presence of Pygospio elegans well. These results coincide with a species distribution modelling study by Folmer et al. (2017). The study also yielded high R-squared values for Scoloplos armiger and low R-squared values for Pygospio elegans biomass and density predictions, even though they were both abundant in the field samples. The biomass and abundance sample data distributions show zero-inflation. The observed/predicted plots in Appendix C and D show that the models struggle to predict the highest and lowest measured data values; especially the 0-values. This bias does not increase or decrease with the runs using multi-seasonal data. This data and prediction behavior was also observed in sample data and species prediction models in the study by Folmer et al. (2017). Zero-inflation of the species data, except Scoloplos armiger and Pygospio elegans, in combination with large differences between training and testing data R-squared values are indicators that the models are not able to learn from the input data and will likely improve with additional sampling data. Due to the observed zero-inflation, it is interesting to consider using more seasonal field data to provide additional available training and testing data per species. This will provide the opportunity to remove the 0-values from the biomass and abundance model training. Then, applying a mask to the biomass and abundance models using the presence-absence model can be tested for mapping. We expect this will increase the performance of the abundance prediction models, and decrease the difference between the training and testing data R-squared values. We decided against removing the 0-values for the present study, as there is already a relatively small amount of sampling data to apply to the models. Removing the 0-measurements would further decrease the input data. We believe this will compromise the performance of the RF models more than including them.

4.3 Macrozoobenthos mapping

Previous studies have mapped and modelled macrozoobenthos distributions using different methods. Features from variational autoencoder models are also increasingly being applied to classification problems. A study by Madhuanand et al. (2023) applied the same variational autoencoder model to Sentinel-2 images to predict environmental and ecological variables in the Dutch Wadden Sea. The features of the VAE model showed distinct and varying patterns as well as layers consisting of solely null values. The models showed an average prediction improvement of 15% with the variational autoencoder features compared to the predictions using multispectral data. The predictions including VAE model features consistently outperformed the multispectral features, which is a different result compared to the present study. The study also found that the prediction capability of environmental variables was higher than for the tested ecological variables, such as total biomass. The VAE model features also did not improve the prediction accuracies of total biomass and species richness as considerably as the environmental variables, which coincides with our findings of the presence-absence model.
There also have been previous efforts to map macrozoobenthos on an intertidal flat of the Westerschelde by using environmental variables, airborne hyperspectral remote sensing and Light Detection and Ranging (LIDAR). The study by Van der Wal et al. (2008) achieved R-squared values of 0.4 for total biomass and 0.43 for species richness predictions based on ground spectra. However, high resolution LIDAR and hyperspectral remote sensing data is less widely available than multispectral data, and costlier to acquire. Their models predict biomass and species richness as a function of the measured environmental variables. The study found a relatively low R-squared value of 0.18 for the biomass predictions of Pygospio elegans, compared to an R-squared of 0.40 for the total observed biomass. The study used 39 observations taken during the summer and autumn of 2006. The best predictions were

achieved by the models using microphytobenthos biomass and sediment characteristics. The predictions of this study are dependent on *in situ* measurements, which is not a feasible method for seasonal monitoring of the Wadden Sea.
A study by Puls et al. (2012) studied the relationship between benthos field data and environmental data in 308 samples from the German Wadden Sea. They modeled benthic community occurrence and structures using a multinomial logistic regression model. This study also used a spatially predetermined testing and training split for the sample data, and found that models using hydrodynamic predictors performed significantly better than with sediment predictors. The accuracy of the predictions, given by Cohen's kappa coefficient, varied between 0.14 and 0.49. Furthermore, a study by Folmer et al. (2017) used SIBES data in combination with six physical predictor variables to create annual species distribution models of macrozoobenthos across the Dutch Wadden Sea. Similar to the present study, the benthos density of Scoloplos armiger is also predicted well by the models of Folmer et al. (2017). Their models also show systematic underprediction for the higher biomass values, and overprediction of low density values. The models of the Folmer et al. (2017) study are again dependent on the temporal resolution of the SIBES sampling data including physical predictor variables. Hence the method can only be applied on an annual scale.

4.4 Recommendations

The results of the present study can be used to improve monitoring of macrozoobenthos, as it can provide a comprehensive and cost-effective supplement to SIBES. Improvements to the presented method will provide the possibility to increase the temporal resolution of macrozoobenthos monitoring in the Dutch Wadden Sea, and also provides interesting insights for potential applications in other intertidal areas globally. Additional sampling data from the study area is required to enhance the proposed prediction models. This will improve the under- and overestimation of high and low values, by increasing the input training and testing data. Additional sample data also offers the opportunity to decrease zero-inflation. Furthermore, additional data allows for modelling seasonal distributions using satellite and VAE model data, with field input data of a different year or season to test the portability of the models without recent field data. Additional field data also provides the opportunity to model other species than the six researched species of the present study. The additional sampling data can then provide enough input to train the random forest models without the 0-values. Furthermore, it is also interesting to apply the models to the entire Wadden Sea or other intertidal areas as the ultimate goals is to propose a multi-seasonal macrozoobenthos distribution monitoring method. This goes especially for the presence-absence model, as it is the best performing model of the present study. Finally, the portability of the models over different satellite of UAV sensors is also worth researching. The use of Planet images requires a license whereas other satellite data such as Landsat data is openly available in for example Google Earth Engine. However, this will likely compromise the spatial resolution of the images and therefore the predictions, and the temporal resolution while it can already challenging to find suitable images to match the sampling campaigns.

5. Conclusions

Benthic macrofauna are an important constituent of the Wadden Sea system and food web. They are annually monitored across the Dutch Wadden as part of the Synoptic Intertidal Benthic Survey (SIBES) by taking field samples during the summer. However, a more time and cost efficient method is desired to increase the temporal resolution of the monitoring to a seasonal scale. The present study proposed a

novel method using seasonal high resolution multispectral satellite and variational autoencoder-extracted features in combination with field data to create random forest prediction models. Seasonal field data was collected over three separate campaigns; April 2022, July 2022 and SIBES data from September 2021. A total of 151 samples were taken from the study area over three seasonal campaigns and subsequently processed in the Texel laboratory of the Dutch Royal Institute for Sea Research (NIOZ). The macrozoobenthic organisms per sample were identified to the highest possible taxonomic level and counted. The laboratory analysis delivered ash free dry biomass data of all operational taxonomic units per sample and their abundances. The ash free biomass and abundance data was used to derive presence-absence data per sampling location. The field data of the three campaigns was used as input training and testing data for seasonal presence-absence, biomass and abundance prediction models for a subset of six different species. Planet multispectral satellite imagery for each sampling campaign was used as input for a variational autoencoder deep learning model (VAE). 64 feature layers per input image were extracted from the VAE and used as additional input feature data for the random forest models, in combination with the original multispectral satellite data.

For the presence-absence model, the predictions using only multispectral features performed equally as well as the predictions including VAE model features. The model achieved high prediction accuracies for the tested species with both feature variables. However, after applying the model to the entire study area the July presence-absence distribution maps show unlikely distributions, which indicates the quality of the satellite image was insufficient. The biomass prediction model did not perform well for all tested species and therefore there was no observable performance difference between the solely multispectral features and VAE-included features. The abundance prediction model also provided weak predictions, but they are improved compared to the biomass model. The best performing predictions are achieved by the predictions of Scoloplos armiger including VAE model features. This species achieved higher R-squared values with the VAE-included runs for every modeled season or combination of seasons compared to the runs using multispectral features. The highest R-squared value is achieved by the run combining April, July and SIBES 2021 data.
In the case of the presence-absence and abundance models, runs with multi-seasonal input data show more stability than the single-season models. Therefore, we conclude the data of multiple seasons can be combined to perform predictions. We observed strong differences in prediction capabilities for different species. Although this may be related to the presence and abundance of the species in the samples for model training, we also observed contrasting performances for the two most abundant species; Scoloplos armiger achieved relatively good predictions, whereas Pygospio elegans achieved relatively poor predictions. Overall, the modelling of the biomass and abundance need improvement. The ash free dry biomass and abundance data was strongly zero-inflated except for Scoloplos armiger and Pygospio elegans. Additionally, we observed large differences between the accuracies of the test and training data predictions. We therefore believe that the performance of the abundance model for the remaining species can be improved by additional field data, so there is more input data and zero-inflation can be reduced or masked. Improvements to the presented method including additional field data will provide the possibility to increase the temporal resolution of macrozoobenthos monitoring in the Dutch Wadden Sea, without the need to annually or seasonally re-execute sampling to execute the predictions. It also provides interesting insights for potential applications in other intertidal areas globally.

References

Beukema, J. J. & Dekker, R. (2020). Half a century of monitoring macrobenthic animals on tidal flats in the Dutch Wadden Sea. *Marine Ecology Progress Series*, *656,* 1-18, DOI: 10.3354/meps13555

Beukema, J. J., Dekker, R., Drent, J. & Van der Meer, J. (2017). Long-term changes in annual growth of bivalves in the Wadden Sea: influences of temperature, food and abundance, *Marine Ecology Progress Series, 573,* 143-156. DOI: https://doi.org/10.3354/meps12122

Bijleveld, A. I., Van Gils, J. A., Van der Meer, J., Dekinga, A., Kraan, C., Van der Veer, H. W. & Piersma, T. (2012). Designing a benthic monitoring programme with multiple conflicting objectives, *Methods in Ecology and Evolution, 3*(3), 526-536, DOI: https://doi.org/10.1111/j.2041-210X.2012.00192.x

Bolam, S. G. & Fernandes, T. F. (2003). Dense aggregations of Pygospio elegans (Claparède): effect on macrofaunal community structure and sediments, *Journal of Sea Research, 43*(3), 171-185. DOI: https://doi.org/10.1016/S1385-1101(03)00007-8

Breiman, L. (2001). Random Forests, *Machine Learning, 45,* 5-32. DOI: https://doi.org/10.1023/A:1010933404324

Compton T. J., Holthuijsen, S., Koolhaas, A., Dekinga, A., Ten Horn, J., Klunder, J., … Van der Meer, J. (2013). Progress report for the 2012 sampling of the synoptic intertidal benthic surveys across the Dutch Wadden Sea (Report No: NIOZ 2018-9), NIOZ for Nederlandse Aardolie Maatschappij BV.

Compton, T. J, Holthuijsen, S., Koolhaas, A., Dekinga, A., Ten Horn, J., Smith, J., … Piersma, T. (2012). Synoptic intertidal benthic survey across the Dutch Wadden Sea: report on data collected from 2008 to 2010 (Report No: 2012.1.SIBES.NIOZ), NIOZ for Nederlandse Aardolie Maatschappij BV.

Compton, T. J., Holthuijsen, S., Koolhaas, A., Dekinga, A., Ten Horn, J., Smith, J., … Piersma, T. (2013). Distinctly variable mudscapes: Distribution gradients of intertidal macrofauna across the Dutch Wadden Sea. *Journal of Sea Research, 82*, 103-116. DOI: http://dx.doi.org/10.1016/j.seares.2013.02.002

Convention and nomination. (n.d.) Wadden Sea. Retrieved October 25, 2022 from https://waddensea worldheritage.org/convention-and-nomination

Crisci, C., Ghattas, B. & Perera, G. (2012). A review of supervised machine learning algorithms and their applications to ecological data, *Ecological Modelling, 240*, 113-122. http://dx.doi.org/10.1016/j.ecolmodel.2012.03.001

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. & Lawler, J. J. (2007). Random forests for classification in ecology, *Ecology, 88*(11), 2783-2792. DOI: 10.1890/07-0539.1

Dairi, A., Harrou, F., Sun, Y. & Khadraoui, S. (2020). Short-Term Forecasting of Photovoltaic Solar Power Production Using Variational Auto-Encoder Driven Deep Learning Approach, *Solar Energy, 105,* 401-413. https://doi.org/10.1016/j.solener.2014.03.018

Drent, J., Bijkerk, R., Herlyn, M., Grotjahn, M., Voß, J., Carausu, M.-C. & Thieltges, D.W. (2017). *Macrozoobenthos.* In: Wadden Sea Quality Status Report. Eds.: Kloepper S. et al., Common Wadden Sea Secretariat, Wilhelmshaven, Germany. Last updated 21.12.2017. Downloaded 5.10.2022. qsr.waddensea-worldheritage.org/reports/macrozoobenthos

Eriksson, B. K., Van der Heide, T., Van de Koppel, J., Piersma, T., Van der Veer, H. W. & Olff, H. (2010). Major Changes in the Ecology of the Wadden Sea: Human Impacts, ecosystem engineering and sediment dynamics, *Ecosystems, 13*(5)*,* 752-764. DOI: 10.1007/s10021-010-9352-3

Flach, E. C. (1992). Disturbance of benthic infauna by sediment-reworking activities of the lugworm Arenicola marina, *Netherlands Journal of Sea Research, 30,* 81-89. DOI: https://doi.org/10.1016/00777579(92)90048-J

Folmer, E., Dekinga, A., Holthuijsen, S., Van der Meer, J., Mosk, D., Piersma, T. & Van der Veer, H. (2017). Species Distribution Models of Intertidal Benthos - Tools for Assessing the Impact of Physical and Morphological Drivers on Benthos and Birds in the Wadden Sea, NIOZ Royal Netherlands Institute for Sea Research

Giere, O. & Pfannkuche, O. (1982). Biology and ecology of marine Oligochaeta, a review. *Oceanography and marine biology: an annual review*, *20*, 679.

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

Godet, L., Fournier, J., Toupoint, N. & Olivier, F. (2009). Mapping and monitoring intertidal benthic habitats: a review of techniques and a proposal for a new visual methodology for the European coasts, *Progress in Physical Geography, 33*(3), 378-402. DOI: 10.1177/0309133309342650

Hartman, O. & Fauchald, K. (1971). Deep-water benthic polychaetous annelids off New England to Bermuda and other North Atlantic areas. Part II, *Marine Biology*, *6,* 1-327

Hartmann-Schröder, G. (1996). Annelida, Borstenwürmer, Polychaeta. Die Tierwelt Deutschlands, 58. Teil. Gustav Fischer Verlag, Jena.

Hayward, P. J. & Ryland, J. S. (1995). Handbook of the Marine Fauna of North-West Europe, Oxford University Press, Oxford.

Honkoop, P. J. C. & Beukema, J. J. (1997). Loss of body mass in winter in three intertidal bivalve species: an experimental and observational study on the interacting effects between water temperature, feeding time and feeding behaviour, *Journal of Experimental Marine Biology and Ecology, 212*(2), 277-297. DOI: https://doi.org/10.1016/S0022-0981(96)02757-8

Horn, S., De la Vega, C., Asmus, R., Schwemmer, P., Enners, L., Garthe, S., … Asmus, H. (2017). Interaction between birds and macrofauna within food webs of six intertidal habitats of the Wadden Sea, *PLoS ONE, 12*(5), 1-23. DOI: https://doi.org/ 10.1371/journal.pone.0176381

Horn, S., Schwemmer, P., Mercker, M., Enners, L., Asmus, R., Garthe, S. & Asmus, H. (2020). Species composition of foraging birds in association with benthic fauna in four intertidal habitats of the Wadden Sea, Estuarine, *Coastal and Shelf Science*, *233*, 106537. https://doi.org/10.1016/j.ecss.2019.106537

Jordahl, K. (2014). GeoPandas: Python tools for geographic data. *URL: Https://Github. Com/Geopandas/Geopandas*.

Kraan, C., Aarts, G., Van Der Meer, J., Piersma, T. (2010). The role of environmental variables in structuring landscape-scale species distributions in seafloor habitats, *Ecology, 91*(6), 1583–1590. DOI: 10.1890/09-2040.1.

Ma, Y., Lin, Y., Nie, Z. & Ma, H. (2020). Structural damage identification based on unsupervised feature extraction via Variational Auto-encoder, *Measurement, 160,* 107811. https://doi.org/10.1016/j.measurement.2020.107811

Madhuanand, L., Philippart, C. J. M., Wang, J., Nijland, W., De Jong, S. M., Bijleveld, A. I. & Addink, E. A. (2023). Enhancing the predictive performance of remote sensing for ecological variables of tidal flats using encoded features from a deep learning model, *GIScience & Remote Sensing, 60*(1), 1 21. DOI: https://doi.org/10.1080/15481603.2022.2163048

Mayer, D. G. & Butler, D. G. (1993). Statistical validation, *Ecological Modelling, 68,* 21-32. DOI: https://doi.org/10.1016/0304-3800(93)90105-2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel. O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research, 12,* 2825-2830.

Planet Team (2017). Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://api.planet.com.

Ryu, J., Choi, J. & Lee, Y. (2014). Potential of remote sensing in management of tidal flats: A case study of thematic mapping in the Korean tidal flats, *Ocean & Coastal Management, 102*(B)*,* 458-470. DOI: 10.1016/j.ocecoaman.2014.03.003

Puls, W., Van Bernem, K. -H., Eppel, D., Kapitza, H., Pleskachevsky, A., Riethmuller, R. & Vaessen, B. (2012). Prediction of benthic community structure from environmental variables in a soft sediment tidal basin (North Sea), *Helgoland Marine Research, 66,* 345-361. https://doi.org/10.1007/s10152-011-0275-y

Reise, K., Baptist, M., Burbridge, P., Dankers, N., Fischer, L., Flemming, B., … Smits, C. (2010). The Wadden Sea – A Universally Outstanding Tidal Wetland, Wadden Sea Ecosystem No. 29. Common Wadden Sea Secretariat, Wilhelmshaven, Germany, page 7-24.

Sijtsma, F.J., Mehnen, N., Angelstam, P. & Muñoz-Rojas, J. (2019). Multi-scale mapping of cultural ecosystem services in a socio-ecological landscape: A case study of the international Wadden Sea Region, *Landscape Ecology, 34,* 1751-1768. DOI: https://doi.org/10.1007/s10980-019-00841-8

Sørensen, T. H., Bartholdy, J., Christiansen, C. & Pedersen, J. B. T. (2006). Intertidal surface type mapping in the Danish Wadden Sea, *Marine Geology, 235*(1-4)*,* 89-99. DOI: https://doi.org/10.1016/j.margeo.2006.10.007

Van der Wal, D., Herman, P. M. J., Forster, R. M., Ysebaert, T., Rossi, F., Knaeps, E., Plancke, Y. M. G. & Ides, S. J. (2008). Distribution and dynamics of intertidal macrobenthos predicted from remote sensing: Response to microphytobenthos and environment, *Marine Ecology Progress Series, 367,* 57-72. DOI: 10.3354/meps07535

Van Hoey, G., Guilini, K., Rabaut, M., Vincx, M. & Degraer, S. (2008). Ecological implications of the presence of the tube-building polychaete *Lanice conchilega* on soft-bottom benthic ecosystems, *Marine Biology, 154,* 1009-1019. DOI: 10.1007/s00227-008-0992-1

Vermeersen, B. L. A., Slangen, A. B. A., Gerkema, T., Baart, F., Cohen, K. M., Dangendorf, S. … Van der Wegen, M. (2018). Sea-level change in the Dutch Wadden Sea, *Netherlands Journal of Geosciences – Geologie en Mijnbouw, 97*(3), 79-127. DOI: https://doi.org/10.1017/njg.2018.7

Wang, Z. B., Elias, E. P. L., Van der Spek, A. J. F. & Lodder, Q. J. (2018). Sediment budget and morphological development of the Dutch Wadden Sea: impact of accelerated sea-level rise and subsidence until 2100. *Netherlands Journal of Geosciences, 97*(3), 183-214. DOI:10.1017/njg.2018.8

Yao, R., Liu, C., Zhang L. & Peng, P. (2019). Unsupervised Anomaly Detection Using Variational Auto Encoder based Feature Extraction, *IEEE International Conference on Prognostics and Health Management,* 1-7. DOI: 10.1109/ICPHM.2019.8819434.

Zerrouki. Y., Harrou, F., Zerrouki, N., Dairi, A. & Sun, Y. (2021). Desertification Detection Using an Improved Variational Autoencoder-Based Approach Through ETM-Landsat Satellite Data, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 202-212. DOI:10.1109/JSTARS.2020.3042760

Appendix A – Random forest models accuracy and R^2 tables

***Table A.1**. Accuracy scores of RF Presence/Absence classifier of testing and training data.*

Seasons	Features	Data	Arenicola marina	Heteromastus filiformis	Lanice conchilega	Macoma balthica	Pygospio elegans	Scoloplos armiger
April	4 band	Test	0.71	0.86	0.64	0.43	0.86	0.71
		Train	0.88	0.76	0.90	0.83	0.90	0.90
	all	Test	0.79	0.79	0.50	0.43	0.79	0.93
		Train	0.90	0.76	0.90	0.90	0.93	0.95
July	4 band	Test	0.57	0.64	0.71	0.79	0.86	0.79
		Train	0.84	0.84	0.73	0.89	0.89	0.89
	All	Test	0.57	0.64	0.57	0.86	0.71	0.79
		Train	0.84	0.87	0.95	0.84	0.92	0.95
April+July	4 band	Test	0.68	0.71	0.64	0.57	0.79	0.86
		Train	0.83	0.81	0.77	0.78	0.89	0.96
	all	Test	0.71	0.71	0.57	0.54	0.71	0.86
		Train	0.91	0.82	0.87	0.78	0.86	0.94
April+July +SIBES21	4band	Test	0.71	0.72	0.71	0.64	0.79	0.86
		Train	0.83	0.84	0.77	0.81	0.89	0.96
	All	Test	0.71	0.72	0.61	0.61	0.71	0.86
		Train	0.85	0.83	0.79	0.85	0.94	0.95

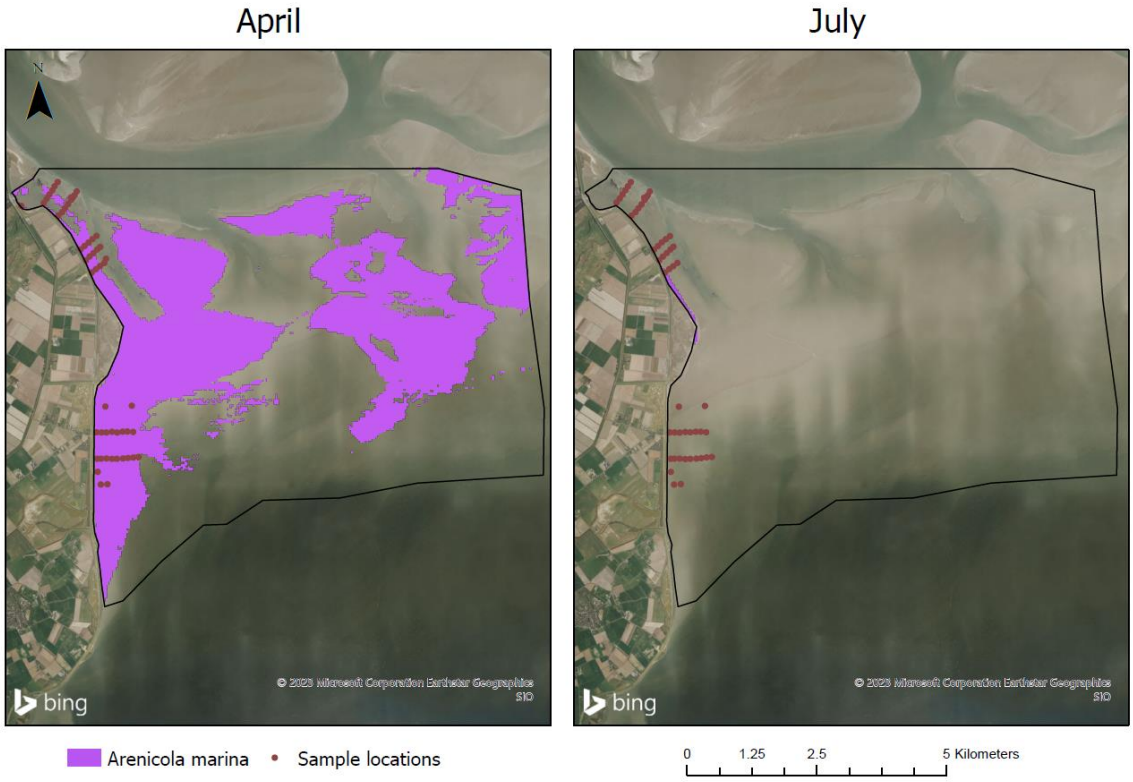
***Table A.2.** R-squared values of RF biomass regressor of testing and training data.*

Seasons	Features	Data	Arenicola marina	Heteromastus filiformis	Lanice conchilega	Macoma balthica	Pygospio elegans	Scoloplos armiger
April	4 band	Test	0.29	-0.15	-0.11	0.43	-0.67	-0.05
		Train	0.42	0.45	0.60	0.39	0.61	0.54
	all	Test	0.21	0.06	-0.01	-0.42	-0.50	-0.14
		Train	0.60	0.51	0.62	0.48	0.76	0.66
July	4 band	Test	-0.50	-0.42	-0.68	-0.46	-0.58	-1.03
		Train	0.64	0.41	0.47	0.46	0.50	0.61
	All	Test	-0.90	0.05	-0.45	-0.80	-0.44	-1.09
		Train	0.66	0.52	0.62	0.62	0.62	0.76
April+July	4 band	Test	0.10	-0.40	0.11	-0.56	-0.24	-0.54
		Train	0.37	0.35	0.44	0.30	0.51	0.61
	all	Test	0.12	-0.21	-0.02	-0.62	-0.06	-0.42
		Train	0.48	0.43	0.52	0.37	0.60	0.70
April+July +SIBES21	4band	Test	0.10	-0.09	-0.03	-0.44	-0.01	-0.87
		Train	0.37	0.28	0.38	0.28	0.55	0.48
	All	Test	0.12	-0.08	-0.06	-0.47	-0.12	-0.35
		Train	0.48	0.35	0.42	0.35	0.65	0.56

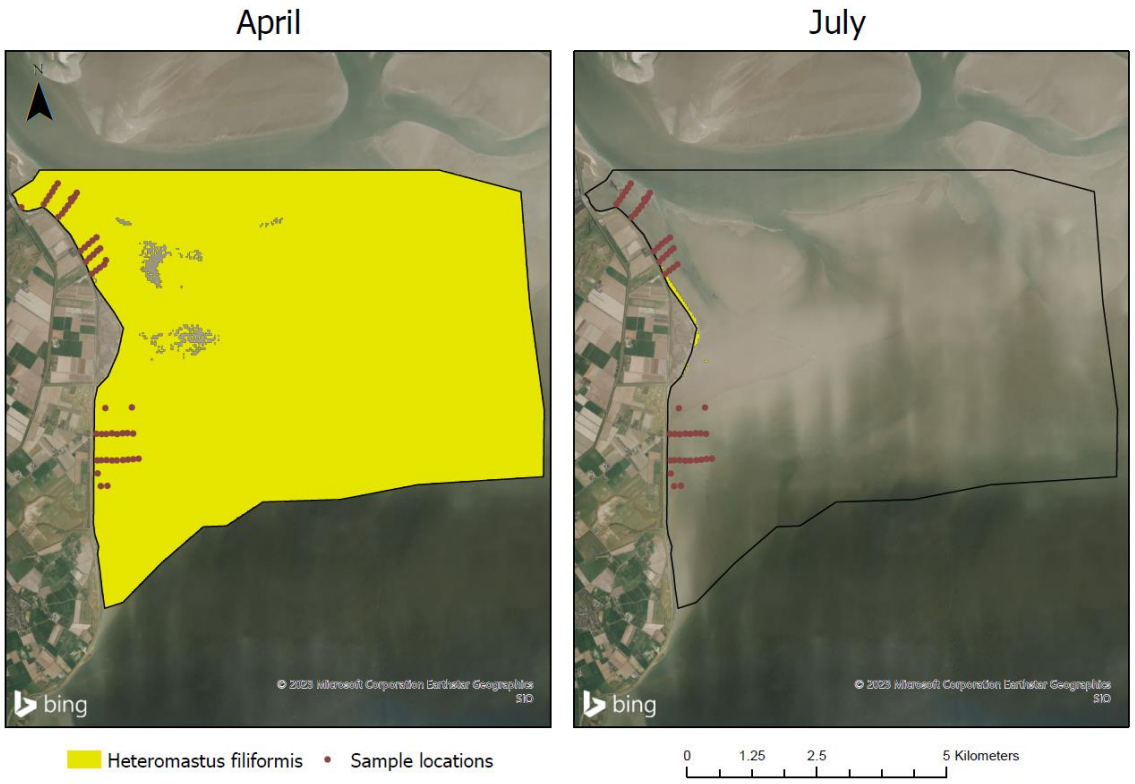
Table A.3. *R-squared values of RF abundance regression model of testing and training data.*

Seasons	Features	Data	Arenicola marina	Heteromastus filiformis	Lanice conchilega	Macoma balthica	Pygospio elegans	Scoloplos armiger
April	**4 band**	**Test**	0.16	0.03	0.16	-0.13	-0.40	0.09
		Train	0.62	0.29	0.62	0.40	0.69	0.78
	all	**Test**	0.06	0.19	0.06	-0.43	-0.34	0.53
		Train	0.72	0.56	0.72	0.49	0.75	0.82
July	**4 band**	**Test**	-0.12	-0.33	-0.12	0.06	0.04	-0.18
		Train	0.51	0.46	0.51	0.54	0.60	0.63
	All	**Test**	-0.30	-0.58	-0.30	0.23	-0.01	0.09
		Train	0.63	0.69	0.63	0.76	0.73	0.78
April+July	**4 band**	**Test**	0.11	-0.01	0.11	-0.04	-0.08	0.43
		Train	0.55	0.39	0.55	0.34	0.53	0.75
	all	**Test**	0.26	-0.36	0.26	-0.04	-0.25	0.50
		Train	0.66	0.66	0.66	0.53	0.64	0.82
April+July +SIBES21	**4band**	**Test**	0.02	0.01	0.02	0.11	-0.19	0.20
		Train	0.54	0.52	0.54	0.45	0.49	0.67
	All	**Test**	0.27	-0.03	0.27	-0.03	-0.30	0.64
		Train	0.67	0.66	0.67	0.55	0.60	0.77

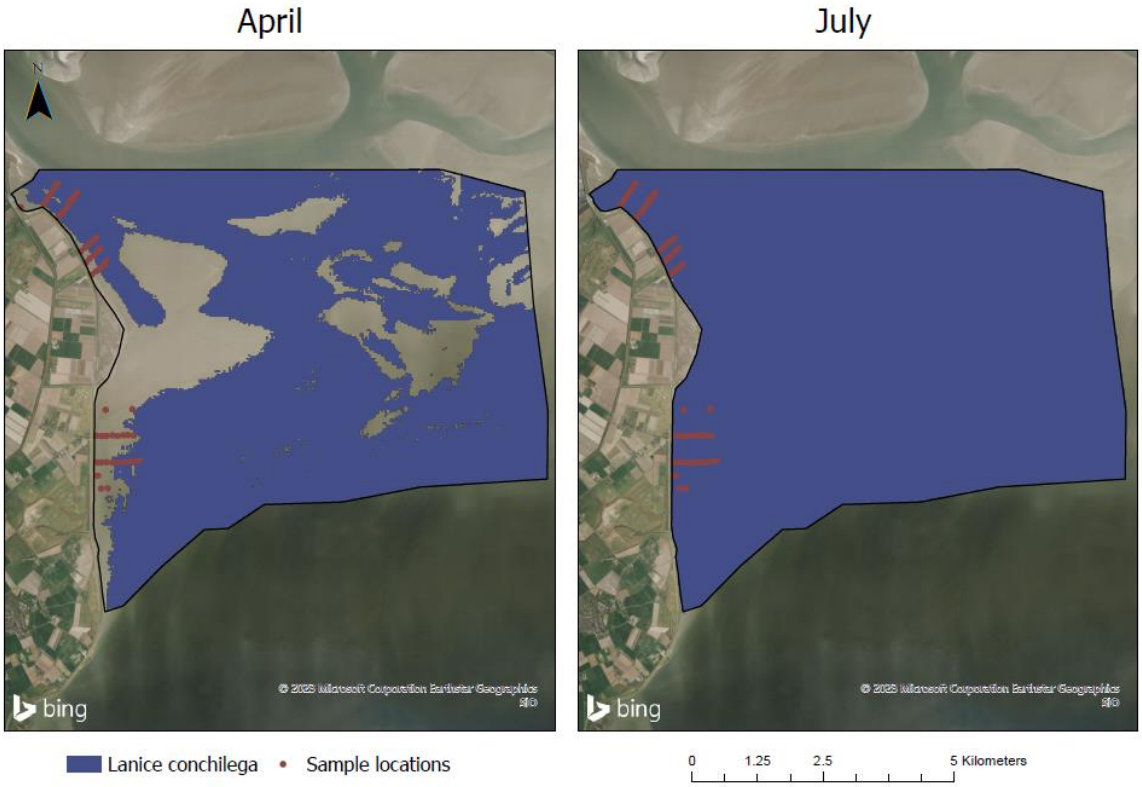
Appendix B – Distribution maps



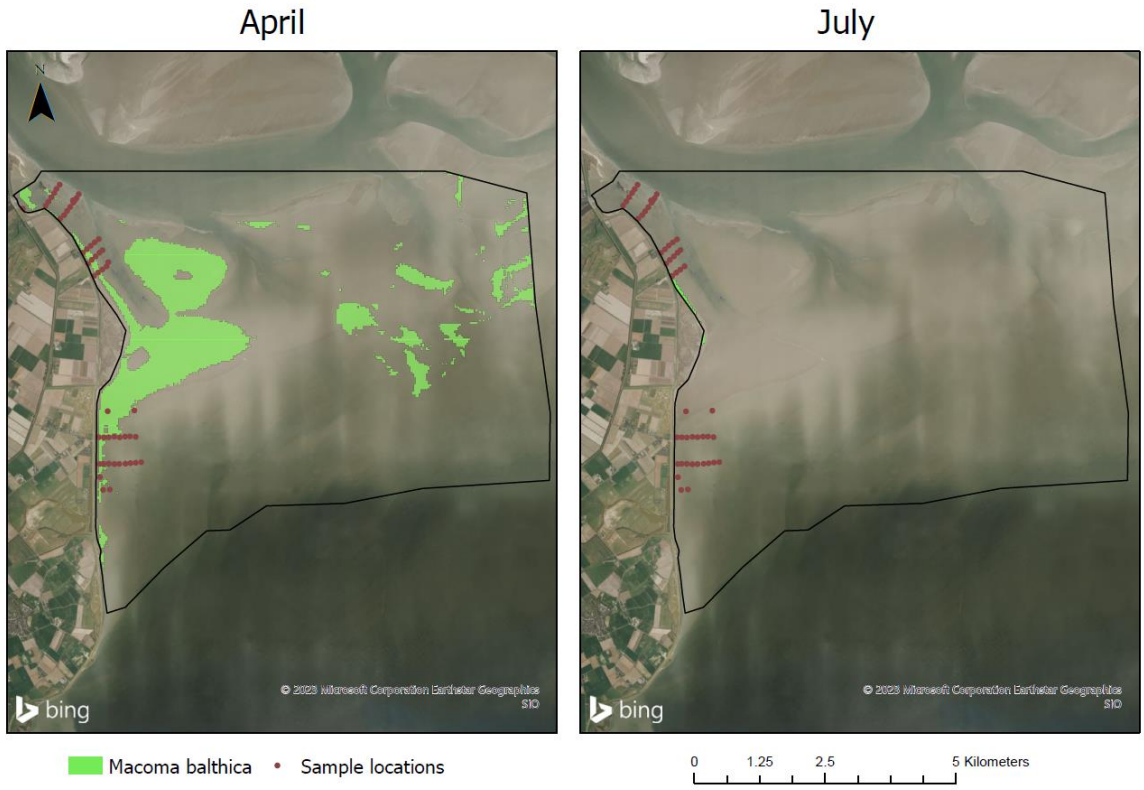
Appendix B.1 Predicted presence distribution of Arenicola marina in the study area for April 2022 (left) and July 2002 (right).



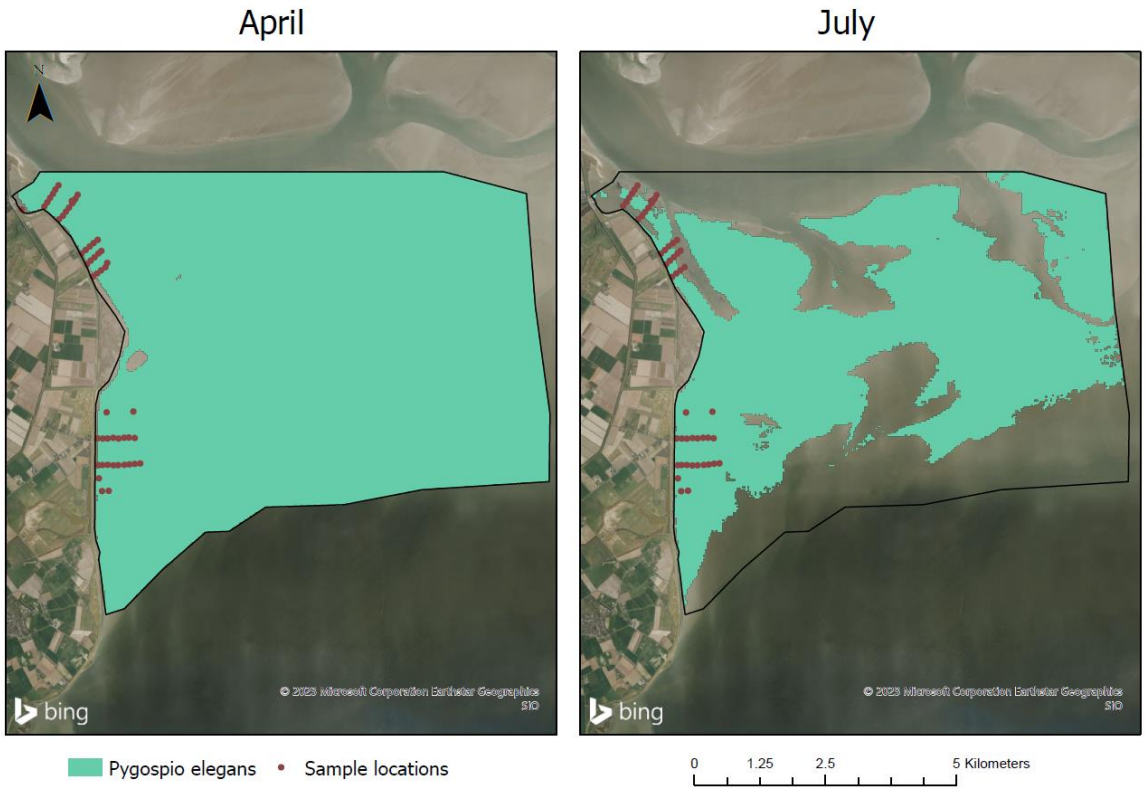
Appendix B.2 Predicted presence distribution of Heteromastus filiformis in the study area for April 2022 (left) and July 2002 (right).



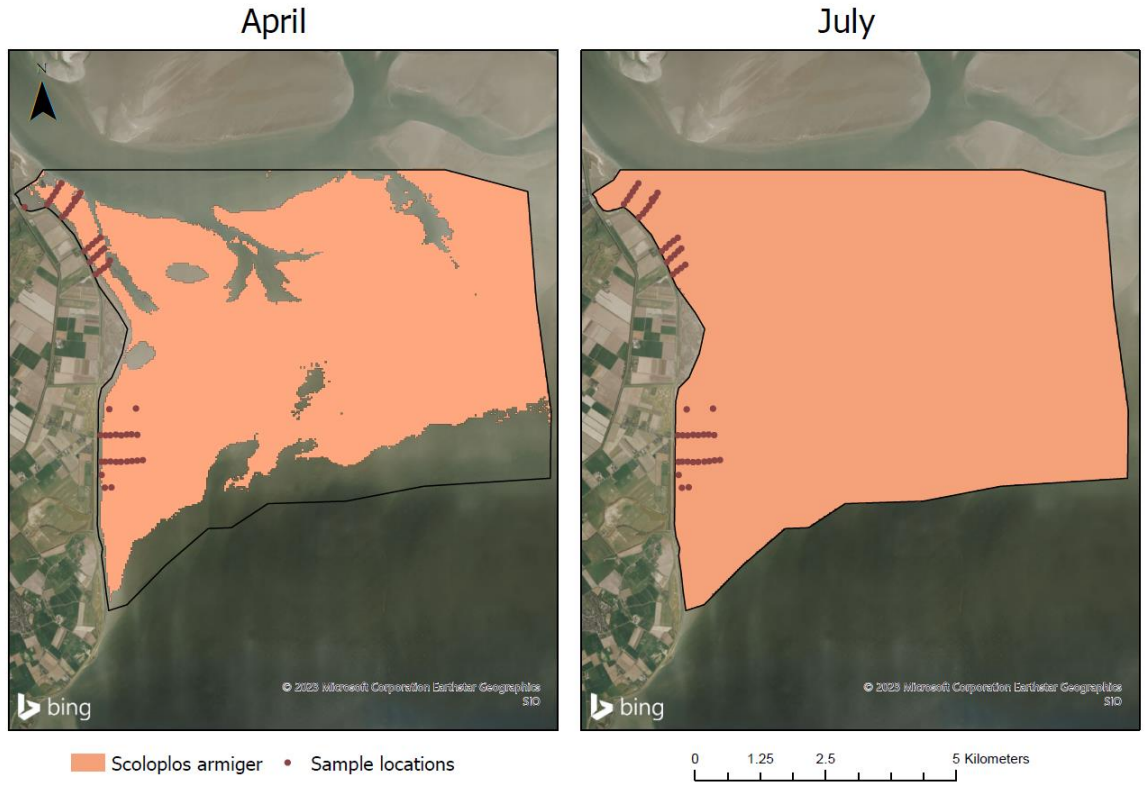
Appendix B.3 Predicted presence distribution of Lanice conchilega in the study area for April 2022 (left) and July 2002 (right).



Appendix B.4 Predicted presence distribution of Macoma balthica in the study area for April 2022 (left) and July 2002 (right).

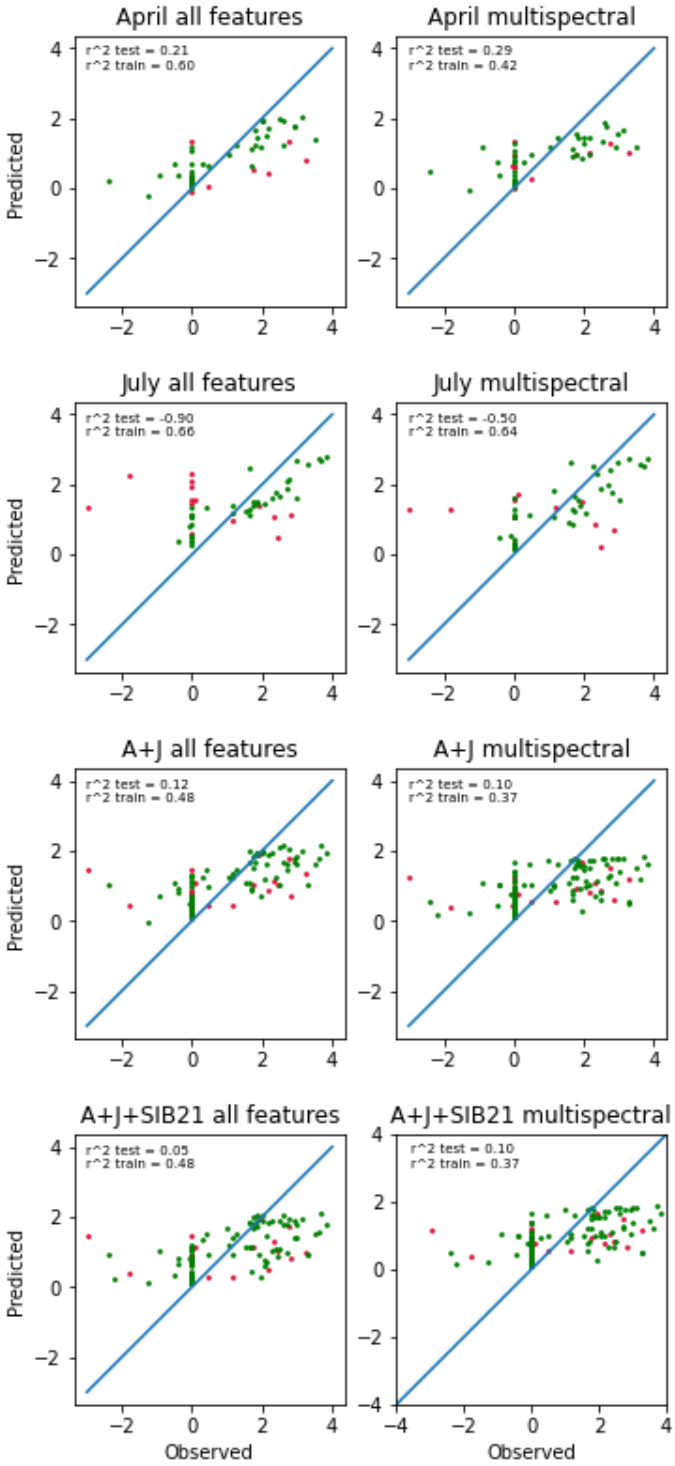


Appendix B.5 Predicted presence distribution of Pygospio elegans in the study area for April 2022 (left) and July 2002 (right).

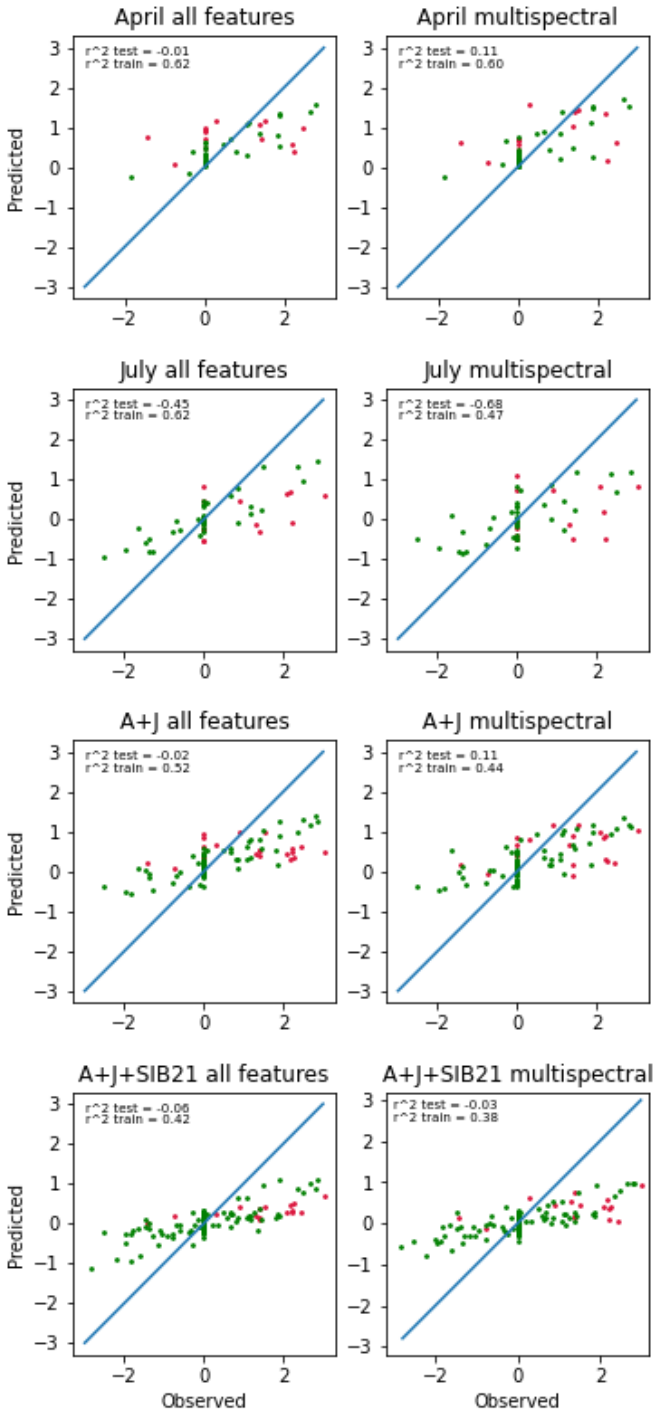


Appendix B.6. Predicted presence distribution of Scoloplos armiger in the study area for April 2022 (left) and July 2002 (right).

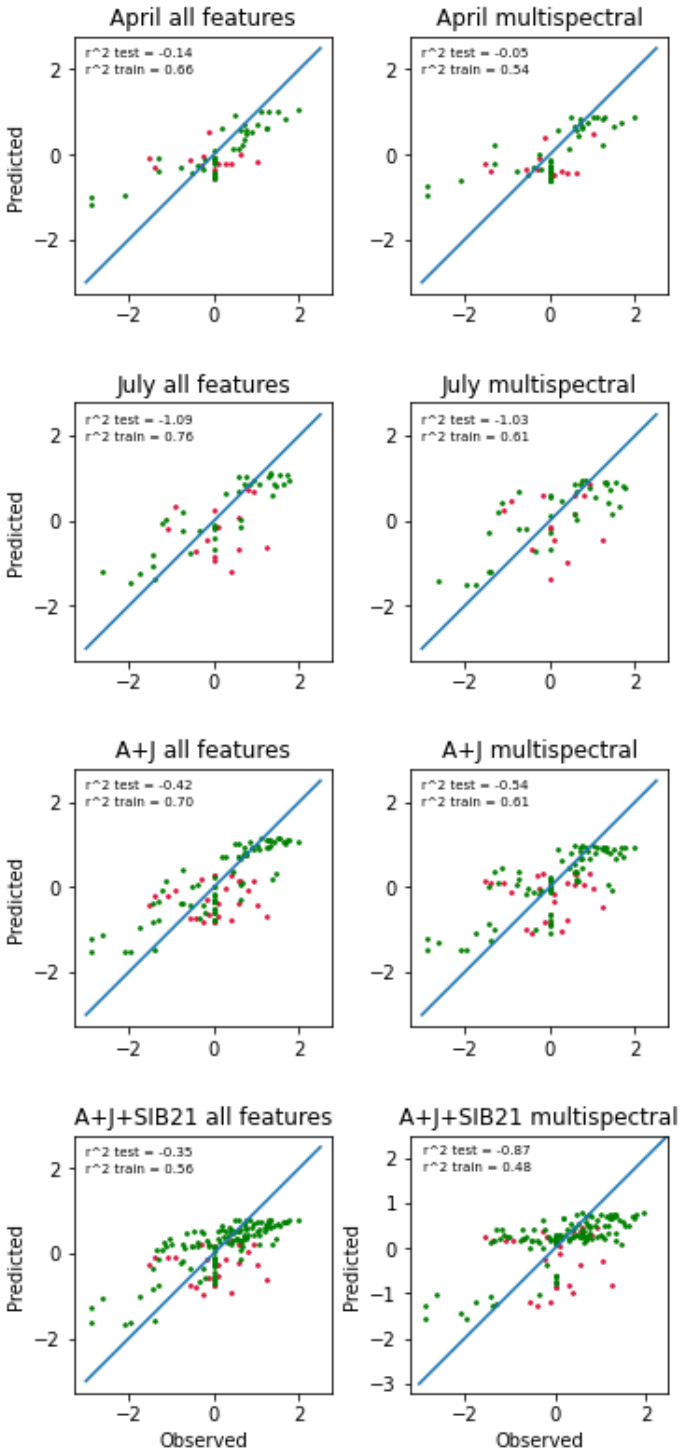
Appendix C – Biomass observed/predicted plots



Appendix B.1 Biomass observed/predicted plots of Arenicola marina: test data in pink, training data in green. The blue line represents an R-squared value of 1.

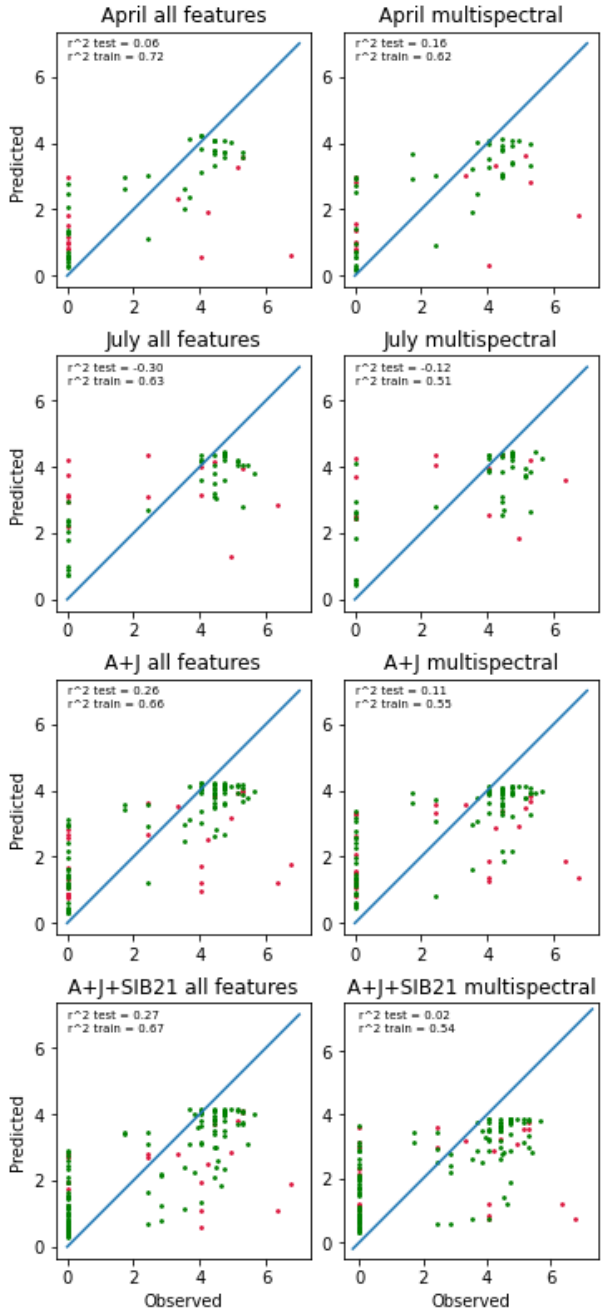


Appendix B.2 Biomass observed/predicted plots of Lanice conchilega: test data in pink, training data in green. The blue line represents an R-squared value of 1.

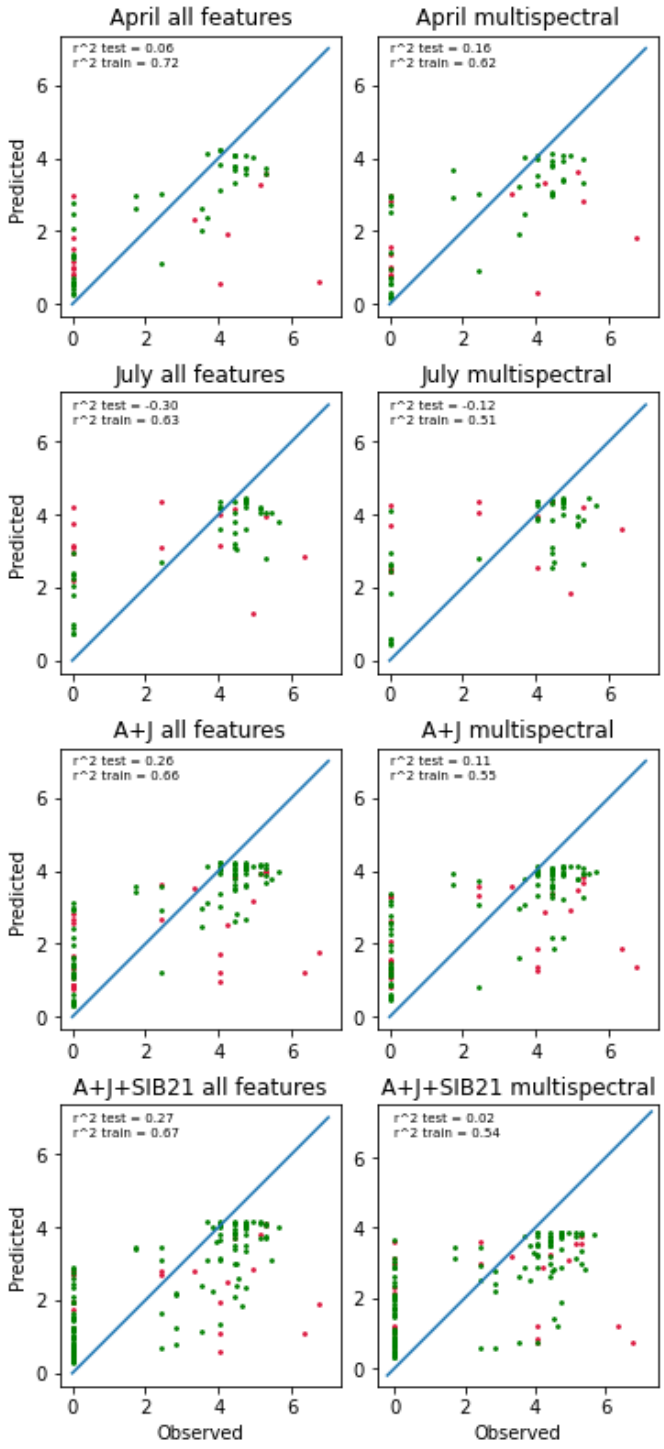


Appendix B.3 Biomass observed/predicted plots of Scoloplos armiger: test data in pink, training data in green. The blue line represents an R-squared value of 1.

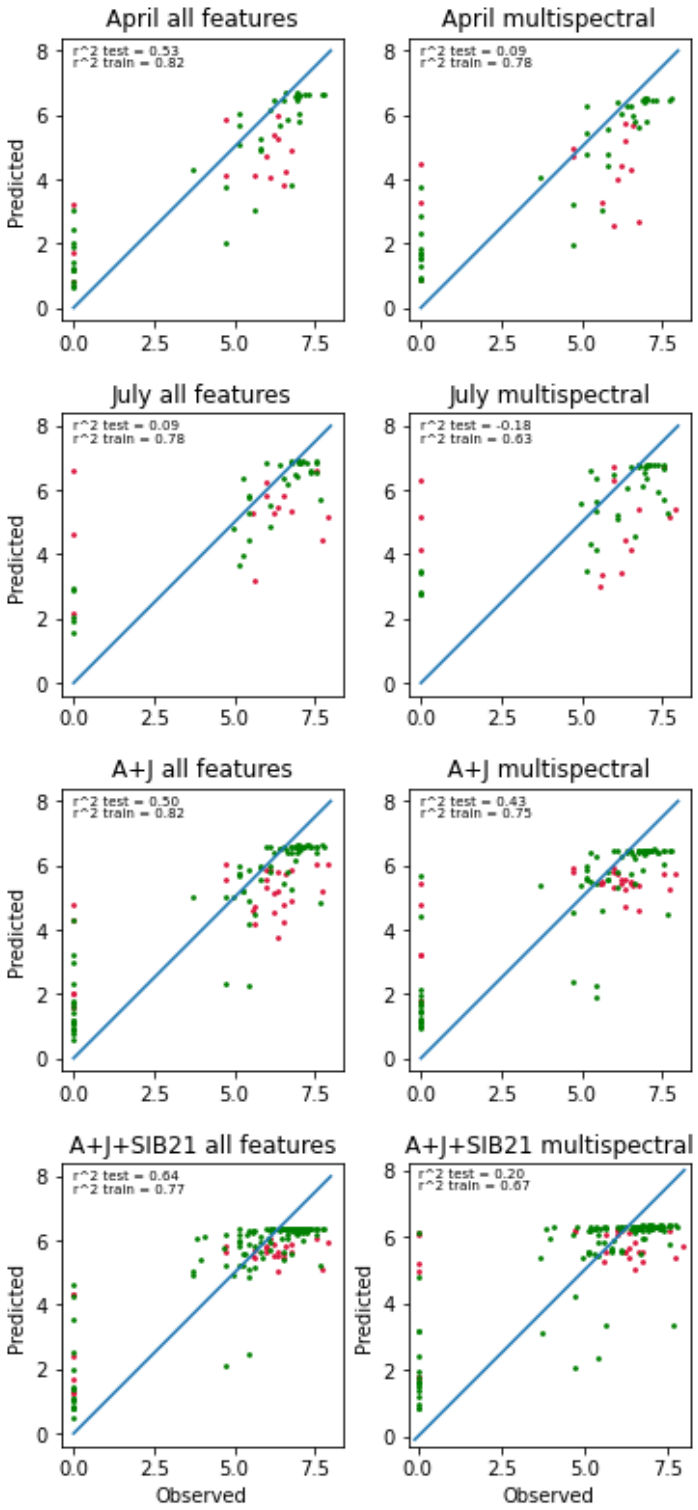
Appendix D – Abundance observed/predicted plots



Appendix C.1 Abundance observed/predicted plots of Arenicola marina: test data in pink, training data in green. The blue line represents an R-squared value of 1.



Appendix C.2 Abundance observed/predicted plots of Lanice conchilega: test data in pink, training data in green. The blue line represents an R-squared value of 1.



Appendix C.3 Abundance observed/predicted plots of Scoloplos armiger: test data in pink, training data in green. The blue line represents an R-squared value of 1.