

Global distribution of muddy coasts using a hybrid classification model

An automated method that employs multispectral satellite imagery and globally available coastal datasets

R.L. Hulskamp

Deltares
Enabling Delta Life

TU Delft

Global distribution of muddy coasts using a hybrid classification model

An automated method that employs multispectral satellite imagery and globally available coastal datasets

by

Romy Hulskamp

Student number: 4309928
romyhulskamp@gmail.com | +31634433401
MSc student Civil Engineering
Specialization in Geosciences and Remote Sensing

to obtain the degree of

Master of Science
in Civil Engineering

at the Delft University of Technology,
to be defended publicly on Tuesday December 7, 2021 at 13:00.

Thesis committee:	Prof. Dr. Ir. S.G.J. Aarninkhof,	TU Delft
	Dr. Ir. A.P. Luijendijk,	TU Delft / Deltares
	Dr. Ir. A. Moreno-Rodenas,	Deltares
	Dr. Ir. S.L.M. Lhermitte,	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Cover: Landsat 4 TM image of Yellow River Delta, China, February 13, 1989
©Earth Observatory | NASA

December 7, 2021

Summary

The coastal area provides important services such as valuable habitats for wildlife, resources for regional development, and buffer zones for the land against natural disasters such as storm surges. But these narrow coastal areas experience pressure from both land and ocean side. In order to regulate sustainable coastal development, protect coastal areas from natural forces, implement ecosystem protection strategies and mitigate the impacts of climate change on coastlines, it is crucial to observe and quantify the changes along coastlines that are vulnerable to these pressures. Among the various coastal environments, the focus of this master thesis is on muddy coasts. A muddy coast is defined as a coastal depositional environment that is usually formed along an unsheltered coastline exposed to low energy conditions and that consists mainly of fine sediments, which are smaller than $63\mu m$ in diameter.

So far, few methods have been developed to detect sandy coasts and the rate of change of sandy coastlines on a global scale, but little is known about the detection and behaviour of muddy coasts. The main objective of this thesis is to develop an automated classification method to identify muddy coasts along all the coastlines of the world. The main objective can be translated into the following research question: *"How can a reliable global mud classification be obtained by analysing the characteristics of muddy coasts using publicly available satellite remote sensing techniques and globally available coastal datasets?"*. To obtain a reliable global mud classification, a method based on the spectral properties of individual mud patches and on the physical geographical characteristics of muddy coasts is proposed.

The spectral properties of individual mud patches are studied at pixel level using multispectral satellite images; a multispectral image classification model is developed based on the spectral properties of individual mud patches at pixel level, that are studied with satellite images. 7500 training points are used to train a global image classification model, that is able to classify each pixel of a multispectral image into seven coastal classes (sand, mud, water, turbid water, vegetation, dry vegetation and other). This model showed high performance, where more than 90 percent of the validation points were classified correctly. 92 percent of the muddy training points were actually recognised as mud by the global image classifier. However, the multispectral image classification also showed some misclassifications and uncertainties. Therefore, the physical geographical characteristics of a muddy coastal system are studied as well using globally available coastal datasets, in order to increase the probability of correctly classifying a muddy coast at global scale. Seven globally available coastal datasets are explored, from which ten variables are extracted: maximum height, variance of height, latitude, absolute latitude, maximum temperature of warmest month, minimum temperature of coldest month, mangrove occurrence, width of tidal flat, width of transition zone and tidal range. The information from the multispectral images and the information from the coastal datasets are used as the input the development of three coastal transect classification models: (1) the multispectral model, using only the information from the multispectral image classification; (2) the coastal model, using only the information from the globally available coastal datasets; (3) the hybrid model, using both the information provided by the multispectral image classification and the information collected from the available global coastal datasets as input, in order to obtain a more reliable and robust classification model that can detect muddy coasts at global scale.

The global coastline is defined by circa 1.8 million transects that measure $1500m$ in length perpendicular to the coastline with a spacing of $200m$ to $450m$. 1868 training transects were labelled and used to train the three coastal transect classification models, which are able to classify all global coastlines into five coastal types: plain beaches, muddy coasts, coastal cliffs, vegetated coasts and other. A model performance assessment showed that the best performing model for the classification of the coastal type of the global coastline is the hybrid model. This global transect classification model uses a total of seventeen coastal features. The six most important features are the occurrence of mud, water and sand pixels along the transect provided by the multispectral image classification model; and the maximum height, height variance and width of transition zone of the transect collected from the coastal datasets. The hybrid model assigned 76 percent of the transects into the correct coastal type. 85 percent of the muddy coasts are correctly recognised by the hybrid classification model. The accuracy of the hybrid model is still increasing significantly, meaning that the model could perform even better if more training transects were added to train the model. The large amount of training and validation data at both pixel and transect level resulted in a reliable and robust global mud classification model. With the addition of the coastal datasets, the model performs better than

when only the multispectral satellite images are used for the classification of muddy coasts.

The hybrid classification model is used to classify all the global coastal transects. This automated model firstly applies the pixel based image classification to the one year composite multispectral image around the transect; secondly, extracts the information of the ten coastal variables from the globally available coastal datasets along the transect; thirdly, analyses the obtained data from the transect and classifies the transect into the coastal type that best matches the observed coastal characteristics of the transect. This automated mud classification model classified nearly 100000 transects randomly spread over the globe from which 12% are classified as muddy coasts; 60% of the classified muddy coasts are located in the tropics.

Contents

Summary i

1 Introduction 1
- 1.1 Problem statement . . . 2
- 1.2 Objectives . . . 2
- 1.3 Research question . . . 3

2 Developing a global mud classifier 4
- 2.1 Methodology . . . 4
- 2.2 Background information on muddy coasts . . . 5
  - 2.2.1 Mud particles . . . 5
  - 2.2.2 Muddy coasts . . . 6
  - 2.2.3 Other coastal environments . . . 10
- 2.3 Multispectral satellite imagery . . . 12
  - 2.3.1 Satellite mission . . . 12
  - 2.3.2 Indices . . . 12
- 2.4 Globally available coastal datasets . . . 13
  - 2.4.1 Global transect system . . . 13
  - 2.4.2 Topographic relief . . . 14
  - 2.4.3 Climate . . . 14
  - 2.4.4 Tides . . . 15
- 2.5 Developing a global hybrid mud classification model . . . 16

3 Multispectral image classification 20
- 3.1 Method . . . 20
  - 3.1.1 Local image classification . . . 20
  - 3.1.2 Cross image classification . . . 21
  - 3.1.3 Global image classification . . . 22
- 3.2 Results . . . 24
  - 3.2.1 Local image classification . . . 24
  - 3.2.2 Cross image classification . . . 26
  - 3.2.3 Global image classification . . . 27
- 3.3 Model performance . . . 35
  - 3.3.1 Overall model accuracy . . . 36
  - 3.3.2 Accuracy per coastal class . . . 37
- 3.4 Conclusions . . . 39

4 Exploring coastal datasets 40
- 4.1 Coastal transect profiles . . . 40
  - 4.1.1 Plain beaches . . . 40
  - 4.1.2 Coastal cliffs . . . 41
  - 4.1.3 Barrier islands . . . 41
  - 4.1.4 Mangrove deltas . . . 44
- 4.2 Assessment of coastal datasets . . . 48
  - 4.2.1 Topographic relief . . . 48
  - 4.2.2 Climate . . . 48
  - 4.2.3 Tides . . . 49
- 4.3 Added value of coastal datasets . . . 52
- 4.4 Conclusions . . . 52

5 Hybrid transect classification **54**
5.1 Method . . . 54
5.1.1 Study area . . . 54
5.1.2 Collecting data . . . 56
5.1.3 Training models . . . 56
5.1.4 Best model . . . 56
5.2 Model performances . . . 56
5.2.1 Feature importance . . . 56
5.2.2 Confusion matrix . . . 57
5.2.3 Overall model accuracy . . . 57
5.2.4 Accuracy per coastal type . . . 59
5.3 Multispectral, coastal and hybrid transect classification . . . 62
5.4 Conclusions . . . 63

6 Regional and global mapping **64**
6.1 Regional mapping of coastal types . . . 64
6.2 Global distribution of muddy coasts . . . 69
6.2.1 Verification . . . 71

7 Discussion **74**
7.1 Possible limitations . . . 74
7.1.1 Top of Atmosphere vs. Surface Reflectance images . . . 74
7.1.2 Individual vs. composite images . . . 74
7.2 Development of hybrid transect classification model . . . 75
7.2.1 Quality of coastal datasets . . . 75
7.2.2 Other coastal datasets . . . 77
7.2.3 Model development . . . 77
7.3 Global upscaling . . . 77
7.3.1 Size of training data . . . 77
7.3.2 Definition and interpretation of coastal types . . . 77
7.3.3 Transect system . . . 78
7.3.4 Ground-truth validation . . . 78

8 Conclusions **79**

9 Recommendations **81**
9.1 Multispectral image classification . . . 81
9.2 Global transect system . . . 81

10 Technical details **82**
10.1 Cloud-free images . . . 82
10.2 Tidal stages . . . 82
10.3 Selection of boxes/points/transects . . . 82
10.3.1 Coastal boxes . . . 82
10.3.2 Training points . . . 82
10.3.3 Training transects . . . 82
10.4 Random Forest classification . . . 83
10.4.1 Smile Random Forest (GEE) . . . 83
10.4.2 Random Forest (Python) . . . 83
10.5 Composite image . . . 83
10.6 Buffered-nearest-neighbour-average . . . 83

Committee **84**

References **85**

1 Introduction

The coastal area is defined as the transition between the continents and the oceans. It spans the land, which is affected by its proximity to the ocean, and the ocean, which is affected by its proximity to the land. The highly dynamic and energetic environment of the coastal area causes the shape of the coast to change constantly. This is called coastal morphodynamics (Bosboom and Stive, 2015).

The coastal area provides important services such as valuable habitats for wildlife, resources for regional development, and buffer zones for the land against natural disasters such as storm surges (Wang et al., 2019). But these narrow coastal areas also experience pressure from both land and ocean sides. On the land side, the coastal area is under pressure due to the increasing human population. The global human population is growing and currently counts more than 7.8 billion people[1]. The coastal area has always attracted people because of its rich opportunities for, among others, habitation, (marine) food production, marine trade and transport, and recreation. A trend of coastal migration is observed; coastal areas have a significantly higher population density than non-coastal areas (Neumann et al., 2015) and are among the most densely populated and developed land areas in the world (Small and Nicholls, 2003). On the ocean side, the pressure of rising sea levels, coastal erosion, more severe storms and subsidence will lead to increasing risks. Recent increases in mean wave energy, extreme coastal wave energy and oceanic wind speeds, combined with rising sea levels, suggest that coastal areas will be exposed to increasing hazards in the coming decades (Li et al., 2018). In order to regulate sustainable coastal development, protect coastal areas from natural forces, implement ecosystem protection strategies and mitigate the effects of climate change on coastlines, it is crucial to observe and quantify the changes along coastlines that are vulnerable to the pressure from land and ocean.

Among the different coastal environments, this thesis focuses on muddy coasts. A muddy coast is defined as a coastal depositional environment consisting primarily of fine sediments, less than $63\mu m$ in diameter, formed along an unprotected coastline exposed to low energy conditions (Healy et al., 2002). These coastal environments are usually associated with rivers characterised by large sediment discharges, such as the Amazon River, the Mississippi River and several Asian rivers. The continuous and abundant supply of fine sediment is a prerequisite for the formation of a muddy open coast, while interaction of tides and waves modulate the shape and extent of intertidal areas (Orseau et al., 2020). Muddy coasts are important coastal systems that have a close relationship with salt marshes and mangroves, create important coastal ecosystems, offer habitats for organisms, provide food production for people's livelihoods and form a natural coastal defence against storms (Murray et al., 2019). This makes muddy coasts very important elements in coastal ecosystems in several aspects including ecological, social and economic.

So far, few methods have been developed to track sandy coasts and the rate of change of sandy coastlines on a global scale (Luijendijk et al. (2018), Mentaschi et al. (2018), Vos et al. (2020)). Figure 1 shows a map of the global distribution of sandy coastlines based on multispectral satellite images made by Luijendijk et al. (2018), with yellow indicating sand and dark brown indicating non-sand. However, little is known about the detection and behaviour of muddy coastlines. Monitoring the morphodynamics of muddy coasts is essential for a better understanding of the evolution of coastal systems (Salameh et al., 2019), for environmental protection, for resource management and for sustainable development of coastal areas (Wang et al., 2019). A brief summary of relevant studies is provided below. Ryu et al. (2002) attempted to determine the waterline of tidal flats in Gomso Bay, Korea, using multispectral Landsat TM data. Liu and Jezek (2004) developed an automated model for coastline extraction by integrating Canny edge detection and locally adaptive threshold methods. A characterisation of surface roughness and sediment texture of tidal flats in the Netherlands using SAR imagery was carried out by Van Der Wal et al. (2005). In another study by van der Wal and Herman (2007), a method for monitoring the grain size of sediments from tidal flats in the Western Scheldt was developed, using information from both SAR and multispectral remote sensing techniques. In "The state of the World's beaches", Luijendijk et al. (2018) presents a global assessment of the occurrence of sandy beaches and the rate of change of the coastline. Murray et al. (2019) made the first analysis of the global distribution and change of tidal flats. Vos et al. (2019) has created an open-source software toolkit that allows the user to obtain time series of the position of the coastline on any sandy coastline worldwide from publicly available satellite images. In another paper, Vos et al. (2020) proposes a new approach to estimate

[1] https://www.worldometers.info/

the slope of the beach based on time series of satellite-derived coastline positions. Recently, de Vries et al. (2021) developed a data-driven approach to analyse diffuse boundaries of mud banks by unmixing the pixels of water and mud from individual Landsat images.

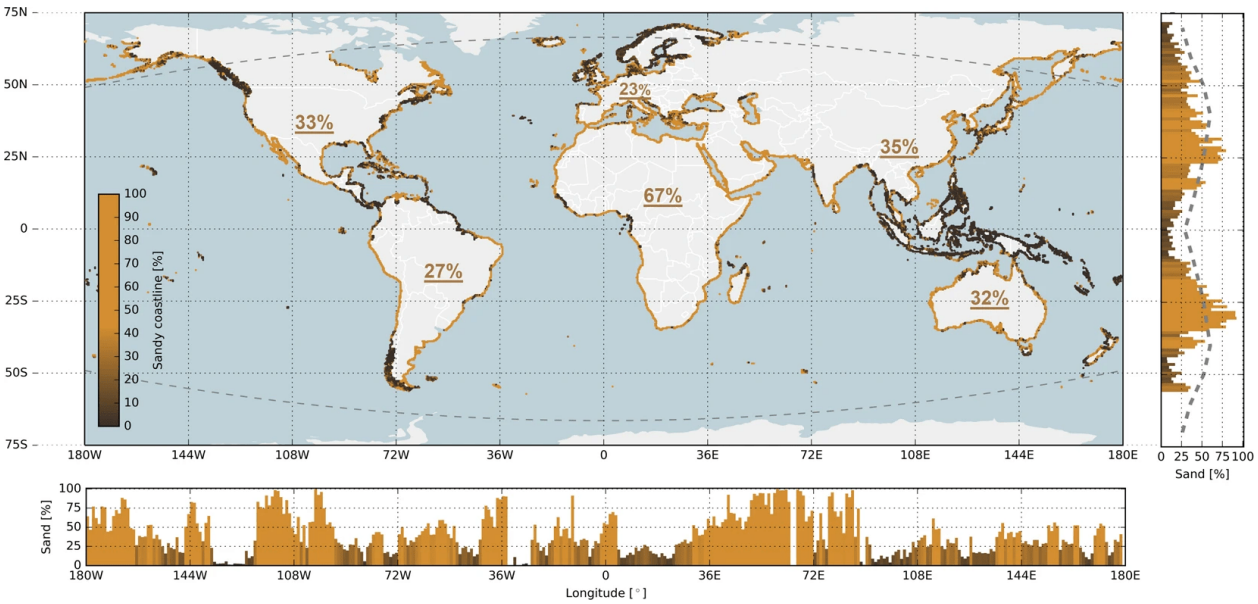


Figure 1: Global distribution of sandy coastlines; the coloured dots along the world's coastline represent the local percentage of sandy coastlines (yellow is sand, dark brown is non-sand). The subplot on the right shows the relative occurrence of sandy coastlines by latitude, with the dashed line representing the latitudinal distribution of sandy coastlines reported by Hayes (1967). The lower subplot shows the relative occurrence of sandy coastlines by longitude. The curved grey dashed lines in the main plot indicate the boundaries of the ice-free coastlines considered in this analysis. The underlined percentages indicate the percentages of sandy coastlines averaged per continent. From Luijendijk et al. (2018).

Among the current knowledge and methods of mud detection, no global assessment to detect muddy coasts and their erosion and accretion rates has been made so far. Sandy beaches have been classified, tidal flat trajectories have been identified and mud in suspension can be detected. Yet, muddy coasts have not been identified at global level. This thesis focuses on the knowledge gap on muddy coasts and will try to provide a reliable mud classification model that can detect muddy coasts at global level.

1.1 Problem statement

Although several studies have been carried out on the classification of muddy coasts on a local scale, often verified by field measurements, no study has yet been conducted on the global classification of muddy coasts. Muddy coasts have important economic, social and ecological functions. Muddy coasts can develop valuable ecosystems, which not only provide a habitat for wildlife, but also offer resources for regional development and a natural defence of the land against severe storms. The increasing pressure on the coasts from both the land side, due to the ever-growing human population, and from the ocean side, due to coastal erosion and climate change, makes it important to monitor the valuable muddy coasts and their coastal dynamics.

1.2 Objectives

The main objective of this thesis is to develop an automated classification method that can identify muddy coasts along all coastlines of the world. The focus of this work will be on distinguishing muddy coasts from other coastal features and from other types of sediment found on the coast (such as sand, rocks, corals, etcetera). Figure 2 illustrates an example of a coastal environment. The features water, turbid water, mud, sand and vegetation are present. The main objective of this thesis is to distinguish mud from sand by developing and applying an automated classification method that is applicable globally. For example, if muddy pixels are recognised on a cross section of the coast (the left vertical red dashed line in Figure 2),

this part of the coast can be classified as muddy. If the classification method is successful on local scale, a final goal of this thesis is to present a global map of the occurrence of muddy coastlines.

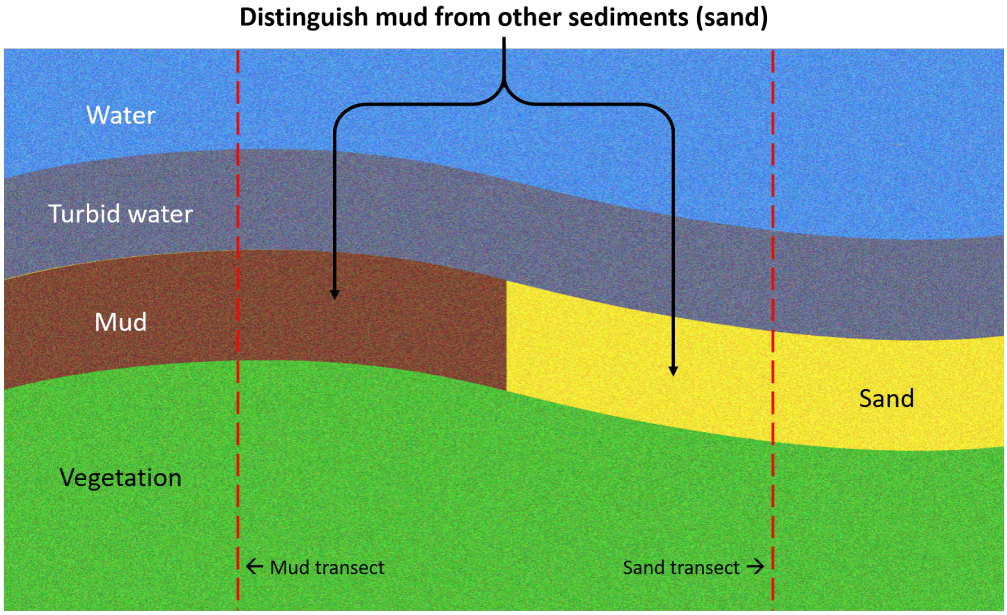


Figure 2: Conceptual sketch of the distinction between mud and other sediments, such as sand, using cross-shore transects. A cross-shore transect where muddy sediments are detected could indicate a muddy coast (left), a cross-shore transect where sandy sediments are detected could indicate a sandy coast (right).

1.3 Research question

The main objective, the development of an automated classification method that can identify muddy coasts at global scale, can be translated into the following research question and breaks down into four sub-questions, which will serve as a guide in answering the main question.

How can a reliable global mud classification be obtained by analysing the characteristics of muddy coasts using publicly available satellite remote sensing techniques and globally available coastal datasets?

1. To what extent can mud be distinguished from other coastal features by the use of publicly and globally available multispectral satellite images?
2. How representative are locally trained image classification methods when applied at global scale?
3. What is the added value of adding globally available coastal datasets in distinguishing mud from other coastal features?
4. What is the global distribution of muddy coasts?

2 Developing a global mud classifier

This section explains the proposed methodology of how a global mud classifier can be developed. First an overview is given of the three phases of the proposed method in section 2.1. After that some background information on muddy coasts is provided in section 2.2, that focuses first on the characteristics of muddy particles and second on the characteristics of muddy coastal systems. Some other coastal environments are also described in order to find out their characteristics and how they differ from muddy coasts. Thereafter, the satellite mission that will be used for analysing the spectral characteristics of muddy particles is mentioned in section 2.3. Next, some globally available coastal datasets are discussed in section 2.4, which may contribute to characterise and distinguish different coastal environments. This section ends with a roadmap for the final development of a global hybrid mud classification model (section 2.5).

2.1 Methodology

To obtain a reliable global mud classification model, a method based on analysing the spectral properties of individual mud patches and on analysing the physical geographical characteristics of muddy coasts as a system, is proposed. Figure 3 gives an overview of the proposed method. The method consists of three phases. During phase 1 and 2 data is collected, modified and processed, which then will be used as an input in phase 3, in which the global mud classification model is developed.

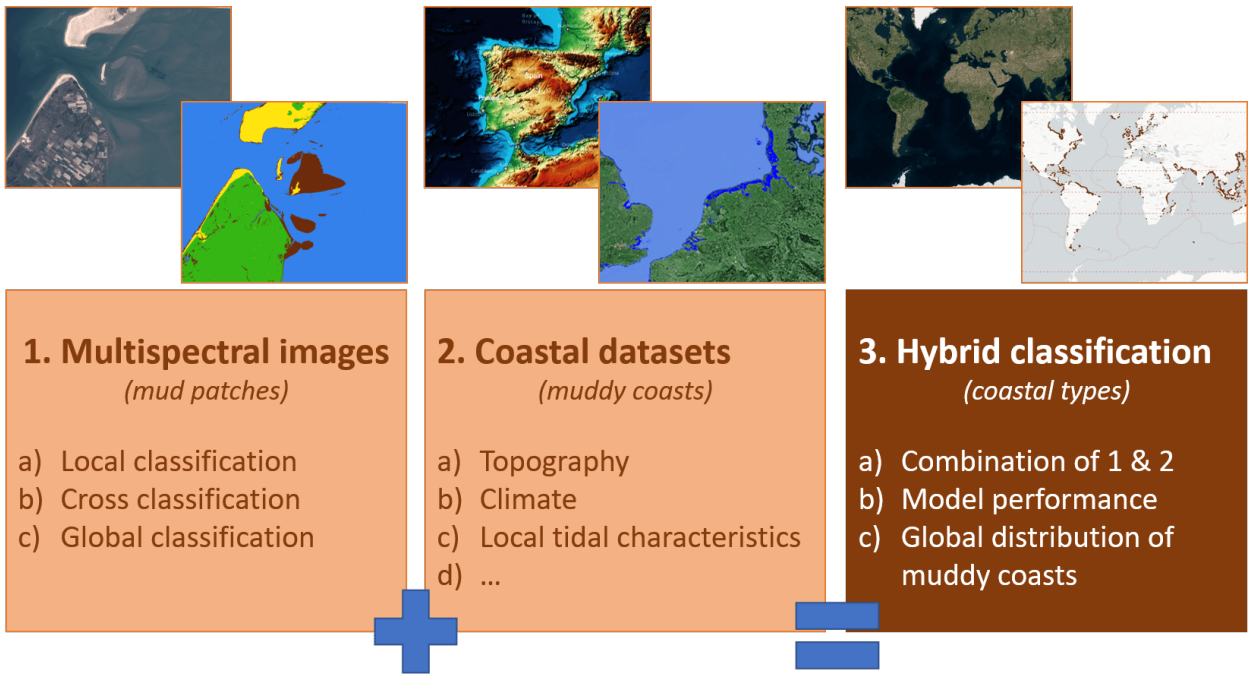


Figure 3: Proposed methodology for developing a reliable automated global mud classifier. Data is collected, modified and processed in phase 1 and 2, which forms the input for phase 3, through which the global mud classifier will be trained and developed.

In the first phase, the spectral properties of individual mud patches found along the globe are studied and an image classification model is created based on their spectral reflectance signature found. In addition, several other coastal features and their spectral properties are studied in order to distinguish them from the muddy features. Three steps are taken during this first phase. Firstly, a locally trained image classifier, which analyses an area in the Dutch Wadden Sea at different tidal stages, is developed to investigate the possibilities of using multispectral satellite images in detecting mud patches. Secondly, the locally trained image classifier is cross validated at two other locations, namely an area on the coast of Suriname and an area on the coast of French Guyana, in order to test the representativeness of a locally trained image classifier at other locations with different conditions. Finally, a global classification model is developed, expanding the

training dataset on a global scale to obtain a more reliable and robust classifier.

In the second phase, globally available coastal datasets will be examined that will be used to test a classification model based on contextual coastal information that characterises muddy coasts as a system. Datasets that are of interest are for example the topography of the land, the dimensions of the intertidal area, the tidal range and the temperature distribution. The datasets are grouped based on the four coastal influencers described in section 2.2.2. It is investigated whether the characteristics of muddy coasts are actually recognisable in these coastal datasets and whether they are distinctive compared to other coastal types.

The third phase uses the information from the multispectral images and the information from the coastal datasets as the input for the development of three coastal transect classification models: (1) the multispectral model, using only the information from the multispectral image classification; (2) the coastal model, using only the information from the globally available coastal datasets; (3) the hybrid model, using both the information provided by the multispectral image classification and the information collected from the available global coastal datasets as input, in order to obtain a more reliable and robust classification model that can detect muddy coasts at global scale. The best performing model will be selected and will classify the entire global coast to obtain the global distribution of muddy coasts.

The methodology of each phase will be explained in detail in the sections 3, 4 and 5.

2.2 Background information on muddy coasts

This section describes some of the key concepts underlying this research; beginning with the definition and properties of individual mud particles in section 2.2.1. Once the definition and properties of individual mud particles are clear, a larger-scale analysis of the properties and behaviour of muddy coasts as a system will be carried out, which will be described in section 2.2.2. Finally, section 2.2.3 briefly discusses some coastal environments, other than the muddy coasts, to obtain information on which characteristics may be distinctive for muddy coasts.

2.2.1 Mud particles

The sediment found on coasts all around the world can come from a variety of sources. A distinction can be made between continental sediments and carbonate sediments. Continental sediments are formed by the weathering of continental rocks. Pieces of shells or remains of marine life together form the carbonate sediments. Mud particles generally originate from continental sediments. This section explains the definition, formation and properties of individual mud particles.

Definition of mud particles Mud is defined as a fluid-sediment mixture consisting of (salt) water, silt, clay and organic material (Bosboom and Stive, 2015). Silt and clay are different soil types. Silt elements are composed of quartz and feldspar. Quartz and feldspar are the most common minerals in the Earth's crust. Silt grains, which are almost spherical in shape, are usually formed by the mechanical weathering of quartz and feldspar. Mechanical weathering is the physical breaking down of rocks into smaller pieces by natural forces. For example, water seeps into the cracks of a rock, then freezes and expands. The crack in the rock is getting wider. The water melts and now more water can seep into the wider crack. The cycle begins again, until the rock disintegrates into several pieces. Further erosion of these pieces will produce almost spherical silt grains. Clay, on the other hand, is created by the chemical weathering of feldspar. Chemical weathering, for example, occurs when water reacts with certain minerals in the rock to form a different product. The reaction of rain with feldspar forms clay, which can be dissolved by further erosion.

The Wentworth classification (Wentworth, 1922), based on the grain size of clastic sediments, defines silt particles as individual particles smaller than very fine sand grains and larger than clay particles. Clay particles are defined as individual particles smaller than silt particles. Based on the grain size of sediments, mud is bounded by a maximum grain size of $63\mu m$, see Figure 4.

Properties of mud particles Now that the definition of the individual mud particles (consisting of silt and clay) is known, the properties of this type of sediment are analysed. The properties of mud particles are important, to understand how multispectral satellite images can be used to detect muddy coasts.

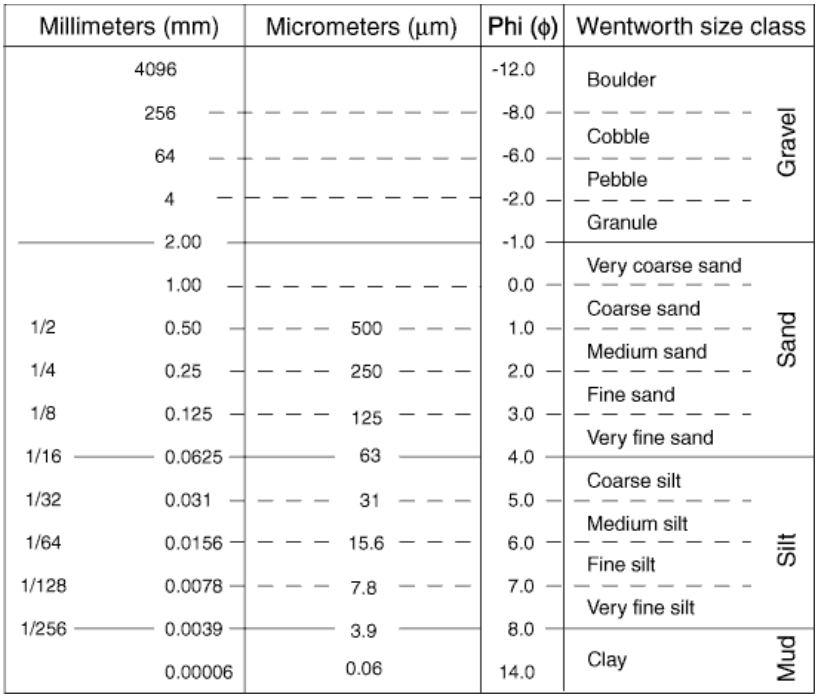


Figure 4: The Wentworth grain size classification. Sediment particles with grain size $\leq 63\mu m$ are defined as mud. Adjusted from Wentworth (1922).

Shape Due to the fact that clay is formed by the chemical weathering of feldspar, clay minerals consist of thin plate-like particles held together by electrostatic forces. Clay minerals are layered metal silicates. When the distance between the particles is small enough, they get stuck in a 'house of cards' structure, this process is called flocculation. Unlike silt particles, which are almost spherical in shape, clay particles have more angular shapes.

Cohesive Clay particles are very small and have a relatively large surface in relation to their volume. This surface is chemically active, which, especially when wet, leads to the typical cohesive characteristics of the bulk form of mud. For example, picking up a handful of mud is easy, because the grains stick together. But if you try to pick up a handful of (fine) sand, the grains slip between your fingers. In addition, the silty nature of mud keeps the soil moist even during low tides, which leads to an even stronger cohesive character.

Settling velocity When a particle falls into still and clear water, it accelerates until it reaches a constant vertical (terminal) velocity called the settling velocity. For very small particles, such as silt and clay, the settling velocity is proportional to D^2, where D is the diameter of the grain. In high concentration mixtures, the settling velocity of a single particle is reduced by the presence of other particles; this is called hindered settling. In very silt-rich environments, sediment concentrations can be considerably high and the mud particles are usually in suspension. The size and shape of mud particles can change quite easily. Flocculation increases the size of individual mud particles. With this increase, the particles become heavier, which will speed up the settling of the particles. On the other hand, the increase in size will also increase the drag force, the resistance the particle 'feels' when settling. This will slow the particle down and may eventually cause deflocculation. An individual mud particle of $D = 10\mu m$ has a settling velocity of approximately $0.1mm/s$ in still and clear water.

2.2.2 Muddy coasts

In addition to the definition and properties of individual mud particles, the larger-scale analysis of the properties and dynamic behaviour of muddy coastal environments is important when studying muddy coasts

with satellite remote sensing techniques and globally available coastal datasets. This section will first describe the definition of muddy coasts. This is followed by a description of some of the characteristics of muddy coasts, including the formation, occurrence and important functions of muddy coastal environments.

Definition of muddy coasts Muddy coasts are defined as the transition zone between the continents an the oceans consisting entirely or largely of muddy sediments, where muddy sediments are defined as in section 2.2.1. Wang et al. (2002) defines a muddy coast as "*A sedimentary-morphodynamic type characterised primarily by fine-grained sedimentary deposits - predominantly silts and clays - within a coastal sedimentary environment. Such deposits tend to form rather flat surfaces, and are often, but not exclusively, associated with broad tidal flats*". This research is limited to coasts along which mud is visible from satellite observations, i.e. over a considerable width normal to the coast and over a considerable distance along the coast. Large areas of muddy coastlines are connected to tropical mangrove systems or moderate salt marshes. Muddy coasts neighbouring mangrove systems cover an area of about 75% of the world's coastline around 25° latitude above and below the equator, according to Flemming (2002).

Properties of muddy coasts The properties and dynamic behaviour of muddy coastal systems are essential to understand, in order to map muddy coasts using satellite imagery and other globally available coastal datasets. What characterises muddy coasts, how do they develop and under what conditions do they often occur? In this paragraph the formation, occurrence and functions of muddy coasts will be discussed.

Formation Certain coastal conditions and processes can influence the morphology of a coastal environment. This paragraph explains four coastal influencers: continental shelf width, topographic relief, climate and tides.

Continental shelf width The process of plate tectonics has (had) an enormous influence on the formation of coastlines and determines the broadest features of the coast. An important element is the configuration of the continental shelf (width and slope of the shelf). Inman and Nordstrom (1971) divided the coasts into three categories, according to their tectonic plate setting. The first type of coasts are the leading edge coasts located on converging plates. These coasts are mostly cliff-shaped, there is a lot of tectonic activity and the continental shelf is narrow and steep. Secondly, there are the trailing edge coasts, which are located on diverging plates, away from the plate boundaries. These coasts are tectonically stable and the continental shelf is wide and gentle. The last class includes the marginal sea coasts. These coasts are tectonically stable as well and are protected from the open ocean by island arcs on converging plate boundaries. Paragraph 2.2.1 explained that it takes a long time for mud particles to settle, because of the small fall velocity and the hindered settling. In order to settle down, mud particles require a relative calm wave environment. Large mid-ocean waves lose energy as they travel over the gently sloping inner continental shelf of a trailing edge coast, consequently these waves do not inhibit deposition of sediment along the coast. As long as the coastal system is supplied with sufficient sediment, the combination of reduced wave energy and facilitation of sediment accumulation allows for the development of large sedimentary features on the wide and flat trailing edge continental margins. Similarly, marginal sea coasts have the potential to develop large sedimentary features due to their wide continental shelves and low wave energy. These coasts are low wave energy coasts because of the gentle slope and shallow waters of the continental shelves in these areas, due to the attenuation of wave energy caused by the shelter provided by nearby or surrounding land masses, island arcs or ice, and due to the limited size of the marginal seas and the consequent limited fetch, which limits the magnitude of the waves that develop locally.

Topographic relief Besides the width of the continental shelf, the topographic relief is a determining factor for the development of muddy coasts. Where mountains are found rivers will form, turning rocks and sediment into smaller particles and transporting them from the continents to the oceans. Muddy coasts are usually associated with long and meandering rivers, which are characterised by a large supply of sediment. Muddy coasts are therefore often found far from the high mountains, so that the rivers have the space and time to weather the rocks into the smallest possible particles.

Climate Another aspect influencing the formation of muddy coasts is climate. Climate is a determining factor in what kind of and to what rate, erosion takes place. The river runoff is determined by the amount of rainfall, which plays an important role in the production and transport of sediments to the coastal environments. The temperature among others, defines the type of weathering that is occurring. As mentioned in section 2.2.1, silt and clay particles are formed in different ways. Mechanical weathering occurs mainly in cold regions, where water can freeze and expand, while chemical weathering occurs mainly in climates with high precipitation and high temperatures. Since the majority of mud particles are formed by chemical weathering, muddy coasts are most abundant in the humid tropical regions. The more chemical weathering can take place, the more mud will be found on the coast.

Tides The last coastal influencer to be discussed in this section is the tidal environment. Tides are very long waves with fairly constant cycles, caused by the gravitational pull of the moon and sun, and the rotation of the Earth. Tides cause the rise and fall of the water level in coastal areas. The tidal range is defined as the difference in height between the high water level and the low water level, see Figure 5. A common approach to the classification of tidal ranges is after Davies (1964). Three categories of tidal ranges are defined; micro-tidal regime with a mean spring tidal range smaller than 2 metres; meso-tidal regime with a mean spring tidal range between 2 and 4 metres; macro-tidal regime with a mean spring tidal range larger than 4 metres. The semi-diurnal or diurnal rise and fall of the water level creates a so-called intertidal area (tidal width in Figure 5) that is exposed at low tide and submerged at high tide. In the absence of waves (or with only very low wave energy) tide dominated coasts develop wide low gradient tidal flats in the intertidal zone and subtidal zone (the area that is only rarely exposed at extremely low tides). Tidal currents in combination with the horizontal movement of the waterline determine the morphology of these tidal flats. As a result of the low energy conditions tide dominated coasts generally consist of relatively fine sediments. The distribution patterns of the sediments are exactly the opposite of those on wave dominated coasts: as the tidal currents increase in strength with greater water depth, the finest sediments are found on the mud flats and in the wetland of the upper intertidal area; the coarser sandy sediments are found in the lower intertidal area and further offshore. From the upper part of the intertidal area to the supra-tidal area (only submerged during spring tides or storm surges), salt marshes are well developed. In tropical to subtropical regions, mangroves occupy the intertidal area. Tide dominated flats occur with large tidal ranges and small wave heights. This is the case for macro-tidal coasts, but tide dominated coasts can also be found for micro-tidal regimes, as long as the wave energy is very low (e.g. in an estuary). It is the relative importance of tides and waves rather than the absolute tidal range that determines the character of the coast.

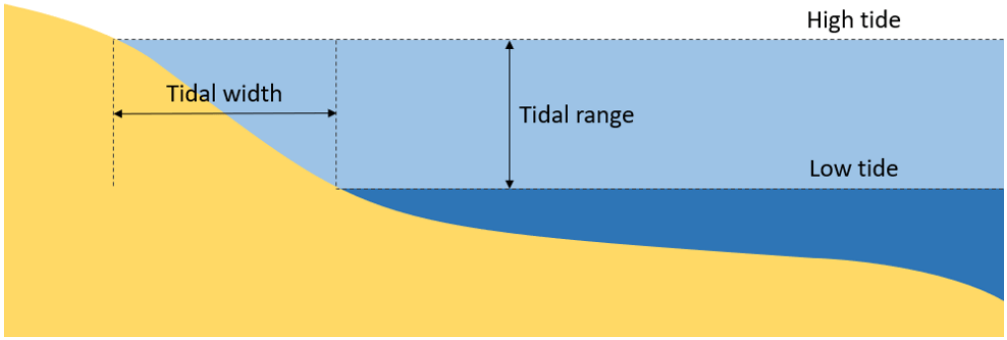


Figure 5: Definition of tidal range and tidal width, vertical and horizontal displacement, respectively, of the sea level due to tides.

Occurrence Hayes (1967) has made a graph showing the relationship between latitude and sediment classes, see Figure 6. Most mud is found in the lower latitudes from approximately 0° to 20°; which are the humid tropics with high temperatures and high precipitation rates where intense chemical weathering can

take place. But mud also occurs at higher latitudes. Sand occurs at all latitudes, with a peak between 20° and 40°. However, up to date no other study has presented the global distribution of muddy coasts.

Examples of extensive muddy coasts can be found on the Guyana coast in South America, where large quantities of mud are transported to the coast by the Amazon River and the Orinoco River (Anthony et al., 2014); near the Yellow River (Murray et al., 2014) and Yangtze River (Nasiha et al., 2019) in China large quantities of mud are found; and near the mouth of the Mississippi River in the USA muddy coasts are present. Another well-known example of a mud dominated area is the Wadden Sea, which stretches from the Netherlands to Denmark. The Wadden Sea consists of large tidal flats and salt marshes that lie between the Wadden Islands and the mainland in the south-eastern part of the North Sea.

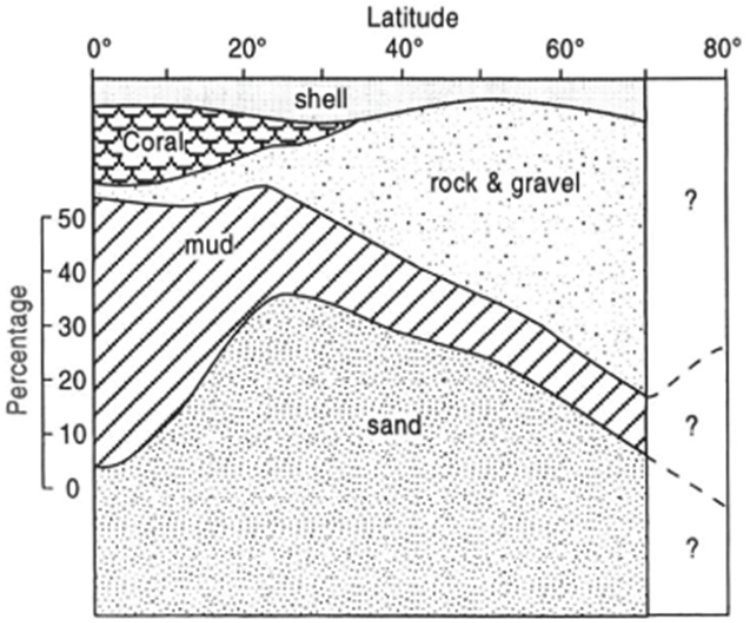


Figure 6: Relative frequency of inner continental shelf sediment types (up to about 60m depth) per latitude. Adjusted from Hayes (1967).

Functions of muddy coasts It is essential to know the distribution and dynamic behaviour of muddy coasts, given the important ecological, social and economic functions they provide. This section gives some examples of important functions of muddy coasts.

A valuable and important function of muddy coasts is the opportunity they offer for coastal ecosystems to develop. Muddy coasts are closely interconnected with salt marshes and mangrove swamps. The calm conditions, which are a prerequisite for the development of muddy coasts, and the silty character of mud, which keeps the soil moist even at low tide, are favourable for the germination of vegetation. The nature of the vegetation is highly dependent on the climatic conditions. Salt marshes occur mainly in moderate climates, at mid-latitude coasts; while mangrove swamps favour tropical and subtropical climates, at low-latitude coasts, see Figure 7. Salt marshes and mangrove swamps can develop into extremely valuable nature reserves. The specific vegetation species provide sheltered habitats for wildlife and places to breed and forage during seasonal migration. Besides, these important coastal systems provide resources for regional development and buffer zones for the land against natural disasters such as storm surges (Wang et al., 2019). The vegetation prevents or inhibits currents and waves from removing sediment by reducing the current velocities and by strengthening the soil with their root systems. Sedimentation is enhanced and the salt marsh or mangrove swamp can expand, a positive feedback system. Extensive salt marshes or mangrove swamps are a sign of a natural estuary that is largely filled with sediment. Salt marshes and mangrove swamps accumulate and stabilise sediments and protect the coastline from erosion and storms.

An example of a valuable nature reserve is the Wadden Sea. In 2009, the Wadden Sea was declared a protected nature reserve by UNESCO. According to UNESCO: *"The Wadden Sea is the largest unbroken*

system of intertidal sand and mud flats in the world, with natural processes undisturbed throughout most of the area. A large temperate, relatively flat coastal wetland environment, formed by the intricate interactions between physical and biological factors that have given rise to a multitude of transitional habitats with tidal channels, sandy shoals, sea-grass meadows, mussel beds, sandbars, mudflats, salt marshes, estuaries, beaches and dunes."[2]

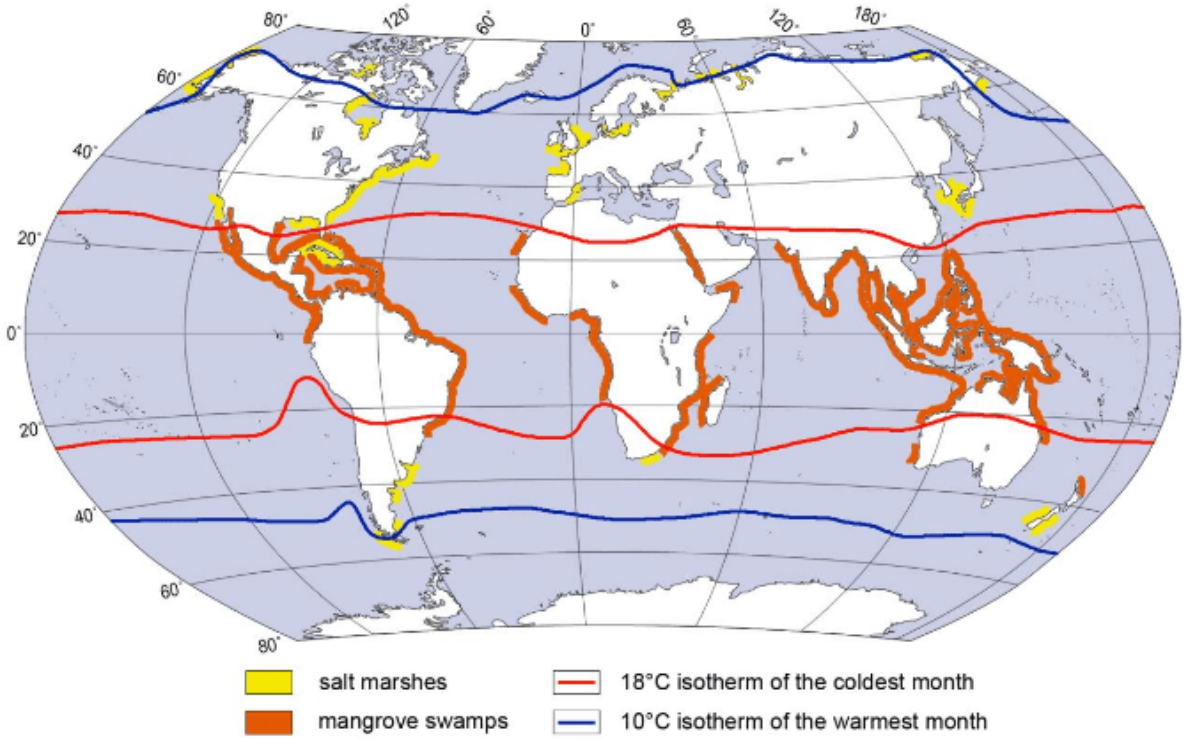


Figure 7: Global distribution of salt marshes and mangrove swamps. Mangrove swamps are found mainly at low-latitude coasts and salt marshes at mid-latitude coasts. From Bosboom and Stive (2015).

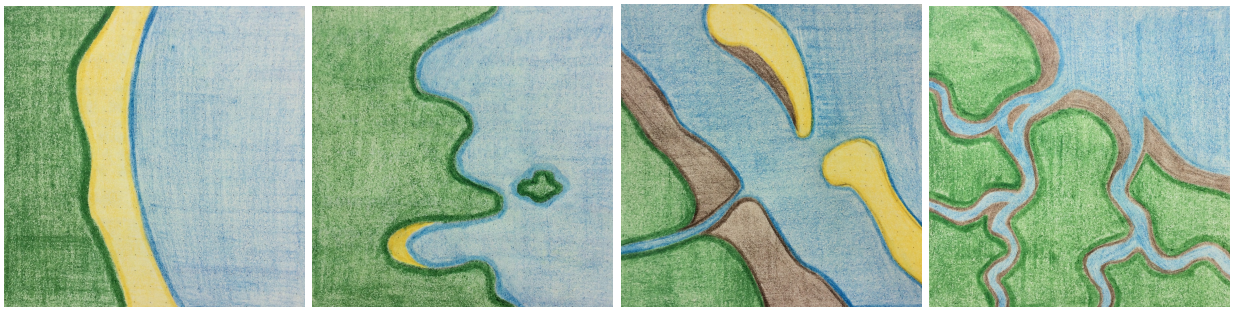
2.2.3 Other coastal environments

In order to examine distinctive features of muddy coasts, this section briefly describes four other coastal environments and some of their characteristics.

Plain beaches Plain beaches are characterised by a distinct stretch of sandy sediment (with a grain size of 0.063 to $2mm$, see Figure 4) at the transition between land and water. Sandy beaches have different configurations (wide or narrow, long or small), develop under a variety of conditions and appear around the entire globe. The sandy coastline is continuously changing and is regularly smooth. The topography close to the coastline (the beach) is expected to follow a gentle slope, further inland the topography is not bound to a specific profile. Dunes can form where the beach is wide enough, there is a large supply of sand and there are prevailing onshore winds. Mangroves are not generally expected, but there is no exclusion of mangroves either. Tidal flats can occur at the transition between land and water; it is expected that for plain beaches the tidal flats are generally narrower than for muddy coasts. Often submarine sandbars are present. Plain beaches can occur at all latitudes, are not restricted to a specific temperature range and can be found under different tidal conditions. A simplified sketch of a plain beach is drawn in Figure 8a.

Coastal cliffs Coastal cliffs are characterised by a sharp transition from land to water; the land-based cliffs disappear very steeply into the water. Along the mainly rocky and jagged coastline, little or no sediment

[2] http://whc.unesco.org



(a) Plain beach (b) Coastal cliff (c) Barrier island (d) Mangrove delta

Figure 8: Overview of four coastal environments analysed in this thesis. Blue, brown, green and yellow represent water, mud, vegetation and sand respectively. Plain beaches are characterised by a distinct stretch of sand; coastal cliffs typically have a high relief, rugged topography and a sharp transition from land to water where no sediment is found; the ocean side of barrier islands is similar to plain beaches, while on the land side the coast often contains wide mudflats and vegetated salt marshes; mangrove deltas typically form in river deltas with abundant sediment supply and are recognised by meandering river patterns, mudbanks and mangroves.

is found; the coastal cliff environment belongs to the erosional type of coasts. There is often a high-energy wave environment where the waves crash against the cliffs. This type of coasts is commonly found at leading edges. Often the land has a high relief and rugged topography. The tidal range is generally not restricted, although it is expected to be (relatively) small compared to the height of the topographical profile. Tidal flats are not expected as no sediment is found near the coastline and the (underwater) topography is typically (very) steep. Due to the lack of sediments and the presence of a high wave energy environment, mangroves are not expected. Coastal cliffs can occur at all latitudes and are not restricted to a specific temperature range, although they are generally more common at high latitudes and lower temperatures. An example of an erosional coast is given in Figure 8b. In this sketch, the land consists of steep sea cliffs and a wave-cut platform is present. There is a small narrow protected beach, which can consist of boulders, cobbles or coarse sand. The erosion is mostly wave induced.

Barrier islands Barrier islands are elongated islands off the mainland, where (sandy) beaches and dunes form on the sea side of the island and mudflats and/or vegetated salt marshes generally develop between the islands and the mainland, as a result of the protection offered by the islands against waves, currents and wind from the open ocean. For the sandy coasts on the ocean side of the island, the characteristics are similar to those of the plain beaches. On the mainland side of the island, the characteristics are different and more similar to those of the muddy coasts. The slope of the topographical profile is usually smaller on this side of the island, due to the low energy environment provided by the protection of the estuary on which this side of the island is located. The tidal range may differ from that observed on the ocean side of the island due to the shape of and the inlets into the estuary. The transition zone from land to water is expected to be wider than for the plain beaches, the same holds for the tidal flats; because the available sediment has a greater chance of settling due to the low energy environment. Mangrove forests or salt marshes may develop, depending on the temperature range and climate of the area. Figure 8c shows a sketch of a barrier island system; two sandy island are present in front of the mainland, between the island and the mainland muddy features are shown.

Mangrove deltas In certain river deltas, where the specific conditions of, among others, low energy environment, abundance of sediment supply and a tropical climate are met, mangrove forest can develop. Mangrove deltas can be recognised by meandering river patterns, mudbanks and mangroves. The topographical profile is expected to be low and to have a very a gentle slope. A mangrove delta environment is not restricted to a certain tidal range. The transition zone from land to water is expected to be wide, due to the abundant supply of sediment from the rivers that can settle on the banks. Mangroves provide a positive feedback system: the mangroves grow because of the calm energy environment and the abundance of sediment, which in turn allows more sediment to settle because of the protection against waves and currents

provided by the mangrove forests. Mangroves grow at low latitudes, in (sub)tropical climates, where the minimum temperature is not much colder than $18°C$, see Figure 7. Figure 8d shows a sketch of a typical mangrove delta with meandering rivers and mudbanks.

2.3 Multispectral satellite imagery

In this section, background information on the satellite mission relevant for this study is given in order to recognise mud patches at pixel level. Subsequently two spectral indices, which can be derived from satellite imagery, are introduced.

2.3.1 Satellite mission

Sentinel 2 (S2) multispectral satellite images, made freely available by the Copernicus program[3], are used in this study. S2 is a wide-swath, high-resolution multispectral imaging mission with a global revisit frequency of 5 days. The S2 MultiSpectral Instrument (MSI) samples thirteen spectral bands: visible and near infrared (NIR) at a spatial resolution of $10m$, red edge (RE) and short wave infrared (SWIR) at $20m$, and atmospheric bands at $60m$. The S2 MSI provides data suitable for assessing the condition and change of vegetation, soil and water cover, as well as for observing inland waterways and coastal areas. In this research, the Top of Atmosphere (TOA) images are studied. This dataset is publicly available since 23-06-2015 to date. Below, in Table 1, the available bands and their corresponding wavelengths are indicated. All bands are used in this study except B9, the water vapour band, and B10, the cirrus band. These bands have no added value in distinguishing mud from other sedimentary classes, as these bands are almost completely absorbed by the atmospheric column. B9 contains information about the amount of water vapour present in the atmosphere and does not contain information about the underlying surface. B10 measures the (cirrus) clouds at each pixel and not the characteristics of the underlying surface.

Name	Pixel size (m)	Wavelength S2A (nm)	Wavelength S2B (nm)	Description
B1	60	443.9	442.3	Aerosols
B2	10	496.6	492.1	Blue
B3	10	560	559	Green
B4	10	664.5	665	Red
B5	20	703.9	703.8	RE 1
B6	20	740.2	739.1	RE 2
B7	20	782.5	779.7	RE 3
B8	10	835.1	833	NIR
B8A	20	864.8	864	RE 4
~~B9~~	~~60~~	~~945~~	~~943.2~~	~~Water vapour~~
~~B10~~	~~60~~	~~1373.5~~	~~1376.9~~	~~Cirrus~~
B11	20	1613.7	1610.4	SWIR 1
B12	20	2202.4	2185.7	SWIR 2

Table 1: Available bands of Sentinel 2 MultiSpectral Instrument. The crossed out bands (B9 and B10) are not used in this study, as these bands are almost completely absorbed by the atmospheric column. S2A and S2B refer to the two identical S2 satellites that operate simultaneously, phased 180° to each other, they measure at minimal different wavelengths.

2.3.2 Indices

Multispectral satellite images can be displayed in natural colour (red, green, blue (RGB)) and in false colour (including infrared information). False colour images are useful for agricultural (crop classification), environmental (vegetation health) and geological applications (soil mapping). Two indices useful for vegetation

[3] https://scihub.copernicus.eu/

and water delineation are explained below. These indices will be added to the set of characteristics (bands), with which the multispectral image classifier will be trained, see section 3.

NDVI The Normalised Difference Vegetation Index (NDVI) is an index developed to identify vegetated areas and distinguish them from bare soil, urban areas, open water, etc. (McFeeters, 1996). The NDVI uses the difference in reflectance signal in the near-infrared (NIR) spectrum and in the red light spectrum for different features. The features with higher NIR reflectance and lower red reflectance are enhanced, while the features with very low NIR reflectance and low red reflectance are suppressed or eliminated. Equation 1 shows how the NDVI is calculated. The values of the NDVI range from -1 to $+1$. Healthy vegetated surfaces strongly reflect NIR wavelengths and strongly absorb red wavelengths; therefore, healthy vegetation will be enhanced and have a high value for the NDVI, close to $+1$.

$$NDVI = \frac{NIR - red}{NIR + red} \tag{1}$$

NDWI The Normalised Difference Water Index (NDWI) is an index developed to delineate open water features and enhance their presence in digital remote sensing images (McFeeters, 1996). The NDWI is derived according to the same principle as the NDVI, but uses different bands, see Equation 2. It enhances features with high green reflectance and low NIR reflectance, while the index eliminates features with low green reflectance and low NIR reflectance. Water surfaces strongly reflect green wavelengths and strongly absorb NIR wavelengths, so these surfaces will be enhanced in the NDWI and have a positive value. Terrestrial vegetation and dry soil, on the other hand, absorb green wavelengths and reflect NIR wavelengths, so these features will have a zero or negative value for the NDWI. The NDWI generally ranges from zero to one, as negative values are usually eliminated.

$$NDWI = \frac{green - NIR}{green + NIR} \tag{2}$$

2.4 Globally available coastal datasets

Having outlined some characteristics of different coastal environments in section 2.2.3, some globally available coastal datasets are selected that could capture these characteristics. This section discusses seven globally available coastal datasets that will be used in the mud classification model. The datasets are grouped according to processes that have (had) an impact on coastal formation and behaviour. Firstly, a global transect system is explained that defines the global ice-free coastline (see section 2.4.1) which will be used to extract the coastal information and later analysed for coastal classification.

2.4.1 Global transect system

For the implementation and classification of the global coastline, a global coastal transect system is used; on all (ice-free) coastlines of the world, transects of $1500m$ in length are placed perpendicular to the global coastline from the 2016 OpenStreetMap (OSM)[4] dataset at $200m$ to $450m$ intervals. The spacing of each transect depends on its location on the globe; closer to the equator, the spacing between them is approximately $450m$, while those closer to the poles are closer together. For the analysis of the coastal information it is important that each transect has the same length, because some coastal variables depend on the distance from the coastline, for example the maximum height of the transect. A transect situated on a coastline next to a mountain range is likely to have a greater maximum height as it moves inland. This global transect system has been modified from the available transect system by Deltares, see Luijendijk et al. (2018).

Approximately 1.8 million transects have been defined that cover the global coastline between 66.5° North and 66.5° South. Available coastal information is extracted every $10m$ along the transects, from which profiles are created and analysed in order to classify the coastline into a particular coastal type. This study focuses on one dimensional classification, therefore the information along the transects is translated into one relevant number.

[4] https://planet.osm.org/

2.4.2 Topographic relief

The topographic relief in the coastal regions can provide information on the coastal environment; e.g. high (variable) coastal topographic relief is often associated with coastal cliffs or rocky coasts. While low (variable) coastal topographic relief is usually related to long and meandering rivers, which provide an abundant supply of sediment to the coasts. In this paragraph one topographic globally available dataset is discussed.

Coastal topography To investigate the coastal topography, the globally available Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT DEM) is used, developed by Yamazaki et al. (2017). This model contains the global terrain elevation in metres at a spatial resolution of $90m$ at the equator. For each coastal transect, the elevation profile of the land is extracted. This is done by determining the elevation for each point along the transect at $10m$ intervals. However, the resolution of the global terrain elevation is coarser; therefore, several points along the transect will be assigned the same value. A moving average filter can be used to smooth the profile. The elevation information from each profile is captured in one relevant number that can be linked to the transect. Therefore, the ***maximum height*** is calculated along the entire profile, i.e. within a range of $1500m$ around the coastline, with the coastline ideally located in the middle at $750m$ (the location of the coastline is often not exactly in the middle of the used transect, as the 2016 OSM coastline location is used and the coastline may have eroded or accreted in the meantime). The maximum height of the land near the coastline can provide information on the type of coast encountered. For example, a muddy coast is usually found in low-lying areas (e.g. river deltas) where abundant sediment is available, as explained in section 2.2.2. Another value used is the ***variance of the height*** along the profile. This value contains information about the dispersion of the height along the transect. If the height profile is relatively flat (e.g. most coastal types where mud is found), the variance will be low. If the elevation profile is relatively steep (e.g. at coastal cliffs), the variance will be higher. These two values, which will be used for the classification of coastal types, do depend on the smoothness of the profile. However, the MERIT DEM is considered sufficiently representative to work with the maximum height and the variance of the height along the transect, as it is resolved at a resolution of $90m$. Therefore, a moving average filter is not applied.

2.4.3 Climate

Climate is a determining factor in what kind of and to what rate erosion takes place (section 2.2.2). The datasets used in this study in order to observe the climatological conditions are latitude, which is closely related to the temperature and the amount of precipitation, the minimum and maximum temperatures, which can be associated to the type of weathering that occurs, and the presence of mangrove forests, which are often found in areas with calm wave environments and can therefore be an indicator of where muddy environments can develop.

Latitude The latitude of each transect is defined as the latitude of the centre of the transect. Two values can be used: the ***latitude***, positive north of the equator and negative south of the equator; or the ***absolute latitude***. The absolute latitude is a more general variable, which does not take into account the differences between the northern and southern hemispheres. The absolute latitude is also used in the figure by Hayes (1967), where the relative frequency of sediment types is plotted against absolute latitude (Figure 6).

Temperatures Temperatures for each transect are extracted from the globally available WorldClim V1 Bioclim dataset (Hijmans et al., 2005). This dataset of bioclimatic variables is derived from monthly temperature observations from weather stations around the world recorded in the period between 1950 and 2000 and has a spatial resolution of approximately $1km$. The ***maximum temperature of the warmest month*** as well as the ***minimum temperature of the coldest month*** are extracted. These temperatures are also used to create the isotherms in Figure 7. Due to the coarse resolution of this dataset, an interpolation has to be carried out. A buffer is drawn around the centre of the transect, which will be enlarged until at least one temperature data point is observed within the buffer; finally, the average of the observed temperature data point(s) within the buffer is calculated and is assigned to the transect. A detailed explanation of this buffered-nearest-neighbour-average is given in section 10.6.

Mangrove presence Giri et al. (2011) have developed a database that shows the global distribution of mangrove forests. The 'Global Mangrove Forests Distribution, 2000' dataset is a compilation of the extent of mangrove forests from the Global Land Survey and the Landsat archive with hybrid supervised and unsupervised digital image classification techniques. The data are available with a spatial resolution of $30m$. The total mangrove area in the year 2000 was estimated at $137,760km^2$ in 118 countries and territories in the tropical and subtropical regions of the world. Mangroves grow at the land-sea transition in tropical and subtropical latitudes where they occur in conditions of high salinity, extreme tides, strong winds, high temperatures and muddy, anaerobic soils (Kathiresan and Bingham, 2001). The highest mangrove forests on earth can be found in Gabon, equatorial Africa, where the top reaches $62.8m$ (Simard et al., 2019). Many mangrove coasts are closely related to muddy coastal environments and can be used as an indicator for muddy coasts. The information from the Global Mangrove Forests Distribution dataset indicates whether or not mangrove forests are present at $30m$ pixel resolution. For processing purposes, each coastal transect is given a True/False for ***mangrove occurrence***. In addition, the percentage of mangrove pixels along the transect is extracted for each transect. However, the percentage of mangrove pixels along a transect does not necessarily imply that there is a greater or lesser chance of the occurrence of mud. Therefore, this percentage will not be used in this study for coastal classification.

2.4.4 Tides

Tide dominated coasts can develop wide tidal flats with a low gradient in the intertidal zone and subtidal zone. To identify the tidal conditions of a particularly coast, this paragraph discusses two datasets that provide information on the tidal width (horizontal displacement of the waterline due to tides) along the coast and one dataset that is used to determine the tidal range (vertical displacement of the waterline due to tides) along the coast.

Intertidal area The Murray Global Intertidal Change Dataset, provided by Murray et al. (2019), contains global tidal flat ecosystem maps at $30m$ pixel resolution generated through a supervised classification of Landsat Archive images. Each pixel was classified as tidal flat, permanent water or other using a globally distributed set of training data. In the analysis, pixels classified as tidal flats represent different types of tidal flat ecosystems, including tidal mudflats, tidal sand flats and wide tidal rock-platforms, while excluding spectral features indicating the presence of vegetation-dominated intertidal ecosystems such as mangroves and vegetated salt marshes. The analysis aimed to identify pixels that are regularly inundated by tides, and can therefore include other intertidal systems where intertidal dynamics can be observed. The dimensions of intertidal areas can contribute to the classification of coastal types, e.g. most coastal cliffs have no intertidal area at all; plain beaches can have intertidal areas, but these are generally less wide than the tidal areas found along muddy coasts, see sections 2.2.2 and 2.2.3. To process the intertidal area data, for each transect the ***percentage of tidal flat pixels*** along the transect is determined. This percentage indicates the width of the tidal flat that overlaps the transect and could potentially distinguish between muddy and other coastal types.

Global Surface Water Another dataset that can be used to estimate the ***width of the transition zone*** from land and water is the Global Surface Water (GSW) dataset, developed by Pekel et al. (2016). This dataset contains global maps showing the frequency with which water was present at pixel level ($30m$ resolution) in the period between 1984 and 2020. A high frequency means that there has always been water, a low frequency means that there has never been water. A frequency of 50% means that half of the time there has been water present and half of the time it has not; this can be interpreted as the intertidal area where dry and wet areas alternate. A water frequency profile is drawn along each transect. From the GSW dataset the width of the frequency profile can be calculated, which is equal to the distance over which a GSW value between 0 and 100% is measured along the transect. In order to remove the nearly stable (converging) start and end of the frequency profile, the 90 percentile of the frequency is used to calculate the desired characteristic, this is the frequency between 5 and 95% with which water was present. As mentioned in section 2.2.2, the width of the intertidal area varies between coastal types. In general, sandy or rocky beaches have a narrower intertidal areas than muddy beaches. This parameter can potentially be used to distinguish between muddy and other coastal types.

Tidal range To determine the vertical displacement of the waterline due to tides, the ***tidal range*** along the coast is estimated. The tidal indicators mean higher high water (MHHW) and mean lower low water (MLLW) from the Deltares Global Tide and Surge Model (GTSM) are used[5]. The MLLW refers to the mean of the lowest tide during one day in the period 1979-2014, the MHHW refers to the mean of the highest tide during one day in the same period. The resolution of the GTSM is approximately $5km$ in the coastal areas. An interpolation is required to extract the variables of interest for each global transect. The interpolation is performed in the same way as for the temperature dataset (see section 2.4.3 and the detailed explanation in section 10.6). A buffer is drawn around the centre of the transect, which is enlarged until at least one tidal data point is observed within the buffer; finally, the average of the observed tidal data point(s) within the buffer is calculated and is assigned to the transect. The tidal range (TR) is calculated by subtracting the MLLW from the MHHW, see Equation 3. The tidal range in itself will potentially not be a distinctive feature for the classification of muddy coasts; it is rather a combination of the tidal range and other coastal features (wave climate, continental shelf configuration, etc.) that determines the morphology and dynamic behaviour of the coastline.

$$TR = MHHW - MLLW \tag{3}$$

2.5 Developing a global hybrid mud classification model

The definition and the characteristics of individual mud particles are known (2.2.1) and the satellite mission that will be used for analysing these characteristics is selected (2.3). The characteristics and behaviour of muddy coastal systems are explained (2.2.2) and the globally available coastal datasets that will be used for analysing these characteristics are selected (2.4). With this information a global mud classification model can be developed. A brief overview follows, consisting of eight steps that are taken to develop a global hybrid classification model that is able to detect muddy coasts at global scale, see Figure 9. The first two steps are based on pixel level and use individual multispectral images; the third step uses the multispectral classification based on pixel level and individual images, that is applied on transect level and composite images; the last five steps are based on transect level and use composite multispectral images. The icons of each step of figure 9 will appear in this report next to the paragraphs explaining the details of this specific step.

Labelling training points To develop a multispectral image classification model that can recognise muddy pixels at global scale, a training dataset is created consisting of 7500 labelled data points (pixel level). These data points are (nearly) randomly distributed around the globe to capture different coastal classes under different conditions. The data points were labelled manually by inspection of individual RGB images taken during the period 01-01-2020 to 31-12-2020. Seven coastal classes have been defined: sand, mud, water, turbid water, vegetation, dry vegetation and other.

Developing multispectral image classification From the 7500 labelled data points, the spectral reflectance properties measured with the S2 MSI are extracted and analysed. An automated random forest classification model is trained with the spectral properties and labels of the data points. This multispectral image classification model can be applied to any pixel at global scale and will label each pixel with the class that best matches the observed spectral properties of the pixel.

Extracting multispectral classified profile With the multispectral image classification model coastal transects will be classified. A composite multispectral satellite image from the year 2020 of the buffer zone around the coastal transect is classified using the multispectral image classification model. As mentioned in 2.4.1, the transects measure $1500m$ in length and data of the classified image is extracted every $10m$ (the resolution of the S2 images is also $10m$). This creates a vector containing the class of each pixel along the transect. Figure 10 shows an example of a multispectral classified profile of a transect located on an atoll island in Fiji. From the S2 RGB composite image (10a) it can be seen that the water is clear, a reef can be observed in front of the island, the island itself is covered with green vegetation. Two coastal classes are

[5] https://publicwiki.deltares.nl/display/BED/References

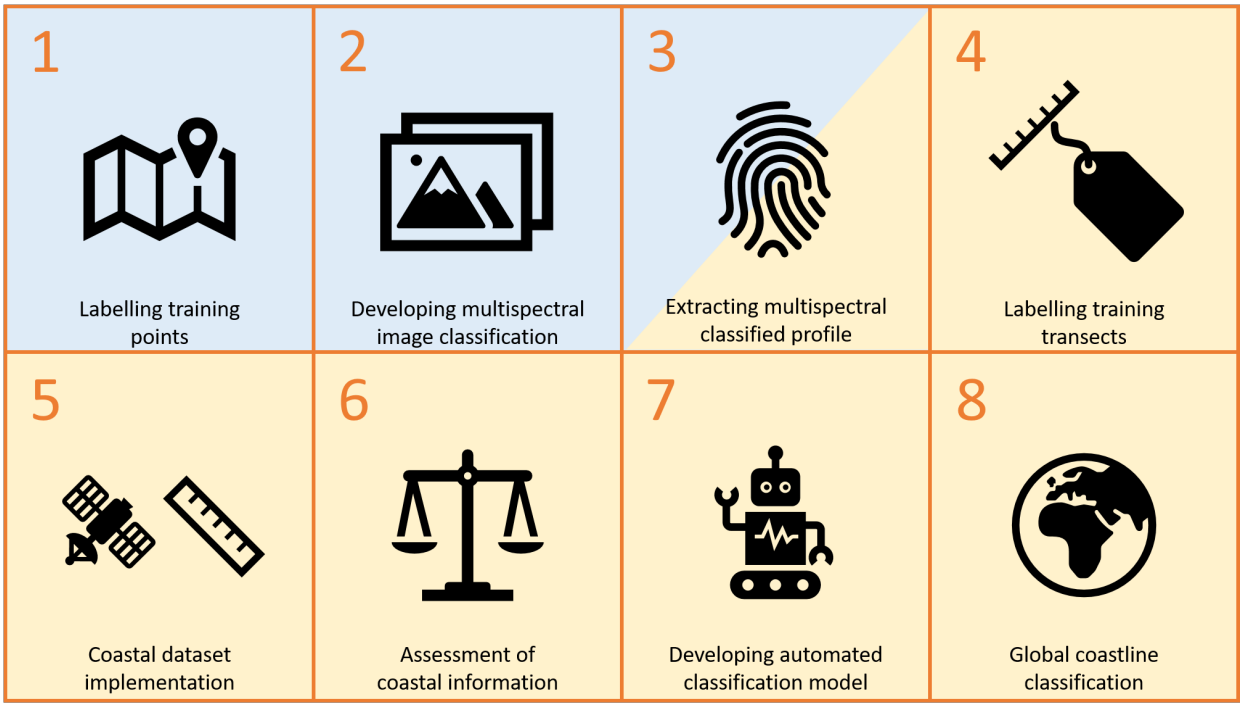


Figure 9: Developing an automated global mud classification model. The blue shaded boxes relate to pixel based classification that uses individual multispectral satellite images, the yellow shaded boxes relate to transect based classification that uses composite multispectral satellite images.

represented in the classified profile (10b); water and vegetation. There are several approaches to convert the classified profile into a representative signature. For each class (sand, mud, water, turbid water, vegetation, dry vegetation, other), every transect is labelled with True/False if the class is present in the profile. For example, if mud pixels are present along the transect (True), there is a possibility that the transect is located on a muddy coast. This could be a first, rough, filter for identifying muddy coasts. Another approach is to calculate percentages of occurrence of each class along every transect. These percentages compare the presence of each class. For example, if more mud pixels than sand pixels are present, there is a possibility that the transect is located on a muddy coast instead of on a sandy coast. The percentages contain more information than the True/False labels. A third option is to analyse the order of the coastal classes of the classified profiles. For example, the order of appearing coastal classes at the transition from land to water on a plain beach could be: water, sand, vegetation; whereas on a muddy coast it could be: water, turbid water, mud, vegetation. However, in this study a one dimensional coastal classification is developed, therefore the order of appearing coastal classes is not used. The percentage of occurrence of each class is used though, because the percentages contain more information than the True/False labels. The multispectral classification profiles will be used as input in the final global coastal transect classification model.

Labelling training transects To develop a final global coastal transect classification model that can recognise muddy coasts at global scale, a training dataset is created consisting of 1868 labelled transects (transect level). These training transects are randomly distributed around the globe to capture different coastal environments under different conditions. The training transects were labelled manually by inspection of composite RGB images from the year 2020. The labelling focuses on the coastline, although the transects measure $1500m$. Contextual information from the image is taken into account, such as the presence of white wash caused by breaking waves, the presence of any textures/patterns in the sediment, the shape of the coastline, etc.. Five coastal types have been defined: plain beaches, muddy coasts, coastal cliffs, vegetated coasts and other.

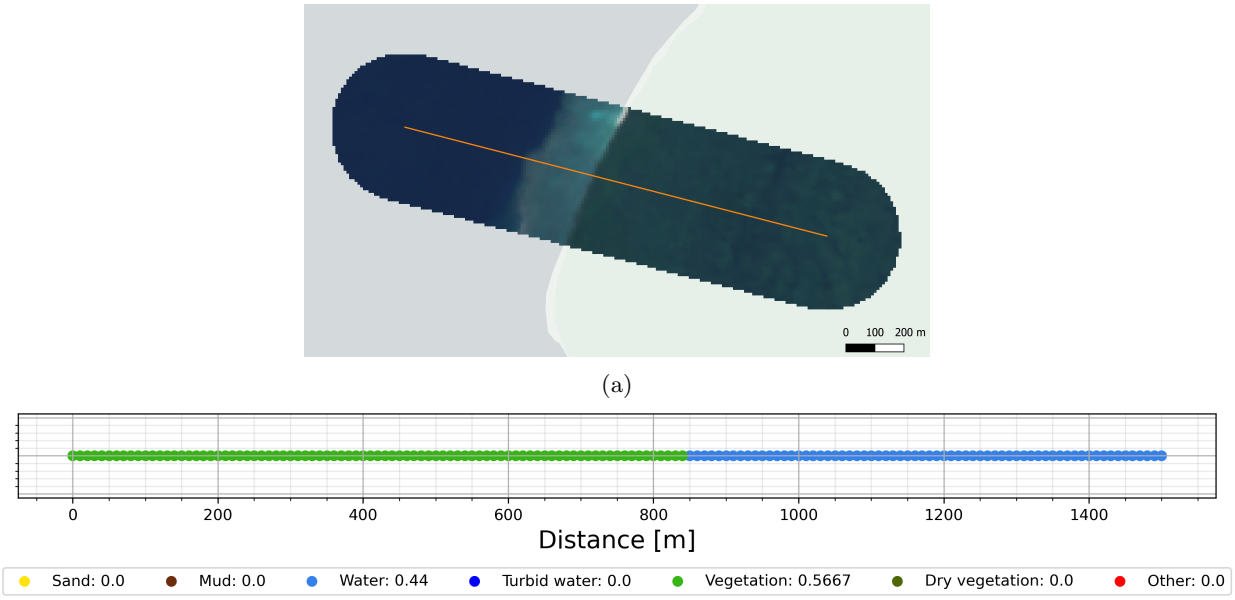


Figure 10: Transect located on an atoll island in Fiji. (a) The S2 RGB composite image of the buffer around the transect, the orange line represents the transect and measures $1500m$; (b) The multispectral classified profile of the transect. The distance along the transect is plotted on the x-axis, $0m$ is at the landward side of the transect (right side of (a)). The numbers in the legend are the percentage of occurrence of each coastal class.

Coastal dataset implementation In theory, a global coastal transect classification model can now be developed using the multispectral image classification profiles of the labelled transects. However, it is expected that the addition of certain globally available coastal datasets can increase the reliability of the model by adding information that cannot be observed with multispectral satellite images. Section 2.4 already described the globally available coastal datasets and some of the variables that can be used to train the transect classification model. The ten coastal variables that will be extracted from the seven coastal datasets for each transect are maximum height, variance of height, latitude, absolute latitude, maximum temperature of warmest month, minimum temperature of coldest month, mangrove occurrence, width of tidal flat, width of transition zone and tidal range. In Figure 11 a conceptual transect profile is sketched of a potentially muddy coast. The horizontal bar at the bottom of the figure represents the multispectral image classification profile; the height profile of the coastal area is sketched, the tidal range is displayed, as well as the mangrove occurrence (green vertical bar), the tidal flat (pink vertical bar) and the transition zone (curved part of the orange line).

Assessment of coastal information With all the collected coastal information, the characteristic 'fingerprints' of each of the five coastal types can be evaluated. This will show how certain variables are distributed for the different coastal types. From this distribution it can be argued which of the variables have distinctive potential.

Developing automated classification model In the final development of an automated transect classification model, three models will be developed and assessed: (1) the multispectral model, using only the information of the multispectral classification profile; (2) the coastal model, using only the information of the globally available coastal datasets; (3) the hybrid model, using both the information of the multispectral classification profile and the globally available coastal datasets.

Global coastline classification The final model with the best performance will be used to classify all the global coastal transects. An automated model will be developed that firstly applies the multispectral

image classification to the composite multispectral satellite image around the transect; secondly, extracts all the coastal variables from the globally available coastal datasets; thirdly, analyses the obtained data from the transect and classifies the transect into the coastal type that best matches the observed coastal characteristics of the transect. Finally, a map will be presented showing the global distribution of muddy coasts.

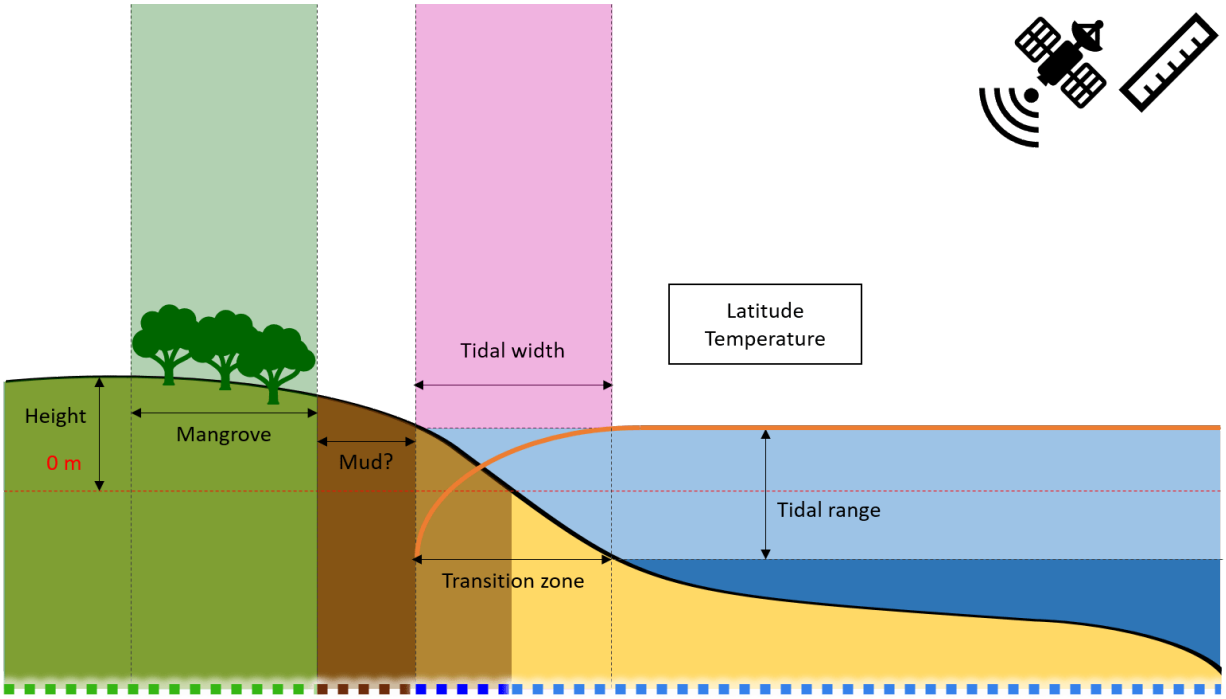


Figure 11: Conceptual transect profile of a potentially muddy coast. The horizontal bar at the bottom of the figure represents the multispectral image classification profile (similar to Figure 10b), where green, brown, dark blue and blue represent vegetation, mud, turbid water and water respectively. The horizontal red dashed line indicates the mean sea level where to the height profile is related. The green and pink vertical bars represent the mangrove occurrence and the tidal flat detection obtained by the coastal datasets. The mean higher high water and the mean lower low water are indicated, which define the tidal range. The transition zone is defined as the curved part of the orange line, which represents the Global Surface Water frequency profile.

3 Multispectral image classification

The spectral characteristics of various coastal classes measured at pixel level across the different bands of the S2 MSI are used to develop a multispectral image classification model. This section consists of three steps, with each step being a continuation of the insights gained in the previous step(s) and each step involving a scaling up of data and study areas of the previous step(s). This section presents the multispectral image classification method on three scaling levels, the results obtained, the performance of the image classification model and the first conclusions that can be drawn from analysing multispectral satellite images.

3.1 Method

The development of the multispectral image classification model will be discussed in detail for each step in this section. The first step involves the development and implementation of a locally trained image classification model. The second step is designed to test how representative a locally trained image classification model is when applied to other local areas. In the final step, a globally trained image classification model is implemented. Each classification model is constructed on the basis of the analysis of the spectral characteristics of mud patches.

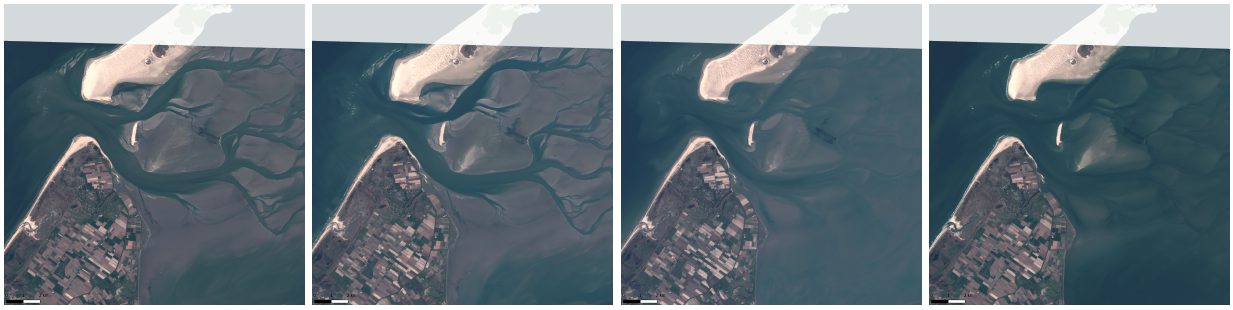
3.1.1 Local image classification

In order to obtain a local multispectral image classification model, the area between the island of Texel and the island of Vlieland in the Dutch Wadden Sea is analysed. It is known from literature that the beaches, especially on the North Sea side of the islands, consist of sandy patches, while the Wadden Sea area (the part between the mainland of the Netherlands and the islands) consists mainly of tidal mud flats (Wolff, 2013). The Wadden Sea is a typical example of a barrier island system. Four S2 cloud free multispectral satellite images were selected with different tidal stages. During this local assessment only four coastal classes are identified in the images: sand, mud, water and vegetation. Forty data points are manually selected and labelled in QGIS Software (Open Source Geospatial Foundation[6]), by a visual inspection of individual RGB images. These data points were used to train an image classification model and resulted in four classified images. The aim of this small-scale classification model is to assess the ability of an image classifier, trained with multispectral satellite images, to detect mud patches.

Study area The study area used in this step is the Dutch Wadden Sea. The Wadden Sea is located in the southeastern part of the North Sea, extends from the Netherlands to Denmark and has a total surface area of approximately $8000km^2$. The Wadden Sea is a shallow coastal area, consisting of barrier islands and a large area of (inter)tidal flats (Wolff, 2013). The Wadden Sea is connected to the North Sea via inlets between the barrier islands, whereby sediment from the North Sea is brought in by the tides. The tidal range in this area varies from 1.4 to $3.4m$ (Postma, 1982), with the tide propagating from the western part of the Wadden Sea to the eastern part. There are also several freshwater inlets. The area is known for its tidal mud flats and tidal salt marshes where millions of migratory birds pass through every year. The exact area of interest is located between the first two Dutch barrier islands, Texel and Vlieland, namely Het Eierlandse Gat (53°12′ N, 4°51′ E), shown on the images of Figure 12.

Method Figure 12 shows four RGB images of the area described above. All images are from the S2 TOA data collection between 01-01-2020 and 31-12-2020. From the image collection, first the cloudy images were filtered, then four cloud-free images with different tidal stages were selected. The tidal stage of each image was determined with the freely available dataset of Rijkswaterstaat Waterinfo[7], from which the astronomical tides measured at Vlieland Haven have been used. From left to right the images show an increasing tide. The training data consists of 40 points, which were manually selected by a visual inspection of the RGB image shown in Figure 12c. Four coastal classes are selected: sand, mud, water and vegetation. From these training points, the properties (the reflectance value for each band) are extracted, as well as the class they to which they are assigned. With these training points, an image classification model is trained and a

[6] http://qgis.org
[7] https://waterinfo.rws.nl/#!/kaart/astronomische-getij/



(a) (b) (c) (d)

Figure 12: Study area of the local image classification model, first step of the multispectral image classification phase. Four individual RGB TOA images of a part of the Dutch Wadden Sea are shown (BOX_187_108), with varying tidal stages, from low tide (left) to high tide (right): (a) tidal stage of $-102cm$; (b) $-59cm$; (c) $-25cm$; (d) $+67cm$.

SmileRandomForest (Breiman, 2001) is implemented. The output of the classifier is the assigning of each pixel to the corresponding class (sand, mud, water, vegetation) that best matches the observed spectral properties of the pixel. This classifier was used to classify the four RGB images, which are shown and discussed in section 3.2.1.

3.1.2 Cross image classification

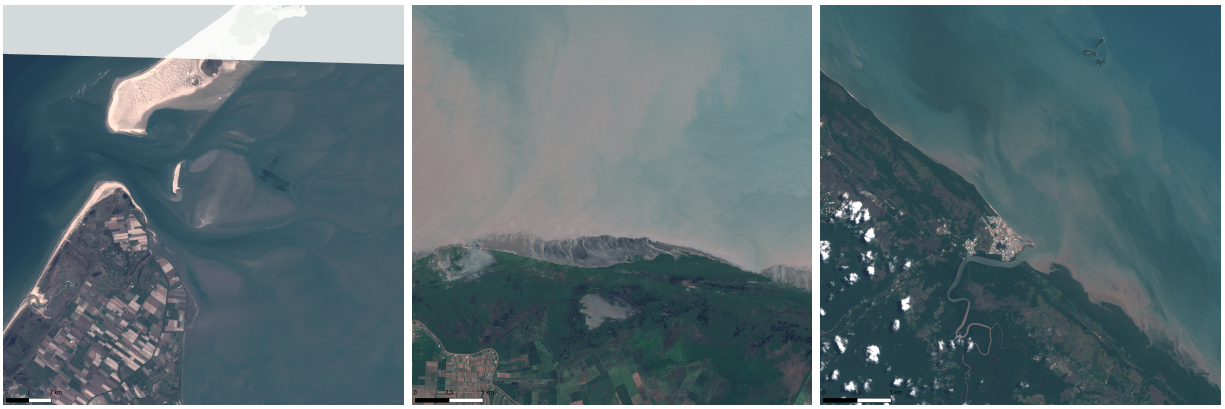
To test the representativeness of a locally trained image classifier, coastal environments in three different countries are analysed in this second step: the Netherlands, Suriname and French Guyana. These regions all contain mud patches according to literature and a visual observation of satellite images. At the three locations training data is collected and based on this training data three different (locally trained) classifiers are developed. These classifiers are used to cross-classify the three locations that are under different conditions. The purpose of the cross classification model is to investigate how well a locally trained classifier performs when applied to another local site (with different characteristics). The performance of the locally trained models is assessed by a visual inspection of the classified images. Finally, the three training datasets are combined to train one (more robust) classifier and to test the performance of the classifier when different conditions are combined.

Study area In this step, three study areas are analysed, which are all known for their (partial) muddy character. The first study area, a part of the Dutch Wadden Sea, is the same area that is used in the development of the local image classification model, which is described in section 3.1.1, see Figure 13a.

The second study area is located on the coast of Suriname, in Bigi Pan national park. This is located about $200km$ west of Paramaribo, the capital of Suriname. This area is part of the Guyana coastline, which is a well-known example of a mud-dominated coast. The Guyana current transports Amazonian sediments (mainly mud sediments) along the north coast of Brazil into French Guyana and Suriname (Augustinus, 1980). The coast of Suriname can be categorised as a low to moderate energy coast. Some low waves are generated by the north-eastern trade winds, which propagate over a shallow, smooth continental shelf (Suriname is located at a trailing edge). The slope of this shelf varies from $1:1600$ to $1:3000$, approximately in the direction of wave propagation. The amplitude of the tidal range is about $3m$ during spring tides and approaches the coast almost perpendicularly. These conditions allow for the development of large intertidal areas, where mudbanks migrate westward and constantly change shape and orientation due to the combined action of waves and currents (de Vries et al., 2021). Along the coastline, on the higher part of the mudflats, extensive mangrove forests have developed. An RGB image of this study area is shown in Figure 13b.

The last study area is located on the coast of French Guyana, around the city of Kourou, which is about $45km$ northwest of the capital of French Guyana, Cayenne. This coast is also part of the Guyana coastline and consisst of an extensive intertidal area with several mudflats (Froidefond et al., 2004). This area was added to the study area, because the ocean water in the RGB image looked different (in colour) from the water in the RGB image of the Surinamese coast, see Figure 13c. However, the conditions are quite similar

to those observed on the Surinamese coast.



(a) (b) (c)

Figure 13: Study areas of the cross image classification model, second step of the multispectral image classification phase. Three individual RGB TOA images are shown of parts of (a) the Dutch Wadden Sea (BOX_187_108); (b) the coast of Suriname (BOX_117_030); (c) the coast of French Guyana (BOX_118_054).

Method Figure 13 shows three RGB images of the study areas described above. All images are taken from the S2 TOA data collection between 01-01-2020 and 31-12-2020, from which the least cloudy image at each location is selected. Training data points are collected from each image. The training data for the Dutch Wadden Sea is the same as is used in the development of the local image classifier (see section 3.1.1) and consist of four coastal classes defined in 40 data points. From the image of the coast of Suriname, 73 training points are selected, consisting of four coastal classes: mud, water, vegetation and a new class 'other', which contains mainly buildings and roads. From the image of the coast of French Guyana, 37 training points are collected, consisting of only three coastal classes: water, vegetation and other, in this case the class 'other' consists mainly of clouds. From these new training points the spectral reflectance properties are extracted and the training points are manually labelled with the corresponding coastal class. For each location, an image classifier is trained with the properties from the training points at that location and a SmileRandomForest is implemented. The output of each classifier labels each pixel with the corresponding class (sand, mud, water, vegetation, other) that best matches the observed spectral properties of the pixel. Three classifiers are developed, one per location, trained with the training data collected at that location. Each classifier is applied to the three study areas (cross classification) to investigate the similarities and differences of the observed spectral reflectance properties of the identified coastal classes between the various locations and conditions. After examining the performance of the locally trained classifier applied to other local environments by a visual inspection of the RGB and classified images, a fourth image classifier will be developed using all training points collected from the three locations defining five coastal classes (sand, mud, water, vegetation, other) combined. This combined image classifier is expected to outperform the locally trained classifiers, according to a visual performance assessment by inspection of the classified images. In section 3.2.2, the results of the cross image classification models are discussed.

3.1.3 Global image classification

1

To develop a global image classification model, a global training dataset is required. 7500 training data points are collected manually at 75 locations around the world by visual inspection of RGB TOA images taken with S2 in the period between 01-01-2020 and 31-12-2020. The training data consists of seven coastal classes: sand, mud, water, turbid water, vegetation, dry vegetation and other. An image classifier is trained with the spectral reflectance properties of these 7500 training points. Several global locations are classified and assessed which are shown in section 3.2.3.

Study area The study area is including the entire ice-free global coastal area in this final step of multispectral image classification. The global coastal area is divided into approximately 24 thousand square boxes ranging from $10x10km$ to $30x30km$, depending on the location on the globe. A total of 75 boxes are selected from which 100 training points per box are collected, attempting to capture different coastal environments on the six continents (excluding Antarctica). In Figure 14 the locations of the 75 locations are indicated with orange dots.

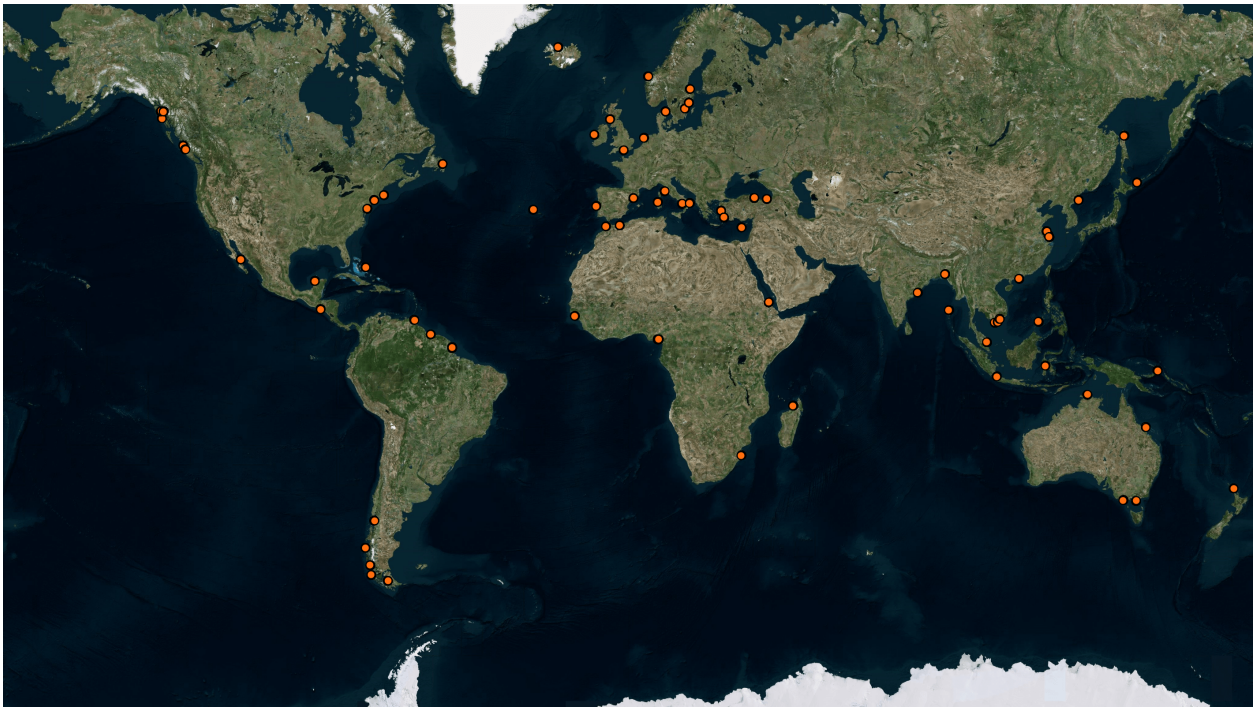


Figure 14: The global distribution of the training points used for the multispectral global image classification. The orange dots represent 75 locations where the training data is collected.

2

Method For every selected box the least cloudy RGB image is selected from the S2 TOA data collection between 01-01-2020 and 31-12-2020. The training data consists of 100 points per box, that are manually selected by visual inspection of the RGB images. The training data consists of seven coastal classes: (1) sand, selected from all types of sandy beaches around the globe; (2) mud, selected from clearly visible and recognisable mudflats; (3) water, containing (clear) water from oceans, lakes and rivers; (4) turbid water, selected in brownish water next to mudflats and other places where brown plumes are clearly visible; (5) vegetation, selected from unambiguous green healthy vegetated areas; (6) dry vegetation, consisting of vegetated areas which are just harvested or dry vegetation which have a brown/yellow colour and (7) other, containing clouds, snow, buildings, rocks/mountains and all other indefinable points. Due to the (nearly) random selection of training points in different coastal environments, the distribution of the manually labelled coastal classes is unequally balanced (the possible consequences for the results obtained with the multispectral global image classifier are discussed at the end of this report in section 7.3.1). From these training points the spectral reflectance properties of each band are extracted, as well as the class they are assigned to. With these training points the global image classifier is trained and a SmileRandomForest is implemented. The global image classifier assigns each pixel a label with the corresponding coastal class (sand, mud, water, turbid water, vegetation, dry vegetation, other) that best matches the observed spectral properties of the pixel. With this global image classifier all coastal areas can be classified. In section 3.2.3 some examples of globally classified images are shown and discussed.

3.2 Results

The results of the three steps of the multispectral image classification method (3.1) are presented and evaluated. First, the local image classification model is discussed and classified images of the Dutch Wadden Sea are shown, together with a spectral signature plot of four coastal classes (sand, mud, water, vegetation). Then the results of the cross image classification model are discussed. Here, some cross-classified images are shown and compared with the images that are classified with a larger and more diverse training dataset. Finally, the global image classification model is discussed. The results obtained are shown in relation to the coastal environment (described in section 2.2.3) in which the analysed location fits.

3.2.1 Local image classification

For the development of the local image classification model, an area in the Dutch Wadden Sea is analysed, as described in section 3.1.1. In this paragraph the obtained results are discussed, classified images of the Dutch Wadden Sea are shown and the spectral signatures of the four coastal classes (sand, mud, water, vegetation) are reviewed.

Classification The RGB images (Figure 15, top images) show that at a lower tidal stage (left figure), more mudflats and mudbanks appear. However, it is not possible to tell from the images whether the mudflats and mudbanks are just below or just above the water level. Another observation from the RGB images is that the shape of the island of Vlieland (top island) changes during the tidal stages. The lower the tide, the more sand appears, especially at the southern side of the island. There is also a difference in the colour of the sand, which can be dry(er) in one (part of the) image and wet(ter) in another. The bottom images in Figure 15 are classified using the classification model trained with the 40 training data points manually selected from the RGB image in Figure 15c. The classified images clearly show the occurrence of mud (brown) at low tide. The water channels (blue) between the banks, especially at the lower tidal stages in Figures 15e and 15f, are well represented. The interior of Texel (the bottom island) is predominantly classified as vegetation (green). Inland, some ditches and canals are also visible, which are classified as water. Part of the mudbanks is classified as sand (yellow) and some vegetation can be recognised within the mudbanks. Near the southern part of Vlieland in the lower tidal stages, a sand- or mudbank is present, but upon visual inspection it is unclear which of the two it is. The classifier uses both the sand and mud class to identify this area as well (Figures 15e and 15f). Furthermore, it seems that the classifier does not distinguish between dry and wet sand. It is also striking that the breaking waves on the North Sea side of Vlieland are classified as mud, as well as the water close to the coastline.

Findings In qualitative terms, this locally trained image classification performed reasonably well for this specific location; the four defined coastal classes are all represented in the classified images in the areas where they are expected.

Spectral signature Figure 16 shows the spectral signatures of the four different coastal classes (sand, mud, water, vegetation). The spectral signature is created by selecting approximately 4000 random pixels from the classified image of each tidal stage. From these total 16000 pixels, the mean reflectance values for each band as well as the $50th$, $70th$ and $90th$ percentiles of the reflectance for each band are calculated. By plotting the reflectance against the wavelengths (bands), the spectral signatures are created. Each class has its own characteristic signature. The idea is to find a distinctive signal (spectral gap) for mud, which can help to distinguish mud from the other coastal classes. Figure 16 shows that the signature of sand (yellow) has the highest reflectance in almost all wavelengths of the four coastal classes considered. In the lower wavelengths (from visible to NIR), the signature of sand shows an increasing tendency, peaking at 0.3 in SWIR 1. The signature of mud (brown) shows a curve that slightly decreases from 0.15 to 0.11 in the visible spectrum; in the RE to NIR spectrum the reflectance is approximately stable around 0.11; for the higher wavelengths (SWIR spectrum) the reflectance of mud decreases to 0.02. The class of water (blue) has a characteristic signature with a relatively small variability, compared to the other coastal classes. The reflectance in the lower wavelengths decreases from 0.12 to almost zero in the higher wavelengths. The signature of vegetation (green) shows a similar behavior to the mud curve in the visible spectrum, then increases rapidly, reaching

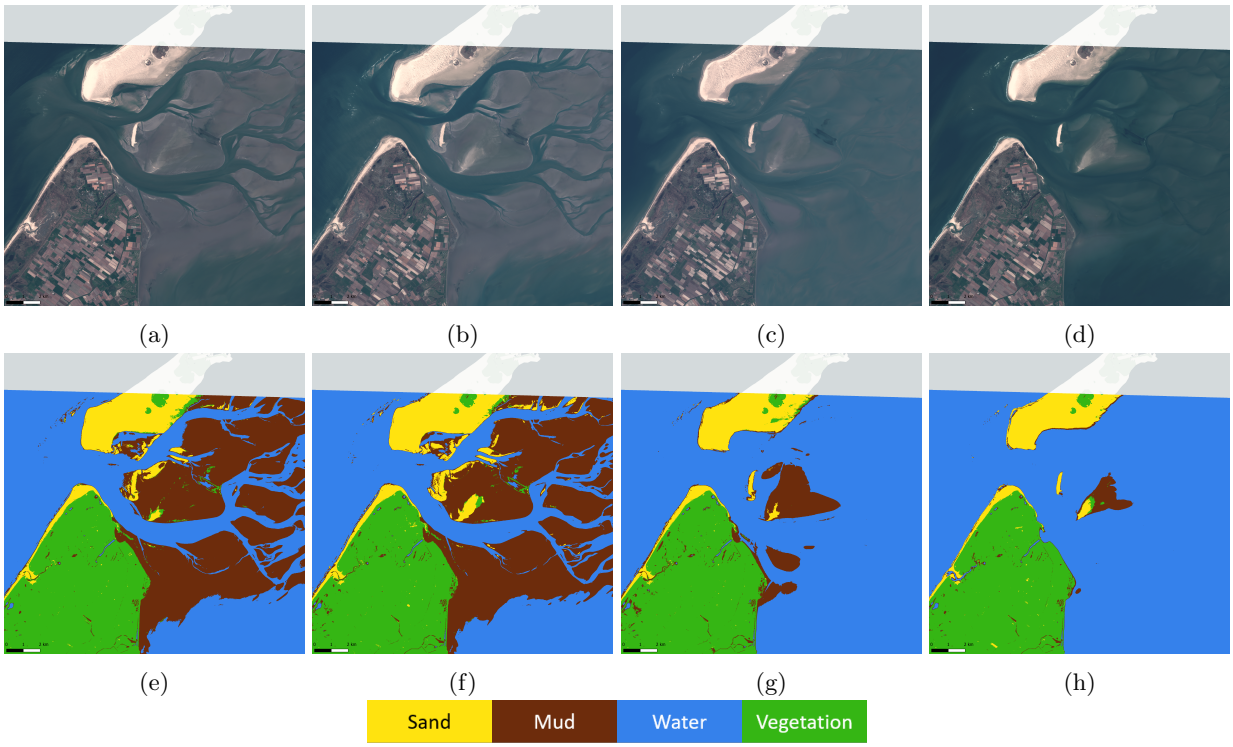


Figure 15: Study area of a part of the Dutch Wadden Sea (BOX_187_108), used in the local image classification model, first step of the multispectral image classification phase. The top figures show four individual S2 TOA RGB images, the bottom figures are the corresponding classified images. The tidal stage increases from low tide (left) to high tide (right). Four classes can be distinguished: yellow for sand, brown for mud, blue for water and green for vegetation.

a peak of 0.29 in the RE4 band. In the higher wavelengths (SWIR spectrum), the reflectance decreases to 0.16.

Findings The spectral signature plot of the local analysis of the Dutch Wadden Sea shows that sand generally has a high reflection in all wavelengths, which is understandable because sand also appears bright in RGB images. On the other hand, mud and water absorb most of the energy of the wavelengths (the wetter something is, the more is absorbed), they usually appear dark(er) in RGB images. Vegetation shows the typical signature where energy of lower wavelengths is absorbed and energy of higher wavelengths is reflected, this is why the NDVI can indicate vegetation. What is also clear from the plot is that the signature of mud stands out between the wavelengths of approximately $700nm$ and $940nm$. The signal shows no overlap with the other classes at those wavelengths, which can be promising to distinguish mud from other coastal classes on a larger scale. However, the analysed data is still too limited in size and spatial coverage to draw any conclusions yet.

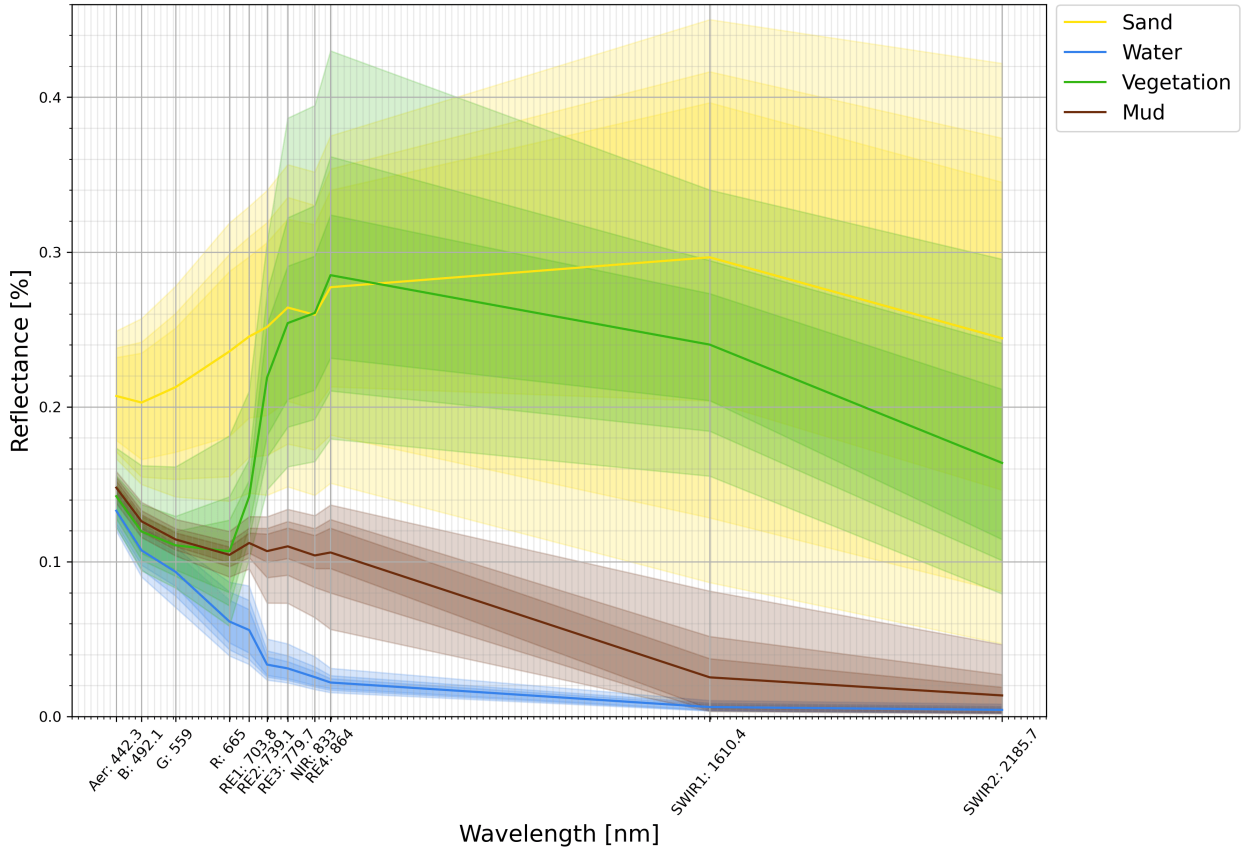


Figure 16: Spectral reflectance signature of sand (yellow), mud (brown), water (blue) and vegetation (green). The wavelength (in nm) of the used S2 bands is plotted on the horizontal axis; the TOA reflectance is plotted on the vertical axis. The filled lines represent the mean reflectance of each class; the shaded areas, from dark to light, represent the $90th$, $70th$ and $50th$ percentiles of the reflectance of each class respectively.

3.2.2 Cross image classification

The cross image classification method uses three different classifiers trained with training points from three locations that are under different conditions. The performance of the locally trained classifiers is compared with the performance of the classifier for which different conditions are combined, by visual inspection of the classified images. The combined classifier is expected to perform better because the spectral reflectance signatures of each coastal class are more generalised and therefore less biased towards one specific location.

Classification The first row of Figure 17 shows the Eierlandsegat between Texel and Vlieland in the Dutch Wadden Sea. A closer look reveals several similarities and differences between the two classified images in Figures 17b and 17c. In Figure 17b, the south-eastern part of Vlieland is mainly classified as sand and vegetation, whereas in Figure 17c the same part of the island also contains water. In Figure 17b the breaking waves are recognised as mud, whereas in Figure 17c it is water. The large mudflat in the middle of the image is recognised in both images, where it has similarities in shape and sandy areas. In Figure 17b there is some vegetation classified within the large mudflat, in Figure 17c this is mainly mud. In both images the water very close to the coastline is classified as mud.

The second row of Figure 17 shows the RGB image of the Surinamese coast as well as two classified images of the same area. From the first classified image (Figure 17e) it is immediately clear that the water in front of the Surinamese coast has different characteristics than the water observed in the Dutch Wadden Sea with which the local classifier is trained; the part that should be water is classified as sand and mud in the image, which is obviously incorrect and not as expected. This is a good example of how a local classifier trained on examples from one region, does not perform well in another region, with a different coastal morphology that is under different coastal conditions. Sand, mud and water are mixed up by the local classifier. The second classified image (Figure 17f) is classified using the combined classifier trained with training data collected at the three different locations described before. Here, the classified image looks more like what is expected based on the RGB image. The mudflats between the vegetated land (green) and the ocean (now blue) are detected by the classifier. There is no sand present on the coast, which is as expected according to the literature and description of this study area in section 3.1.2.

For the third location, the coast of French Guyana, the results of the classified images are shown in the third row of Figure 17. The locally classified image (Figure 17h) shows some muddy areas, where in reality there are not only mudflats present but also a lot of suspended sediment, as shown by the brownish plumes visible in the ocean in the RGB image (Figure 17g). In addition, the clouds are classified as sand, because this type of class was not yet included in the locally trained classifier of the Dutch Wadden Sea. In the combined classified image, these errors have been (roughly) resolved; the mudflats are now only present very close to the coastline as expected based on the RGB image and the clouds appear as 'other', the newly added coastal class.

Findings In summary, the cross classification method shows that not all sand, mud, water and vegetation that is found on the globe have the same signature as the sand, mud, water and vegetation found in the study area of the Dutch Wadden Sea. Each class shows (slightly) different characteristics at different locations, due to different conditions. Therefore, it is recommended to use a larger training data set, collected from many different coastal areas on the globe. The cross classification method also shows that when combining the training data from three different locations, the classified images look more like what is expected from the RGB images and have fewer errors and less discontinuities.

3.2.3 Global image classification

This section discusses the results obtained with the global image classification model. 7500 training data points collected from 75 locations around the globe are used to train the classifier. It is expected that the use of more training data points will provide more robust and reliable results. The results of the global image classification model are discussed on the basis of the different coastal environments described in section 2.2.3 (plain beaches, coastal cliffs, barrier islands and mangrove deltas) to provide insight into the performance of the classifier per coastal environment. Do the coastal environment show the expected characteristics? Are the coastal environments recognisable when classified using the seven coastal classes? Three individual RGB images and their corresponding classified images are shown and discussed for each coastal environment. Afterwards, two examples of misclassification are shown, in order to gain more insight how a mud classifier can be improved. This section ends with an overview of (in)correctly classified features by the global image classification model.

Plain beaches Plain beaches are characterised by a clear strip of sandy sediment at the transition between land and water.

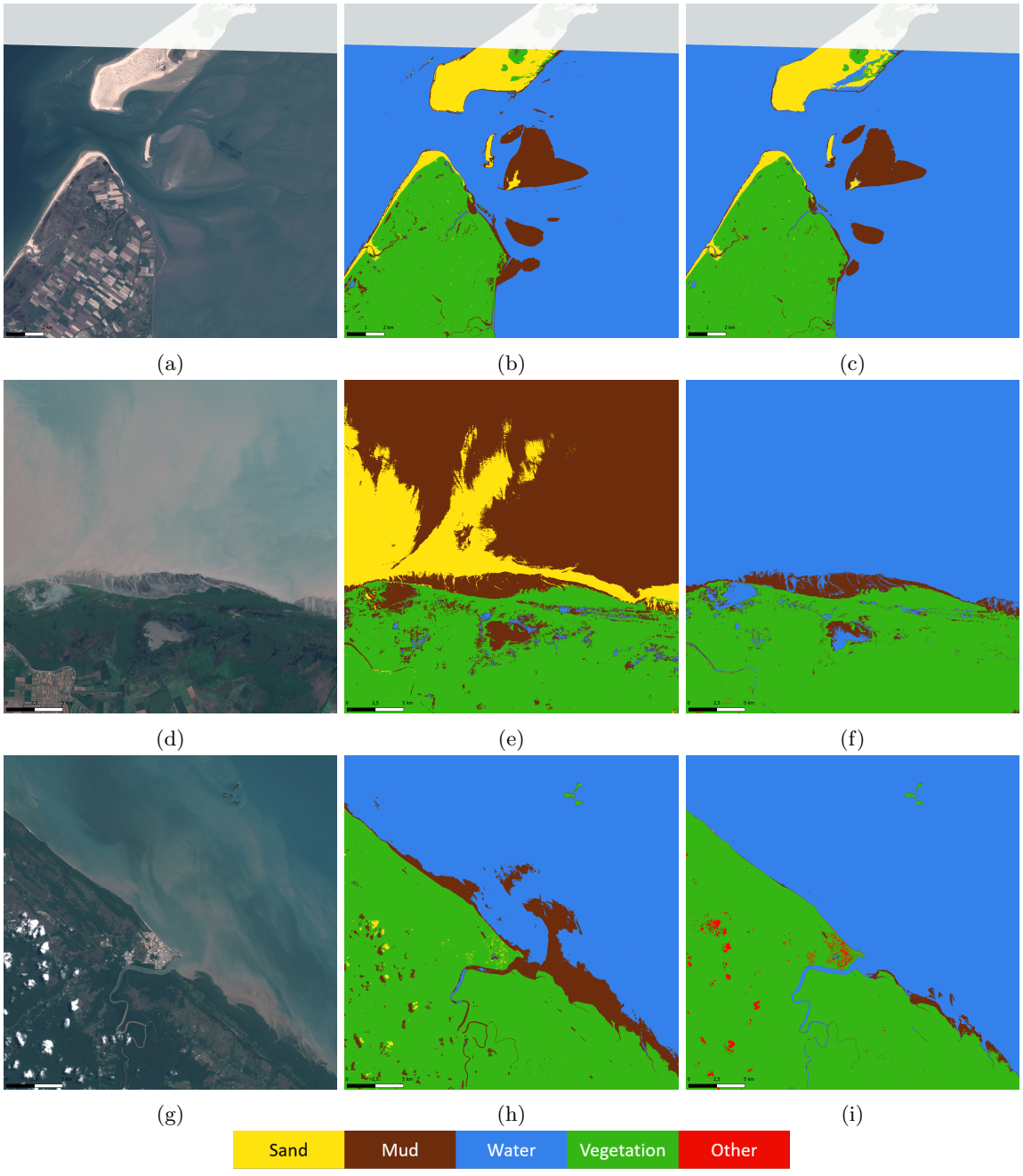


Figure 17: Individual S2 TOA RGB images and their corresponding classified images for three study areas: Dutch Wadden Sea (BOX_187_108) (first row), Suriname (BOX_117_030) (second row), and French Guyana (BOX_118_054) (third row). First column shows individual S2 TOA RGB images, second column shows images classified by the local image classifier trained with data collected from the Dutch Wadden Sea, third column shows images classified by the combined image classifier trained with data collected from the three areas. Five coastal classes can be distinguished: yellow for sand, brown for mud, blue for water, green for vegetation and red for other.

Results Figure 18 shows three individual RGB images of distinct plain beach environments in Brazil, Mexico and Poland. The first location, Brazil (Figure 18a), shows an easily recognisable sandy beach. The land consists mainly of vegetation and a few buildings are present. In the ocean, white wash (breaking waves) is present as a straight line some metres away from the coastline. Behind the white wash there is just water. In the classified image of the Brazilian coast (Figure 18d), the beach is classified as sand (yellow), the land is classified as both types of healthy and dry vegetation (green and dark green) and also contains the class other (red) where buildings and roads are present. The water close to the coast is classified as turbid water (dark blue) and as water (blue) further away from the coast. There are also some red dots (other) present at the location of the breaking waves. Figure 18b shows part of the Mexican coastline. A clear beach is present, the landward side consists mainly of vegetation and a (dry) river. Waves are breaking very close to the coastline. The water looks clear and blue. The classified image (Figure 18e) labels the beach as sand (yellow), the mainland as both vegetation types ((dark) green), the river as sand (this could be a dry river bed). The water is primarily labelled as water (blue), but also contains some red dots (other). Close to the coastline, some strips of the class other (red) and turbid water (dark blue) are present, representing breaking waves. The last example shows part of the coastline of Poland (Figure 18c). The same pattern as in Figures 18a and 18b can be recognised: clear blue water, a strip of a plain sandy beach, behind the beach some vegetation and on the mainland some villages are present. The classified image (Figure 18f) shows the characteristics correctly. There are no or only minor deviations.

Findings For the plain beaches, the main characteristic in the RGB satellite image is the strip of (bright) sand. In the three examples given, the sandy beaches are correctly classified by the global image classifier. A feature that does not have its own coastal class is the white wash caused by the breaking waves. This feature is however recognised by the model and often classified as both classes turbid water and other. Overall, the model produces classified images as expected.

Coastal cliffs In this coastal environment, coastal cliffs end very steeply in the water. No or little sediment is found along the mainly rocky coastline. There is often a high-energy wave environment where the waves crash against the cliffs.

Results Figure 19 shows three examples of a coastal cliff environment in New Zealand, Australia and the United States of America. Several predicted coastal classes can be found when looking at the transition from land to water in each image. In Figures 19a and 19d it can be seen that in the top part of both images the land (green) turns into water (blue) without any other classes being involved. No breaking waves are present in this part of the image. In the bottom part of both images breaking waves are observed, these are classified as other (red) and turbid water (dark blue). Also on the landward side of this part of the coast some interesting things are happening, namely some shaded parts of the cliffs are classified as other (red). In Figures 19b and 19e coastal cliffs are found, as well as (protected) sandy beaches between the cliffs. The transition from land to water near the cliffs shows the coastal classes vegetation, other, turbid water and water. A lot of white wash is present in the RGB image along the entire coastline, due to the high-energy wave environment. The transition from land to water near the (protected) beaches contains the coastal classes sand, turbid water, other and water. The classes turbid water and other in these sections can be explained by the presence of breaking waves as well, which have no label of their own. In Figures 19c and 19f, parts of the cliffs in the image are classified as mud and other. This may be due to the shaded area of the cliff itself. In addition, there is a lot of white wash, which is again classified as other and turbid water.

Findings The examples of the coastal cliffs show the expected patterns. There are some minor misclassifications, especially where land or water are in the shaded area of the cliffs. Also in these examples, white wash is usually classified as other or turbid water, just like the examples of the plain beaches. In general, coastal cliffs can be recognised in the classified images as areas where no sediment (mud or sand) is found and where in some cases breaking waves are present which are indicated as turbid water and other.

Barrier islands Barrier island systems can be recognised by elongated islands off the mainland. On the ocean side of the islands, (sandy) beaches and dunes may be present (similar to the plain beaches). Between

the islands and the mainland, mudflats and/or vegetated salt marshes generally develop, as a result of the protection offered by the islands against waves, currents and wind from the open ocean. It is expected that both sandy areas and muddy patches will be represented on the classified images.

Results In Figure 20 three examples are shown of barrier island systems located in Mozambique, United States of America and the Netherlands. In the individual RGB images the barrier islands are clearly visible and are separated by rivers, deltas or estuaries from the mainland. On the seaward side of the islands (sandy) beaches, breaking waves and (turbid) water can be seen. The landward side of the islands shows vegetation, mud(flats) and (turbid) water. The different features of the seaward side and the landward side of the islands are recognised by the classifier, both the sand (yellow) and the mud (brown) class are present. In Figures 20a and 20d, however, it seems like that part of the water with a high concentration of sediment in suspension is incorrectly labelled as mud. In Figures 20b and 20e, the area between the islands and the mainland consists largely of vegetation (salt marshes), in some parts mud patches are found. In Figures 20c and 20f, the mudflats between the island and the mainland (which is not shown in the image, but is present in the south of the island) are labelled correctly.

Findings According to these examples of barrier islands, the global image classifier is able to distinguish between muddy and sandy parts. The general multispectral observations from the satellite images are correctly classified as expected based on the knowledge about the behaviour of barrier islands. A point of attention is the confusion between mud and turbid water by the classifier.

Mangrove deltas From above mangrove deltas can be recognised by meandering river patterns, mud banks and vegetation. The main coastal classes that are expected to be encountered are mud, vegetation and (turbid) water.

Results Figure 21 shows three examples of mangrove deltas in Venezuela, Brazil and Guinea. Each of the individual RGB images shows islands or sections of land covered with vegetation and separated by waterways and meandering rivers. Mud banks are often present at the transition from land to water. The water of the rivers and the oceans to which they flow appears brownish in the RGB images in Figures 21a and 21b. The water in Figure 21c is much more blue (clear), except for the areas very close to the coast. The classified images behave as expected. Figure 21d shows the mud banks at the expected locations. In the southern end of the rightmost island, a small sandy area can be seen, which also appears bright in the RGB image. The water in this classified image is classified as clear water, whereas in the RGB image it looks brownish (turbid water). Figure 21e shows well the distinction between the mud banks and the turbid water surrounding them. In figure 21f the meandering rivers are classified as mud (brown), while they are expected to be (turbid) water and even appear as clear water in the RGB image. On the southwestern side (the open ocean side) there is a narrow sandy beach, which can be recognised in both the RGB and the classified image of Guinea.

Findings Qualitatively, the mangrove deltas are very well represented in the classified images. The main features of meandering river patterns, mud banks and vegetation are recognised by the classifier, although there is confusion between mud and turbid water (as is the case with the examples of the barrier islands).

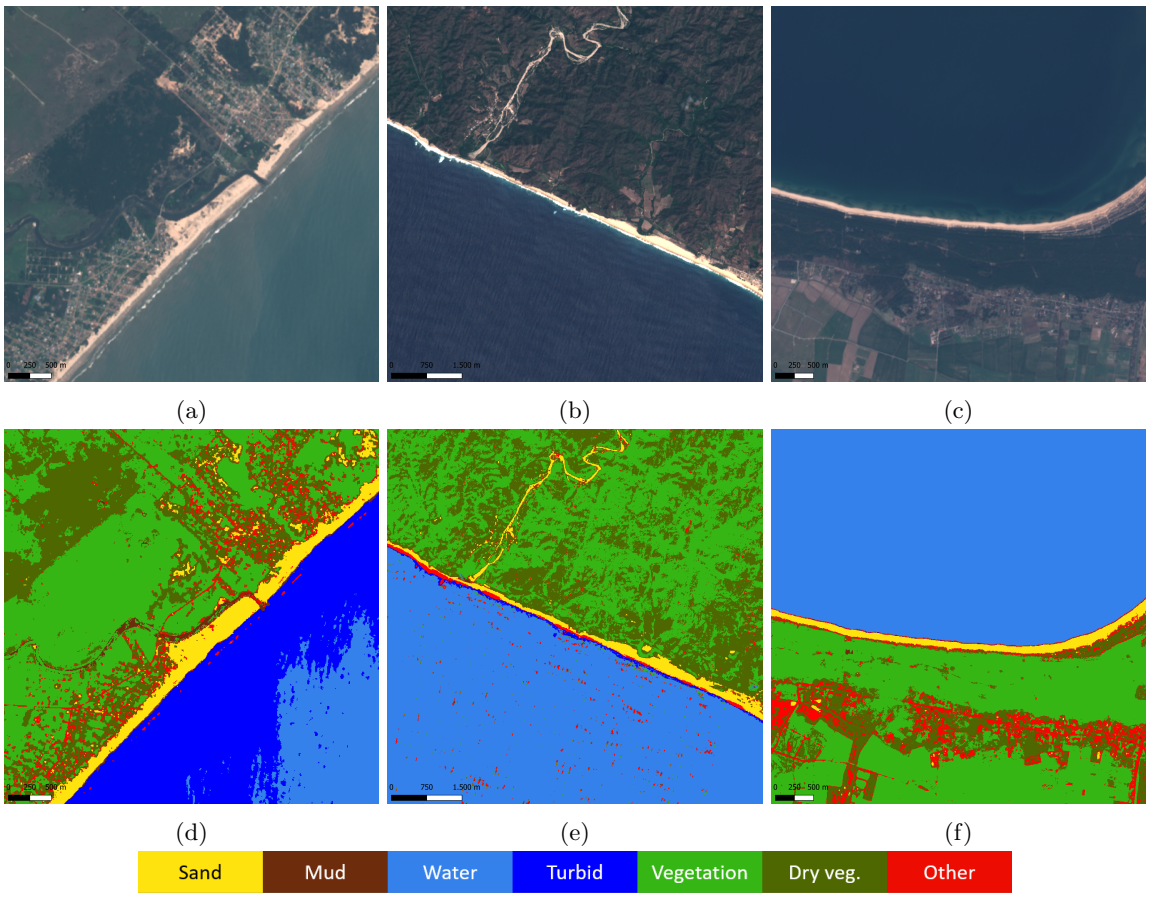


Figure 18: Three examples of plain beaches; first row: individual S2 TOA RGB images; second row: classified images with global image classification model. (a) and (d) Brazil (BOX_074_000); (b) and (e) Mexico (BOX_137_001); (c) and (f) Poland (BOX_188_001).

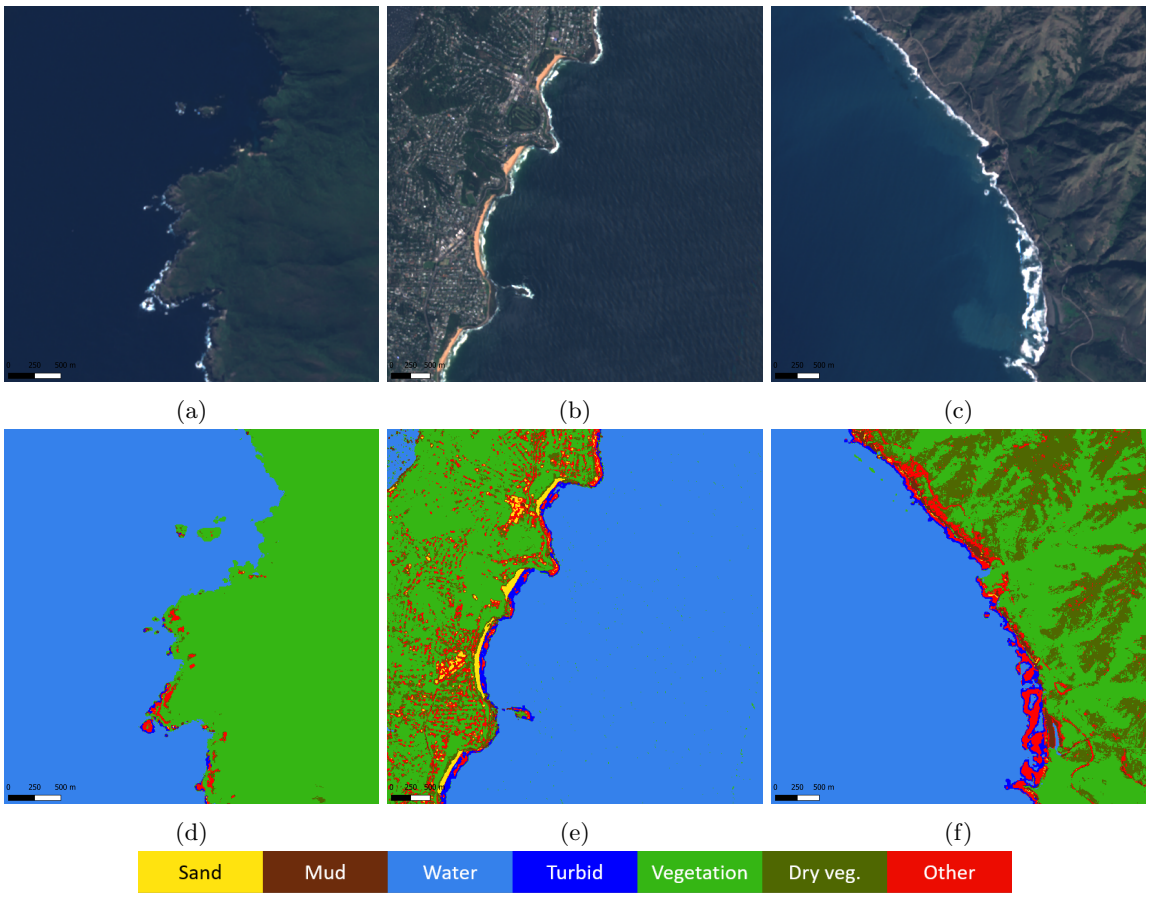


Figure 19: Three examples of coastal cliffs; first row: individual S2 TOA RGB images; second row: classified images with global image classification model. (a) and (d) New Zealand (BOX_064_000); (b) and (e) Australia (BOX_085_001); (c) and (f) United States of America (BOX_158_000).

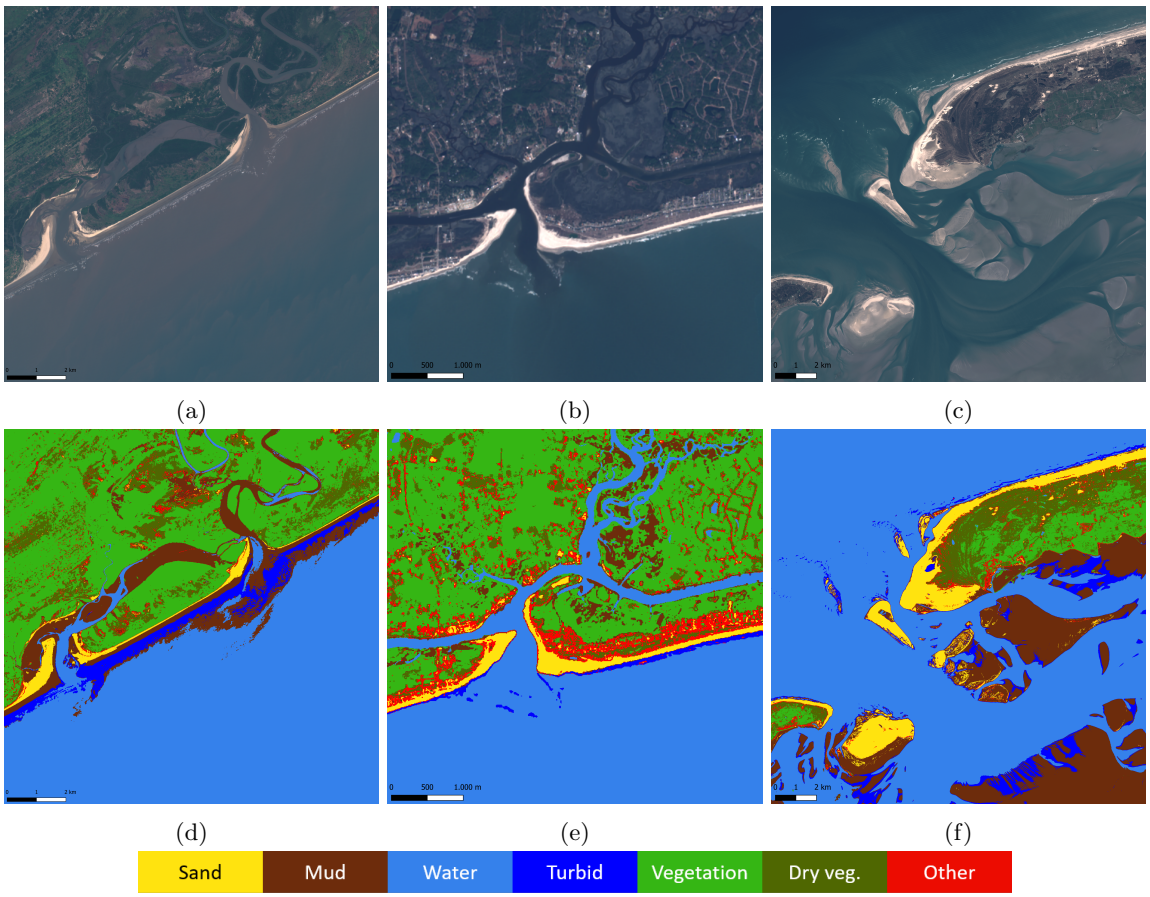


Figure 20: Three examples of barrier islands; first row: individual S2 TOA RGB images; second row: classified images with global image classification model. (a) and (d) Mozambique (BOX_101_002); (b) and (e) United States of America (BOX_160_001); (c) and (f) Netherlands (BOX_187_112).

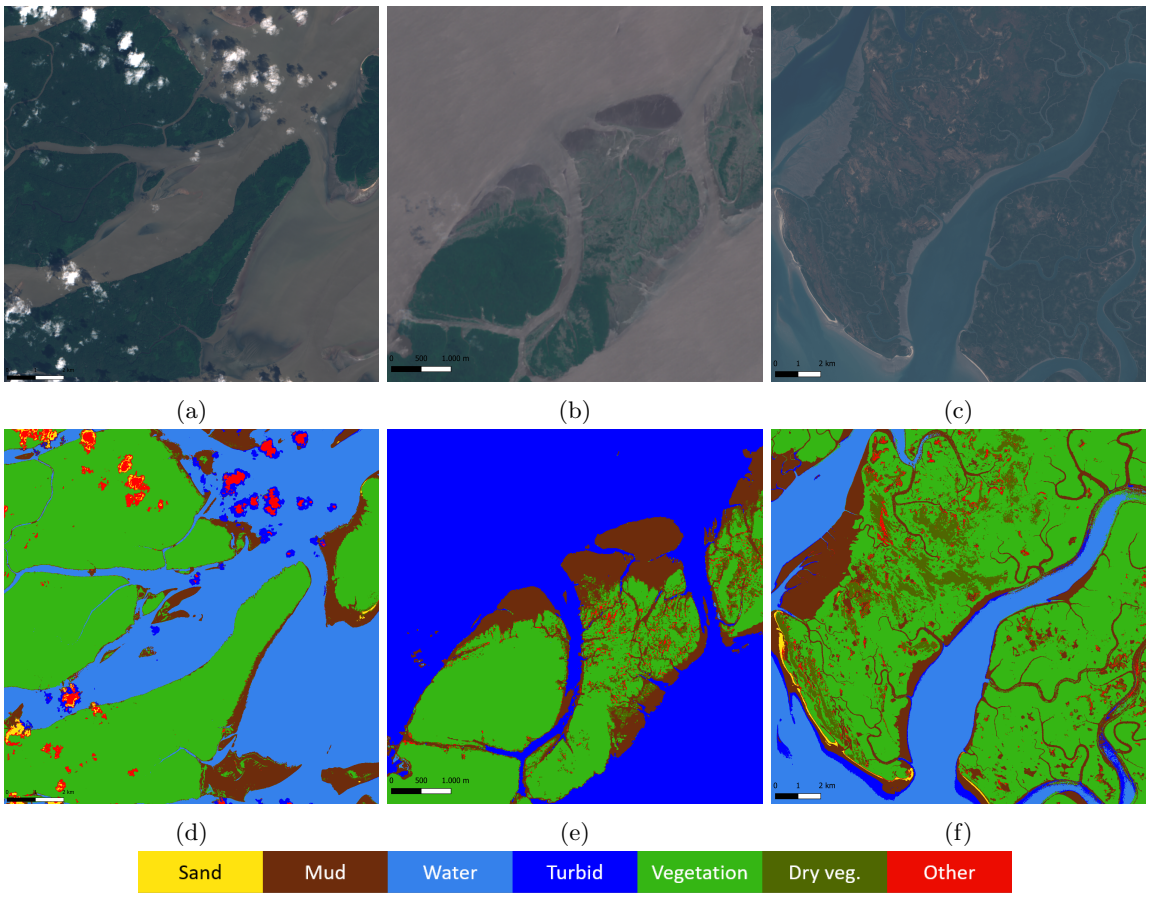


Figure 21: Three examples of mangrove deltas; first row: individual S2 TOA RGB images; second row: classified images with global image classification model. (a) and (d) Venezuela (BOX_117_004); (b) and (e) Brazil (BOX_118_001); (c) and (f) Guinea (BOX_120_100).

Misclassifications Two examples in Figure 22 show that dark materials or dark sediments are confused with mud by the global image classification model. In Figures 22a and 22c a part of the island of Tenerife (Spain's Canary Islands) is shown. Tenerife is a volcanic island that is formed by a continuous series of eruptions over millions of years. In the RGB image, a small volcanic crater can be seen (in the middle of the image), surrounded by dark coloured volcanic material that reaches to the coastline. This dark material is recognised by the global image classifier as mud, which is incorrect. The second example shows a coastline of the northern island of New Zealand shown in Figures 22b and 22d. A beach is visible at the outflow of the river and a very narrow beach in the upper part of the image next to the vegetated land, that both consist of dark sand. Breaking waves can be observed along the entire coastline in the RGB image from which it can be inferred that this is not a muddy coast. However, the global image classification model recognises part of the beach as mud.

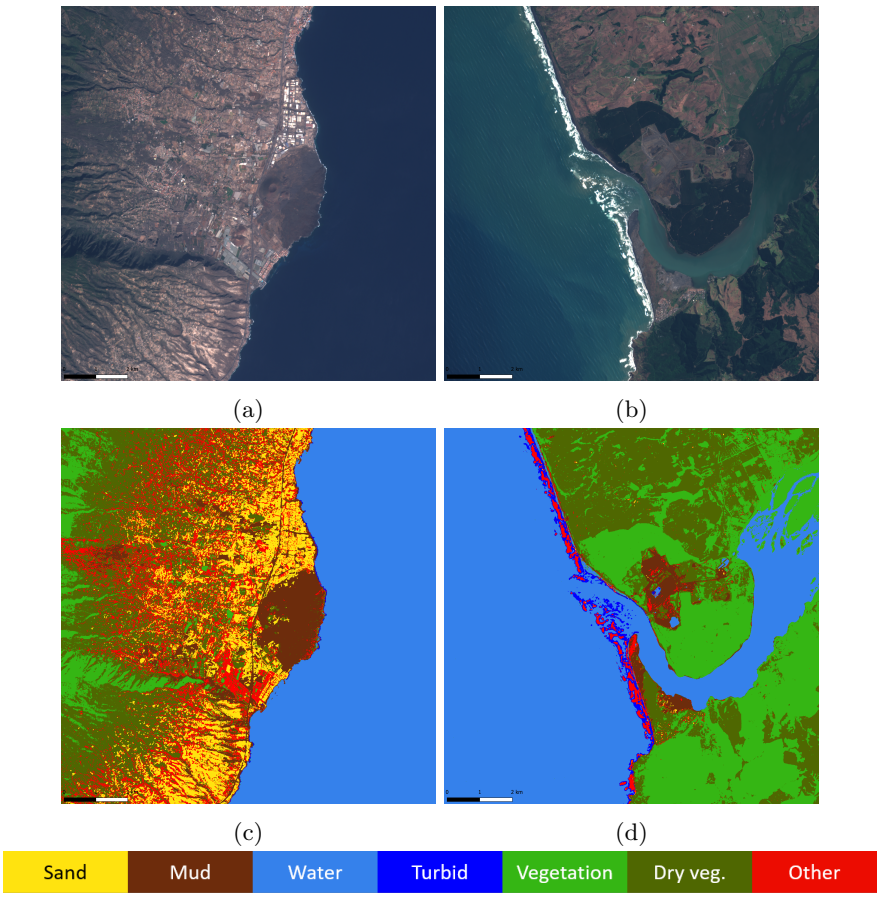


Figure 22: Two examples of misclassifications; first row: individual S2 TOA RGB images; second row: classified images with global image classification model. (a) and (c) Spain (BOX_142_084); (b) and (d) New Zealand (BOX_064_268).

Summary Table 2 summarises the findings on the performance of the global classification model per coastal environment, when analysing the examples of the classified images given above.

3.3 Model performance

A model performance analysis is carried out on the 7500 manually selected training points, where for each point a coastal class is assigned by visual inspection of individual S2 TOA RGB images. The model performance analysis tests the robustness, repeatability and durability of the global multispectral image classifier. Multiple values can be used to assess the performance. In the next paragraphs the overall accuracy of the model is assessed, as well as the f1-score for each coastal class separately.

Coastal environment	Correctly classified features	Incorrectly classified features
Plain beaches	• Sandy beaches are very well represented	• Breaking waves (white wash) do not have their own class and are therefore (incorrectly) classified as turbid water and other
Coastal cliffs	• No sediment (sand or mud) is found along the rocky cliffs	• Shaded areas of the cliffs are difficult to classify correctly by the classifier and are mainly classified as mud and other • Breaking waves (white wash) do not have their own class and are therefore (incorrectly) classified as turbid water and other
Barrier islands	• Classifier is able to distinguish between sandy and muddy parts	• Confusion between mud and turbid water
Mangrove deltas	• Mudbanks are labelled correctly by the classifier • Classifier is able to distinguish between sandy and muddy parts	• Confusion between mud and turbid water
Misclassifications		• Volcanic material and dark sediment is recognised as mud

Table 2: Overview of (in)correctly classified features by the global image classification model per coastal environment. The classifier is able to distinguish between muddy and sandy features (in one image), confusion still exists in shaded areas, turbid water and dark coloured materials.

3.3.1 Overall model accuracy

The assessment is performed on the global image classification model and relates the overall accuracy of the model against the number of training points. It is expected that the accuracy of the classification model will increase if the number of points used for training is increased. In general, accuracy increases rapidly when there are few training points and more are added, but at some point adding more training points will no longer significantly increase accuracy (and may even decrease it, due to overfitting). This can be demonstrated with the performance plot.

The accuracy of the model is obtained by inspecting the confusion matrix of the classifier, an example of a confusion matrix is shown in Figure 23. For this confusion matrix, 6750 training points were used to train the global image classifier, which is tested on 750 validation points. The horizontal axis shows the predicted label, while the vertical axis shows the true label. On the diagonal are the true positives (TP) and the true negatives (TN); i.e. when the predicted label of a particular point is equal to the true label of that particular point. To the right of the diagonal are the false positives (FP) and below the diagonal are the false negatives (FN); i.e. when the predicted label is not equal to the true label. From the values of the confusion matrix the overall accuracy of the model can be obtained using Equation 4; the number of correctly classified validation points (the sum of the diagonal) divided by the total number of validation points.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

From the confusion matrix it can also be deduced which coastal classes are often confused with each other. Figure 23 shows that the coastal class ‘other’ is frequently confused with the class ‘vegetation’ and ‘dry vegetation’. The class ‘dry vegetation’ is regularly confused with the class ‘vegetation’. Furthermore, all classes show little or no confusion with each other. In addition, the confusion matrix shows that the number of training points is not evenly distributed among the different coastal classes.

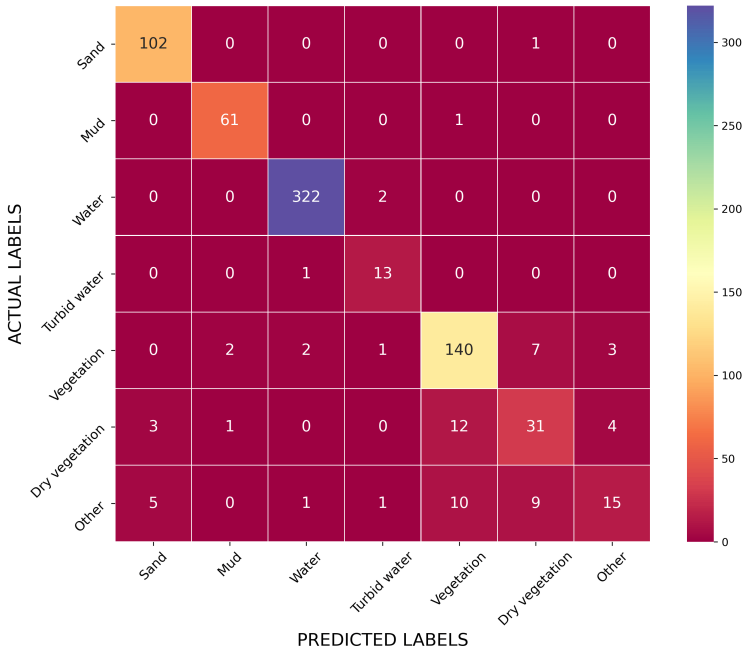


Figure 23: Confusion matrix of 750 validation points classified with the global image classification model trained with 6750 training points. The training dataset of the global image classification model is not equally distributed among the coastal classes, as a result the numbers on the diagonal differ significantly.

The performance analysis is carried out on the 7500 manually selected training points, that are labelled by visual inspection of individual S2 TOA RGB images, from which 750 points are randomly selected to be used as validation points. The remaining points are used for training the classification model and are divided in smaller groups. Firstly, a model is trained with one group of training points. This model classifies the 750 validation points and the overall accuracy of the model is calculated, using Equation 4. Next, a new model is trained with two groups of training points, the same validation points are classified with the new model and the overall accuracy is calculated. This is done until all training points are used. This process is repeated 100 times with different randomly selected groups (cross validation) to average out errors and biases. Figure 24 shows the performance plot for the global image classification model, with the number of training points on the x-axis and the overall model accuracy on the y-axis. The solid line represents the mean accuracy, the shaded area represents the $90th$ percentile of the accuracy. It can be seen that an increase in the number of training points from zero to approximately 1000 shows a significant increase in accuracy, afterwards the accuracy stabilises. The red cross marks when the accuracy exceeds 90 percent. This is when the model has been trained with approximately 2650 training points. From the analysis of this performance plot it can be concluded that the use of more than 2650 training points does not significantly increase the accuracy of the model.

An overall model accuracy of 90% means that 90 percent of the validation points are correctly labelled when the global multispectral image classification model is applied.

3.3.2 Accuracy per coastal class

A value to assess the robustness, repeatability and durability of the global image classification model for each separate coastal class is the f1-sore. The f1-score is a trade-off between the precision (how many points are recognised by the classifier) and the recall of a model (how many points are wrongly assessed by the classifier), see Equations 5, 6 and 7.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

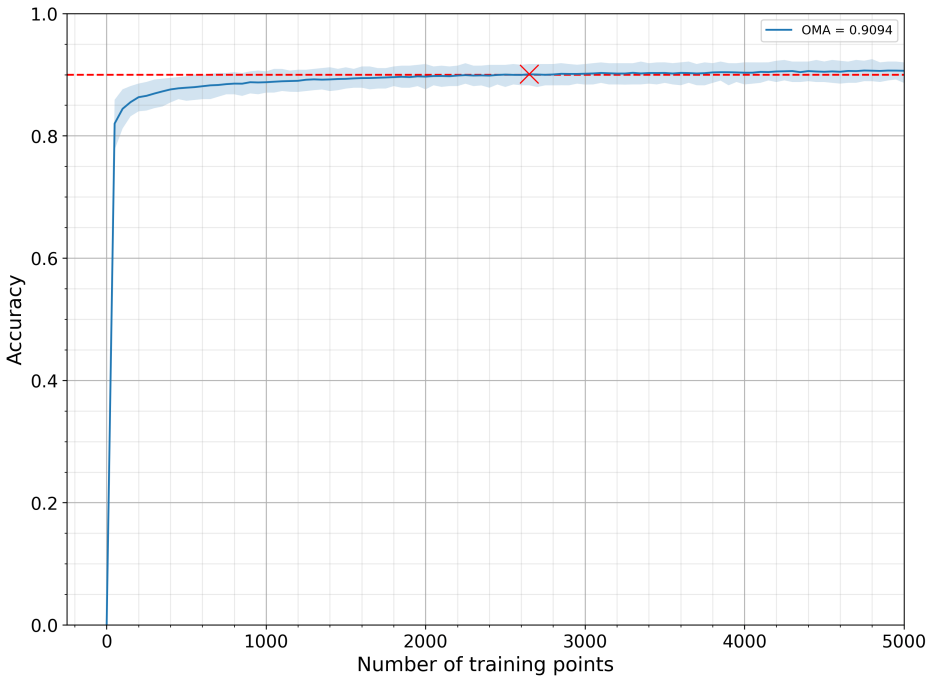


Figure 24: Overall model accuracy (OMA) of the global multispectral image classification model when classifying seven coastal classes. The number of training points used to train the model is shown on the x-axis and the corresponding accuracy (Equation 4) on the y-axis. The solid line represents the mean accuracy, the shaded area represents the $90th$ percentile of the accuracy. The red dashed line shows the 90% accuracy level. The red cross shows at which number of training points the accuracy of the model exceeds 90%.

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{7}$$

This analysis is performed in the same way as the overall model accuracy assessment described above; it uses the confusion matrix to calculate the f1-score and it is executed 100 times. Figure 25 plots the f1-score for each of the seven coastal classes against the number of training points used for that class. It can be seen for each coastal class that an increase in the number of training points leads to an increase in the f1-score and that after using a certain number of training points the f1-score stabilises. For the class water only 100 training points are needed to achieve a sufficient and stable f1-score, for other classes more training points are needed to stabilise the accuracy score. The graph shows that the f1-score for the classes sand, mud, water and vegetation (nearly) stabilises for the given amount of training points. The f1-score for the other classes (turbid water, dry vegetation and other) still increases and converges significantly. This shows that for these classes, adding more training points can improve the performance. What also stands out in the graph is that the $90th$ percentiles for the classes turbid water, dry vegetation and other are relatively large compared to the $90th$ percentiles for the other four classes; in other words, the variability is larger. This may be due to the less 'strict' definition of these classes. For example, the class other includes roads, buildings, clouds, snow and undefined features. Nevertheless, these are not the classes of greater importance.

The class which is of importance in this study, mud, still shows an increase in performance. With 580 muddy training points, the f1-score reaches about 92%, which means that 92% of the pixels with the mud-label are actually recognised as mud by the global multispectral image classification model.

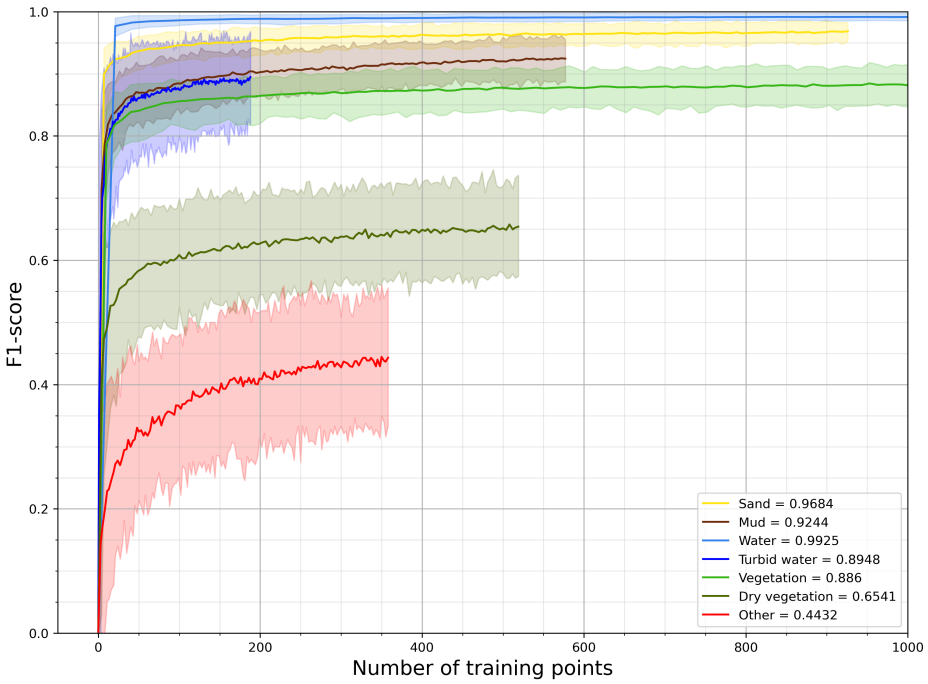


Figure 25: Accuracy per coastal class when using the global multispectral image classification model. The number of training points is shown on the x-axis and the f1-score on the y-axis for classes sand, mud, water, turbid water, vegetation, dry vegetation and other. The solid lines represent the mean f1-scores for each class, the shaded areas represent the $90th$ percentiles of the f1-score for each class.

3.4 Conclusions

The local image classification model shows that the coastal feature 'mud' can be distinguished on pixel level by a locally trained classifier using individual multispectral satellite images acquired by S2. By applying the local image classification model to other places on Earth, it is shown that the spectral reflectance properties of the pixels of the defined coastal classes differ at various locations and under various conditions because they were not correcly recognised by the classifier, and that the combination of more training points at more locations improves the performance of the classification model. The global image classification model is therefore trained with 7500 training points spread over 75 locations on the globe. The overall model accuracy of the global model exceeds 90 percent, which means that 90 percent of the validation points are classified correctly. The individual accuracy for the coastal classes of greater importance also shows high and (nearly) stabilised values. The muddy class reaches 92 percent accuracy, which means that 92 percent of the muddy training points are actually classified as mud by the global image classifier. However, it is clear from the two examples of misclassification that the global multispectral classification model does not (always) correctly recognise muddy pixels and confuses mud with dark-coloured material.

4 Exploring coastal datasets



The multispectral image classification model classifies individual pixels of satellite images into seven coastal classes, among which the class 'mud' represents a classification of the soil type. However, it is desirable to classify coastal environments in order to monitor the morphodynamics of muddy coasts and understand their behaviour, so that these vulnerable environments can be protected and developed in a sustainable manner. The multispectral image classification model can be applied to the transects of the global coastline and classify the pixels along the transects. The overall accuracy of the model is 90% and the accuracy of the classification of muddy pixels is 92%. However, the examples in section 3.2.3 still showed some uncertainties and misclassifications of muddy pixels. There is confusion between mud and turbid water, some shaded areas of coastal cliffs are classified as mud, and often dark coloured-material (such as dark sand or volcanic materials) are recognised as mud. When observing the RGB images, using common knowledge and contextual information, one can recognise that these coasts are not muddy coasts. Is it possible to add coastal information that will help a classifier to recognise these kind of errors? The next phase of the proposed methodology to obtain a reliable global mud classification model is to analyse the physical geographical characteristics of a muddy coastal system by using globally available coastal datasets, in order to increase the probability of correctly classifying a muddy coast at global scale.

In this section, some interesting globally available coastal datasets related to different coastal environments will be investigated. The knowledge of physical geographical characteristics of a muddy coastal system, as described in section 2.2.2, will be compared to the available datasets, in order to eliminate certain areas from being a muddy coast (such as dark-sanded beaches or volcanic areas) and to increase the probability of correctly classifying a muddy coast. This section will implement the globally available coastal datasets, described in section 2.4, and will create transect profiles for different coastal environments in section 4.1, similar to the conceptual transect profile sketched in Figure 11. The distribution of the coastal datasets among the different coastal environments is assessed in section 4.2. Section 4.3 shows an example of the added value of the coastal datasets. The conclusions obtained are described in section 4.4.

4.1 Coastal transect profiles



In section 2.4, seven globally available coastal datasets are described, from which ten coastal variables are defined, that could potentially help to detect muddy coasts, since they could capture some of the physical geographic characteristics of muddy coasts. The results of the exploration of coastal datasets will be discussed using the four coastal environments described in section 2.2.3, which were also used to discuss the results of the global multispectral image classification model in section 3.2.3. For each coastal environment, a transect profile is shown including the multispectral image classification profile and the information obtained from the coastal datasets. This approach allows for an intuitive, visual and contextual inspection of the coastal datasets and gives a first idea of the added value of using different coastal datasets.

4.1.1 Plain beaches

Plain beaches are characterised by a distinct stretch of sandy sediment at the transition between land and water. The topography close to the coastline (the beach) is expected to follow a gentle slope, further inland the topography is not bound to a specific profile. Dunes may be present and can be recognised as a peak in the topographical profile close to the beach. Mangroves are not generally expected, but there is no exclusion of mangroves either. Tidal flats can occur at the transition between land and water; it is expected that for plain beaches the tidal flats are generally narrower than for muddy coasts. The Global Surface Water (GSW) frequency profile is expected to be at least 0% where the beach ends on the landward side and 100% in the open ocean. The width of the transition zone depends on many factors, such as the width of the beach, the wave action zone, the tidal range, etc.. For plain beaches, the width of the transition zone (width of GSW profile) is generally expected to be narrower than for muddy coasts. Plain beaches can occur at all latitudes, are not restricted to a specific temperature range and can be found under different tidal conditions.

Results The same region of the Brazilian coastline that was used in section 3.2.3 is analysed. Figures 26a and 26b show a clearly recognisable sandy plain beach. Along the coastline, transects are drawn with

a length of $1500m$ and a spacing of approximately $415m$. Information from the coastal datasets for the transect marked in red is shown in Figure 26c.

The height profile shows a peak close to the coastline, inland the height decreases slightly and then increases again; the peak close to the coastline could indicate a dune profile. The maximum height reaches approximately $8m$ within about $700m$ of the coastline. No mangroves are present along this transect. The tidal flats extend over a width of $100m$, the tidal flat is located where the beach turns into water. The transition between land and water according to the GSW profile coincides with the location where sand pixels are found. This transition zone is $100m$ wide (between $5\% - 95\%$ of the GSW profile). The transition zone observed by the GSW profile is positioned slightly landward with respect to the tidal flats observed by Murray et al. (2019). The transect is located at $-34°$ latitude (in the southern hemisphere), the temperature varies from $8°C - 27°C$ and the tidal range is $0.52m$.

4.1.2 Coastal cliffs

Coastal cliffs are characterised by a sharp transition from land to water. The topographical profile is (very) steep near the coastline and the maximum height is expected to be higher than for the other coastal types. The tidal range is generally not restricted, although it is expected to be (relatively) small compared to the height of the topographical profile. A sharp transition from land to water is expected for the GSW profile, because the height profile is almost vertical in the transition zone; the width of this transition is expected to be (very) small or zero, at least narrower than for the plain beaches. Tidal flats are not expected as no sediment is found near the coastline and the (underwater) topography is typically (very) steep. Due to the lack of sediments and the presence of a high wave energy environment, mangroves are not expected. Coastal cliffs can occur at all latitudes and are not restricted to a specific temperature range, though they are generally more common at high latitudes and lower temperatures.

Results The same region of the coastline of New Zealand that was used in section 3.2.3 is analysed. Figures 27a and 27b show a sharp transition between water and land. Along the coastline, transects are drawn with a length of $1500m$ and a spacing of approximately $340m$. Information from the coastal datasets for the transect marked in red is shown in Figure 27c.

The height increases rapidly within $300m$ of the coastline, where it reaches a maximum height of about $70m$. The tidal range for this region is small ($0.29m$) and is not visible in the plot due to its proportion to the topographic profile. The GSW profile shows a sudden transition from land to water where the transect crosses the coastline. However, the total width of the transition zone, as defined in section 2.4.4 and shown in the graph as orange shaded vertical bar(s) (i.e. between 5% to 95% of the GSW frequency profile), is much wider than expected and equals $170m$. Moreover, no tidal flats or mangroves were observed. The transect is located at $-47°$ latitude (in the southern hemisphere). The temperature varies from $1°C$ to $18°C$.

4.1.3 Barrier islands

Barrier islands are elongated islands off the mainland, where (sandy) beaches and dunes form on the ocean side of the island and mudflats and/or vegetated salt marshes generally develop between the islands and the mainland. For the sandy coasts on the ocean side of the island, the characteristics are very similar to those of the plain beaches. On the mainland side of the island, the characteristics are different. The slope of the topographical profile is usually smaller on this side of the island, due to the low energy environment provided by the protection of the estuary on which this side of the island is located. The tidal range may differ from that observed on the ocean side of the island due to the shape of and the inlets into the estuary. The transition zone from land to water is expected to be wider than for the plain beaches. The same holds for the tidal flats, because the available sediment has a greater chance of settling due to the low energy environment. Mangrove forests or salt marshes may develop, depending on the temperature range and climate of the area.

Results Figures 28a and 28b show part of Terschelling, an island in the Dutch Wadden Sea. This island was also used as an example in section 3.2.3; now the middle of the island is analysed. The north of the island faces the (open) North Sea, the south faces the protected Wadden Sea (and the mainland). The

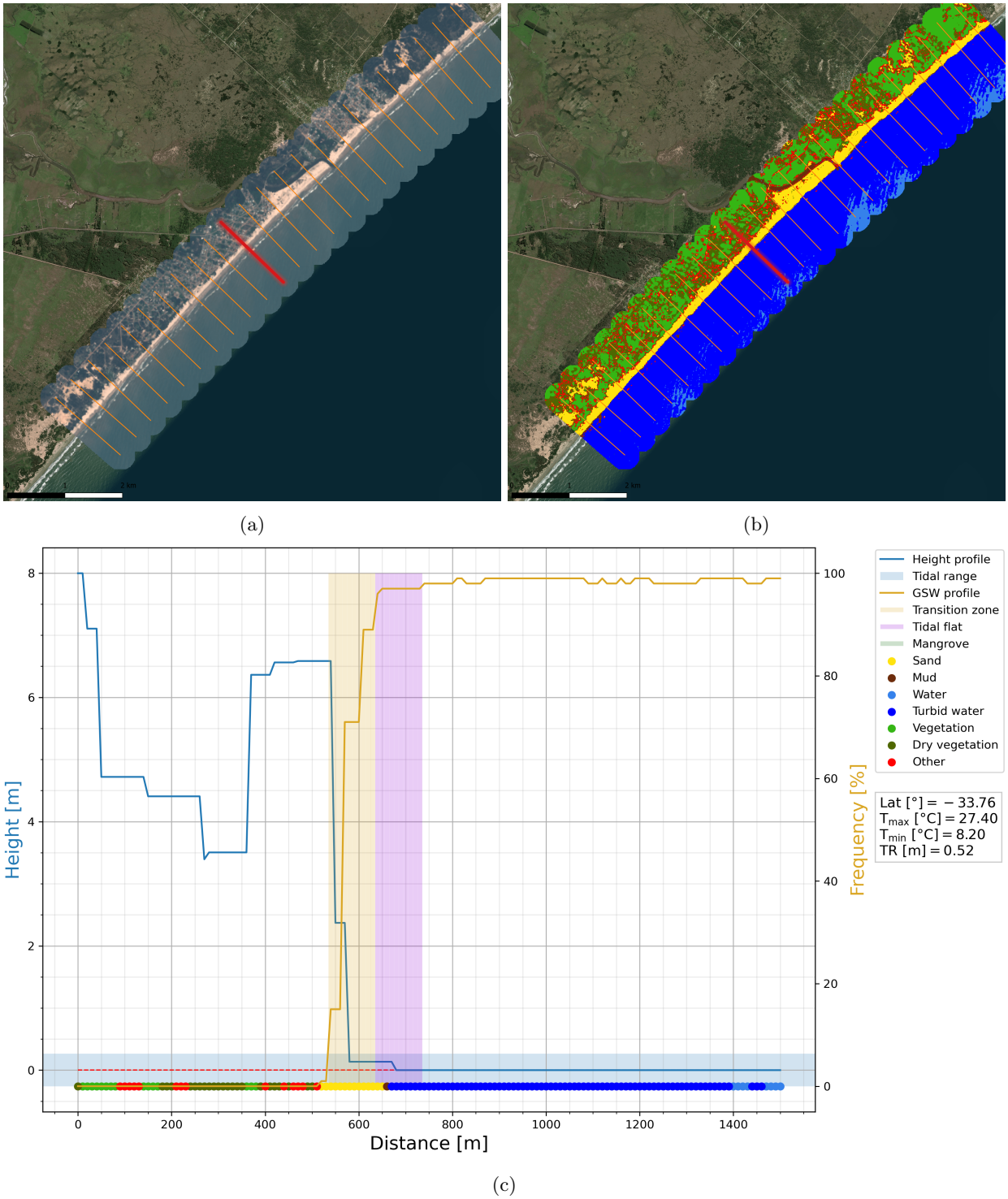


Figure 26: Example of a plain beach in Brazil. (a) individual S2 TOA RGB images around the transects; (b) classified images around the transects classified with global multispectral image classifier; (c) transect profile of the transect marked in red (BOX_074_000_10). The horizontal axis is the length [m] along the transect, where $0m$ is the landward end and $1500m$ the seaward end. The left y-axis represents the height [m] above mean sea level to which the height profile is related, as well as the tidal range, the right y-axis represents the frequency [%] with which water is present (GSW profile). The multispectral image classification is shown at the bottom of the graph with coloured dots referring to the coastal class of each pixel along the transect. The transition zone (orange), tidal flat occurrence (purple) and mangrove occurrence (green) are plotted as a vertical shaded bar, the tidal range (blue) is plotted as a horizontal shaded bar.

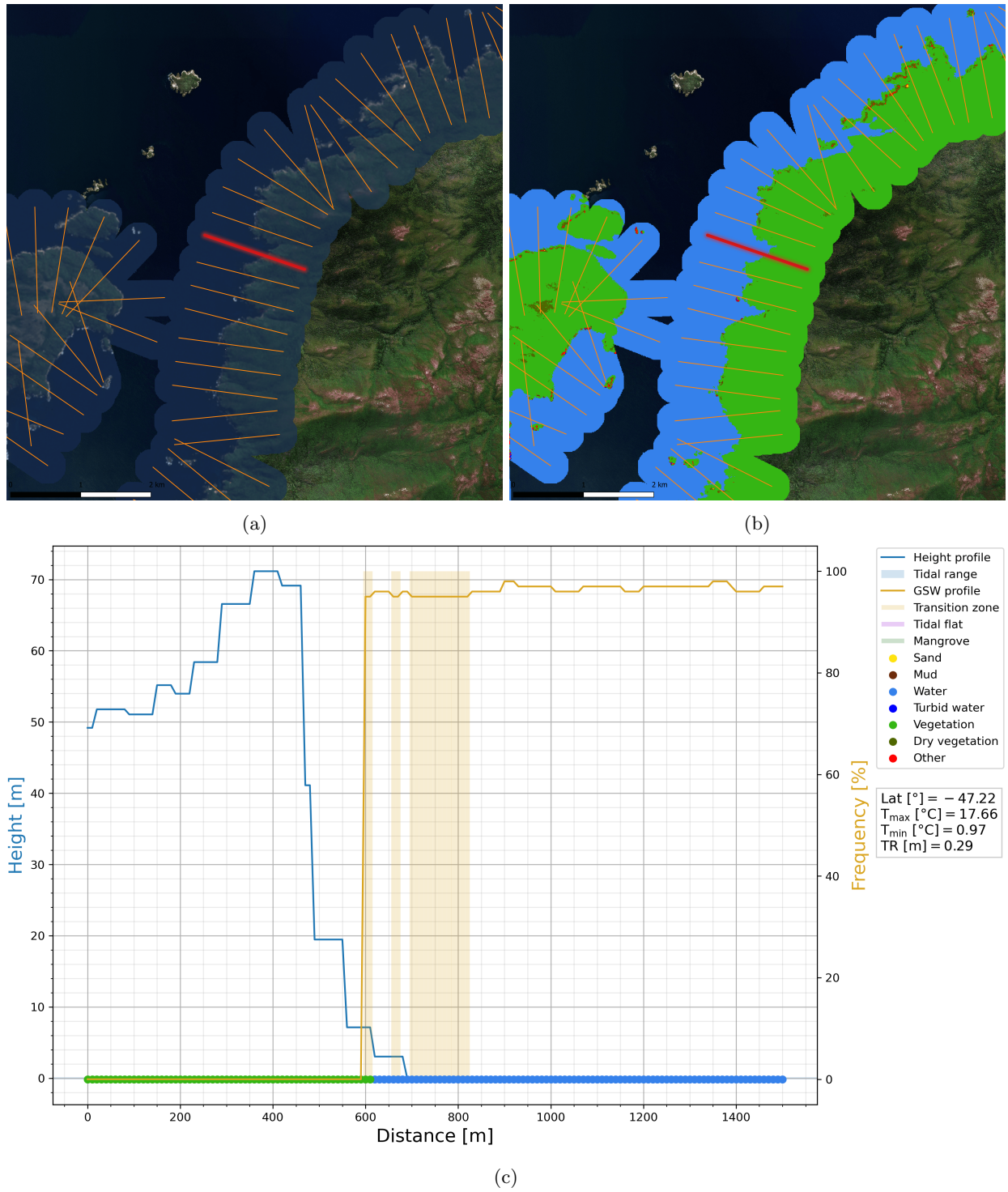


Figure 27: Example of a coastal cliff in New Zealand. (a) individual S2 TOA RGB images around the transects; (b) classified images around the transects classified with global multispectral image classifier; (c) transect profile of the transect marked in red (BOX_064_000_33). The horizontal axis is the length [m] along the transect, where $0m$ is the landward end and $1500m$ the seaward end. The left y-axis represents the height [m] above mean sea level to which the height profile is related, as well as the tidal range, the right y-axis represents the frequency [%] with which water is present (GSW profile). The multispectral image classification is shown at the bottom of the graph with coloured dots referring to the coastal class of each pixel along the transect. The transition zone (orange), tidal flat occurrence (purple) and mangrove occurrence (green) are plotted as a vertical shaded bar, the tidal range (blue) is plotted as a horizontal shaded bar.

difference between the northern and southern coastline can already be observed in these two images; sandy beaches in the north, mudflats in the south. Along the coastline, transects are drawn with a length of $1500m$ and a spacing of approximately $300m$. Two transects are analysed that cross the two different coastlines. The information from the coastal datasets for the transects marked in red is shown in Figure 28c (the sandy coastline) and Figure 28d (the muddy coastline).

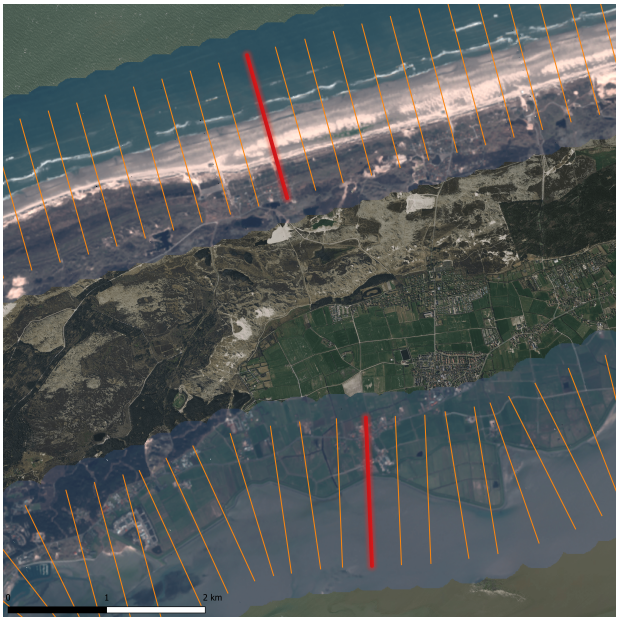
The topographic profile of the sandy coastline reaches a maximum height of $10m$ within $400m$ from the coastline at the points where the sandy beach becomes vegetation. The topographic profile of the muddy coastline only reaches a maximum height of just under $2m$ within $200m$ from the coastline. The tidal range for both transects is approximately $1m$, with the tidal range for muddy transect being slightly greater. The GSW profile for the sandy transect is stable where water pixels are classified and gradually decreases to zero from the point where the first sandy pixel is classified. The width of the transition zone is $130m$. The GSW profile for the muddy transect behaves differently. For the entire part of the transect where muddy pixels are classified, the GSW frequency is lower than 95%, which means that this part is never completely wet (water), but also never really dry (land). This confirms the behaviour of a mudflat. At the intersection with the coastline, the GSW profile shows a sudden jump to zero, this could indicate the dyke present here, as can be observed in Figure 28a. The total width of the transition zone is $870m$, approximately seven times wider than the sandy transect. Both transects show the presence of a tidal flat; for the sandy transect, the tidal flat is lies slightly more seaward than the GSW transition zone and has a width of $70m$. The tidal flat coincides with the location where both sandy and turbid water pixels are classified. The tidal flat of the muddy transect coincides with the GSW transition zone observed in the transect and is $810m$ wide. No mangroves were found in either transect as the transects are not located in a (sub)tropical climate. The transects are located at 53° latitude (in the northern hemisphere). The temperature varies from $0°C - 20°C$.

4.1.4 Mangrove deltas

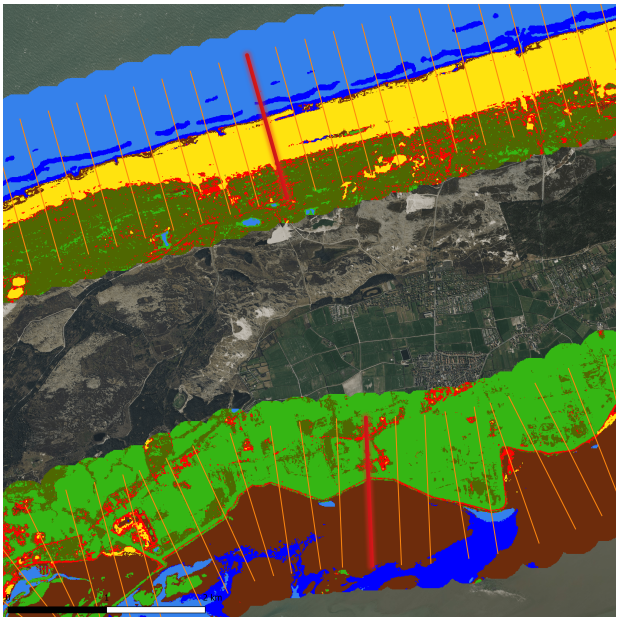
Mangrove deltas are low-lying deltas with meandering river patterns, mud banks and vegetation (mangrove forests). The topographical profile is expected to be low and with a gentle slope. A mangrove delta environment is not restricted to a certain tidal range, although tidal ranges are generally larger for mud dominated areas than for other coastal environments. The transition zone from land to water is expected to be wide, due to the abundant supply of sediment from the rivers that can settle on the banks; the same applies for the tidal flats. Obviously, mangrove forests are expected. Mangroves provide a positive feedback system: the mangroves grow because of the calm energy environment and the abundance of sediment, which in turn allows more sediment to settle because of the protection against waves and currents provided by the mangrove forests. Mangroves grow at low latitudes, in (sub)tropical climates, where the minimum temperature is not much colder than $18°C$, see Figure 7.

Results The same region of the coastline of Guinea that was shown in section 3.2.3 is analysed. Figures 29a and 29b show islands or sections of land that are covered with vegetation and separated by waterways and meandering rivers. Mud banks are visible at the transition from land to water. Along the coastline, transects are drawn with a length of $1500m$ and a spacing of approximately $490m$. Information from the coastal datasets for the transect marked in red is shown in Figure 29c.

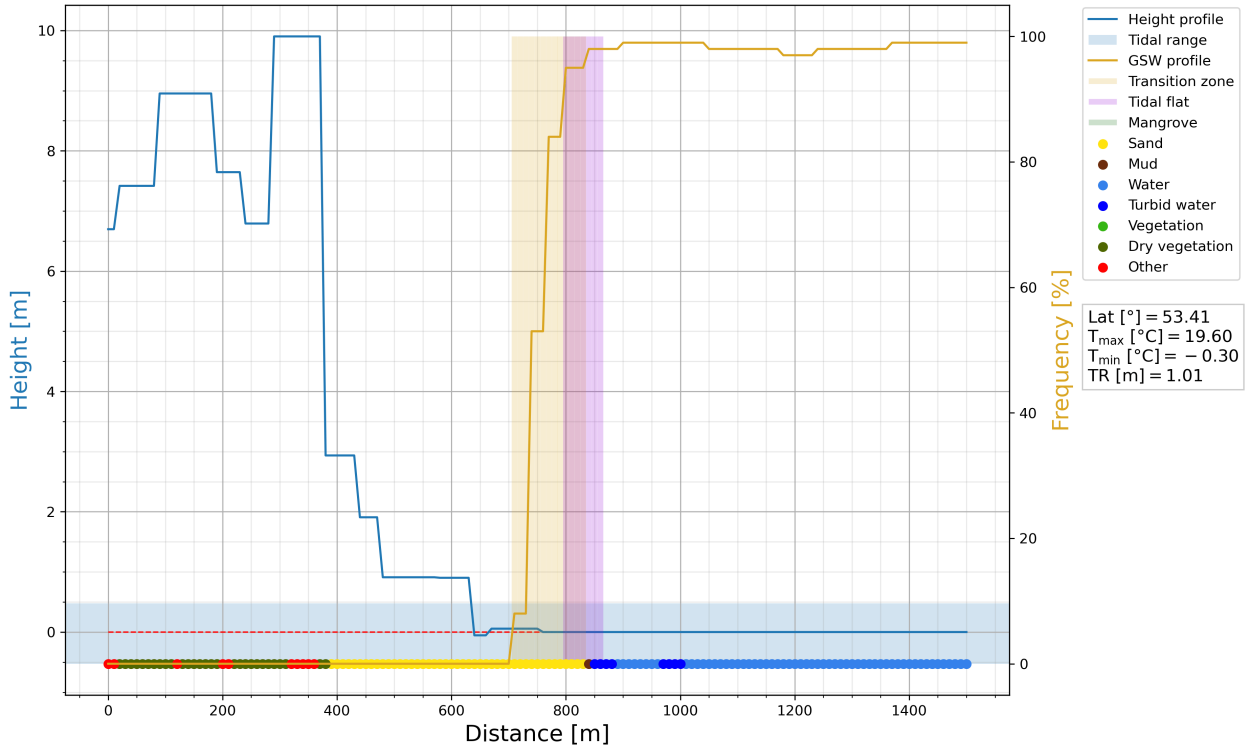
The transect profile meets the expectations described above fairly well. The topographic profile remains very low in elevation, reaching a maximum height of $3.3m$ within $400m$ of the coastline and equalling zero where the vegetation changes to mud flats. The tidal range is $1.11m$. The GSW profile is stable at the location where water pixels are classified and gradually decreases to zero at the location where mud pixels are classified. The transition zone coincides with the classified mud pixels and has a width of $230m$. A tidal flat as defined by Murray et al. (2019) is not present. On the landward side just adjacent to the mudflats, mangroves are present and measure $110m$ in width. The point where the DEM profile and the GSW profile drop to zero coincides with the point where mangroves begin to grow and mudflats end. The transect is located at 11° latitude (in the northern hemisphere) and the temperature ranges from $19°C$ to $39°C$; these variables satisfy the conditions for mangrove growth.



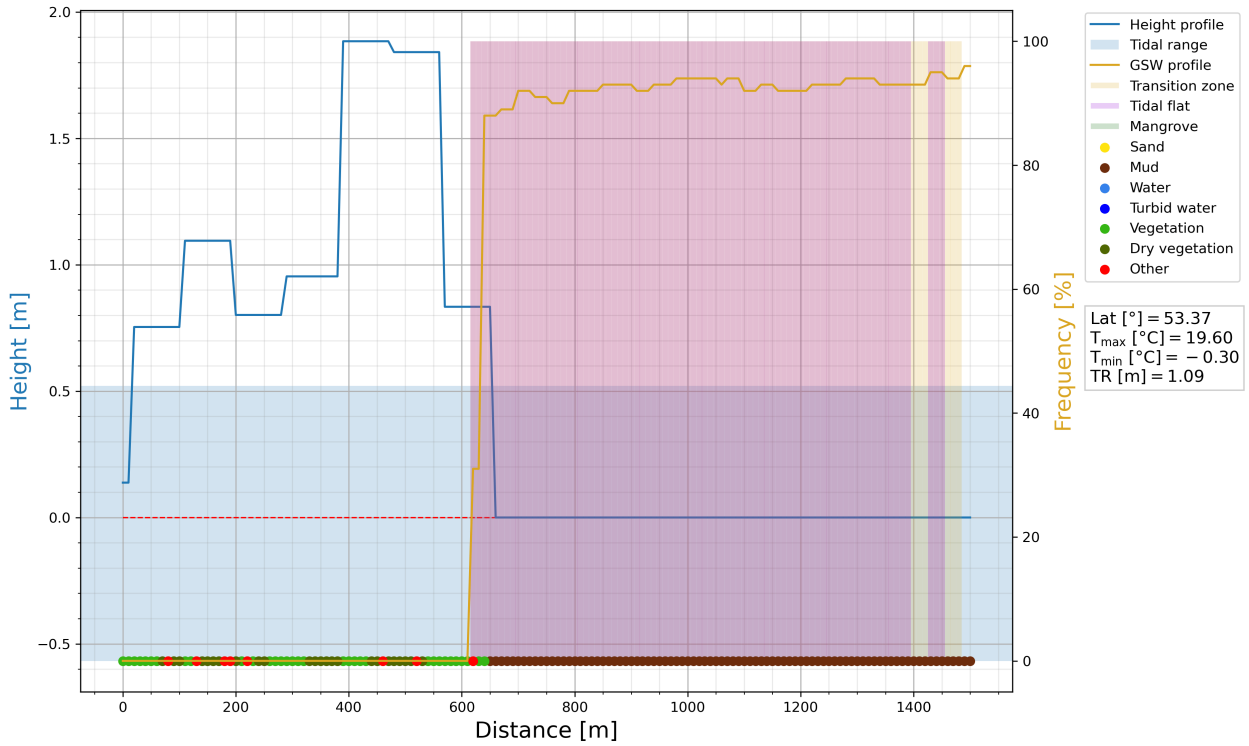
(a)



(b)

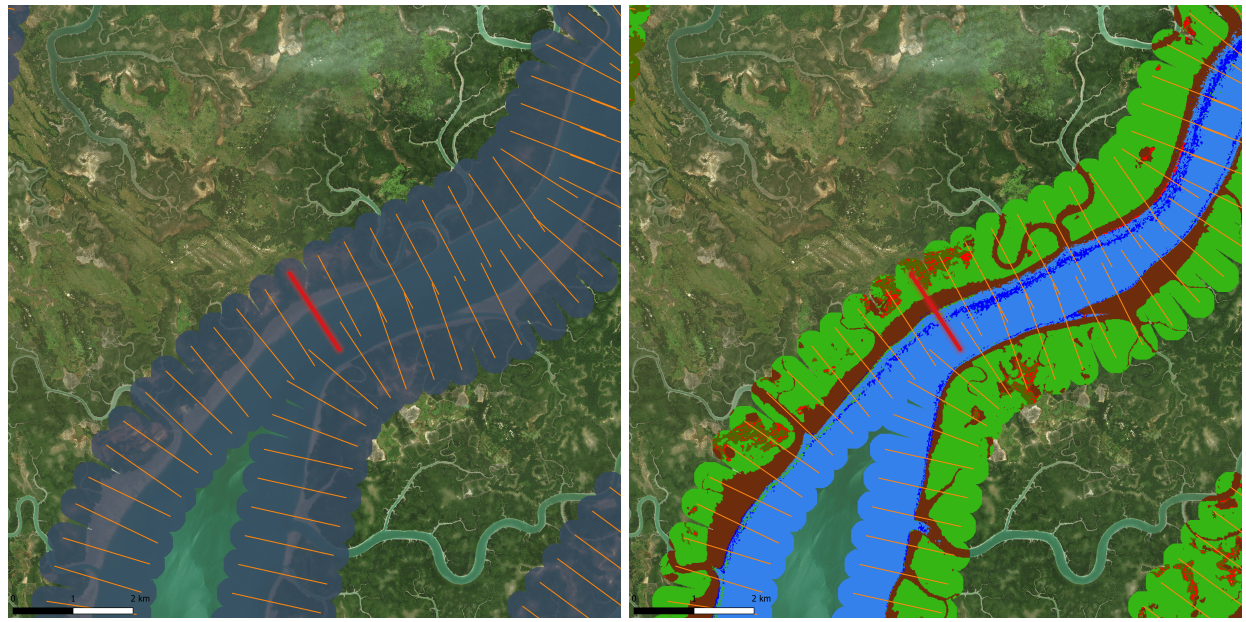

Height [m]
Frequency [%]
Distance [m]
0
200
400
600
800
1000
1200
1400
0
2
4
6
8
10
0
20
40
60
80
100
Height profile
Tidal range
GSW profile
Transition zone
Tidal flat
Mangrove
Sand
Mud
Water
Turbid water
Vegetation
Dry vegetation
Other
Lat [°] = 53.41
T_max [°C] = 19.60
T_min [°C] = − 0.30
TR [m] = 1.01

(c)

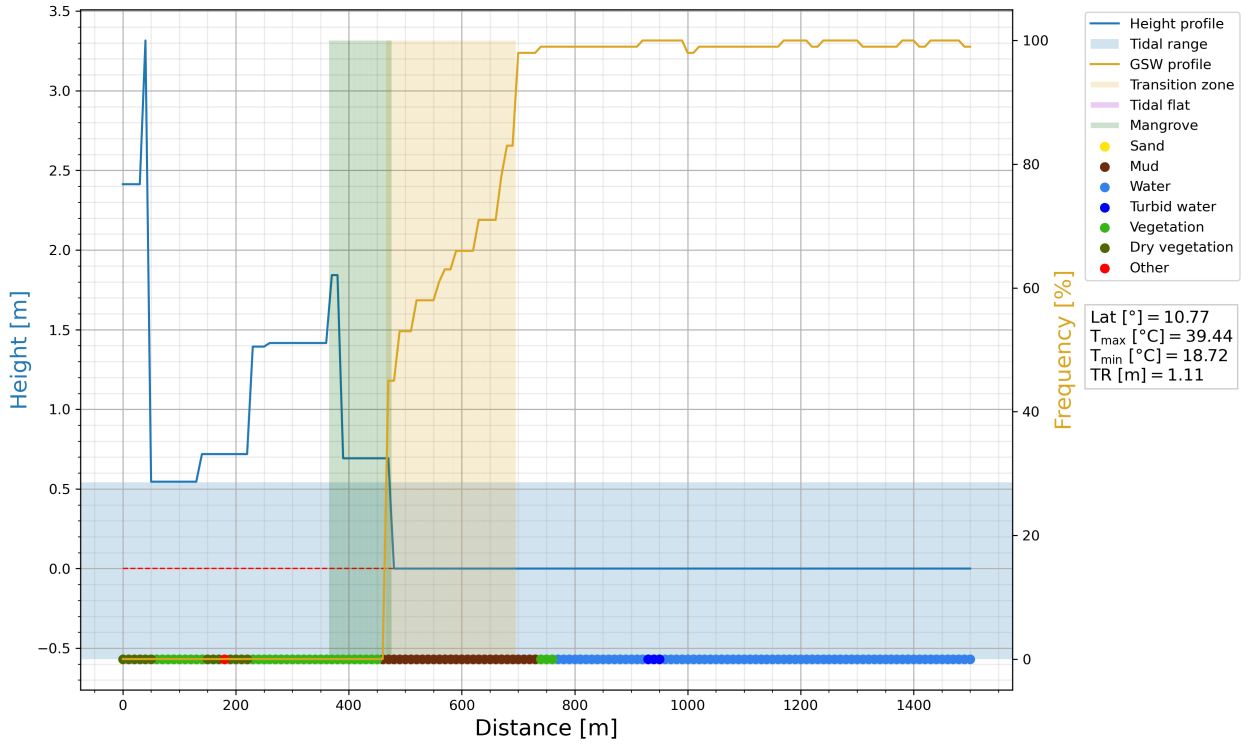


(d)

Figure 28: Example of a barrier island in the Netherlands. (a) individual S2 TOA RGB images around the transects; (b) classified images around the transects classified with global multispectral image classifier; (c) transect profile of the upper transect marked in red (BOX_187_112_17); (d) transect profile of the bottom transect marked in red (BOX_187_112_79). The horizontal axis is the length [m] along the transect, where $0m$ is the landward end and $1500m$ the seaward end. The left y-axis represents the height [m] above mean sea level to which the height profile is related, as well as the tidal range, the right y-axis represents the frequency [%] with which water is present (GSW profile). The multispectral image classification is shown at the bottom of the graph with coloured dots referring to the coastal class of each pixel along the transect. The transition zone (orange), tidal flat occurrence (purple) and mangrove occurrence (green) are plotted as a vertical shaded bar, the tidal range (blue) is plotted as a horizontal shaded bar.



(a) (b)



(c)

Figure 29: Example of a mangrove delta in Guinea. (a) individual S2 TOA RGB images around the transects; (b) classified images around the transects classified with global multispectral image classifier; (c) transect profile of the upper transect marked in red (BOX_120_100_134). The horizontal axis is the length [m] along the transect, where $0m$ is the landward end and $1500m$ the seaward end. The left y-axis represents the height [m] above mean sea level to which the height profile is related, as well as the tidal range, the right y-axis represents the frequency [%] with which water is present (GSW profile). The multispectral image classification is shown at the bottom of the graph with coloured dots referring to the coastal class of each pixel along the transect. The transition zone (orange), tidal flat occurrence (purple) and mangrove occurrence (green) are plotted as a vertical shaded bar, the tidal range (blue) is plotted as a horizontal shaded bar.

4.2 Assessment of coastal datasets

6

Section 4.1 showed that some physical geographical characteristics of the different coastal environments are clearly visible in the transect profiles and that in some cases these characteristics have a distinctive potential from other coastal environments. However, only four examples of different coastal environments are given above. In this section, the distribution of each coastal dataset per coastal type is shown in density plots to give a more general insight into the behaviour of the coastal datasets for different coastal types. Based on these plots, it will be discussed whether the characteristics per coastal type meet expectations.

The histograms in Figure 30 show the distribution of each coastal dataset for four coastal types. The distribution is based on 1868 manually labelled transects in the coastal types: plain beaches, muddy coasts, coastal cliffs and vegetated coasts. The labelling of transects is discussed in more detail in section 5.1.1. Each coastal dataset is briefly discussed in this section.

4.2.1 Topographic relief

Maximum height The distribution of the maximum height (Figure 30a) shows that most muddy coasts have a very low maximum height, the peak of the density estimation is relatively narrow and peaks at about $5m$. Most muddy coasts do not exceed a maximum height of about $20m$. Plain beaches also have a relatively low maximum height (peaking at about $10m$), but the density estimation shows more dispersion than for the muddy coasts. Plain beaches with maximum heights exceeding $80m$ occur occasionally. The maximum heights of the coastal cliffs and the vegetated coasts show a large variability and can be any height that is showed in the graph ($0m - 100m$). The distribution of the maximum height is as expected: generally low maximum heights for muddy coasts with low variability; for coastal cliffs a high variability of maximum heights (due to the less 'strict' conditions for coastal cliffs).

Height variance The distribution plot of the height variance (Figure 30b) shows that there is a very large dispersion for all coastal types, except for the muddy coasts. Only the muddy coasts show a peak in the plot. This peak is at the lower values, which means that the topographical profile of the coastal area varies little (gentle slope), while higher values mean that the profile varies a lot and has steeper slopes. The height variance is related to the maximum height and therefore shows a roughly similar pattern.

4.2.2 Climate

Latitude The latitude distribution plot (Figure 30c) shows that the different coastal types are more common in some places around the world than in others. Remarkably, the transects used for making this plot are not evenly distributed, there are more transects in the northern hemisphere (right side of the plot) than in the southern hemisphere (left side of the plot). The northern hemisphere contains more land and therefore more coastline as well. In fact, the transects are not evenly distributed around the globe anyway, as the equator has a larger circumference than other latitudes, and thus more transects are automatically located there. However, this distribution plot shows how each coastal type is distributed relative to itself and not relative to other coastal types.

Plain beaches peak between 20° N and 45° N, but occur at all latitudes. Muddy coasts show a distinct peak at about 3° N. This is consistent with the knowledge that many muddy coasts occur at lower latitudes. However, muddy coasts also occur at the higher latitudes of the northern hemisphere. Coastal cliffs are most common at higher latitudes and show a dip in the graph near the equator. Vegetated coasts occur everywhere and show two almost equal peaks where they are most frequent; one near the equator at about 5° S and one at about 55° N.

Absolute latitude The distribution based on absolute latitudes (Figure 30d) is a more general approach to indicate the most common locations of certain coastal types. Plain beaches occur at all latitudes; muddy coasts are most common at lower latitudes; vegetated coasts occur most at low and high latitudes and less at the mid-latitudes; coastal cliffs are most frequent at higher latitudes and are nearly absent at low latitudes. This graph can be compared with the graph of Hayes (1967), see Figure 6. Again, mud is most common at the lower latitudes close to the equator. Sand occurs everywhere with a peak around 25°. Rock and gravel are most common at the higher latitudes, peaking around 70° latitude.

Minimum and maximum temperatures Both temperature profiles (Figures 30e and 30f) show that coastal cliffs occur mainly in colder regions, while muddy coasts occur mainly in warmer regions. This is as expected, as muddy particles mainly form by chemical weathering in humid tropical climates (section 2.2.2; whereas large rocks break down into smaller rocks in the colder regions by mechanical weathering.

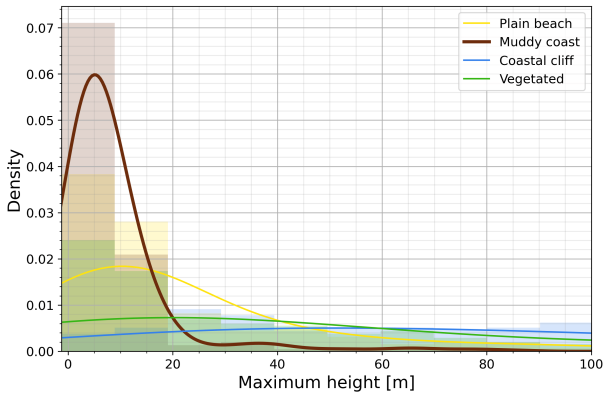
Mangrove occurrence According to the distribution plot of mangrove occurrence (Figure 30g), most mangroves grow in muddy coastal environments. Mangroves (almost) never occur near coastal cliffs. However, this graph is a bit misleading because there are only two values available for the input: True (1) or False (0). Therefore, the distribution plot is not the best way to examine the relationship between mangroves and different coastal types.

4.2.3 Tides

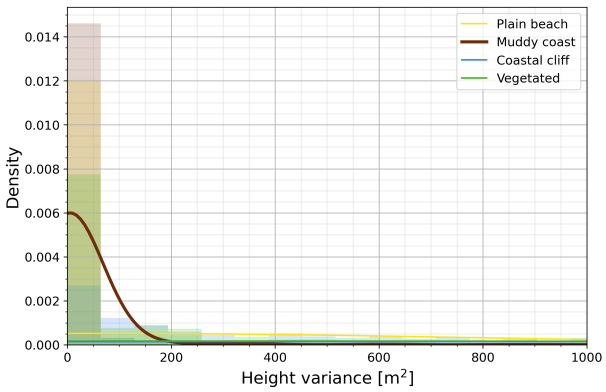
Width of tidal flats The width of the tidal flats (Figure 30h) shows a distribution that meets expectations fairly well. The coastal cliffs have no or very narrow tidal flats. In muddy coastal environments tidal flats are the most common and are also the widest. At the plain beaches and the vegetated coasts tidal flats occur more often than at coastal cliffs, these tidal flats are generally narrow, but are wider than those along the coastal cliffs.

Width of transition zone The distribution plot of the transition zone width (Figure 30i), shows that the muddy coasts have the widest transition zones, which can easily occupy half the length of the transect. The plain beaches, coastal cliffs and vegetated coasts, all, show a peak at a width of about $70m$, however the dispersion of the width of the transition varies for those coastal types. The plain beaches show the least dispersion, the coastal cliffs the most. The profile of the coastal cliffs is not as expected. The transition zone was expected to be very narrow (due to the steep topographic profiles), but shows large variability. This can be partially explained by Figure 27c, the GSW profile extends to just below 95% in front of the coastline and is therefore included in the total width of the transition zone. This is not necessarily the part of interest; rather, it is the part where the sudden jump occurs. Another explanation is that in many coastal cliff environments islands and/or fjord-like features are present along one transect. The GSW profile then exhibits multiple peaks and troughs, resulting in the presence of multiple transition zones, all contributing to the total width of the transition zone. However, the way the dataset is used in this study is limited.

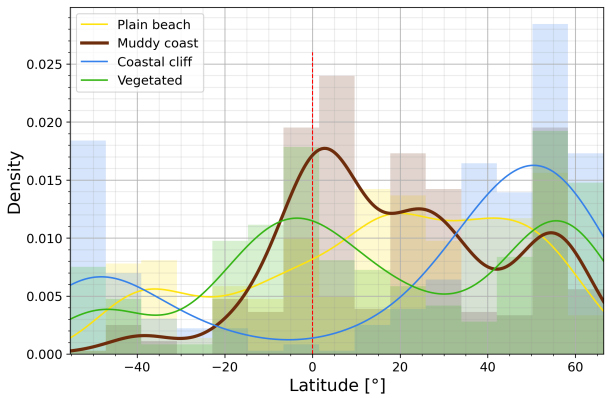
Tidal range The tidal range plot (Figure 30j) shows that the muddy coasts have a fairly distinct profile compared to the three other coastal types. Most muddy coasts have a tidal range of about $0.7m$ and quite a few extremes of about $2.2m$, while most other coasts have a smaller tidal range (approximately $0.2 - 0.4m$). This is as expected according to the description given in 2.2.2, which explains that (muddy) tide dominated flats occur often at larger the tidal ranges (although often in combination with small wave heights). The tidal range defined in this study in Equation 3 does not correspond to the classification of tidal ranges proposed by Davies (1964), which is based on the average high spring tide and average low spring tide. These are more extreme water levels and hence result in larger tidal ranges. However, the approach in this study can still be useful and shows a clear distinction for muddy coasts.


Plain beach
Muddy coast
Coastal cliff
Vegetated
Density
Maximum height [m]

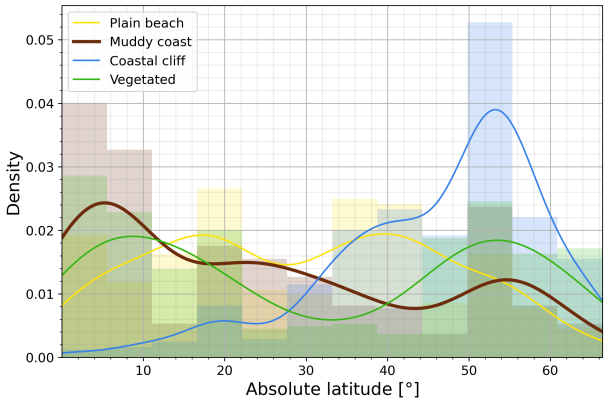
(a)


Plain beach
Muddy coast
Coastal cliff
Vegetated
Density
Height variance [m^2]

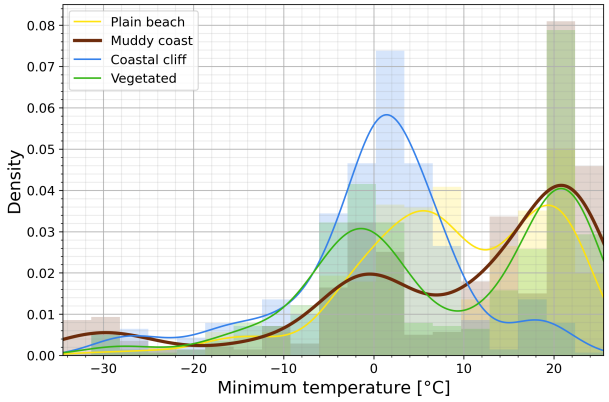
(b)


Plain beach
Muddy coast
Coastal cliff
Vegetated
Density
Latitude [°]

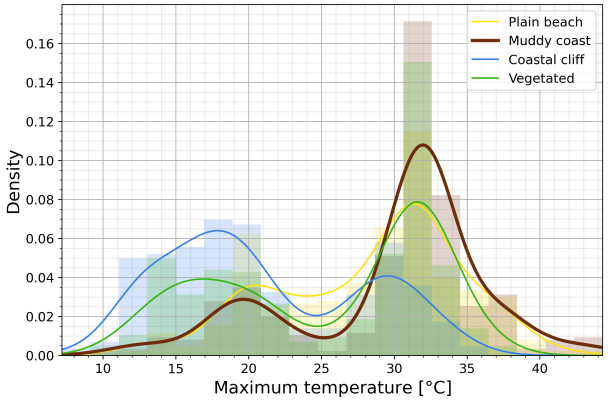
(c)


Plain beach
Muddy coast
Coastal cliff
Vegetated
Density
Absolute latitude [°]

(d)


Plain beach
Muddy coast
Coastal cliff
Vegetated
Density
Minimum temperature [°C]

(e)


Plain beach
Muddy coast
Coastal cliff
Vegetated
Density
Maximum temperature [°C]

(f)

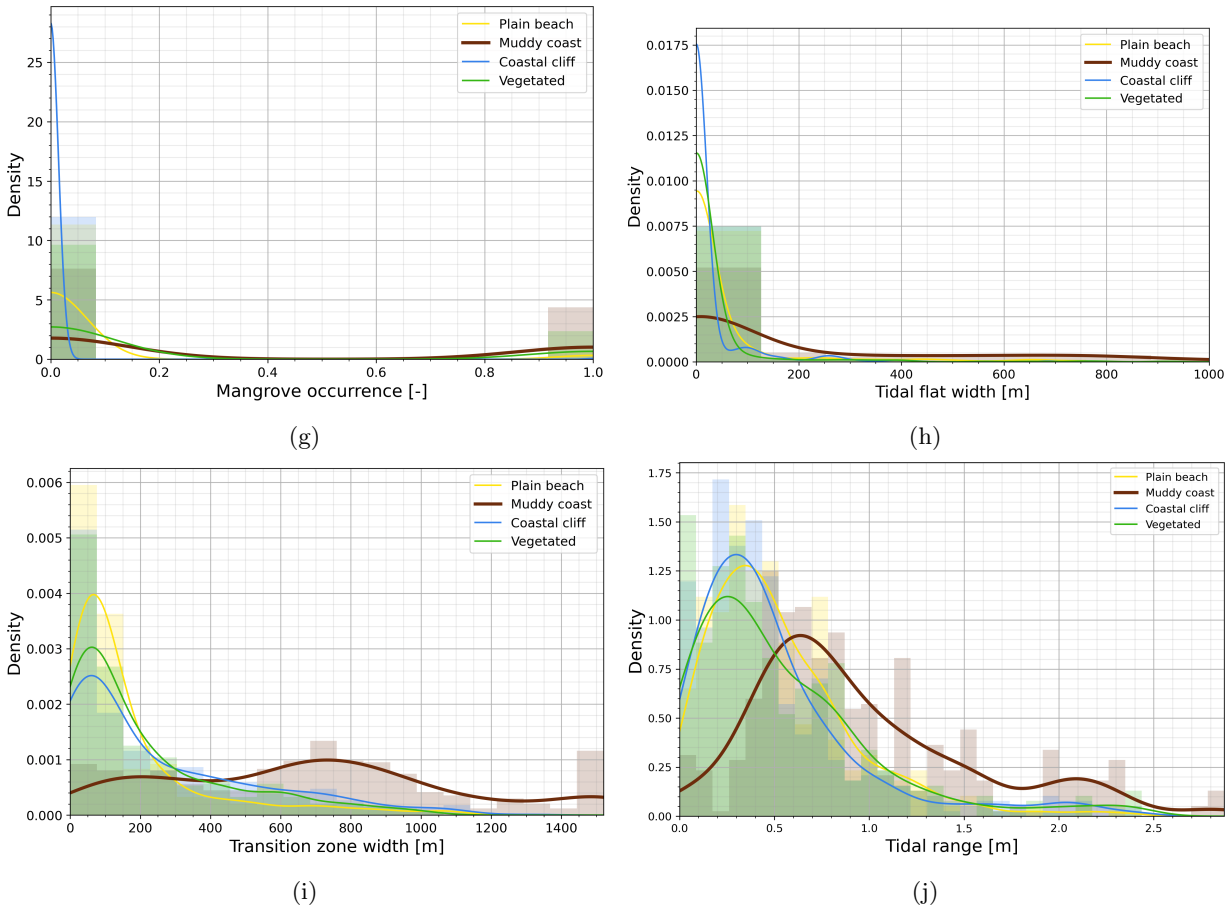


Figure 30: Distribution plots for the ten global coastal variables explored in this study, based on 1868 labelled transects at global scale. The vertical shaded bars show the (true) densities, the solid lines show the estimated densities for the four coastal types: plain beach, muddy coast, coastal cliff and vegetated coast. (a) Maximum height; (b) height variance; (c) latitude; (d) absolute latitude; (e) minimum temperature of coldest month; (f) maximum temperature of warmest month; (g) mangrove occurrence; (h) width of tidal flat; (i) width of transition zone; (j) tidal range.

4.3 Added value of coastal datasets

Section 4.2 explained that the distribution of coastal datasets among the various coastal types meets the expected patterns and that the selected coastal datasets are able to capture the physical geographical characteristics of the different coastal types. In section 3.2.3 two examples were shown where muddy pixels were recognised by the global multispectral classification model, but it was known that these 'muddy pixels' were dark-coloured sand or dark-coloured volcanic material. In order to recognise these kind of misclassifications, it was proposed to add contextual information by analysing physical geographical coastal characteristics.

In Figure 31 the same coastal area on Tenerife is analysed as is shown in Figure 22a. For the red marked transect, that crosses the 'muddy pixels', the transect profile is created that shows the information obtained with the coastal datasets. For this case the maximum height exceeds $80m$ and the slope of the height profile is relatively steep. No tidal flats are observed and the transition from land to water is quite steep and sudden; the width of the transition zone is narrow: $110m$. These characteristics could potentially fit to the coastal cliff environment, see the distribution plots in Figure 30.

This plot shows that by analysing only the spectral reflectance properties of muddy particles with multispectral satellite images, this transect might be classified as muddy coast. However, when analysing the physical geographical characteristics of muddy environments with globally available coastal datasets, this transect might better match with the characteristics of coastal cliffs.

4.4 Conclusions

In this section, ten coastal variables are analysed, obtained from seven globally available coastal datasets. The transect profiles of four different coastal environments in section 4.1, show that the behaviour of the coastal variables used in this study are consistent and follow expected patterns, therefore they can be considered as reliable coastal variables.

The coastal variables are distributed as expected (for muddy coasts) and confirm the physical geographical characteristics (of muddy coasts). The height profiles from the MERIT DEM (Yamazaki et al., 2017) are analysed and show that for muddy coasts both the maximum height and the variance of the height profile is low and distinctive from other coastal types. The distribution of the (absolute) latitude of the location of the coastline showed that most muddy coasts are found around the equator in the tropics, which is in line with Figure 6 from Hayes (1967) and with the knowledge of the formation of muddy coasts (section 2.2.2). The minimum and maximum temperatures for each transect extracted from the WorldClim V1 Bioclim dataset (Hijmans et al., 2005) are usually higher for muddy coasts than for other coastal types (more chemical weathering takes place at higher temperatures, section 2.2.2). The mangrove forests defined by the Global Mangrove Forests Distribution (Giri et al., 2005) are higher positively correlated with muddy coasts than with other coasts. However, not all transects where mangroves grow are also 'visible' muddy coasts, as the mangroves can cover the entire mudbanks that lie beneath them and therefore cannot be observed from above with satellite images. The intertidal areas observed by the Murray Global Intertidal Change Dataset (Murray et al., 2019) for most muddy coasts are wide, as well as the width of the transition zone defined with the Global Surface Water dataset from Pekel et al. (2016). This can be explained by the typical configuration of the continental shelf where muddy coasts often occur (section 2.2.2) and the abundance of sediment that is supplied. The last coastal variable that is explored in this study, is the tidal range determined from the mean higher high water and mean lower low water from the Deltares Global Tide and Surge Model. Tide dominated flats occur most often with large tidal ranges (section 2.2.2), which can be seen in the tidal range distribution of muddy coasts as well. Section 4.3 showed that the addition of contextual coastal information could potentially increase the probability of correctly classifying a muddy coast at global scale.

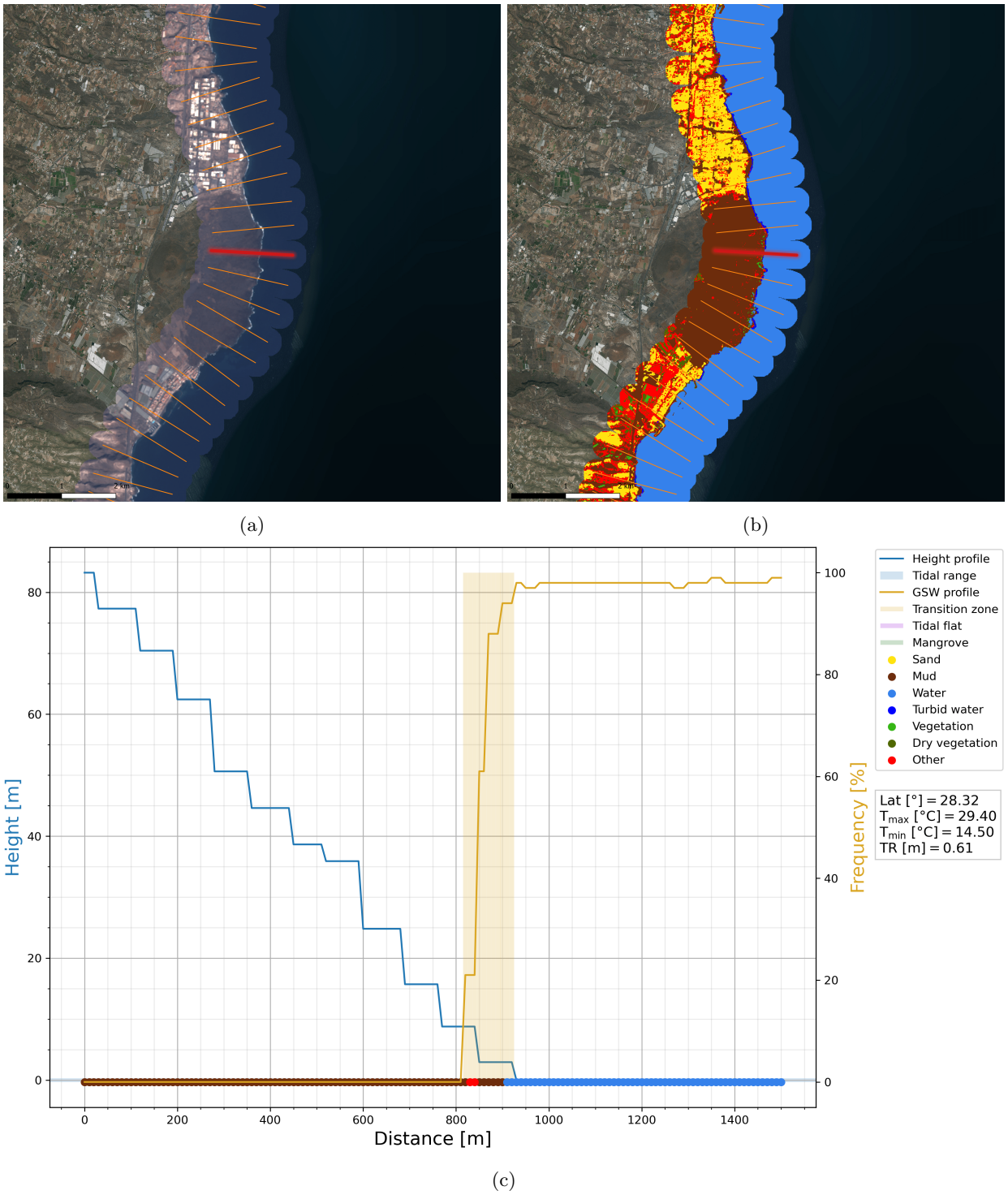


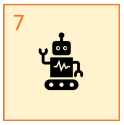
Figure 31: Example of a coastline consisting of volcanic material in Tenerife, Spain. (a) individual S2 TOA RGB images around the transects; (b) classified images around the transects classified with global multispectral image classifier; (c) transect profile of the upper transect marked in red (BOX_142_084_57). The horizontal axis is the length [m] along the transect, where $0m$ is the landward end and $1500m$ the seaward end. The left y-axis represents the height [m] above mean sea level to which the height profile is related, as well as the tidal range, the right y-axis represents the frequency [%] with which water is present (GSW profile). The multispectral image classification is shown at the bottom of the graph with coloured dots referring to the coastal class of each pixel along the transect. The transition zone (orange), tidal flat occurrence (purple) and mangrove occurrence (green) are plotted as a vertical shaded bar, the tidal range (blue) is plotted as a horizontal shaded bar.

5 Hybrid transect classification

In this section, the reliability, robustness and performance of correctly classifying muddy coasts at global scale with an automated global mud classification model will be investigated. The aim of this model is to detect muddy coasts at transect level and will therefore be called the hybrid transect classification model. The hybrid transect classification model is obtained by combining the information of the spectral reflectance properties of mud at pixel level provided by the multispectral satellite images (section 3) and the information of the physical geographical characteristics of mud at larger-scale level collected with the globally available coastal datasets (section 4). By combining these two sources of processed information it is expected that a more robust and reliable model for the classification of muddy coasts at global scale will be obtained than by using only one source of information. For example, from multispectral satellite images one can recognise various coastal features, such as sand, water and vegetation at pixel level, but coastal features such as topography and tidal conditions cannot be observed from that image, whereas this kind of information is also expected to have added value for the recognition of certain coastal types, as is shown in section 4.3. The hybrid model will take this added knowledge into account and is therefore expected to have a higher performance in correctly classifying muddy coasts and to be more robust and reliable.

Section 5.1 first explains how the hybrid transect classification model is developed. Then the performance of the hybrid model is assessed in section 5.2 and compared to the performance of the two other models, that use only the information from the multispectral image classification or that use only the information from the globally available coastal datasets. Subsequently, in section 5.3 the three models are put in practice to analyse the performance on the area of Tenerife, the example used before in section 3.2.3 for the multispectral image classification and in section 4.3 for the coastal datasets. This section ends with a summary of the results found and selects the best performing global mud classification model in section 5.4. The best performing model will be used for the global classification of muddy coasts and the results of the global mapping will be shown in section 6.

5.1 Method



The hybrid transect model for classification of global muddy coasts uses the transect system, explained in section 2.4.1. 1868 transects randomly distributed around the world are labelled. The labelling distinguishes five coastal types: plain beaches, muddy coasts, coastal cliffs, vegetated coasts and others. The labelling is done manually by visual inspection of composite TOA RGB images acquired by S2 in the year 2020. The label is assigned to the entire transect ($1500m$ in cross shore length), but the manual assessment of labelling focuses mainly on (the small area around) the coastline. Finally, a random forest classifier is trained with the properties of the 1868 labelled transects.

5.1.1 Study area



The study area of the global hybrid transect classification model covers the entire (ice-free) global coastline between the Arctic and Antarctic circles (between $+66.5°$ and $-66.5°$ latitude). The global coastal area is divided into approximately 1.8 million transects of $1500m$ wide in the cross shore direction and spaced at $200 - 450m$ along the coast, depending on the location on the globe, as explained in section 2.4.1. For training purposes of the global transect classification models, a total of 1868 transects have been randomly selected in order to capture different coastal types under different conditions on the six continents (excluding Antarctica). In Figure 32, the locations of the 1868 transects are marked with orange dots.

The labelling of the training transects is done manually by visual inspection of one year composite TOA RGB images acquired by S2 in the year 2020. Around every selected transect, a buffer of $250m$ wide is drawn for which the composite image is assembled. The transect can belong to one of the following coastal types:

1. *Plain beaches* - A clear strip of sandy sediment must be present at the coastline;
2. *Muddy coasts* - A tidal flat/mud bank must be present at the coastline and must visible on the satellite image (this excludes muddy coasts that are covered (from above) by mangrove forests);
3. *Coastal cliffs* - This coastal type is assigned to coasts that show large rocky features along the coastline that can be observed from the satellite image;

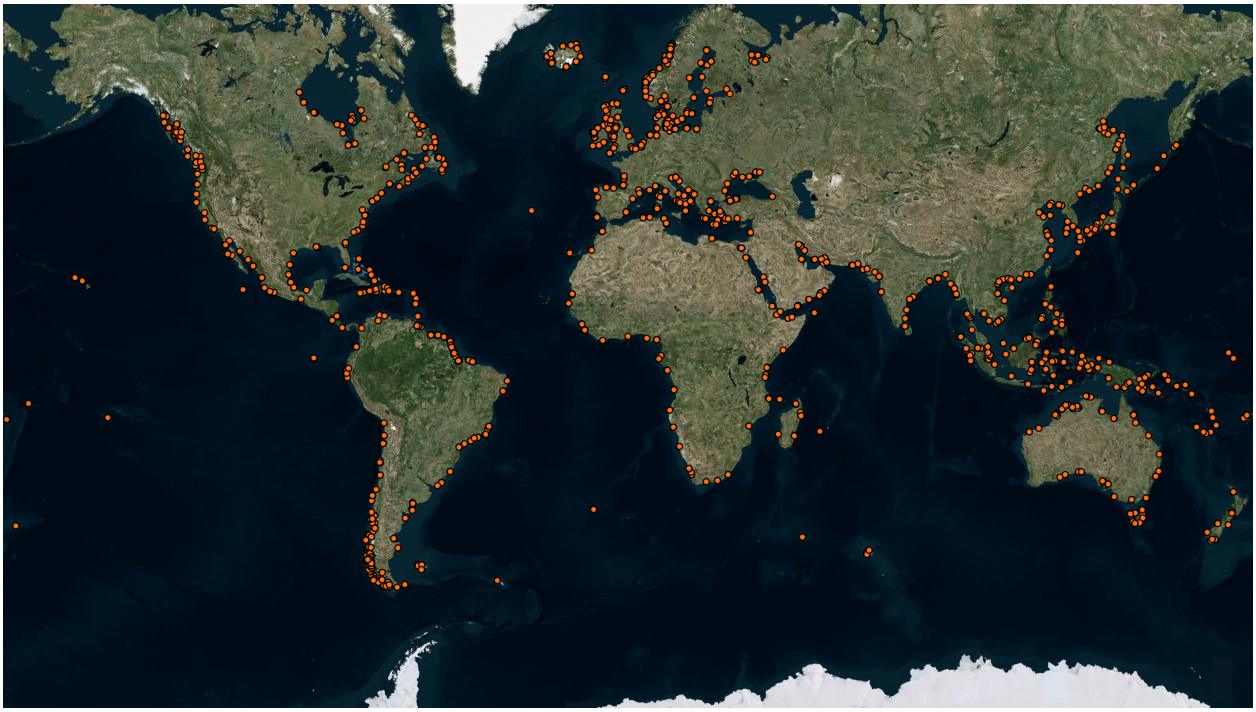
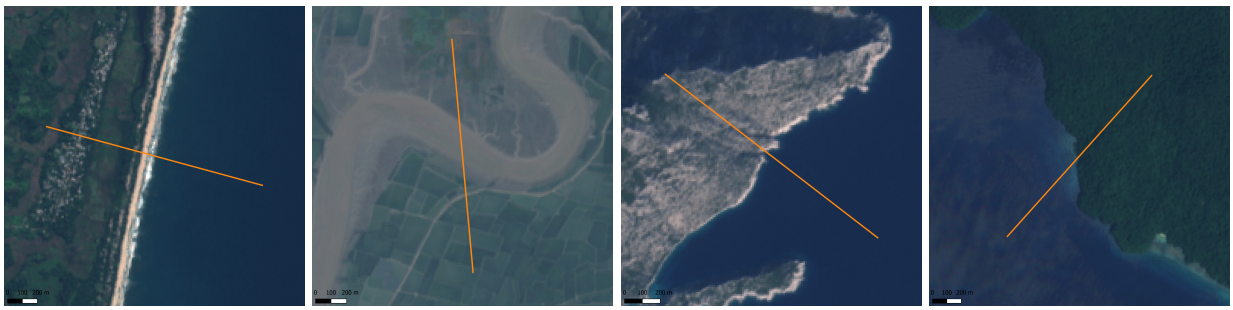


Figure 32: The global distribution of the locations of the training transects used for training the three global transect classification models (multispectral, coastal and hybrid). The orange dots represent the 1868 transects that have been manually labelled and whose training data has been collected.

4. *Vegetated coasts* - The coastline is covered with vegetation, on the satellite image no sediment/beach is observed between water and land;
5. *Other* - All indefinable coasts, including coasts covered with ice/snow, anthropogenic structures that shape the coastline such as harbours and dykes, etc..

In Figure 33 four examples of coastal transects are shown that cross different coastal types and are labelled manually by a visual inspection of the underlying one year composite TOA RGB image.



(a) (b) (c) (d)

Figure 33: Four examples of coastal transects that are assigned to four different coastal types: (a) plain beach (Madagascar); (b) muddy coast (Bangladesh); (c) coastal cliff (Turkey); (d) vegetated coast (Indonesia). The image of each figure is the one year composite TOA RGB image acquired by S2.

5.1.2 Collecting data

Information from the labelled transects is obtained both from the composite TOA RGB satellite images, as well as from the globally available coastal datasets. The information along the transect profile is processed as described in section 2.5 for the multispectral image classification and in section 2.4 and 4.1 for the coastal datasets. The processed information provides one uniform value per coastal feature (i.e. the percentage of sand pixels along the transect or the maximum height of the topographic profile, etcetera) for training the (one dimensional) classification model.

5.1.3 Training models

A supervised random forest classifier (Breiman, 2001) with 15 decision trees is trained using the uniform values per coastal feature for each labelled transect. The output of the classifier assigns each transect a predicted label with the corresponding coastal type (plain beach, muddy coast, coastal cliff, vegetated coast, other) that best matches the observed coastal characteristics of the transect. With this classifier all coastal transects can be classified into a certain coastal type.

Several classification models are developed, using different subselections of coastal features, in order to achieve the highest accuracy in the classification of coastal transects. A performance analysis is carried out on the different classification models. Three models are developed and assessed:

1. *Multispectral model* - This classification model is trained using all multispectral features, namely the percentages of pixels of each class along the entire transect (i.e. sand, mud, water, turbid water, vegetation, dry vegetation and other).
2. *Coastal model* - This classification model is trained using all the variables of the globally available coastal datasets (i.e. maximum height, variance of height, latitude, absolute latitude, minimum and maximum temperature, occurrence of mangrove, width of tidal flats, width of transition zone and tidal range).
3. *Hybrid model* - This model is trained using all the multispectral features selected by the multispectral model and all the variables selected by the coastal model.

5.1.4 Best model

The robustness, durability and repeatability of the three classification models of global transects are tested. The performance of the different models is compared in section 5.2. The model with the best performance will be used to classify the coastal types of all global transects. Based on this classification, a global map of mud occurrence will be presented and some statistics will be shown and discussed in section 6.

5.2 Model performances

In this section, the results of the three different global transect classification models (multispectral, coastal and hybrid) are discussed. This is done on the basis of the performance of the models, with different accuracy values being assessed. A performance analysis is carried out on the manually labelled training transects, where a coastal type is assigned to each training transect by visual inspection of S2 TOA RGB 1 year composite images of the year 2020. The performance analysis tests the robustness, repeatability and durability of the classification models. The performance of the three models is assessed by calculating the overall model accuracy, as well as the accuracy for the muddy coasts only. The best performing model will eventually classify all transects. The result of the global classification will be shown in section 6.

5.2.1 Feature importance

The three different global transect classification models use a different subselection of coastal features. In order to make a subselection where each feature adds valuable information to gain a higher performance, the importance of each feature is studied. The importance of features differs when a different subselection of features is used. For each model (multispectral, coastal and hybrid) a feature importance plot is made. To

obtain a classification model with the highest performance and the smallest number of features, a comparison is made of how much accuracy each feature adds. First, the feature with the highest importance is used to train the model, then the feature with the second highest importance is added, until all features are added in order of importance. The overall model accuracy plot is then used to assess whether a feature should be included or not; if it does not increase accuracy, the feature is not included.

Figure 34 shows the feature importance plot of each of the models, where the different coastal features are ordered according to their importance (upper row), together with the overall model accuracy plot of each model, that is computed every time a new coastal feature is added (bottom row). From these plots it can be seen that for every model all added coastal features increase the accuracy (slightly). Therefore all features are included. The three final models obtained in this step are used in the next sections. The following features (in order of importance) are selected for each model:

1. *Multispectral model* - Water, mud, vegetation, turbid water, sand, dry vegetation and other.
2. *Coastal model* - Maximum height, width of transition zone, height variance, tidal range, minimum temperature, absolute latitude, maximum temperature, latitude, width of tidal flats and mangrove occurrence.
3. *Hybrid model* - Mud, maximum height, height variance, water, sand, width of transition zone, absolute latitude, turbid water, minimum temperature, tidal range, maximum temperature, latitude, vegetation, other, dry vegetation, width of tidal flats and mangrove occurrence.

5.2.2 Confusion matrix

The accuracy of each transect classification model is obtained by inspecting their confusion matrices, which are shown in Figure 35. Each classification model is trained with 75% of the 1868 labelled transects; 25% of the labelled transects were used for validation of the models. In Figure 35 he horizontal axis shows the predicted label, while the vertical axis shows the actual label of the transect. On the diagonal are the true positives (TP) and the true negatives (TN); i.e. when the predicted label of a particular transect is equal to the true label of that particular transect. To the right of the diagonal are the false positives (FP) and below the diagonal the false negatives (FN); i.e. when the predicted label is not equal to the true label.

The confusion matrices show the misclassifications between the five different coastal types. From the confusion matrices, it can be observed that all three classification models show a fairly large confusion between coastal cliffs and vegetated coasts. This can be explained by the fact that in some cases the only difference between these two coastal types is the narrow appearance of rocky features on the shoreline; apart from that, the landward and the seaward side may appear exactly the same. The confusion is therefore quite logical. However, neither of these coastal types is the coastal type of interest, that is the muddy coast. Focusing on the muddy coasts; for the multispectral model the confusion of muddy coasts with all other coastal types is relatively high, especially with the vegetated coasts. This can be explained by the presence of (mangrove) forests covering the sedimentary beaches at the coastline. Some beaches can be very narrow and therefore cannot be observed from above with satellite images. For the coastal model the most confusion of muddy coasts is also with the vegetated coasts, followed by the second most confusion with the plain beaches. The same applies to the muddy labels in the confusion matrix of the hybrid model. What is also striking is that for the coastal data model and the hybrid model there is (nearly) no confusion between muddy coasts and coastal cliffs, which can be explained by the distribution plots in section 4.2, where muddy coasts and coastal cliffs show often the most distinction in each coastal variable. Finally, it can be seen that the confusion matrix of the hybrid model overall shows the least confusion between coastal types.

5.2.3 Overall model accuracy

The accuracy assessment is performed on the three global classification models (multispectral, coastal and hybrid) and relates the overall model accuracy, when classifying five coastal types (plain beaches, muddy coasts, coastal cliffs, vegetated coasts and other), to the number of used training transects. The accuracy of the classification models is expected to increase as the number of used training transects is increased. In general, accuracy increases rapidly when there are few training transects used and more are added, but

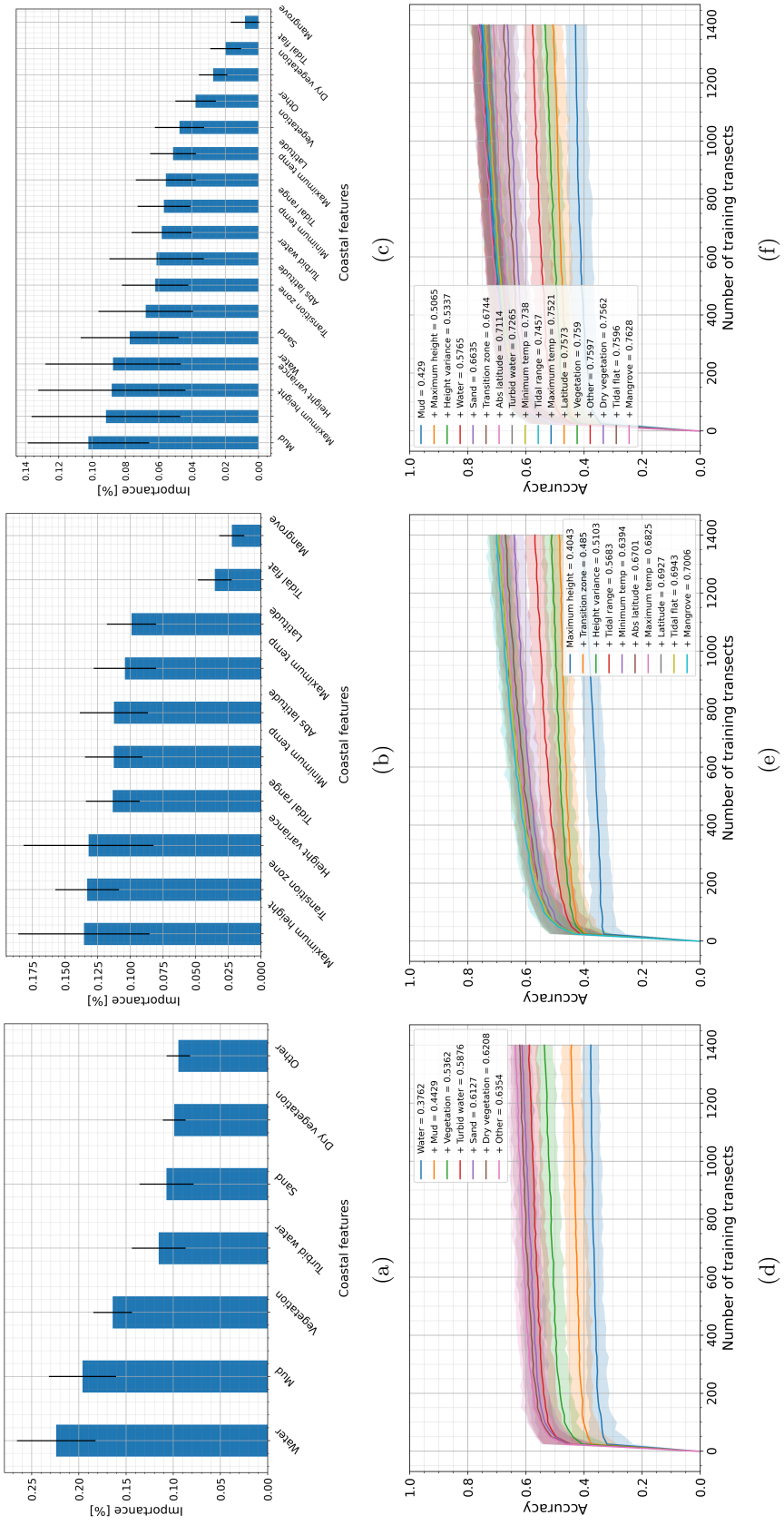
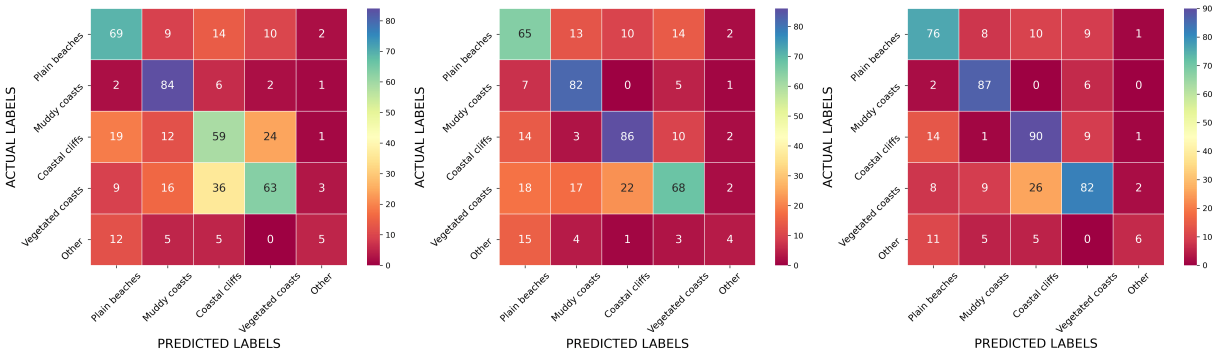


Figure 34: Selecting features to be used in each transect classification model; (a) and (d) Multispectral model; (b) and (e) Coastal model; (c) and (f) Hybrid model. Upper row, importance of coastal features used for developing the classification model in descending order. The blue bars represent the mean of the importance for each feature, the black lines represent the standard deviation of the importance of each feature. Bottom row, overall model accuracy plot for the transect classification model, for classifying five coastal types, when adding the features in order of importance to the previous feature(s). The number of used training transects is on the x-axis and the overall model accuracy (Equation 4) on the y-axis. The solid line represents the mean accuracy, the shaded area represents the $90th$ percentile of the accuracy. The legend shows which coastal feature is added to the previous one and the accuracy reached.



(a) (b) (c)

Figure 35: Confusion matrices of 468 validation transects classified with the (a) multispectral, (b) coastal and (c) hybrid transect classification model trained with 1490 labelled transects spread globally. The predicted labels are on the x-axis, the actual labels on the y-axis.

at some point adding more training transects will no longer significantly increase accuracy (and may even decrease it, due to overfitting). This can be demonstrated with the performance plot.

From the values of the confusion matrix, the overall accuracy of each model can be calculated using equation 4 described in section 3.3.1; the number of correctly classified validation transects (the sum of the diagonal) divided by the total number of validation transects.

The performance analysis is carried out on 1868 manually labelled transects (442 transects were labelled as plain beach, as muddy coast, as coastal cliff, as vegetated coast; 100 transects were labelled as other), from which 468 transects are randomly selected (but equally distributed among the coastal types) to be used for validation. The remaining transects are used for training the classification models and are divided into smaller groups. First, a model is trained with one group of training transects. This model classifies the 468 validation transects and the overall accuracy of the model is calculated. Then a new model is trained with two groups of training transects, the same validation transects are classified with the new model and the overall accuracy is calculated. This is done until all training transects have been used. This process is repeated 100 times with different randomly selected groups (cross validation) to average out errors and biases.

Figure 36 shows the overall model accuracy for the three global transect classification models. It can be seen that an increase in the number of used training transects from 0 to 200 shows a significant and rapid increase in accuracy for all models. After that, the accuracy does increase at a smaller rate. The accuracy profile of the multispectral model shows the smallest increasing rate for the given number of used training transects and reaches an overall model accuracy of 64%. The coastal model shows a higher accuracy, that increases over the whole trajectory with the largest rate of the three models; by adding more training transects than that are used now, the accuracy of this model may increase significantly. Now it stops at an accuracy of 70%. The hybrid model performs best according to the overall accuracy plot, reaching an accuracy of 76%. The performance of the hybrid model also may continue to increase significantly after the given number of used training transects. The $90th$ percentiles show that the models are fairly consistent when sampled 100 times. The width of the $90th$ percentile shows the variability of the model, which is about 7% for each model.

From the analysis of this overall model accuracy plot, it can be concluded that the hybrid model has the highest accuracy. The accuracy of all models is still increasing, so adding more training transects will likely improve performance.

5.2.4 Accuracy per coastal type

A value to assess the robustness, repeatability and durability of the global transect classification models for each individual coastal type is the f1-sore. The f1-score is a trade-off between the precision (how many transects are correctly recognised by the classifier) and the recall of the model (how many transects are wrongly assessed by the classifier), see section 3.3.2 and Equations 5, 6 and 7.

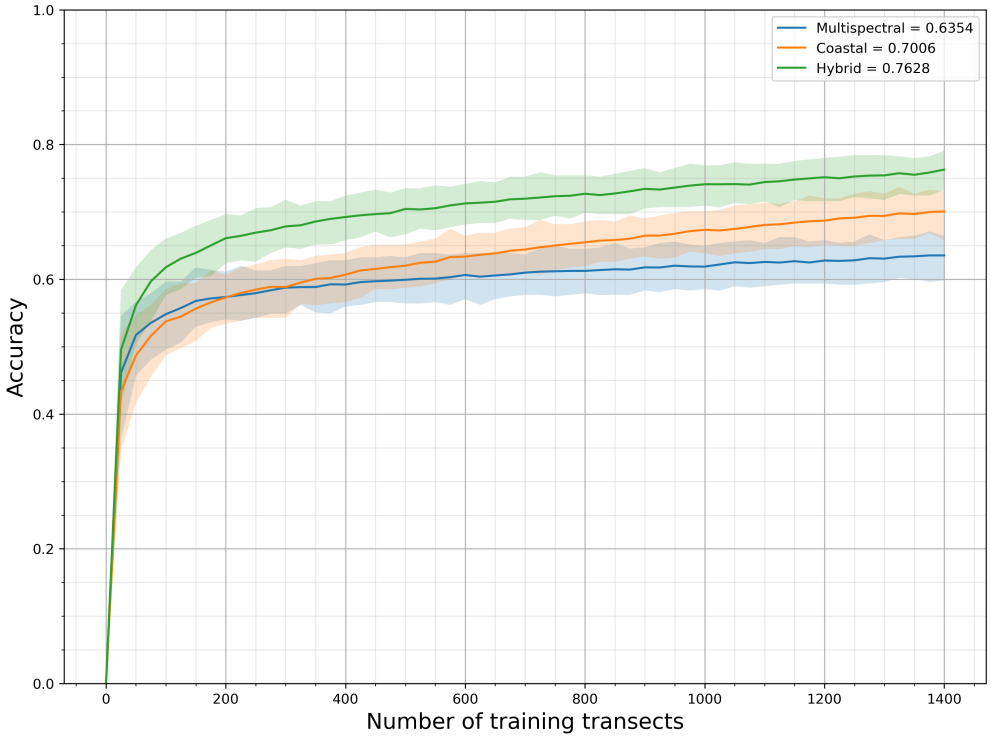


Figure 36: Overall model accuracy plot for three global transect classification models (multispectral, coastal and hybrid) that classify transects into five coastal types. The number of used training transects is on the x-axis and the overall model accuracy (Equation 4) on the y-axis. The solid line represents the mean accuracy, the shaded area represents the $90th$ percentile of the accuracy. The number behind the transect classification models in the legend are the final accuracies when 75% of the training transects are used to train the model and 25% of the training transects are used to validate the model.

This analysis is performed in the same way as the overall accuracy assessment described before; it uses the confusion matrix to calculate the f1-score and it is executed 100 times, to cross validate the f1-score. For the model that showed the highest overall model accuracy in section 5.2.3 (the hybrid model), the f1-score of each of the five coastal types is inspected in order to better understand why the classifier is inhibited in accuracy when classifying the global coastline. The f1-score for the muddy coasts is assessed for each model, as that is the coastal type that is the focus of this research and the coastal type that will ultimately be classified on a global scale.

For the hybrid model, the accuracy of each of the five coastal types is calculated and shown in Figure 37. This Figure shows that a same number of training transects are used for each coastal type (except for 'other'). Each coastal type shows a rapid increase when less number of training transects are used. After increasing the training data, the accuracy of the different coastal types are beginning to converge and flatten the increase in accuracy. All coastal types reach an accuracy between 71% and 85%. The muddy coasts show the highest accuracy. The vegetated coasts have the lowest accuracy (except for 'other'), this might be due to the fact that the characteristics of this coastal type is highly variable. For example, a vegetated coast can be either a mangrove forest next to the ocean or a mountain fully covered with pine trees next to the ocean. These coasts require two completely different conditions (in terms of topography, climate and tides), but they both belong to the vegetated coasts.

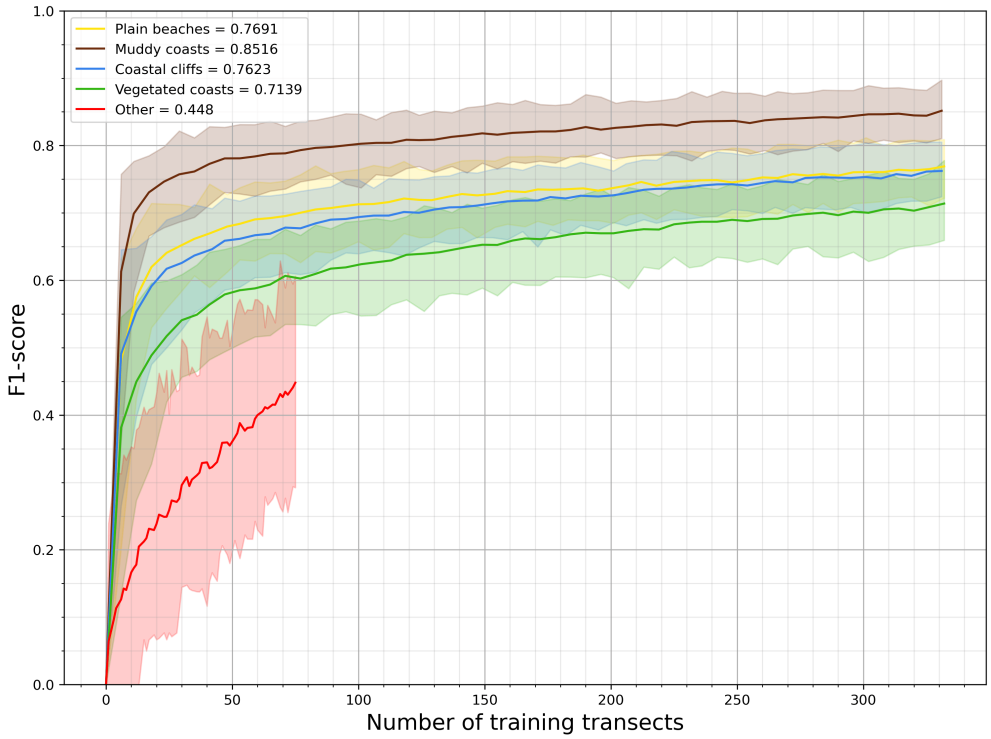


Figure 37: Accuracy plot for each of the five coastal types, when using the hybrid global transect classification model. The number of used training transects is on the x-axis and the f1-score (Equation 7) on the y-axis. The solid lines represent the mean f1-scores, the shaded areas represent the $90th$ percentiles of the f1-scores. The numbers behind the coastal types in the legend are the final f1-scores when 75% of the training transects per each coastal type are used to train the model and 25% of the training transects per each coastal type are used to validate the model.

Figure 38 shows the accuracy (f1-score) of each global transect model for the classification of muddy coasts plotted against the number of used muddy training transects. It can be seen that for all models an increase in the number of used training transects leads to an increase in the accuracy. The three models show the same behaviour as described in section 5.2.3 and shown in Figure 36. The multispectral model has the lowest accuracy of 78%. The coastal model shows a slower increase at the beginning of the profile, but has the strongest increasing slope at the end. This model performs in second place and reaches an accuracy of 81%, which is slightly higher than the multispectral model. The hybrid model performs best

for the classification of muddy coasts with an accuracy of 85% (and still rising for increasing the number of used muddy training transects). The $90th$ percentile is about $10\% - 12\%$ for each model and proves that the models perform relatively equally when the muddy training and validation transects are mixed. This indicates the robustness of the models.

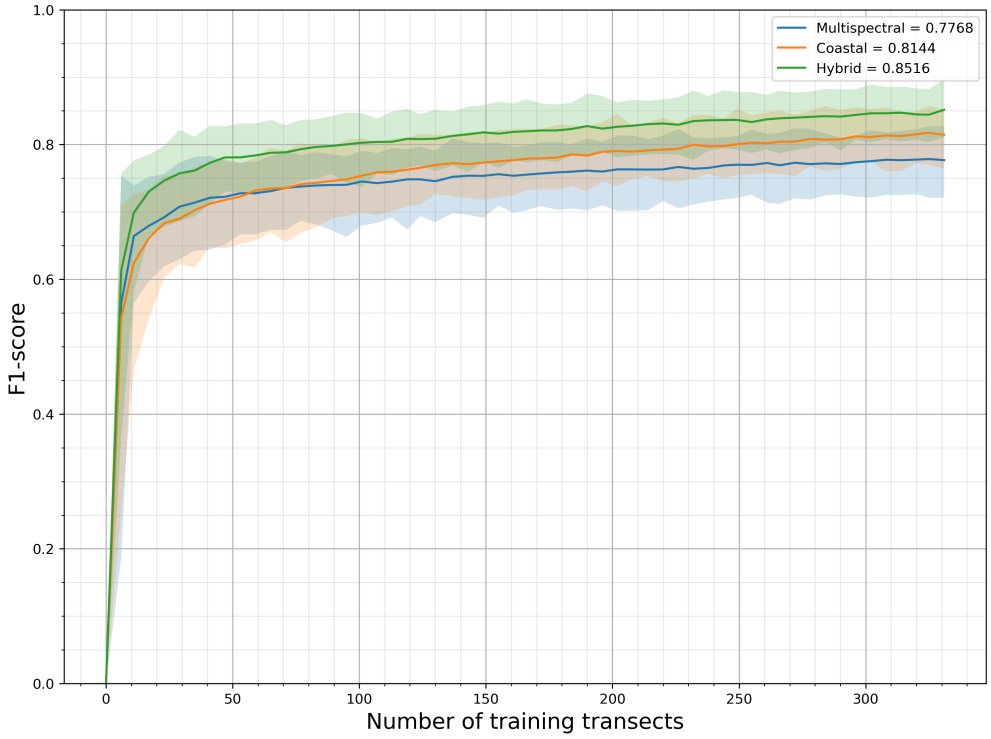


Figure 38: Accuracy plot for one of the five coastal types, namely the muddy coasts, when using each of the three global transect classification models (multispectral, coastal and hybrid). The number of used training transects is on the x-axis and the f1-score (Equation 7) on the y-axis. The solid lines represent the mean f1-scores, the shaded areas represent the $90th$ percentiles of the f1-scores. The numbers behind the transect classification models in the legend are the final f1-scores when 75% of the muddy training transects are used to train the model and 25% of the muddy training transects are used to validate the model.

From the analysis of the accuracy of classifying muddy coasts, it can be concluded that the hybrid model has the highest accuracy. The accuracy of this model is still increasing, so adding more muddy training transects will likely improve performance.

5.3 Multispectral, coastal and hybrid transect classification

The statistical analysis of the model performances, shows that the hybrid model has the highest accuracy in both the overall model accuracy when classifying five coastal types, as well as for the classification of muddy coasts. How do the three models perform in practice? Again, the example of Tenerife is analysed. In Figure 39 the same area of Tenerife, Spain, is shown as analysed before. The coastal transects are classified by each of the three models (multispectral (39a), coastal (39b) and hybrid (39c)). The multispectral transect model classifies the transects that cross the volcanic area as muddy coasts, because the volcanic material is classified as muddy pixels with the multispectral image classification model. The coastal transect model classifies the same transects as nearly all coastal cliffs, because (among others) the height profile showed a high maximum height (see Figure 31c). Two of the transects that cross the volcanic area are classified as plain beach. The hybrid transect model classifies all the transects crossing the volcanic area as coastal cliffs. Apparently the model 'knows' that this coast cannot be a muddy coasts because the information obtained by the coastal datasets do not match with the physical geographical characteristics of muddy coasts, even though the pixels along the transects are all recognised as mud. The two transects that were classified as

plain beach by the coastal model are classified as coastal cliff by the hybrid model. This can probably be explained by the fact that there are no sandy pixels recognised along the transect; the coastal model does not have this information. This example shows that the hybrid transect model also performs the best in practice, because it combines information of both 'worlds'.

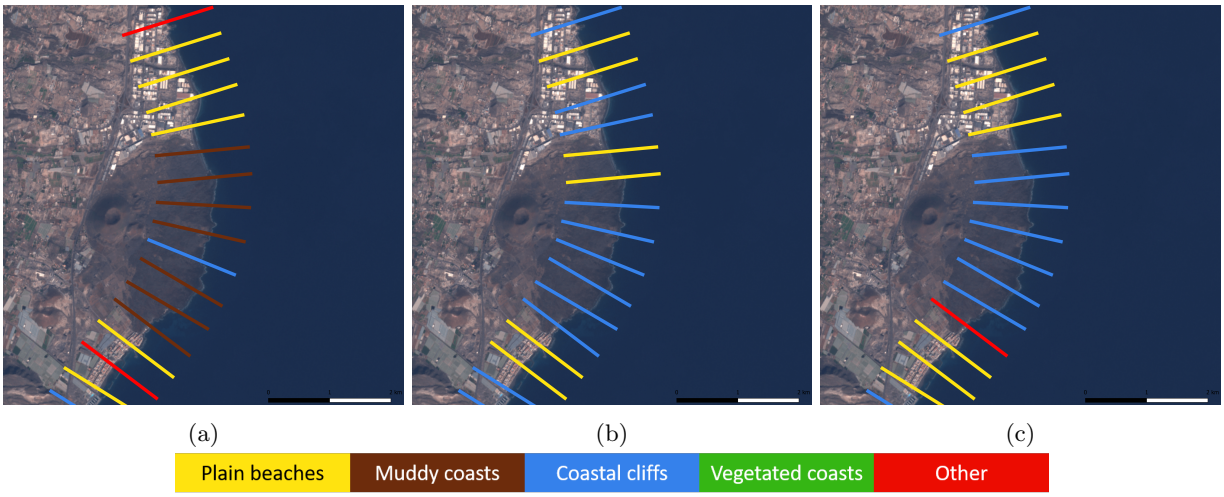


Figure 39: Example of a coastline consisting of volcanic material in Tenerife, Spain. The coastal transects are classified with the (a) multispectral, (b) coastal and (c) hybrid transect classification model.

5.4 Conclusions

The best performing model for the classification of the coastal type of each transect is the hybrid model. This global transect classification model is obtained by combining the pixel information provided by the multispectral image classification and the geographical information collected from the available global coastal datasets, for which a total of seventeen coastal features are used. The six most important coastal features are the occurrence of mud, water and sand pixels along the transect from the multispectral image classification; and the maximum height, height variance and width of transition zone of the transect from the coastal datasets. This shows that both pixel level input and larger scale input are important for detecting muddy coasts.

The most confusion among muddy coasts, according to the hybrid model, is with vegetated coasts. This is probably because many muddy coasts occur close to mangrove forests, where the mud banks are sometimes covered by these forests, sometimes not. Therefore, the geographical characteristics of these coasts are very similar, with the exception of (narrow) mud banks visible on multispectral images (the pixel characteristics). The overall accuracy of the hybrid model is the highest of the three models and reaches 76% when 1400 training transects are used to train the model. This means that out of 100 transects, 76 transects are assigned to the correct coastal type. Besides, the overall model accuracy plot (Figure 36) shows that the accuracy of the hybrid model is still increasing significantly, meaning that the model could perform even better if more labelled transects were added to train the model. The accuracy of the hybrid model per coastal type reaches 85% for muddy coasts and is the highest of the three models. This means that 85 out of 100 muddy coasts are correctly recognised by the hybrid classification model. Again, the accuracy still increases significantly, which means that by adding more training transects the model could perform even better. The model recognises the muddy coasts best of all five coastal types.

The example in section 5.3 shows that the combination of both the information from the multispectral image classification model and the information from the globally available coastal datasets resulted in the most reliable and robust transect classification. In addition, the example confirms that the hybrid model has improved on the multispectral model by adding the information from the coastal datasets.

The hybrid model, as trained and developed in this section, will be used to classify muddy coasts on a global scale. The global map with the occurrence of muddy coasts will be presented in the next section.

6 Regional and global mapping

The conclusion in section 5.4 tells that the hybrid transect classification model performs best in terms of overall model accuracy and best in terms of accuracy for the coastal type muddy coasts. Therefore, this model is applied to classify the global coastal transects into the five coastal types. The global coastline is divided into about 1.8 million transects with a spacing of $200m - 450m$. First, the model automatically constructs the cloud-free 1 year composite S2 TOA multispectral image of the $250m$ area around the transect of the year 2020. Second, this composite multispectral image is classified using the multispectral image classification model developed in section 3.1.3, where each pixel is labelled with the coastal class (sand, mud, water, turbid water, vegetation, dry vegetation, other) that best matches the observed spectral properties of the pixel. The third step consists of extracting the information from the multispectral image classification and the information from the globally available coastal datasets along the transect with a spacing of $10m$. The coastal variables, defined in section 2.4, are computed and saved for the transect. The coarser coastal datasets are interpolated in this step and the required variable is calculated and stored. After data collection, the hybrid transect classification model is applied. The model classifies the transect into the coastal type that best matches the observed characteristics of the transect.

The hybrid transect classification model will be applied on a global scale and the global distribution of muddy coasts will be presented in section 6.2. But before that, several regional cases are presented in section 6.1, to gain confidence in the hybrid transect classification model.

6.1 Regional mapping of coastal types

Six regional cases are presented that give an insight into how the hybrid transect classification model performs. Several predominantly muddy environments have been selected, showing various contrasts between muddy coasts and other coastal types. Figure 40 shows the locations of the six regions on the world map.

The first location is the Dutch Wadden Sea, which is a well-known example of a barrier island system, see Figure 41. The open coastline (northwest in the image) consists of sandy plain beaches, the inner sides of the islands and the coastline of the mainland consist mainly of muddy coasts. This example shows that the classifier is able to distinguish between sandy and muddy beaches, which are close together in the image but occur under different characteristics.

The next regional case is located in Bangladesh at the mouth of the Tetulia River, see Figure 42. This system shows the large amount of mud banks alternating with vegetated coasts (mainly mangrove forests). In addition, few sandy beaches are observed located primarily on the outer edges of the land facing the open sea. This example shows that the classifier can recognise mud banks next to vegetation (classified as muddy coasts) and coasts where the vegetation covers the mud banks from above (classified as vegetated coasts).

The area around Shanghai (China), where the Yangtze river flows into the ocean, is known for its vast mudflats. This example, see Figure 43, shows the contrast between the muddy coasts and the coasts where the city of Shanghai extends. The transects where the city extends to the coastline are classified as plain beaches. This can be explained by the fact that buildings with bright roofs are often recognised as sandy pixels by the multispectral image classifier and therefore the transects are classified as plain beaches rather than 'other'. However, the distinction between muddy coasts and the city is fairly well displayed.

On the South American continent, a regional case study is conducted on the coastline near Paramaribo (Suriname), see Figure 44. This area is part of the well-known extensive muddy Guyana coastline that has been discussed before in section 3.1.2. This case study shows that nearly the entire coastline consists of muddy beaches. Two exceptions are some transects located in the Paramaribo estuary and some transects located along the coastline where the mangrove forests have expanded seaward, causing the transects to not intersect the coastline transition (now classified as vegetated coasts).

The fifth example shows another area in the Netherlands; the Western Scheldt and the Eastern Scheldt, both of which are estuaries, see Figure 45. The Eastern Scheldt (upper estuary in the figure) is an estuary that can be closed by the Eastern Scheldt Storm Surge Barrier and contains mainly mudflats. The Western Scheldt (lower estuary in the figure) is the estuary of the Scheldt river that is open to the North Sea and shows alternating transects of muddy coasts and plain beaches. This is in line with ground-truth studies, (Van Der Wal et al., 2005) and (van der Wal and Herman, 2007), that have shown that the Wester Scheldt estuary consists of a mixture of sandy and muddy particles.

The last example shows a strongly eroding coast near Demak (Indonesia), see Figure 46. Due to land subsidence and relative sea level rise, large parts of the region have been completely flooded, which is clearly visible in the figure. The classified transects alternate between plain beaches, muddy coasts, vegetated coasts and other. The plain beaches are mainly located where bright buildings are recognised; muddy coasts are located north of the flooded land, where mud is still being supplied and conditions allow mud to settle; vegetated coasts are mainly found where the land is largely flooded and only some vegetation is left above the sea; other coasts are located where the transects consist entirely of water pixels (in the flooded region) or next to the city in the south of the image, where there are buildings or infrastructure.

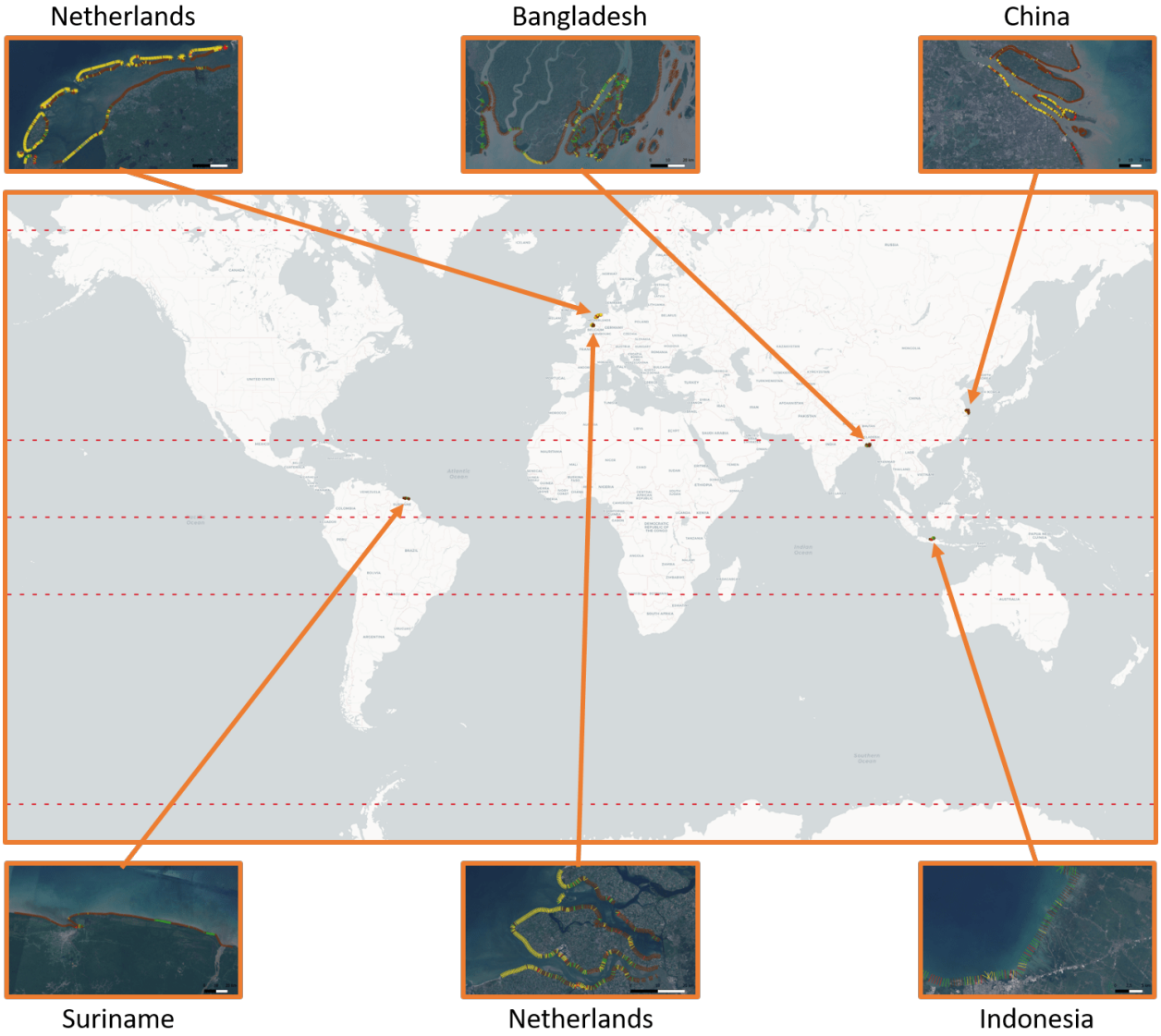


Figure 40: Locations of six regional cases of predominantly muddy environments, that are classified with the hybrid transect classification model.



Figure 41: Wadden Sea, Netherlands, example of a barrier island environment, where sandy beaches are found at coasts facing the open ocean and muddy coasts are found at coasts between the islands and the mainland. The transects on the coastline were classified using the hybrid transect classification model that classifies each transect into one of the five coastal types shown in the legend.

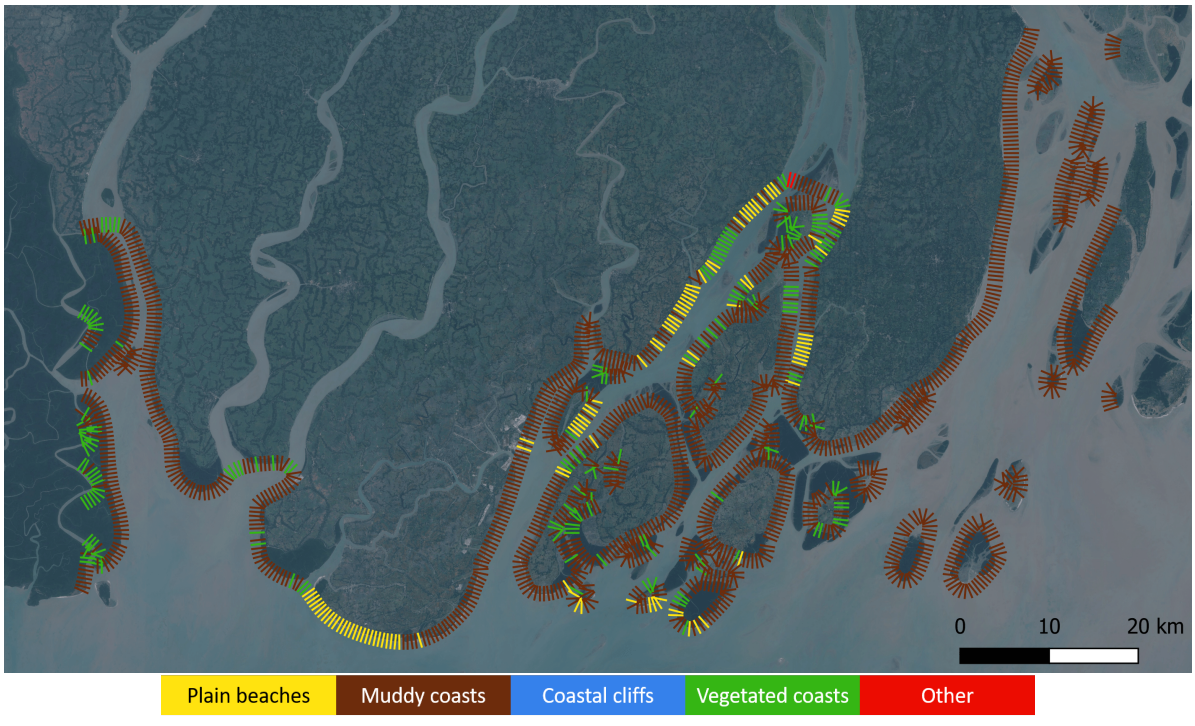


Figure 42: Tetulia river mouth, Bangladesh, example of a mangrove delta environment. The transects on the coastline were classified using the hybrid transect classification model that classifies each transect into one of the five coastal types shown in the legend.

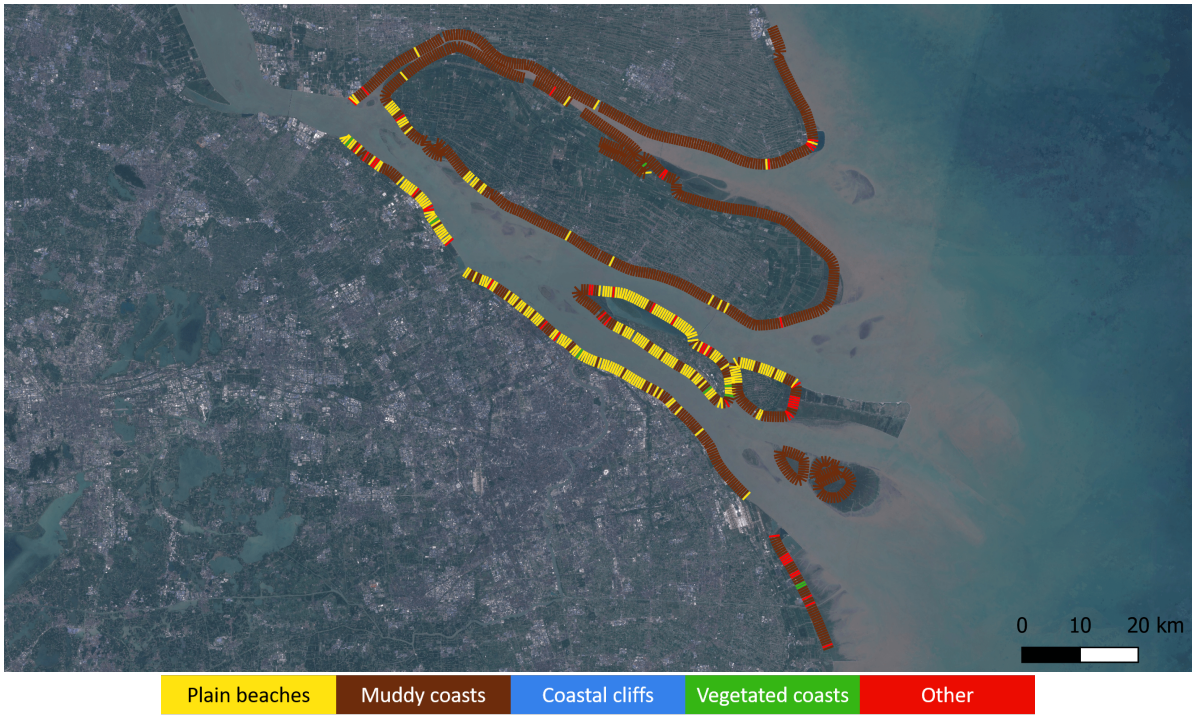


Figure 43: Shanghai, China, example of an environment with wide mudflats and a city extending to the coastline. The transects on the coastline were classified using the hybrid transect classification model that classifies each transect into one of the five coastal types shown in the legend.



Figure 44: Paramaribo, Suriname, example of an environment with wide mudflats. The transects on the coastline were classified using the hybrid transect classification model that classifies each transect into one of the five coastal types shown in the legend.

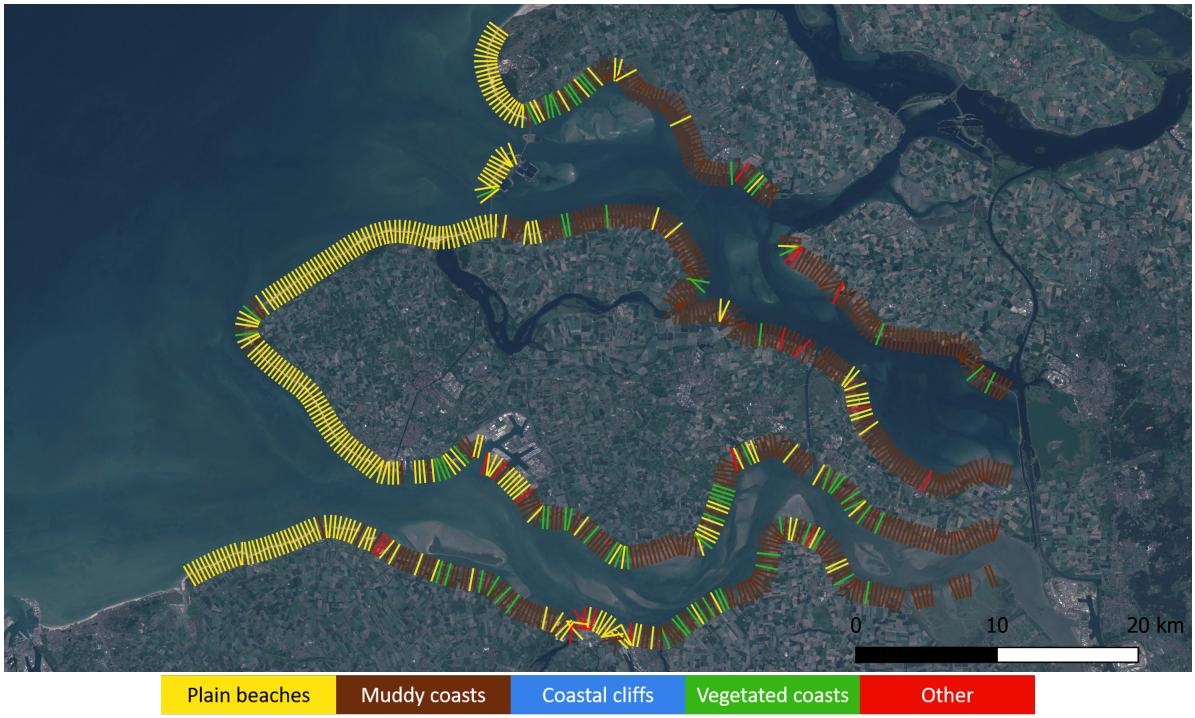


Figure 45: Western Scheldt and Eastern Scheldt, Netherlands, example of an estuary environment where muddy and sandy patches alternate. The transects on the coastline were classified using the hybrid transect classification model that classifies each transect into one of the five coastal types shown in the legend.

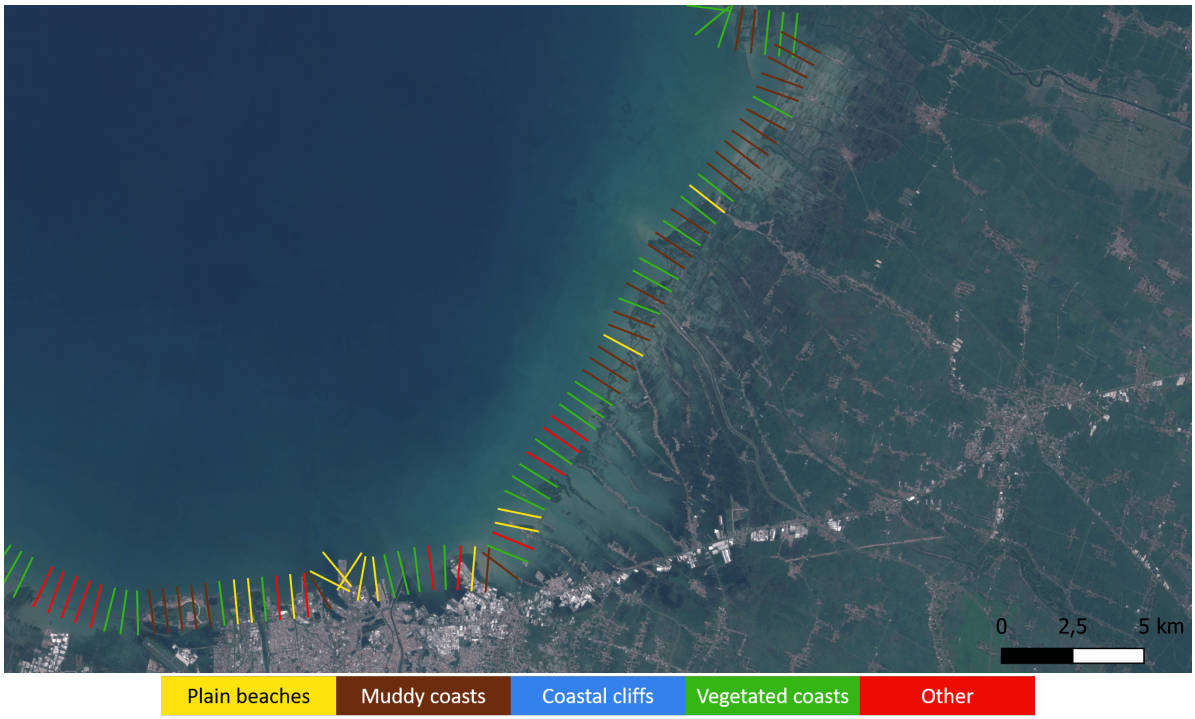


Figure 46: Demak, Java, Indonesia, example of an eroding coast where large areas of the land are flooded. The transects on the coastline were classified using the hybrid transect classification model that classifies each transect into one of the five coastal types shown in the legend.

6.2 Global distribution of muddy coasts

8

Having gained confidence at the regional scales in the performance of muddy coastline classification by the hybrid model, the global upscaling has been carried out. During this research, it was not possible in terms of time to classify all 1.8 million transects. However, currently the model is running and the transects are being classified globally. By now 100000 transects have been classified. These are randomly distributed on a global scale and already give a rough overview of the global distribution of the five coastal types.

Table 3 shows and compares the number of classified transects per coastal type for each transect model. The three models classify approximately a similar number of transects per coastal type; this is a measure of how robust and reliable all three models are. Of the 100000 transects classified, 12% are muddy coasts; 24% are plain beaches (according to Luijendijk et al. (2018), 31% of the world's ice-free coasts are sandy); coastal cliffs cover 27% of the world's coasts; 35% of transects are classified as vegetated coasts. The remaining 2% are classified as other. 7069 of the 11813 classified muddy transects are located between 25°N and 25°S, that is 60%. Flemming (2002) estimated that muddy coasts neighbouring mangrove systems cover an area of about 75% of the world's coastline around 25° above and below the equator.

	Multispectral model	Coastal model	Hybrid model	% of total classified transects
Plain beaches	25183	25871	**23832**	**24 %**
Muddy coasts	12689	13094	**11813**	**12 %**
Coastal cliffs	27855	29203	**27227**	**27 %**
Vegetated coasts	31433	29661	**35321**	**35 %**
Other	2840	2171	**1807**	**2 %**
Total	100000	100000	**100000**	**100 %**

Table 3: Overview of number of classified transects per coastal type for the three transect classification models (multispectral, coastal and hybrid), from a total of 100000 globally classified transects.

Figure 47 shows the global distribution of the transects classified as muddy coast by the hybrid model. This map provides a rough overview; some well-known muddy coasts stand out immediately. Guyana's coastline is known for its broad muddy sedimentary features and is clearly reflected on the global map. In Africa, two major river deltas where a lot of muddy sediments are deposited are also clearly visible on the map, namely the river delta that spans Gambia, Senegal and Guinea-Bissau; and the river delta in Nigeria and Cameroon. A large part of the eastern coastline of China is characterised by wide mudflats between the Yellow river and the Yangtze river, these mudflats are also very well depicted as muddy coasts. On closer inspection, the Wadden Sea in the Netherlands, Germany and Denmark are densely covered in muddy transects.

Coasts where muddy transects are not expected are often not classified as muddy coasts either. For example, the western coast of South America, which lies on a leading edge of two continental plates (red thin line in the figure), is known for its high-energy environment with steep coastal features (coastal cliffs) and a high mountain range relatively close to the coast; no conditions for mud to develop and deposit. On the world map, muddy transects are absent along this coast. Another example is the coast of the Mediterranean Sea and the coast of the Western Sahara, where there is a lack of large sediment supply by large rivers. Here too, muddy coasts are (nearly) not classified.

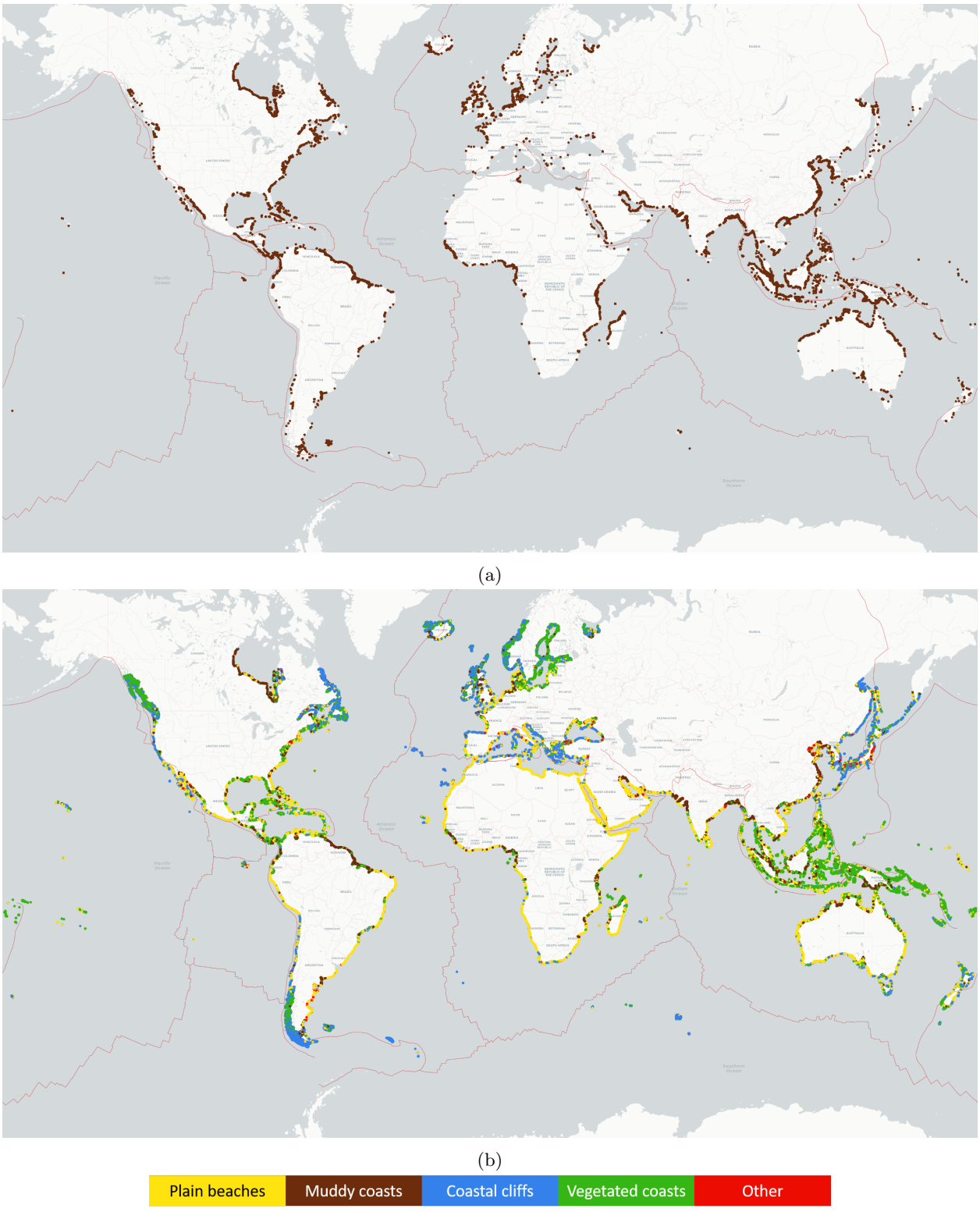


Figure 47: Global maps of 100000 classified transects. Top: muddy coasts (brown); bottom: all coastal types (legend). The horizontal dashed red lines represent, from top to bottom, the Arctic Circle (+66.5°), the Tropic of Cancer (+23.5°), the Equator (0°), the Tropic of Capricorn (−23.5°) and the Antarctic Circle (−66.5°). The thin red lines represent the edges of the tectonic plates.

6.2.1 Verification

With 100000 transects being classified as muddy coasts by the hybrid model, a verification can be done on the amount of muddy pixels that are often required to recognise a coast as a muddy coasts (by the hybrid model, but also by myself who labelled in first instance the muddy transects). Figure 48 confirms that on muddy coasts the most muddy pixels occur. On most muddy coasts only less than 5% of the pixels along the transect are muddy, however this implies that there is still a muddy beach of mudflat/bank present of $75m$ wide.

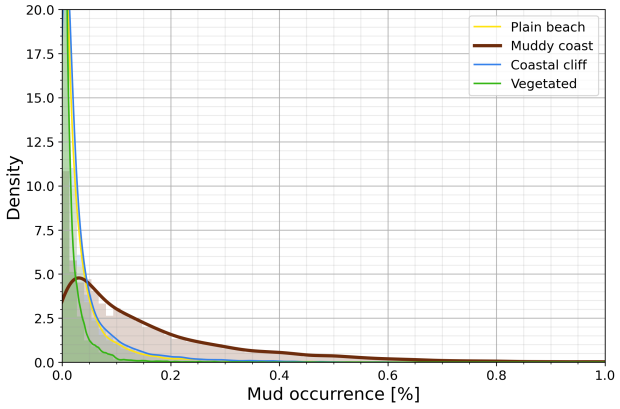


Figure 48: Distribution plot of the occurrence of muddy pixels based on 100000 classified transects (by the hybrid model) at global scale. The vertical shaded bars show the (true) densities, the solid lines show the estimated densities for the four coastal types: plain beaches, muddy coasts, coastal cliffs and vegetated coasts.

A final global distribution map of muddy coasts with the same formatting as Figure 1 from Luijendijk et al. (2018) is shown in Figure 49, where the relative occurrence of muddy coasts by latitude and longitude is plotted. Since not all 1.8 million coastal transects have been classified, the subplot on the right and the lower subplot might be a little distorted. The occurrence of muddy coastlines by latitude can be compared to the latitudinal distribution of muddy coastlines reported by Hayes (1967), see also Figure 6. Muddy coastline occur most often in the tropics, where Hayes (1967) reported a peak as well. However, the occurrence of muddy coastlines does not exactly correspond to the proposed distribution of Hayes (1967).

In Figure 50 the occurrence of three coastal types (muddy coasts, plain beaches and coastal cliffs) are plotted against latitude and compared to the latitudinal distribution of three sediment types (mud, sand, rock and gravel) according to Hayes (1967). The main pattern of global distribution is fairly well followed by the three of them.

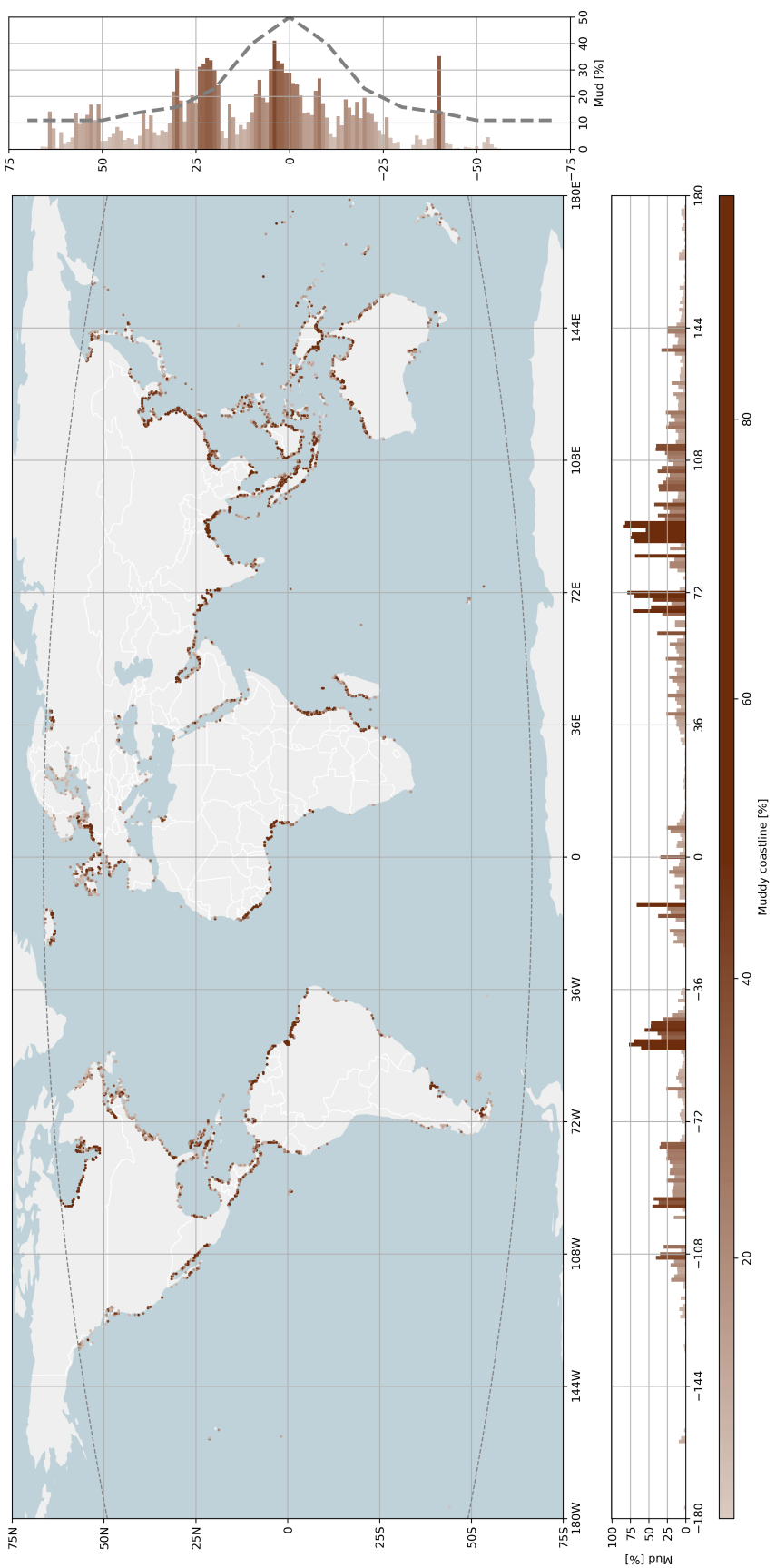


Figure 49: Global distribution of muddy coastlines based on 100000 classified transects by the hybrid model; the coloured dots along the world's coastline represent the local percentage of muddy coastlines (dark brown is mud, light is non-mud). The subplot on the right shows the relative occurrence of muddy coastlines by latitude, with the dashed line representing the latitudinal distribution of muddy coastlines reported by Hayes (1967). The lower subplot shows the relative occurrence of muddy coastlines by longitude. The curved grey dashed lines in the main plot indicate the boundaries of the ice-free coastlines considered in this analysis.

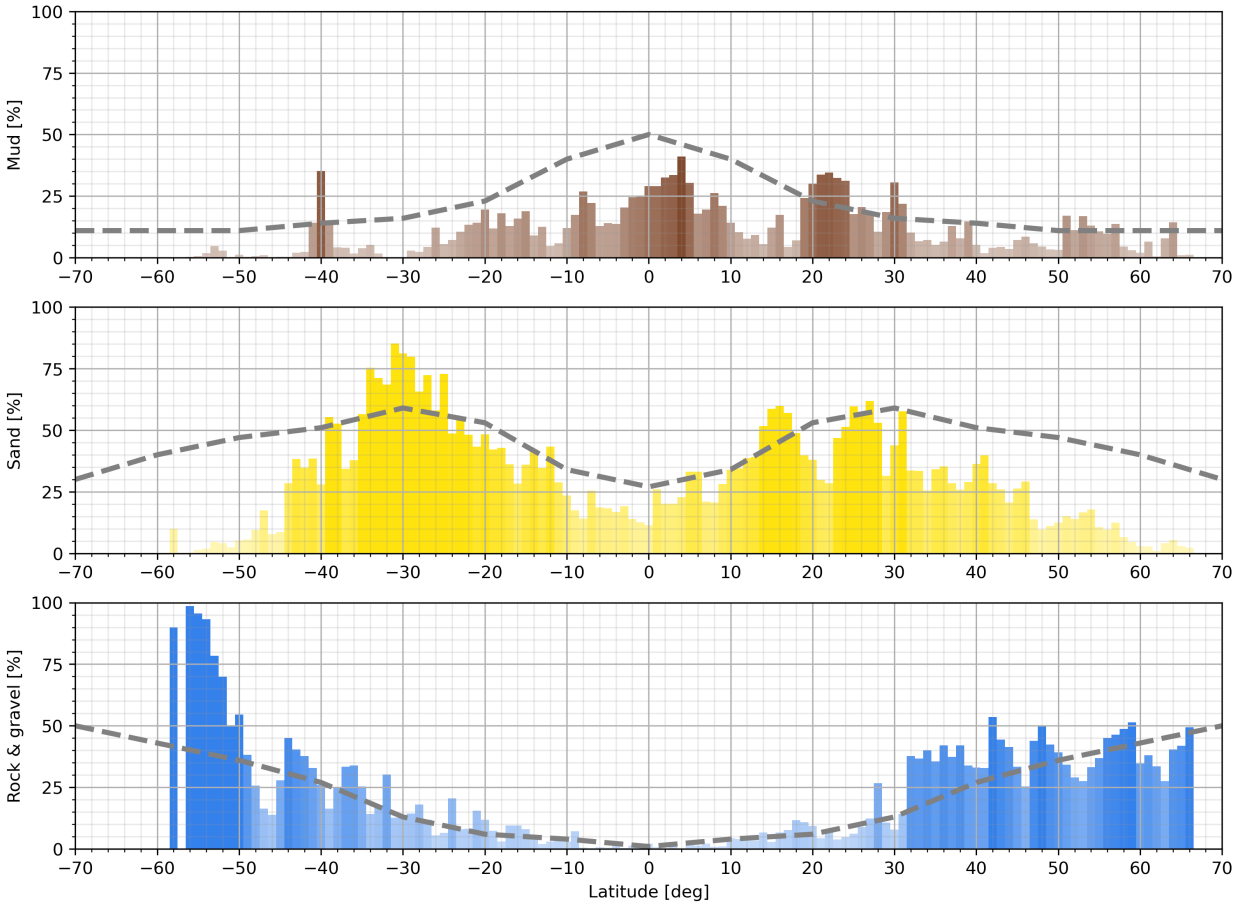


Figure 50: Global distribution of muddy coasts, plain beaches and coastal cliffs based on 100000 classified coastal transects by the hybrid model. The relative occurrence of the coastal types are plotted against latitude, with the dashed lines representing the latitudinal distribution of sediment types: mud (top), sand (middle), rock and gravel (bottom) reported by Hayes (1967).

7 Discussion

This section contains a critical reflection on a number of limitations, assumptions and approaches used in this study. In the first section (7.1), two possible limitations related to the use and processing of multispectral satellite images are mentioned. Next, the quality of the coastal datasets used, the potential of adding more coastal datasets and the approach to model development are discussed in section 7.2. This section end with a reflection of some of the assumptions made with regard to global upscaling, section 7.3.

7.1 Possible limitations

7.1.1 Top of Atmosphere vs. Surface Reflectance images

During this study cloud-free images of the S2 TOA collection are used. However, the SR images of S2 are also analysed to investigate their influence on the global multispectral image classification model. This dataset differs in that the images have been preprocessed by applying a filter to remove atmospheric disturbances. The main differences with the TOA images are in the aerosol band (B1) and the blue band (B2). It is expected that the multispectral image classification model trained with the SR images will perform better because the atmospheric disturbances have been filtered out; atmospheric measurements are not relevant to the spectral characteristics of the coastal earth surface that are investigated in this study. However, the global multispectral image classification model trained with the SR images did not show major deviations compared to the model trained with the TOA images. The importance of each spectral band for the development of the model, see Figure 51, shows that the the blue and aerosol band are the least important in the TOA model and the second least important in the SR model. The overall accuracy of the SR trained model is marginally lower, see Figure 52. The SR RGB images appear slightly brighter than the TOA RGB images and the SR classified images appear more pixelated than the TOA classified images. There is no convincing argument (yet) to change the use of the TOA images to that of the SR images. Therefore, the cloud-free S2 TOA images are used for all image analyses and for all classification models developed in this study.

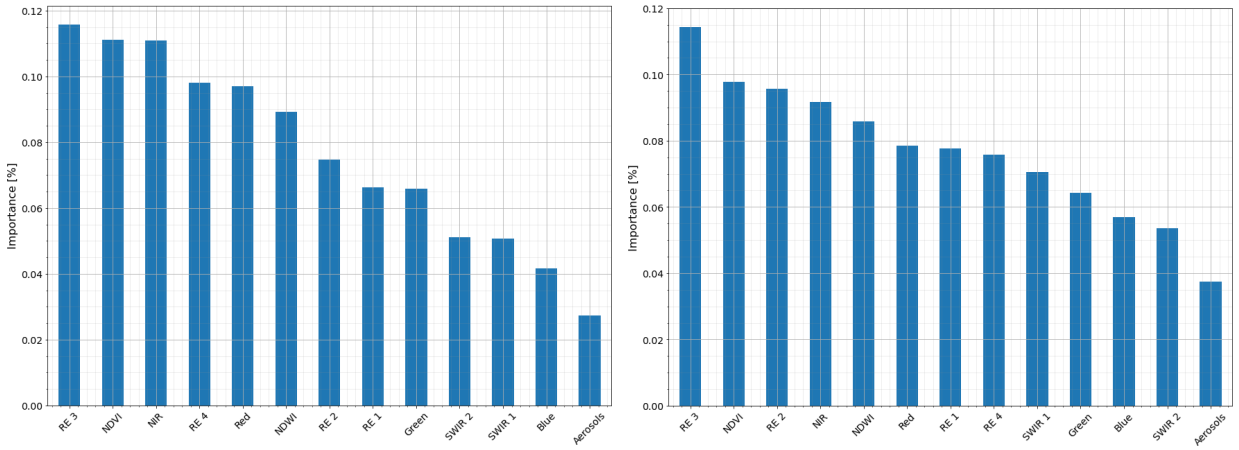


Figure 51: Importance of spectral bands, in descending order, used for developing the (a) TOA and (b) SR multispectral image classification model.

7.1.2 Individual vs. composite images

In the part of the study concerning the multispectral image classification, individual S2 TOA images are used, that capture the least cloudy moment per location during the period 01-01-2020 to 31-12-2020. The training points are labelled by inspecting the RGB images of these snapshots and the spectral reflectance properties of each pixel are extracted from these individual snapshots. It is assumed that the spectral reflectance properties obtained from these snapshots for pixel-based training represent the true characteristics for the

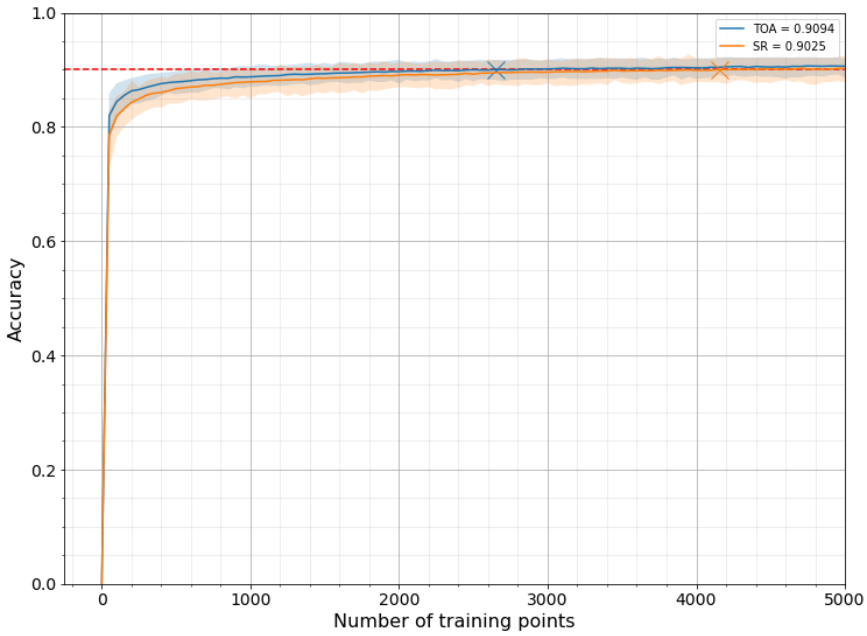


Figure 52: Overall model accuracy plot for two multispectral image classification models: (a) TOA; (b) SR that classify pixels into seven coastal classes. The number of used training points is on the x-axis and the overall model accuracy (Equation 4) on the y-axis. The solid line represents the mean accuracy, the shaded area represents the $90th$ percentile of the accuracy. The number behind the image classification models in the legend are the final accuracies when 5000 of the training points are used to train the model. The blue and the orange crosses show where the TOA model and the SR model exceed 90% accuracy.

coastal class they are assigned to and are a valid way of obtaining their spectral properties. However, for the coastal type classification, which is implemented at transect level, capturing one moment in time can give a distorted impression. For example, if an image is taken at high tide, the coastline may not contain a beach as it may be submerged, while at low tide the same coastline may contain a muddy beach and exposed mudflats that are now emerged. When 1-year composite images are used, which are obtained by stacking all cloud-free images from that year and calculating the average reflectance value of each pixel, the tidal (daily/monthly) and seasonal variations are filtered out. Therefore, it is assumed that 1-year composite images give no or a less distorted picture, suggesting a more robust, representative coastal type of that location.

7.2 Development of hybrid transect classification model

7.2.1 Quality of coastal datasets

In the development of the hybrid transect classification model seven globally available coastal datasets are implemented in addition to the information obtained by the multispectral image classifier. However, the variables of some coastal datasets are based on multispectral satellite images as well. For example, the Murray Global Intertidal Change Dataset (Murray et al., 2019) is trained with multispectral satellite images, from which pixels are allocated to the classes tidal flat, permanent water and other. The quality and added value of this dataset is debatable, as it is based on the same principle as the multispectral image classification model derived in this study. Therefore, it may be considered to use this dataset as a verification or reference of the global distribution of muddy coasts, rather than including it in the hybrid transect classification model. In Figure 53 the global distribution of tidal flats is shown. The green, blue and purple areas are the areas where tidal flats are observed and show great similarities with the global muddy coast distribution shown in Figure 47a. The latitudinal distribution of tidal flats show peaks at approximately the same latitudes as the peaks of the latitudinal distribution of muddy coasts (right plot in Figure 49).

Another example of a coastal dataset whose quality is questionable is the Global Surface Water dataset developed by Pekel et al. (2016). This coastal dataset has not been validated in terms of the implementation used in this study. Section 2.4.4 explains how the frequency with which water was present over a 32-year

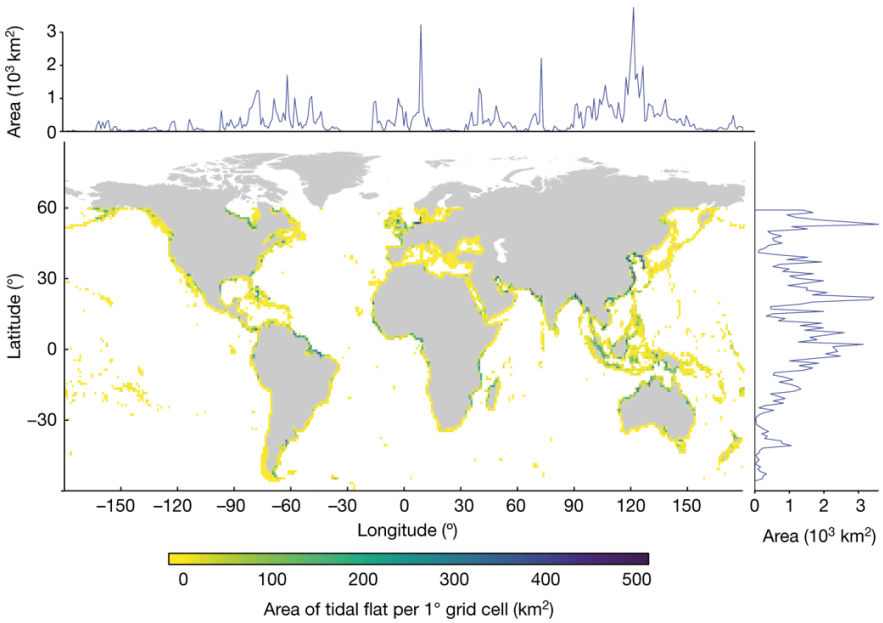


Figure 53: The global extent of tidal flats mapped as km^2 per 1° grid cell, 2014-2016. Summaries of latitudinal and longitudinal extents of tidal flats are shown above and to the right. From Murray et al. (2019).

period is interpreted as the intertidal area where dry and wet surfaces alternate. However, this dataset was developed to map long-term changes in global surface water occurrence and document multi-decadal trends, which provides insights into, among others, the impacts of climate change and trends in sea level rise. The dataset has been validated against these long-term statistics and not against the indication of intertidal areas. This might affect the results of the hybrid transect classification model, as it may contain incorrectly interpreted data. In Figure 54 the distribution of the width of the transition zone (derived from the GSW dataset) among the different coastal types is shown; muddy coasts clearly have the widest transition zones most often. However, quite often the coastal cliffs have fairly wide transition zones as well, which is not as expected. Therefore, the quality assessment of the coastal datasets should be validated based on how the data is implemented in this study, before it can be added to the global hybrid transect classification model that classifies muddy coasts.

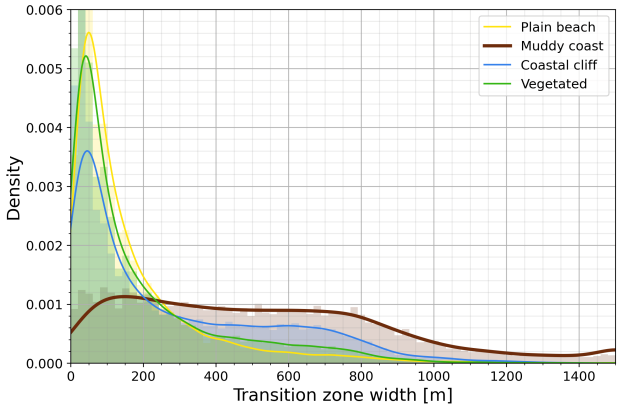


Figure 54: Distribution plot of the width of the transition zone based on 100000 classified transects (by the hybrid model) at global scale. The vertical shaded bars show the (true) densities, the solid lines show the estimated densities for the four coastal types: plain beaches, muddy coasts, coastal cliffs and vegetated coasts.

7.2.2 Other coastal datasets

In this study, a limited set of seven coastal datasets were used to investigate the physical geographical characteristics of muddy coasts on a global scale, which improved the results of the hybrid transect classification model. By adding more coastal datasets that have (had) an influence on the development and configuration of the coastline, the accuracy may become even higher. The most promising coastal datasets to investigate in further research are in my opinion: the beach face slope, which can provide information on the type or size of the sediment present on the beach; the wave conditions, which can indicate what type of energy climate (high/moderate/low) the coastline is exposed to, which is related to the type of sediment that can settle; the near shore slope, which can provide information on how the waves will propagate to the coastline and whether and what kind of sediment can settle.

7.2.3 Model development

For the three transect classification models (multispectral, coastal, hybrid), it is examined if the overall model accuracy increased when coastal variables are added to the classification model in order of importance. If the overall model accuracy did not show a decrease, the coastal variable is used in the final model. It turned out that all variables increase the accuracy of the model, so all variables were included. However, the trade-off between added value and extra computation time, extra data storage of overfitting is not considered. Moreover, the hybrid transect model now uses several variables that originate from the same datasets (e.g. maximum height and height variance from the MERIT DEM dataset). Consequently, these datasets have more weight in the development of the transect classification model and in the classification of the global coastlines. This can result in a biased classification (model). E.g. an exception in the elevation profile results in two exceptions in the variables that are included in (the development of) the model; while the number of muddy pixels along the transect results in only one variable considered. In this example, two of the three variables are then exceptions and a muddy coast may be erroneously not recognised as a muddy coast. Therefore, a classification model with weighted coastal variables could improve the classification, if the most explanatory (important) coastal variables were given more weight and if variables coming from one dataset were given one weight together.

7.3 Global upscaling

7.3.1 Size of training data

The size and distribution of training data affects the performance of a model. The multispectral image classification model uses a large number of training data (7500 training points), but the overall model accuracy already nearly stabilises when one third of the training points are used to train the model, see Figure 24. When looking at the coastal class accuracy plot (Figure 25), it is clear that the training points are unequally distributed among the coastal classes. For some coastal classes the accuracy still increases when more training points are added, while for others the accuracy is already (nearly) stabilised. This might have been a limiting factor for the performance of the multispectral image classification model. Therefore, it is recommended to use an equal distribution in terms of training size per coastal class and/or to increase the number of training points until a stable accuracy has been reached. The insight of equally distributed training data is applied to the development of the three transect classification models, see Figure 37. However, the model performance plots (Figures 36 and 37) still show increasing profiles for both the overall model accuracy as well as for the accuracy per coastal type. Therefore, it is recommended to use a larger set of labelled transects in order to higher the performance of the final hybrid transect model.

7.3.2 Definition and interpretation of coastal types

A muddy coast is defined in this study as a transect where a tidal flat/mud bank must be present at the coastline and must visible on the satellite image; this excludes muddy coasts that are covered (from above) by mangrove forests or other types of vegetation. The labelling of the transects is done with the best of intentions, but it is a personal interpretation of how the coasts look from above from 1-year composite S2 RGB satellite images with a spatial resolution of $10m$ and to which coastal type they belong. In addition, muddy coasts can occur in various conditions, revealing slightly different characteristics. Now these are all

grouped together, generalising the differences in types of muddy coasts. Muddy coasts are composed of different proportions of silt, clay and even sand particles. The same applies to sandy coasts, which may also contain silt or clay elements. This makes it difficult to draw a strict line as to whether a coast belongs to the muddy or the sandy coasts. Moreover, there are coasts that exhibit both characteristics and where both sandy and muddy pixels are recognised by the multispectral image classification model. To which coastal type should these coasts be manually labelled and automatically classified? One suggestion would be to have several persons label the coastal transects to remove biases of personal interpretation or to use all available information per transect (multispectral image information and coastal dataset information) rather than just the RGB images to label the transects.

7.3.3 Transect system

In this study, the coastlines are approached through a global transect system. This means that data along the entire transect is collected,aggregated to one single value per coastal variable and used to train the model and classify the coastline. However, for the coastline classification it is not (always) relevant what the land and the sea ends of the transect look like. Therefore, this transect system can be improved by analysing a smaller, but relevant, section around the coastline. However, the coastline positions of the year 2016 (from the OSM 2016 database) used in this study do not exactly coincide with the position of the coastlines of the year 2020 analysed in this study. The position of the coastline may have shifted, due to erosion, accretion or perhaps anthropegenic phenomena. Consequently, some transects may contain more land or more water information, which can bias the classification of the coastline, as the information along the entire transect is used. To solve this problem, an updated coastline position can be used, that is accurate for the year 2020. Or, instead of using one single value per coastal feature (one dimensional), the entire profile along the transect can be analysed, which contains even more information, to perform a two dimensional coast analysis.

7.3.4 Ground-truth validation

The ground truth validation in this research is based on literature studies. Real ground truth data were searched for, that contain grain size samples from specific beaches. However, since these datasets are difficult to find and the sample location and time are often impractical to compare with the information obtained by satellite images or coastal variables; validation of the samples at pixel level (ground truth level) is not worthwhile. Therefore, ground truth validation is performed based on literature studies. There are plenty of articles that describe stretches of coastline with (often) some detailed information on the sediments encountered, see Flemming (2002). However, since no ground truth measurements of particle sizes of beaches were used in this study, the classification model cannot determine whether the muddy coasts truly consist of mud particles smaller than $63 \mu m$ in diameter.

8 Conclusions

This research aimed at estimating a global distribution of muddy coasts; to monitor their coastal dynamic behaviour in the future and protect these valuable ecosystems. The main objective of this thesis is to develop an automated classification method that is applicable globally, that is able to distinguish muddy coasts from other coastal features and from other types of sediment found on the coast. A final goal of this thesis is to present a validated global occurrence of muddy coastlines.

In this section the final conclusions will be summarised. This will be done by firstly answering the sub-questions, as formulated in section 1.3. Finally a conclusion will be given that answers the main research question.

To what extent can mud be distinguished from other coastal features by the use of publicly and globally available multispectral satellite images? In section 3.1.3 a method is implemented that has developed an multispectral image classification model, which is able to classify seven coastal classes; namely sand, mud, water, turbid water, vegetation, dry vegetation and other. This model is trained with 7500 training data points (pixel level) randomly spread over the world, that are manually labelled by a visual inspection of individual cloud-free S2 TOA RGB images acquired during the period between 01-01-2020 and 31-12-2020. The obtained global multispectral image classification model has an overall model accuracy of more than 90%, meaning that 9 out of 10 pixels are correctly classified in the coastal class they belong to. The accuracy of correctly classifying muddy pixels reaches also more than 90%. This multispectral global image classification model is therefore considered to be able to distinguish mud patches from other coastal features based on satellite images acquired by S2.

How representative are locally trained classification methods when applied on a global scale? In section 3.1.2 a method is described that tests the representativeness of a locally trained multispectral image classifier when applied to another local site. This method shows that the multispectral locally trained image classifier (that is trained with the spectral properties of the coastal classes observed in the Dutch Wadden Sea) does not perform well at another location (the coast of Suriname in the example). The coastal classes have (slightly) different spectral signatures at different locations at the globe under different conditions. Therefore, the local muddy pixels from the Dutch Wadden Sea (used as first training location) are not sufficiently representative to be used as the only training data to develop a multispectral image classification model that can be applied at larger scale. This is understandable as beaches are composed of several sediments consisting of particles with different sizes and different origins, (Wentworth, 1922). Beaches may contain silt, clay and sand particles in a variety of proportions. There is no hard line in the natural composition of beaches, rather there is a fluent transition of different proportions of sediment that, taken together, appear and behave as a mud or sand beach. Consequently, muddy coasts at different locations may emit other spectral reflectance signals, which are not recognised as 'mud' by a locally trained multispectral image classifier.

What is the added value of adding globally available coastal datasets in distinguishing mud from other coastal features? Seven available coastal datasets are analysed in section 4, from which ten coastal variables are extracted. In section 5.2 and 5.3 three different transect classification models are compared. The first model (multispectral model) is trained with only the information obtained by multispectral image classification that analyses the spectral characteristics of mud at pixel level. The second model (coastal model) is trained with only the information obtained by the seven coastal datasets that analyses the geographical characteristics of mud at larger scale. The last model (hybrid model) is trained with both information obtained by multispectral image classification as well as by the seven coastal datasets. The overall model accuracy is compared for the three models, as well as the accuracy of the classification of muddy coasts only. It follows that the overall model accuracy of the coastal model is 6.5% higher than for the multispectral model and the accuracy the classification of muddy coasts of the coastal model is 4% higher than for the multispectral model. When adding the globally available coastal datasets to the multispectral model (= hybrid model) the overall model accuracy increases with 13% relative to the multispectral model and reaches 76%. The accuracy of the classification of muddy coasts increases with 7.5%

relative to the multispectral model and reaches 85%. Therefore, it can be concluded that the addition of globally available coastal datasets to the multispectral image classification improves the transect classification model significantly.

In this approach, a global coastal transect system is used to obtain the variables from the satellite images and the coastal datasets. In this study, a uniform value for each coastal variable used to train the transect classification model. However, more information is stored in the transects, namely the profiles of each coastal variable. These 'sub-transect' profiles could be used in further research to assess, for example, the width of the muddy beach or the configuration of the transition from land to water, derived from the order of occurrence of coastal classes (i.e. water-sand-vegetation or turbid-mud-vegetation)).

What is the global distribution of muddy coasts? The global distribution of muddy coasts is showed in Figure 49. This map shows 11813 classified muddy transects (in brown) from the 100000 classified transects in total. Approximately 12% of the classified transects are muddy coasts, of which 60% occur in the tropics. The latitudinal distribution of classified muddy coasts follows fairly well the latitudinal distribution pattern proposed by Hayes (1967) in the right plot of the figure.

How can a reliable global mud classification be obtained by analysing the characteristics of muddy coasts using publicly available satellite remote sensing techniques and globally available coastal datasets? In this research a reliable global mud classification is obtained by analysing the spectral reflectance properties of muddy patches with multispectral satellite images, and by analysing the physical geographical characteristics of muddy coasts as a system with globally available coastal datasets obtained from satellite remote sensing techniques, ground measurements and global models.

The spectral reflectance properties of muddy patches are determined by analysing the spectral reflectance signals from 7500 randomly selected training points at global scale. The training points were defined by $10x10m$ pixels acquired from cloud-free individual S2 TOA multispectral images. The multispectral image classification model trained with the reflectance information of seven coastal classes showed high performance (over 90% accuracy) in the classification of coastal classes at pixel level. The physical geographical characteristics of muddy coasts as a system are determined by analysing the information from the global coastal datasets of 1868 randomly selected coastal training transects at global scale. The training transects measure $1500m$ in length perpendicular to the coastline and contain information every $10m$ along the transect. Each coastal variable exhibits profiles that can be related to the geographical characteristics of muddy coastal systems.

Using the information from the multispectral image classification and the globally available coastal datasets, a hybrid transect classification model was developed that was trained with 1868 training transects selected randomly around the globe. This hybrid model showed high performance (of 85% accuracy) in the classification of muddy coasts at transect level. In addition to the classification of muddy coasts, the model is also able to distinguish plain beaches, coastal cliffs and vegetated coasts, which is an unintended but beneficial side effect.

The large amount of training and validation data at both pixel and transect level results in a reliable and robust mud classification model. With the addition of the coastal datasets, the model performs better than when only the multispectral satellite images are used for the classification of muddy coasts.

9 Recommendations

In this section, two important recommendations are mentioned that will most likely be implemented (by myself) in a future paper publication regarding possible improvements and clarification of the model (development). Both recommendations are concrete and feasible.

9.1 Multispectral image classification

Primarily based on the possible limitations related to the use and processing of the multispectral satellite images and related to the size and distribution of the training data, mentioned in the discussion of section 7.1.1, 7.1.2 and 7.3.1, the multispectral image classification model can be improved and clarified by using 1 year composite S2 SR images and use an equally distributed training dataset. The multispectral image classification model might be improved with three adjustments. Notwithstanding the fact that the SR images do not necessarily perform better than the TOA images (but also not worse), I think that the use of the SR images is more useful and justified because the atmospheric disturbances, which have no added value for the classification of surfaces such as mud coasts, have already been filtered out. The use of one year composite images is recommended when the training points for the multispectral image classifier are labelled. This is because the hybrid transect model uses one year composite images as well, so by using the same input settings for manually labelling the training points, as output settings for classifying the pixels of an image into coastal classes, the most accurate outcome will be created, containing the least errors. The third improvement for the multispectral image classifier will be to use an equally distributed training dataset among the different coastal classes, for which the accuracy per coastal class has (nearly) stabilised. This will provide a more reliable classification, because some classes are now defined by (very) few training points for which the accuracy did not stabilise and converge, and therefore could contain biases.

9.2 Global transect system

In this study, the coastal transect system that is used stems from the OSM 2016 database, while the satellite images that are used are originate from 2020. This causes that the coastal transects do not cross the coastline exactly in the middle of the transect due to the shifting of the coastline (erosion/accretion), therefore some transects may contain more land and others may contain more water, which can bias the classification (model). An example of a shifted coastline can be found in Figure 44, where the green transects (vegetated coasts) are transects that do not reach the waterline anymore due to accretion of mangroves. This will have influence on the global distribution of muddy coasts. An easy and feasible solution is to update the OSM database to a version from 2020 or to use satellite images from 2016 so that the middle of the transects coincide (more accurate) with the exact position of the coastline. Subsequently the vegetated transects in the example mentioned above, will probably be recognised as muddy transects. Another beneficial effect is that when this is realised, a smaller and more relevant area around the coastline can be analysed, which may improve the coastal classification, as the irrelevant features further away from the area of interest are now excluded from the analysis (such as cities, islands in front of the coast, muddy or sandy features land inwards, etc.). For further research it is then recommended to focus on two dimensional classification by using spatial information instead of relying in the linear cross-section.

10 Technical details

10.1 Cloud-free images

For getting the (most) cloud-free images, a function from Google Earth Engine is used (Gorelick et al., 2017), that filters the selected images on their cloudiness. The CLOUDY_PIXEL_PERCENTAGE is an image property provided by the S2 MSI defined as: the granule-specific cloudy pixel percentage taken from the original metadata. A limit is set on 15%, which means that the image tile, in which the area of interest is located, contains less than 15% cloudy pixels. For the individual images, the image collection of the year 2020 is then sorted on the least CLOUDY_PIXEL_PERCENTAGE, the image tile with the least amount of cloudy pixels is used for further processing in this study. For the composite images, all the image tiles that have than 15% cloudy pixels are used to make the composite image.

10.2 Tidal stages

The tidal stage information that is used in the local multispectral image classification model, is taken from *Rijkswaterstaat*, which is an operational agency of the Ministry of Infrastructure and Water Management in the Netherlands. Every 10 minutes the water level at several locations in the Netherlands is measured relative to NAP (Normaal Amsterdams Peil). In this research the water level measurements are taken from Vlieland Haven, which is located near to the area of interest for which is it used in this study. Four least cloudy satellite images were selected from the year 2020. From those images the acquisition time was obtained, which then can be matched to the water level measurements of *Rijkswaterstaat*.

10.3 Selection of boxes/points/transects

10.3.1 Coastal boxes

In the global multispectral image classification model a total of 75 coastal boxes are selected (section 3.1.3), from which the training points are collected. The boxes are selected by applying the random selection function in QGIS, that selects a random subset of a given vector layer. From the 24009 coastal boxes a random subset of 100 boxes was created, which were inspected and categorised on their appearance: into boxes that contain sand, mud, mainly vegetation and mainly water. Because there were not enough boxes that contained mud a new random subset of 100 boxes was created, which were inspected and categorised in the same way. Until 20 boxes were found that contained mud, 25 that contained sand, 25 that contained mainly vegetation and 25 that contained mainly water.

10.3.2 Training points

Training points were selected at different stages in this research. Firstly, training points were selected from one image for the local image classifier (section 3.1.1). This selection was manual and tried to capture the four coastal classes (sand, mud, water, vegetation). The training points that were selected for the cross image classifier (section 3.1.2), was also manual and tried to capture five coastal classes (sand, mud, water, vegetation, other). The selection of globally spread training points for the global image classifier (section 3.1.3), was done by applying the random extraction of points inside the borders of a given vector layer, a function in QGIS. This function extracted 100 random points per coastal box, that where then labelled manually into seven coastal classes (sand, mud, water, turbid water, vegetation, dry vegetation, other). In this way a total of $100 * 75 = 7500$ training points were manually labelled. The random selection of training points resulted in an unequal distribution of points among the seven coastal classes.

10.3.3 Training transects

For the transect classification models 1868 transects at global scale were selected. This was done by first selecting a random subset of coastal boxes with the SAMPLE function in *Python*, within each coastal box 5 transects are selected by applying the same SAMPLE function to the transects that are located withing that coastal box. These transects were inspected manually and labelled into the five coastal types. In total 3114 transects were labelled, from which 442 were assigned to the muddy coasts. For the training transects

an equal distribution among the coastal types is followed, therefore the same amount of transects (442) that cross the plain beaches, coastal cliffs and vegetated coasts were randomly selected in *Python* with the SAMPLE function from the 3114 labelled transects. 100 transects were selected for the coastal type 'other'.

10.4 Random Forest classification

10.4.1 Smile Random Forest (GEE)

The multispectral image classification models use a SMILERANDOMFOREST (Breiman, 2001) classifier that is available through GEE. The random forest has 15 decision trees and is trained with the 7500 training points from which the spectral reflectance values of the bands are extracted, as well as their coastal class label. The output mode is set to CLASSIFICATION that assign the class number to each pixel that is classified by the model.

10.4.2 Random Forest (Python)

The transect classification models use the RANDOMFORESTCLASSIFIER that is available through *sklearn*[8] which is a free software machine learning library for Python. This random forest classifier does the same as the random forest classifier that is available through GEE, however the advantage of building a random forest in Python is that it is easy to implement and that you can play around with some variables. The random forest has 15 decision trees and is trained with the 1868 training transects from which the coastal variables are extracted, as well as their coastal type label. The classifier assigns the coastal type number to each transect that is classified by the model.

10.5 Composite image

The transect classification models use the 1-year composite satellite image around the transects. This image is composed of all the images from the year 2020 that contain less than 15% cloudy pixels. From these images the mean spectral reflectance values for each band are computed in the percentile range. The percentile range is defined between the $14th$ and $15th$ percentiles. This is done with the function in *Python*: EE.REDUCER.INTERVALMEAN(14,15) that is applied to the image collection of the year 2020.

10.6 Buffered-nearest-neighbour-average

Three variables of the coastal datasets need to be interpolated. This interpolation is done in the following way: a buffer is drawn around the centre of a transect, which is enlarged until at least one data point of the coastal dataset is observed within the buffer; finally, the average of the observed tidal data point(s) within the buffer is calculated and is assigned to the transect. The size of the buffer is increasing from $10 - 100 - 1000 - 5000 - 10000 - 25000 - 50000$ to $75000m$ around the centre point of the transect. For each buffersize it is checked whether it overlaps with the coarse coastal dataset. If at least one pixel overlaps that buffersize is used. Subsequently, from all points that overlap, the values of the coastal dataset are extracted from which the average is calculated and used as input data for the transect. If the buffer does not contain any values of the dataset, the buffersize is increased.

[8] https://scikit-learn.org/stable/index.html

Committee

The thesis committee consist of:

Prof. Dr. Ir. S.G.J. Aarninkhof,	TU Delft
Dr. Ir. A.P. Luijendijk,	TU Delft / Deltares
Dr. Ir. A. Moreno-Rodenas,	Deltares
Dr. Ir. S.L.M. Lhermitte,	TU Delft

References

E. J. Anthony, A. Gardel, and N. Gratiot. Fluvial sediment supply, mud banks, cheniers and the morphodynamics of the coast of South America between the Amazon and Orinoco river mouths. Geological Society, London, Special Publications, 388(1):533 LP – 560, 1 2014. doi: 10.1144/SP388.8. URL http://sp.lyellcollection.org/content/388/1/533.abstract.

P. G. Augustinus. Actual development of the chenier coast of suriname (South America). Sedimentary Geology, 26(1-3):91–113, 1980. ISSN 00370738. doi: 10.1016/0037-0738(80)90007-X.

J. Bosboom and M. J. F. Stive. Coastal dynamics I. Delft Academic Press, Delft, 2015.

L. Breiman. Random Forests. Machine Learning, 45(1):5–32, 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

J. L. Davies. A morphogenic approach to world shorelines. Zeitschrift für Geomorphologie, 8(5):127–142, 1964. doi: 10.1127/zfg/mortensen/8/1964/127. URL http://dx.doi.org/10.1127/zfg/mortensen/8/1964/127.

J. de Vries, B. van Maanen, G. Ruessink, P. A. Verweij, and S. M. de Jong. Unmixing water and mud: Characterizing diffuse boundaries of subtidal mud banks from individual satellite observations. International Journal of Applied Earth Observation and Geoinformation, 95(November 2020):102252, 2021. ISSN 03032434. doi: 10.1016/j.jag.2020.102252. URL https://doi.org/10.1016/j.jag.2020.102252.

B. W. Flemming. Geographic distribution of muddy coasts, 2002.

J. M. Froidefond, F. Lahet, C. Hu, D. Doxaran, D. Guiral, M. T. Prost, and J. F. Ternon. Mudflats and mud suspension observed from satellite data in French Guiana. Marine Geology, 208(2-4):153–168, 2004. ISSN 00253227. doi: 10.1016/j.margeo.2004.04.025.

C. Giri, E. Ochieng, L. L. Tieszen, Z. Zhu, A. Singh, T. Loveland, J. Masek, and N. Duke. Global Mangrove Forests Distribution, 2000, 2005. URL https://doi.org/10.7927/H4J67DW8.

C. Giri, E. Ochieng, L. L. Tieszen, Z. Zhu, A. Singh, T. Loveland, J. Masek, and N. Duke. Status and distribution of mangrove forests of the world using earth observation satellite data. Global Ecology and Biogeography, 20(1):154–159, 2011. doi: https://doi.org/10.1111/j.1466-8238.2010.00584.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1466-8238.2010.00584.x.

N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202:18–27, 2017. ISSN 00344257. doi: 10.1016/j.rse.2017.06.031. URL https://doi.org/10.1016/j.rse.2017.06.031.

M. O. Hayes. Relationship between coastal climate and bottom sediment type on the inner continental shelf. Marine Geology, 1967. ISSN 00253227. doi: 10.1016/0025-3227(67)90074-6.

T. R. Healy, Y. Wang, and J.-A. Healy. Muddy coasts of the world. Elsevier, Amsterdam, 2002.

R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25(15):1965–1978, 2005. ISSN 08998418. doi: 10.1002/joc.1276.

D. L. Inman and C. E. Nordstrom. On the Tectonic and Morphologic Classification of Coasts. The Journal of Geology, 79(1):1–21, 1971. doi: 10.1086/627583. URL https://doi.org/10.1086/627583.

K. Kathiresan and B. L. B. T. A. i. M. B. Bingham. Biology of mangroves and mangrove Ecosystems. volume 40, pages 81–251. Academic Press, 2001. ISBN 0065-2881. doi: https://doi.org/10.1016/S0065-2881(01)40003-4. URL https://www.sciencedirect.com/science/article/pii/S0065288101400034.

N. Li, Y. Yamazaki, V. Roeber, K. F. Cheung, and G. Chock. Probabilistic mapping of storm-induced coastal inundation for climate change adaptation. Coastal Engineering, 133(January):126–141, 2018. ISSN 03783839. doi: 10.1016/j.coastaleng.2017.12.013. URL https://doi.org/10.1016/j.coastaleng.2017.12.013.

H. Liu and K. C. Jezek. Automated extraction of coastline from satellite imagery by integrating Canny edge detection and locally adaptive thresholding methods. International Journal of Remote Sensing, 25 (5):937–958, 2004. ISSN 01431161. doi: 10.1080/0143116031000139890.

A. Luijendijk, G. Hagenaars, R. Ranasinghe, F. Baart, G. Donchyts, and S. Aarninkhof. The State of the World's Beaches. Scientific Reports, 8(1), 12 2018. ISSN 20452322. doi: 10.1038/s41598-018-24630-6.

S. K. McFeeters. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, 17(7):1425–1432, 1996. ISSN 13665901. doi: 10.1080/01431169608948714.

L. Mentaschi, M. I. Vousdoukas, J. F. Pekel, E. Voukouvalas, and L. Feyen. Global long-term observations of coastal erosion and accretion. Scientific Reports, 8(1):1–11, 2018. ISSN 20452322. doi: 10.1038/s41598-018-30904-w.

N. J. Murray, R. S. Clemens, S. R. Phinn, H. P. Possingham, and R. A. Fuller. Tracking the rapid loss of tidal wetlands in the Yellow Sea. Frontiers in Ecology and the Environment, 12(5):267–272, 2014. ISSN 15409309. doi: 10.1890/130260.

N. J. Murray, S. R. Phinn, M. DeWitt, R. Ferrari, R. Johnston, M. B. Lyons, N. Clinton, D. Thau, and R. A. Fuller. The global distribution and trajectory of tidal flats. Nature, 565(7738):222–225, 2019. ISSN 14764687. doi: 10.1038/s41586-018-0805-8. URL http://dx.doi.org/10.1038/s41586-018-0805-8.

H. J. Nasiha, P. Shanmugam, and R. Sundaravadivelu. Estimation of sediment settling velocity in estuarine and coastal waters using optical remote sensing data. Advances in Space Research, 63(11):3473–3488, 2019. ISSN 18791948. doi: 10.1016/j.asr.2019.02.023. URL https://doi.org/10.1016/j.asr.2019.02.023.

B. Neumann, A. T. Vafeidis, J. Zimmermann, and R. J. Nicholls. Future coastal population growth and exposure to sea-level rise and coastal flooding - A global assessment. PLoS ONE, 10(3), 2015. ISSN 19326203. doi: 10.1371/journal.pone.0118571.

S. Orseau, N. Abascal Zorrilla, N. Huybrechts, S. Lesourd, and A. Gardel. Decadal-scale morphological evolution of a muddy open coast. Marine Geology, 420(November 2019):106048, 2020. ISSN 00253227. doi: 10.1016/j.margeo.2019.106048. URL https://doi.org/10.1016/j.margeo.2019.106048.

J. F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward. High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633):418–422, 2016. ISSN 14764687. doi: 10.1038/nature20584.

H. Postma. Hydrography of the Wadden Sea : movements and properties of water and particulate matter. Ecology of the Wadden Sea, 1, 1982. URL http://ci.nii.ac.jp/naid/10029457474/en/.

J. H. Ryu, J. S. Won, and K. D. Min. Waterline extraction from Landsat TM data in a tidal flat a case study in Gomso Bay, Korea. Remote Sensing of Environment, 83(3):442–456, 2002. ISSN 00344257. doi: 10.1016/S0034-4257(02)00059-7.

E. Salameh, F. Frappart, R. Almar, P. Baptista, G. Heygster, B. Lubac, D. Raucoules, L. P. Almeida, E. W. Bergsma, S. Capo, M. D. De Michele, D. Idier, Z. Li, V. Marieu, A. Poupardin, P. A. Silva, I. Turki, and B. Laignel. Monitoring Beach Topography and Nearshore Bathymetry Using Spaceborne Remote Sensing: A Review. Remote Sensing, 11(19), 2019. ISSN 20724292. doi: 10.3390/rs11192212.

M. Simard, L. Fatoyinbo, C. Smetanka, V. H. Rivera-Monroy, E. Castañeda-Moya, N. Thomas, and T. Van der Stocken. Mangrove canopy height globally related to precipitation, temperature and cyclone frequency. Nature Geoscience, 12(1):40–45, 2019. ISSN 17520908. doi: 10.1038/s41561-018-0279-1. URL http://dx.doi.org/10.1038/s41561-018-0279-1.

C. Small and R. J. Nicholls. A global analysis of human settlement in coastal zones. Journal of Coastal Research, 19(3):584–599, 2003. ISSN 07490208. doi: 10.2307/4299200.

D. van der Wal and P. M. Herman. Regression-based synergy of optical, shortwave infrared and microwave remote sensing for monitoring the grain-size of intertidal sediments. Remote Sensing of Environment, 111 (1):89–106, 2007. ISSN 00344257. doi: 10.1016/j.rse.2007.03.019.

D. Van Der Wal, P. M. Herman, and A. Wielemaker-Van Den Dool. Characterisation of surface roughness and sediment texture of intertidal flats using ERS SAR imagery. Remote Sensing of Environment, 98(1): 96–109, 2005. ISSN 00344257. doi: 10.1016/j.rse.2005.06.004.

K. Vos, K. D. Splinter, M. D. Harley, J. A. Simmons, and I. L. Turner. CoastSat: A Google Earth Engine-enabled Python toolkit to extract shorelines from publicly available satellite imagery. Environmental Modelling and Software, 122:104528, 2019. ISSN 13648152. doi: 10.1016/j.envsoft.2019.104528. URL https://doi.org/10.1016/j.envsoft.2019.104528.

K. Vos, M. D. Harley, K. D. Splinter, A. Walker, and I. L. Turner. Beach Slopes From Satellite-Derived Shorelines. Geophysical Research Letters, 47(14), 2020. ISSN 19448007. doi: 10.1029/2020GL088365.

Y. Wang, T. Healy, and Members of SCOR Working Group 106. Definition, properties, and classification of muddy coasts. Muddy Coasts of the World: Processes, Deposits and Function, (June 2014):9–18, 2002.

Y. Wang, Y. Liu, S. Jin, C. Sun, and X. Wei. Evolution of the topography of tidal flats and sandbanks along the Jiangsu coast from 1973 to 2016 observed from satellites. ISPRS Journal of Photogrammetry and Remote Sensing, 150(June 2018):27–43, 2019. ISSN 09242716. doi: 10.1016/j.isprsjprs.2019.02.001. URL https://doi.org/10.1016/j.isprsjprs.2019.02.001.

C. K. Wentworth. A Scale of Grade and Class Terms for Clastic Sediments. The Journal of Geology, 30(5): 377–392, 1922. ISSN 0022-1376. doi: 10.1086/622910.

W. J. Wolff. Ecology of the Wadden Sea: Research in the past and challenges for the future. Journal of Sea Research, 82:3–9, 2013. ISSN 13851101. doi: 10.1016/j.seares.2013.03.006. URL http://dx.doi.org/10.1016/j.seares.2013.03.006.

D. Yamazaki, D. Ikeshima, R. Tawatari, T. Yamaguchi, F. O'Loughlin, J. C. Neal, C. C. Sampson, S. Kanae, and P. D. Bates. A high-accuracy map of global terrain elevations. Geophysical Research Letters, 44(11): 5844–5853, 2017. ISSN 19448007. doi: 10.1002/2017GL072874.