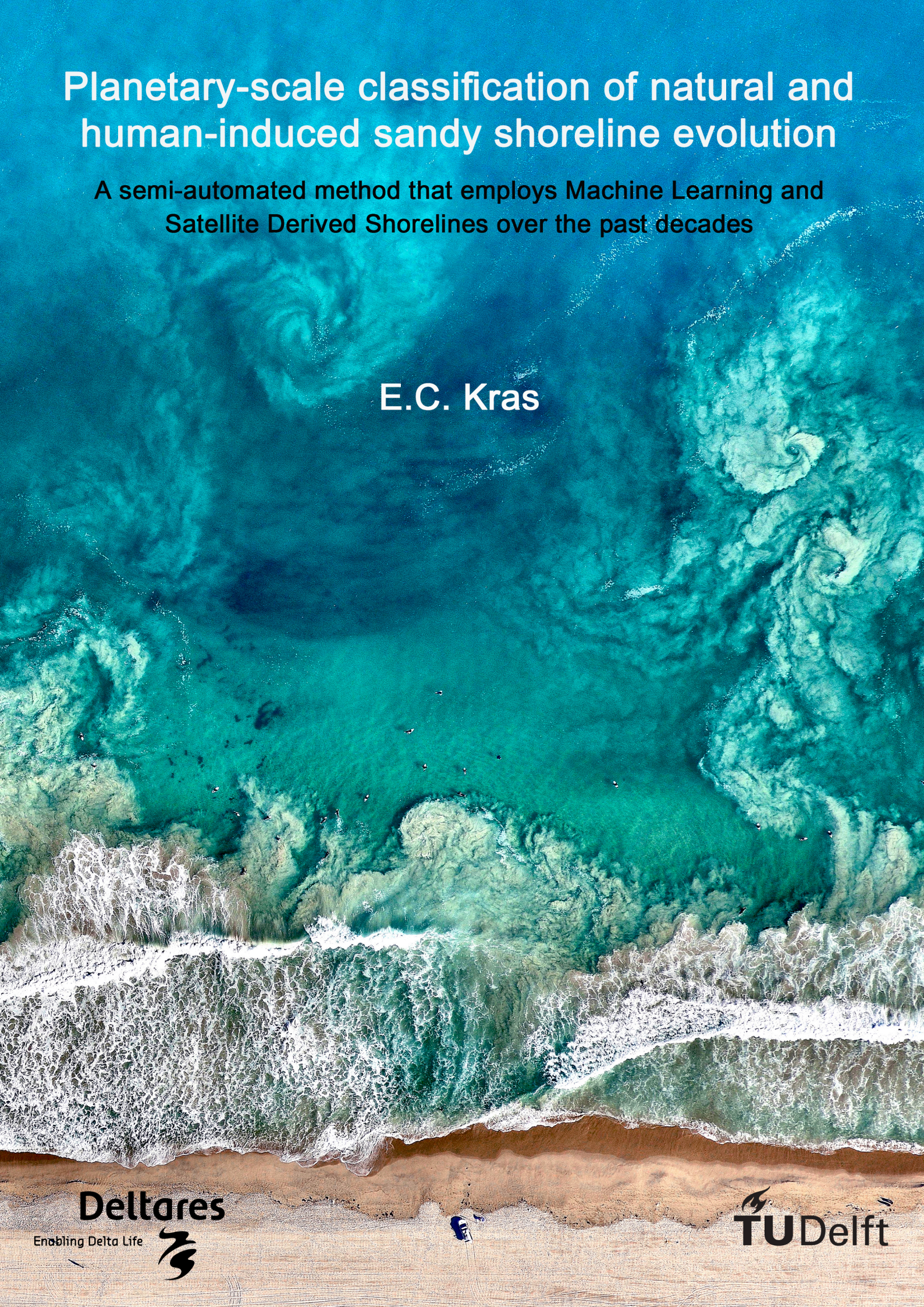
Planetary-scale classification of natural and human-induced sandy shoreline evolution

A semi-automated method that employs Machine Learning and Satellite Derived Shorelines over the past decades

E.C. Kras



Deltares
Enabling Delta Life

TU Delft

Planetary-scale classification of natural and human-induced sandy shoreline evolution

A semi-automated method that employs Machine Learning and Satellite Derived Shorelines over the past decades

by

E.C. Kras

Student number: 4276469
etiennekras@gmail.com | +31624511139
MSc student Hydraulic Engineering
Specialization in Coastal Engineering

to obtain the degree of

Master of Science
in Civil Engineering

at the Delft University of Technology,
to be defended publicly on Tuesday August 27, 2019 at 15:00.

Thesis committee:	Prof. dr. ir. S. G. J. Aarninkhof,	TU Delft
	Dr. ir. S. de Vries,	TU Delft
	Ir. A. P. Luijendijk,	TU Delft / Deltares (supervisor)
	Ir. W. P. de Boer,	TU Delft / Deltares
	Dr. ir. J. A. A. Antolinez,	Deltares

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

Cover: Aerial photograph of swirling waves offshore Perth, Australia

Deltares
Enabling Delta Life

TU Delft

Preface

This thesis is the result of somewhat more than ten months of research at Deltares and concludes my masters in Hydraulic Engineering, with a specialization in Coastal Engineering, at the Faculty of Civil Engineering and Geosciences at the Delft University of Technology. The topic of this thesis is referred to represent the very thin line between the Coastal Engineering and Data Science practice by employing Machine Learning on satellite imagery, and therefore hopefully expresses my interest in exploring new scientific research areas. I experienced doing research, especially on the aforementioned topic, as very challenging but also rewarding and satisfying. For this reason, I am very grateful for the opportunity to work in the professional environment Deltares has been offering.

Anyone with sufficient background knowledge in the field of Civil Engineering should be able to understand the applied approach, research methods, results and conclusions presented in this report as the Data Science part is still fairly superficial and extensively explained. Both the applications of satellite imagery and Machine Learning are becoming increasingly popular and are constantly subject to improvements. Consequently, I encourage the people reading this report to focus on the applied approach and research methods.

I would not have been able to finish this research without the help of many others. First of all, I would like to thank Sierd de Vries for putting me into contact with Arjen Luijendijk after the thesis ideation meeting of about a year ago. Your feedback during meetings was also highly appreciated. A special thanks goes to my supervisor Arjen Luijendijk, your ideas, interpretations, expert insights and critical comments helped me to raise the level of this research significantly. Besides this, I would like to thank Wiebe de Boer for the feedback on my report and the critical comments on the applied methods and results. Thank you José Antolinez for introducing me to, guiding me through and feeding my interest in the topic of Data Science. I also want to express my appreciation to professor Stefan Aarninkhof, for chairing the graduation committee and providing feedback during meetings. I want to thank Gerben Hagenaars as well, who was part of Deltares till about eight months ago. Your help during the start-up of my research was of great importance.

Last but not least, I want to thank my family, friends and fellow students for the support concerning my complete masters study. Your influence on my overall development and performance is certainly not to be forgotten. Enjoy reading!

Etiënne Kras
Delft, August 2019

Abstract

Today's coastal zones are densely inhabited as the majority of the world's population lives in these attractive areas. The shorelines in coastal zones are shaped by complex spatial and temporal variable interactions between natural forcings like changes in mean sea-level, tides, wave and wind conditions, and storm surges. Besides, natural hazards such as coastal erosion, tropical cyclones, hurricanes, typhoons, floods, salt intrusion and tidal surges threaten a major part of the world's population. Furthermore, climate change is likely to increase the risk of natural hazards. As a response, humans changed the world's shorelines and the forcing-driven processes that work on them, to increase protection against hazards and keep supporting their activities. These human interventions are deployed discontinued in time, disperse spatially and might result in negative consequences leading to human-induced hazards.

This research focuses on one particular hazard to coastal communities, coastal erosion, which is extended to a term referred to as shoreline evolution by incorporating coastal accretion as well. Up till now, detailed local-scale studies are able to expose human and natural drivers of shoreline evolution and provide a possibility to make a step towards intentional rather than accidental coastal engineering. Digital imaging and, more recently introduced, satellite imagery proved to be a promising new technology to measure and monitor shoreline evolution at bigger temporal and spatial scales. Nevertheless, the opportunity to develop a model that exposes the drivers of shoreline evolution on a planetary scale remains unexplored. This is due to the required computational effort as well as the large variability in coastal systems around the world. With an ever-increasing data availability, data-driven models incorporating Machine Learning (ML) proved to be an efficient alternative approach to heavy computing classical process-driven models in civil engineering practice. Next to this, a coastal classification can be used as a means to inventory the aforementioned variability. Therefore, the research objective in this study is to explore the possibility of exposing and classifying the drivers of shoreline evolution on a planetary scale, by employing ML on satellite imagery.

In order to fulfill the research objective, various methods have been developed, employed and molded into a multi-step framework at which the annual dataset behind the Shoreline Monitor of Luijendijk et al. (2018) forms the foundation. This dataset incorporates an assessment of the occurrence of sandy beaches and a change detection method called Satellite Derived Shorelines (SDS) to represent shoreline evolution on a planetary scale, every 500 m (transect) along the coast for the past 33 years. As the SDS method was found to be inaccurate for non-sandy beaches, the remainder of this research only considers sandy shoreline evolution. Firstly, a feature extraction and quantification algorithm is developed and tested on local scales. Here, a feature refers to multiple shoreline evolution transects obviously differing from their neighbors with change rates larger or smaller than 0.5 $m/year$. Secondly, the aforementioned algorithm is deployed on a global scale. This allowed to generate planetary-scale datasets and a map with prograding and retreating shoreline evolution features, whereas the drivers behind these features may not be known yet. Hereafter, various human and natural drivers are manually related to features on a local scale, as it was already proved that the spatiotemporal variability within features is largely in line with drivers listed in literature. Subsequently, an unsupervised clustering method (Gaussian Mixture Model, GMM) is employed to automatically group similarities in spatiotemporal variable characteristics in shoreline evolution on a global scale. After this, locally identified drivers of shoreline evolution are correlated to automatically grouped features of the previous step. This allowed to generate planetary-scale driver correlated maps with similar prograding and retreating shoreline evolution features. Finally, the obtained semi-automated insight in globally mapped feature-driver correlations is considered to statistically derive a planetary-scale classification of natural and human-induced sandy shoreline evolution using spatial and temporal characteristics.

In this research, 33 continuous shorelines are created for (combined) continents or islands with an area larger than 25.000 km^2. Here, the largest area is referred to as the combined continent of Afro-Eurasia, whereas the smallest area is related to the island of Sicily (Italy). For all continuous shorelines, approximately 985.000 transects (around 390.000 km of shoreline) from the dataset behind the Shoreline Monitor (44% is actually sandy) are included in the feature extraction and quantification algorithm. This resulted in two global datasets with

3033 prograding and 2121 retreating sandy shoreline evolution features (5154 in total) with an offset towards larger long-term shoreline changes. In other words, it is found that a combined total of approximately 22.000 km of shoreline is either prograding or retreating. Twelve characteristics (four spatial and eight temporal) are found to be useful to explain spatiotemporal variability within a shoreline evolution feature. This implies that, for any place, country or continent on earth, we are able to extract spatial and temporally variable shoreline evolution features that may impose hazards to coastal communities.
Manually analyzing nearly 7000 km of combined shoreline resulted in a total of 212 human as well as natural driver-related shoreline evolution features (126 prograding and 86 retreating). It is found that features containing very large and irregular change rates are primarily caused by unnatural drivers. Employing the GMM for 5154 extracted and quantified features resulted in an optimal total of seventeen groups with similarities. Thirteen of these groups, around 80% of all features, are validated to contain multiple correlated drivers of shoreline evolution. Therefore, we can show statistical similarities in characteristics between an extracted feature and any other feature in the world. This tells something about the state of the shoreline evolution, allows to couple literature from other similar locations and enables us to reveal global distributions that were not known before.
By elaborating on the statistically derived classification, it is found that natural and human-induced shoreline evolution accounts for approximately 16 and 25% of the total of globally extracted and quantified features respectively. By looking at mean long-term shoreline changes on continent level, it is found that Asia's features are mainly due to (in)direct human influences. 57% of the features is present in a regime with complex and combined (compound) influences. It is also known what this compound regime consists of in terms of potential drivers. For example, on average, the continent of Africa consists of more natural influences. But, whenever humans interfere, mean long-term shoreline changes are generally larger due to the lack of knowledge or economic resources. In contrast to the continent of Africa, Europe and Oceania show more human influences that have smaller mean effects. Consequently, we can indicate the most probable or relevant driver of the extracted shoreline evolution features as well as whether the features are located in the human, natural or compound influence regime.

All the outcomes above can support detailed local-scale investigations and therefore provide an enhanced opportunity to make a step towards intentional rather than accidental coastal engineering. Hence, it is concluded that the developed and applied methods that employ ML on the annual dataset behind the Shoreline Monitor can be used to expose and classify (in)direct human and natural influences on sandy shoreline evolution using spatial and temporal characteristics on a planetary scale. Nonetheless, almost 60% of the shoreline evolution features are located in the compound regime, which still requires (rather than supports) local-scale investigations to determine the correct influence or driver. This outcome is primarily due to three major limitations incorporated in the research methods, which are referred to as the exclusion of variability in natural processes and location dependency, multiple driver influences, and oscillatory drivers of shoreline evolution.

More research is required to elaborate on the opportunities that can enhance insight in the compound regime, to improve the obtained results and advance the applicability of this study. One of the main recommendations is to investigate the possibility of increasing the spatial and temporal resolution in the Shoreline Monitor dataset. This might allow to resolve all three major limitations outlined above.

Contents

Acronyms and Abbreviations

Acronym / abbreviation	Description
AI	Artificial Intelligence
AIC	Akaike Information Criterion
AICc	corrected Akaike Information Criterion
BIC	Bayesian Information Criterion
CART	Classification And Regression Tree
DIVA	Dynamic Interactive Vulnerability Analysis
EM	Expectation-Maximization
ENSO	El Niño - Southern Oscillation
EOF	Empirical Orthogonal Functions
EPSG	European Petroleum Survey Group
ESA	European Space Agency
GDP	Gross Domestic Product
GEE	Google Earth Engine
GM	Global Model
GMM	Gaussian Mixture Model
GRIP	Global Roads Inventory Project
HBASE	Human Built-up And Settlement Extent
IT	Information Technology
LECZ	Low Elevation Coastal Zone
LM	Local Model
MAE	Mean Absolute Error
ML	Machine Learning
MLE	Maximum Likelihood Estimation
NaN	Not a Number
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
NDWI	Normalized Difference Water Index
OSM	OpenStreetMaps
PCA	Principal Component Analysis
PDF	Probability Density Function
QGIS	Quantum Geographic Information System
RMSE	Root Mean Square Error
SDS	Satellite Derived Shoreline
SLR	Sea-Level Rise
std	standard deviation
UAE	United Arab Emirates
USA	United States of America
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
WDPA	World Database on Protected Areas
WGS	World Geodetic System
WPI	World Port Index

1 Introduction

Misinterpretation of (problematic) shoreline evolution signals hampers informed decision making and the design and implementation of engineering interventions to mitigate hazards. Here, shoreline evolution considers both the phenomena of coastal erosion / accretion over a longer timespan and is driven by complex and combined (compound) human and natural alterations to the sediment budget, including sources and sinks of beach sediment, or the processes that work on them (Stive et al., 2002 and Bird and Lewis, 2015).

Humans try to understand and predict compound interactions that lead to shoreline evolution by building models. These can be summarized in process-driven or data-driven models. Process-driven models, also referred to as physical models, are based on a set of mathematical equations that focus primarily on the involved processes. Data-driven models build relationships between input and output data without focusing too much on the underlying processes.

Up till now, detailed local-scale studies or models are able to expose human and natural drivers of shoreline evolution and provide a possibility to make a step towards intentional rather than accidental coastal engineering. Digital imaging and, more recently introduced, space-borne remote sensing techniques referred to as satellite imagery proved to be a promising new technology to measure and monitor shoreline evolution at bigger temporal and spatial scales (Aarninkhof et al., 2003 and Hoepffner and Zibordi, 2009). Moreover, computer-related problems regarding global assessments were solved by the introduction of the Google Earth Engine (GEE) (Donchyts et al., 2016; Pekel et al., 2016 and Gorelick et al., 2017). Nevertheless, the opportunity to develop a model that exposes the drivers of shoreline evolution on a planetary scale remains unexplored. This is, amongst others, due to the large variability in coastal systems or zones around the world. A coastal classification can be used as a means to inventory this variability (Finkl, 2004 and Bosboom and Stive, 2015).

1.1 The coastal zone

"A coastal zone forms the interface between land and sea, delineated as the part of the land affected by its proximity to the sea, and the part of the sea affected by its proximity to the land"
~ K. Mangor in "Shoreline Management Guidelines", 2004

Today's coastal zones are densely inhabited as the majority of the world's population lives in these attractive areas (Bush et al., 2001; Small and Nicholls, 2003; Al-Tahir and Ali, 2004 and Seto and Shepherd, 2009). McGranahan et al. (2007) found that 10% of the world's population and 13% of the world's urban population are situated in the Low Elevation Coastal Zone (LECZ), defined as the area less than 10 m above sea-level. According to Neumann et al. (2015), the LECZ exhibits high rates of population growth (see Figure 1.1). This is caused by urbanization, which is driving populations towards the coast (McGranahan et al., 2007), resulting in the shortage of land in the LECZ (see Figure 1.2).

The world's shorelines are shaped by location dependent interactions between natural forcings like changes in mean sea-level, tides, wave and wind conditions, and storm surges. The spatial and temporal variability within these forcings is subject to climatic effects like solar radiation and temperature distribution, atmospheric circulation and wind patterns, and oceanic circulation (Bosboom and Stive, 2015). With increasing spatial and temporal scale, the relative importance of climatic effects on the spatiotemporal variability within aforementioned forcings becomes more relevant considering coastal erosion / accretion in heavily populated zones.

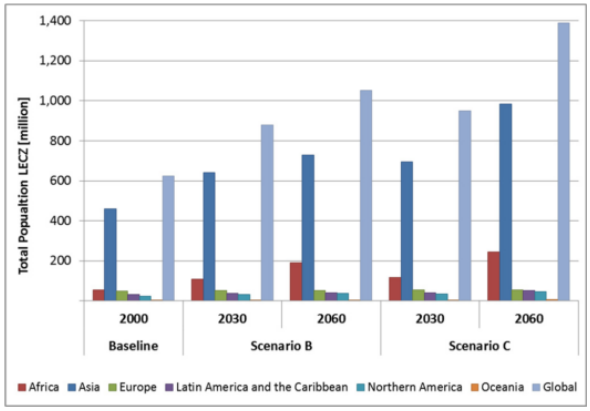


Figure 1.1: Graphical presentation of the population within the Low Elevation Coastal Zone (LECZ) in the year 2000. Besides this, scenario B (C) shows the population at the lowest (highest) end of forecasts for the year 2030 (2060). (source: Neumann et al., 2015)

Figure 1.2: Photograph of wind waves coming close to homes in the Hawke's Bay village of Haumoana, New Zealand. There is a shortage of land in this Low Elevation Coastal Zone (LECZ), due to human settlement. (source: `www.tvnz.co.nz`)

The natural forcing imposes, besides the hazard of coastal erosion, various other natural hazards that threaten a major part of the world's population such as tropical cyclones, hurricanes, typhoons, floods, salt intrusion and tidal surges (Neumann et al., 2015). As a response, human populations changed the world's shorelines and the forcing-driven processes that work on them, to increase protection against hazards and keep supporting their activities (LaGro, 2005). These human interventions are deployed discontinued in time, disperse spatially and might result in negative consequences leading to human-induced hazards. Anthoff et al. (2006) found that the existing population, land area, and economic activity within the 10 m LECZ will not be threatened by Sea-Level Rise (SLR) due to climate change in the 21^{st} century. Nonetheless, Ahsan et al. (2016) states that the rising sea-level (Church and White, 2011) will increase the risk of hazards.

1.2 Shoreline evolution and its drivers

A variety of performed local-scale studies focus on the increasing hazard of coastal erosion. Some studies focused on the impact-related prediction of the magnitude of coastal erosion (Addo et al., 2008 and Ranasinghe, 2016) or on the assessment of (future) impacts of coastal erosion on human settlements, as a result of climate change (Hinkel et al., 2013 and Cowan, 2017). However, in some cases, impacts related to coastal accretion are also relevant. For instance when considering recreational activities, siltation of harbors due to sediment bypassing (de Bruijn, 2018) and degrading ecosystems (Airoldi and Bulleri, 2011).

Other studies focused on causes related to one specific erosional driver like storm surge (Dolan et al., 1978), coastal squeeze (Pontee, 2011) and coastal structures (Vaidya et al., 2015), while Bird and Lewis (2015) focused on multiple drivers of coastal erosion like the reduction of sediment supply by inland dunes, submergence and increased wave attack, alongshore losses of beach sediment and the interception of longshore drift by breakwaters. Different from the studies above, Wang (2018a) performed a study to explore the possibility of exposing drivers of sandy shoreline evolution through satellite imagery by using Satellite Derived Shorelines (SDS). Considering local-scale cases, Wang (2018a) proved that spatiotemporal variability in sandy shoreline evolution caused by various types of natural and human drivers (seasonal variability, climate variability, land subsidence, port breakwaters and nourishments), is largely in line with the conclusions listed in literature.

1.3 Research objective and scope

As data availability is ever-increasing, data-driven models are emerging as an popular alternative approach to classical process-driven models in civil engineering practice (Reich, 1996). Moreover, the application of Machine Learing (ML) techniques in data-driven models have proven to increase its efficiency even more. The research objective in this study is to explore the possibility of exposing and classifying the drivers of shoreline evolution on a planetary scale, by employing ML on satellite imagery. Here, it is opted to reveal the correlation of a driver to planetary-scale distributions of ML derived groups of similarities in spatiotemporal variability in shoreline evolution.

The Shoreline Monitor of Luijendijk et al. (2018), an application that uses Satellite Derived Shorelines (SDS) and the GEE, represents a planetary-scale dataset with around 2.2 million shore-normal 500-m spaced transects containing sandy / non-sandy labels and annual shoreline changes over the past decades. It is stated that the Shoreline Monitor, considering the proof of Wang (2018a) for a higher spatial and temporal resolution, has the potential to be further utilized to fulfill the research objective. By using the Shoreline Monitor, spatiotemporal variability can be captured in multiple sandy shoreline evolution signals / transects (referred to as features from now on). These features can be coupled to drivers afterwards.

The focus in this research is on the large spatial (planetary) and long-term temporal scale. For feasibility purposes, oscillatory drivers of shoreline evolution like seasonal and climate variability are excluded from the scope. Next to this, the analysis ignores the location as well as natural processes and focuses on persistent or episodic human and natural drivers that stand at the basis of possibly hazardous shoreline evolution trends. Examples of drivers of shoreline evolution that are included are human interventions like nourishments, port breakwaters or land reclamations and natural phenomena like sand bar migration, tidal inlets or storms. Based on the scope, the below-presented hypothesis is formulated. The hypothesis stands at the basis of the research methods.

'A persistent or episodic driver of shoreline evolution shows approximately similar characteristics in spatiotemporal variability everywhere on earth.'

1.4 Research question

The research objective is translated into a research question and three supportive sub-questions, all formulated below. Within the sub-questions, key words are underlined. These key words are used to refer to the sub-questions in the remainder of this research.

To what extent can the annual dataset behind the Shoreline Monitor be used to derive a planetary-scale classification of sandy shoreline evolution when employing Machine Learning?

1. *On a planetary scale, how can shoreline evolution features from the Shoreline Monitor be extracted and quantified?*
   This sub-question focuses on developing an algorithm to extract features, which contain multiple sandy shoreline evolution signals, from the Shoreline Monitor. In an 'along-the-shoreline'-analysis, the spatiotemporal variability of these features is captured and various characteristics are quantified.

2. *What is the correlation between feature similarities and drivers of sandy shoreline evolution, and what is the validity of these correlations?*
   This sub-question elaborates on sub-question one by correlating similarities in feature quantifications to drivers of sandy shoreline evolution, using a preprogrammed Machine Learning algorithm and manual insight. Moreover, the correctness of these correlations is validated.

3. *What is the derived planetary-scale classification of sandy shoreline evolution?*
   In this sub-question, a planetary-scale classification of sandy shoreline evolution is derived from analyzing the correlations of sub-questions two.

1.5 Research outline

This report consists of seven main chapters. Chapter 2 outlines the existing knowledge in the research area in relation to the research objective stated in Chapter 1. Special attention is given to the Shoreline Monitor from Luijendijk et al. (2018), as this forms the foundation of this research. The following chapter, Chapter 3, consists of three sections in which the research methods for each sub-question are elaborated on. Chapter 4 consists of three sections in which each individual sub-question is answered, based on the methods explained in the previous chapter. The discussion in Chapter 5 contains a reflection upon the research methods and results. Limitations of the research, mainly focusing on assumptions, decisions and restrictions, as well as the applicability of present and potential outcomes are outlined. Based on the results of the three sub-questions, an answer to the research question is given in Chapter 6. Moreover, some concluding remarks related to the outcomes are presented. Finally, Chapter 7 contains concrete and feasible comments on possible further research and improvements. A graphical representation of the report layout is shown in Figure 1.3.

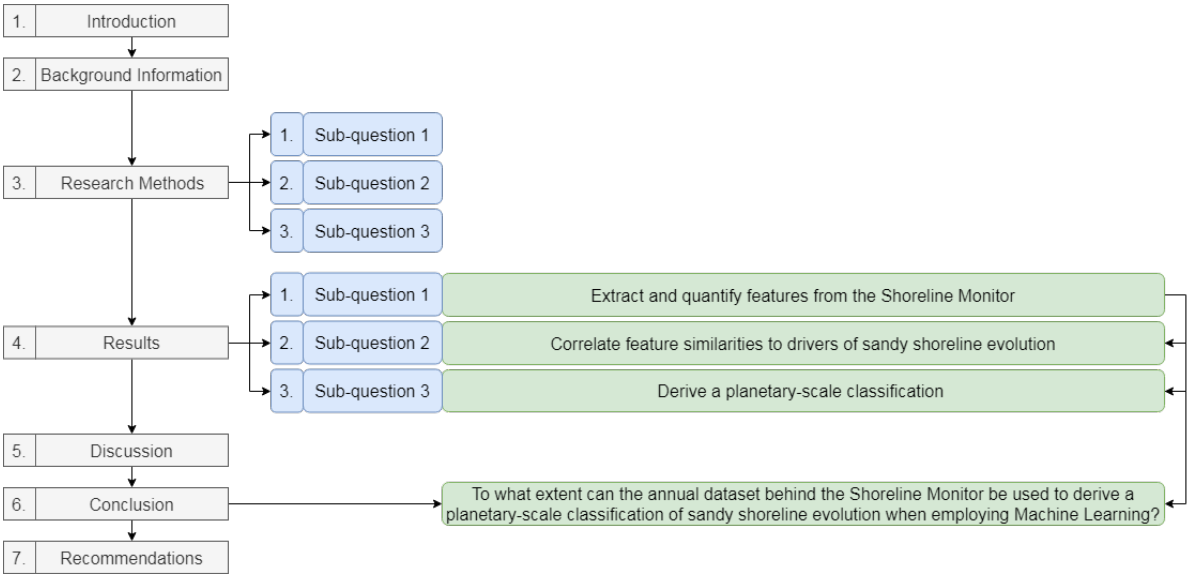


Figure 1.3: The report structure with its seven main chapters (in grey). The chapters research methods and results have structured sections based on the sub-questions (in blue). The three sub-questions (indicated by their key words, see Section 1.4) and the research question are shown in green if they are answered in a particular section or chapter.

2
Background Information

This chapter provides an overview of the existing knowledge in relation to the research objective stated in Section 1.3. The first section of this chapter (2.1) describes the basis of this research, the Shoreline Monitor with its assumptions and outcomes, in detail. Section 2.2 explains the different human and natural drivers of shoreline evolution that can be revealed by analyzing signals in the Shoreline Monitor dataset. Section 2.3 deals with the principle of ML and the selected data-driven algorithm that is used to build relationships between the shoreline evolution signals and the considered drivers of shoreline evolution. At the end of the introduction of every section, key points summarize the most important background information.

2.1 The Shoreline Monitor

Change detection, the process of identifying differences in the state of an object or phenomenon by observing it at different times (Singh, 1988), is a particularly interesting application of satellite imagery. Colwell et al. (1983) stated that timely and accurate change detection of earth's surface features provides the foundation for an increase in the understanding of relationships and interactions between human and natural phenomena to better manage and use resources, which is supported by findings in Almonacid-Caballer et al. (2016).

The Shoreline Monitor is a dataset that incorporates an assessment of the occurrence of sandy beaches and a change detection method called Satellite Derived Shorelines (SDS). The Shoreline Monitor represents shoreline evolution on a planetary scale, every 500 m along the coast for the past 33 years, by using a transect system and the GEE platform (Luijendijk et al., 2018). The GEE platform, the transect system, the sandy beach detection and the SDS detection method are outlined in the sections below.

Key points

- The Shoreline Monitor, the foundation of this research, is a dataset that incorporates the GEE, a 500-m spaced shore-normal transect system, a sandy beach detection method and a change detection method called SDS to provide a sandy or non-sandy labeled annual shoreline evolution signal (spanning at most 33 years) for every of the 1.8 million transects in the world;
- The SDS detection algorithm proved to be suitable to study shoreline evolution signals with changes bigger than 0.5 $m/year$ for the period of 1984 - 2016 and experiences difficulties representing correct shoreline evolution signals for non-sandy labeled transects;
- The shore-normal defined transect system as well as the SDS detection method with linear regression result in various frequently occurring inaccuracies in the Shoreline Monitor, even when only considering the sandy labeled transects.

2.1.1 The Google Earth Engine

The GEE is a cloud-based platform for planetary-scale analyses that bring Google's computational capabilities to bear on a variety of high-impact societal issues for non remote-sensing experts (Gorelick et al., 2017). The two main issues requiring expert knowledge are addressed as the employment of image correction techniques and the software skills or computing power to process global datasets (Gorelick et al., 2017).

The Earth Engine platform solves the aforementioned hurdles regarding the usage of satellite datasets (Gorelick et al., 2017). Gorelick et al. (2017) stated that the GEE provides easy access to high-performance computing resources for processing very large geospatial datasets, without having to suffer the Information Technology (IT) pains. Moreover, the Earth Engine contains a freely available analysis-ready data catalog, where

satellite data is already corrected and preprocessed Gorelick et al. (2017). Examples of corrected and preprocessed satellite datasets are ESA Sentinel 2 and NASA Landsat 5, 6 and 8 missions (Gorelick et al., 2017 and Hagenaars et al., 2018). Besides providing access to Google's resources and datasets, the GEE also provides a publicly available frontend application called Timelapse. Timelapse is a zoomable video that shows earth's surface changes over the past 35 years (`https://earthengine.google.com/timelapse/`).

2.1.2 Transect system

The transect system behind the Shoreline Monitor is created by generating 1 km transects (500 m land- and seaward) with 500 m spacings perpendicular to the highest resolution (z8) generalized shorelines from 2016 OpenStreetMap (OSM) (Luijendijk et al., 2018). The OSM shorelines (30374 pieces in total) are presented in longitudinal and latitudinal units in reference systems World Geodetic System (WGS) 84 or its equivalent European Petroleum Survey Group (EPSG) 4326 (OpenStreetMap, 2019). These reference systems are used for, for instance GPS systems. However, OSM shorelines can also be obtained in reference systems WGS 84 / Pseudo-Mercator or its equivalent EPSG 3857 (units in m). EPSG 3857 is frequently used as web projection for, for instance Google Maps, Bing and ArcGIS (OpenStreetMap, 2019).

In total, approximately 2.2 million transects are defined to represent the total ice-free shoreline. By summing straight intercepts between the 2.2 million transects, this shoreline length is reported to have a length of 1.1 million km (Luijendijk et al., 2018). To speed up computation time, the transects have been split and labeled with id's according to their position in the world. A detailed description of this structure can be found in Appendix A. Eventually, around 1.8 million transects remain after applying latitude filters (60°N and 50°S) (Luijendijk et al., 2018).

van Leeuwen (2018) assessed the influence of a manual revision of the unique and small-scale inaccuracies of the global transect system on the Shoreline Monitor's results. Besides this, various frequently occurring inaccuracies are outlined and referred to as weather-induced disturbances, estuaries and bays, deltas and river mouths, shallow shorelines, and man-made structures. Different solutions to these inaccuracies, related to the global transect system, are listed below and were recommended to be employed in the scripting architecture behind the Shoreline Monitor (van Leeuwen, 2018).

1. Employing a built-in fault detector to consecutive transects;
2. Plotting transects on SDS instead of on the OSM shoreline;
3. Adjusting the transect length, either dynamically or statically.

2.1.3 Sandy beaches

In order to assess the global sandy beach occurrence, the work-flow presented in Figure 2.1 is followed. After collecting all ESA Sentinel-2 satellite images for the year 2016, a composite 10 m resolution image is constructed. This image is created using an image composite technique, based on the work of Donchyts et al. (2016), which reduces the impact of cloud cover and shadows on the image quality (Hagenaars et al., 2018; Wang, 2018a and Luijendijk et al., 2018).

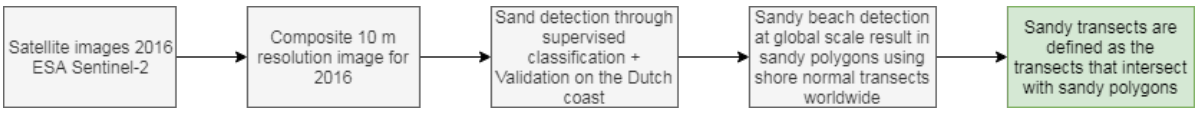


Figure 2.1: The work-flow to assess the worldwide occurrence of sandy beaches. The final result is shown in green, where the remaining steps that are needed to come to the final result are shown in grey. (source: adjusted from Luijendijk et al., 2018)

Sandy beaches are detected locally by applying a pixel-based supervised classification called Classification And Regression Tree (CART) (Luijendijk et al., 2018). This classification algorithm proved to give the highest percentage of true positives on different training and testing locations at the Dutch coast (Luijendijk et al., 2018). The CART algorithm is part of a decision tree, which recursively partitions the space such that the samples with the same labels are grouped together (Pedregosa et al., 2011). According to Luijendijk et al. (2018), the CART algorithm is upscaled to the global domain and joined with the shore normal transect system, to automatically define 'sandy', 'non-sandy' and 'undetermined sediment composition' transects. The 'sandy'

label comprises of quartz, carbonates sands and gravel. This global sandy beach assessment was visually validated to have 96% accuracy (Luijendijk et al., 2018).

A planetary-scale visualization of the result of the sandy beach detection method is shown in Figure 2.2. Luijendijk et al. (2018) stated that approximately 31% of the world's ice-free beaches / transects are sandy. The latitudinal distribution of sandy beach occurrence is largely in line with the distribution reported by Hayes (1967).

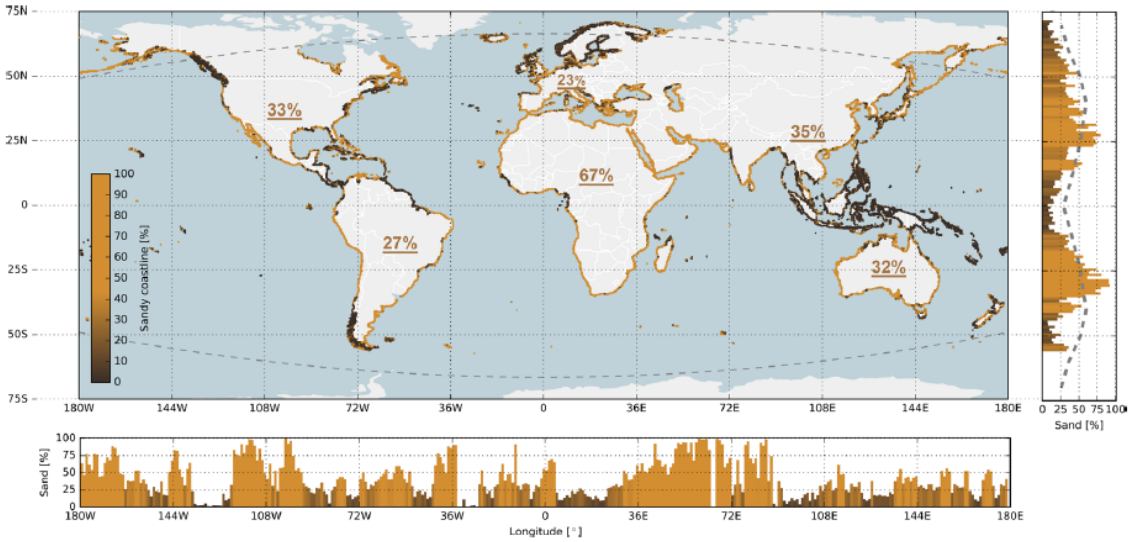


Figure 2.2: Global distribution of sandy shorelines where the colored dots along the world shoreline represent the local percentage of sandy shorelines (yellow is sand, dark brown is non-sand). Per degree longitude or latitude, represented by the subplot on the bottom and the right respectively, the relative occurrence of sandy beaches is shown as well. The percentage of sandy beaches per continent is shown in the map. Latitude filters are shown with dashed lines in the main plot, while the Hayes distribution is shown with a dashed line in the right subplot. (source: Luijendijk et al., 2018)

2.1.4 Shoreline detection

A coastal parameter suitable for monitoring shoreline evolution is the position of the shoreline (Hagenaars et al., 2018). However, as the 'true' shoreline position is rather challenging to obtain, a series of coastal state indicators have been used as proxies for practical purposes (Boak and Turner, 2005). Originally, two types of proxies are used to derive the shoreline position: the subjective visible high-water line or wet/dry boundary from coastal imagery and the more objective intersection of a tidal datum (mean sea-level or mean high water) with the shoreline profile (Boak and Turner, 2005). Recently, with the introduction of various image processing techniques, a third type of proxy or coastal state indicator (not visible to the human eye) was introduced (Boak and Turner, 2005). This indicates a shift from subjective to objective shoreline detection (Boak and Turner, 2005).

The combined application of various image processing techniques resulted in the development of the SDS detection algorithm (Hagenaars, 2017). Hagenaars et al. (2018) found that the accuracy of SDS detection, using the Normalized Difference Water Index (NDWI), is within sub-pixel precision (15 m) if composite images are created. A visualization of the SDS based on one composite image is shown in Figure 2.3, for the example case of the Sand Engine at the Dutch Coast (Hagenaars et al., 2018). Consequently, SDS proved to be suitable to study shoreline evolution on a planetary scale if shoreline changes are bigger than the sub-pixel precision over the timespan of 1984 - 2016 (33 years), i.e. $+/-$ 0.5 $m/year$. Next to that, the shoreline evolution found by satellite imagery, which usually spans 30 to 40 years, might be representative for even longer trends and is therefore indicative for future behavior (Dolan et al., 1978).

Figure 2.3: An example of the employment of the Satellite Derived Shorelines (SDS) detection algorithm for the Sand Engine study site. The resulting SDS is shown in black. (source: Hagenaars et al., 2018)

The work-flow presented in Figure 2.4, to derive annual planetary-scale shoreline changes, is followed in Luijendijk et al. (2018). This work-flow is comparable to the work-flow in Figure 2.1, considering the first two steps. However, there are two main differences in the first two steps. Firstly, instead of using ESA Sentinel-2 images for only 2016, all satellite images of the years 1984 - 2016 of different NASA satellite missions are retrieved. Secondly, instead of creating one 10 m resolution composite image for 2016, yearly composite images are created with 30 m resolution. The yearly composites are created using an optimal averaging period of 192 days, considering the 16-day revisiting time for a satellite sensor (Luijendijk et al., 2018). Luijendijk et al. (2018) stated that due to the averaging period, the average water level in the images corresponds to mean sea-level provided that all images are cloud free. As a result, seasonal variability in wave and beach characteristics is averaged out, where variability in longer temporal scales is still present (Luijendijk et al., 2018). While the Shoreline Monitor provides annual shoreline positions, it is also possible to generate biweekly (16-day) composite images when using a moving averaging period of 192 days (Luijendijk et al., 2018).

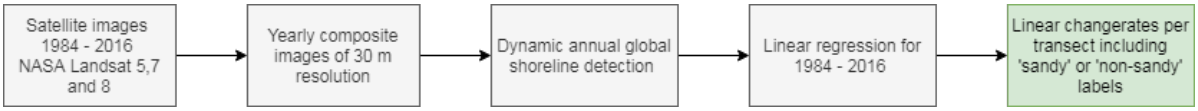


Figure 2.4: The work-flow to derive global shoreline changes for the years 1984 - 2016. The final result is shown in green, where the remaining steps that are needed to come to the final result are shown in grey. (source: adjusted from Luijendijk et al., 2018)

Global annual shorelines are obtained by using the SDS detection algorithm on the yearly composite images. In case of the detection of several shoreline vectors, induced by lakes and rivers, only the most seaward shoreline position is analysed (Luijendijk et al., 2018). Merging and computing the intersections of the global shore-normal transect system and the annual shorelines results in at most 33 shoreline positions relative to the transect origin for the timespan of 1984 - 2016 (Hagenaars et al., 2018 and Luijendijk et al., 2018).

By applying a linear regression to the shoreline positions in each transect, an easy-to-interpret persistent shoreline trend is obtained (Hagenaars, 2017). Within the linear regression, SDS positions more than three times the standard deviation are considered outliers and are not taken into account. Approximately 60% of the 2.2 million transects show an uncertainty bandwidth smaller than 50% to the linear regression (Luijendijk et al., 2018). After applying linear regression, transects are omitted if the resulting shoreline evolution trend is bigger than 100 $m/year$, contains less than 5 out of 33 shoreline positions or has a temporal coverage smaller than seven years (Luijendijk et al., 2018). Results of the above-presented method were validated on four different case studies. Combining the sandy beach occurrence assessment and the shoreline detection results in the Shoreline Monitor's publicly available frontend application which can be accessed through this link: `shorelinemonitor.deltares.nl`.

The results of the study in Luijendijk et al. (2018) are mainly focused on the sandy transects, as it is acknowledged that the methods used in the Shoreline Monitor experience difficulties in muddy environments with mild foreshore slopes. The remainder of this research therefore only considers sandy shoreline evolution. Luijendijk et al. (2018) found that roughly 81% of the original number of sandy transects remain after applying the filters outlined in the paragraph above. 24% of the sandy beaches are eroding at rates exceeding 0.5 $m/year$, while 28% are accreting and 48% are stable (Luijendijk et al., 2018). Besides these general figures, hot spots comprising of at least 5 km sandy shorelines with all transects either eroding or accreting with change rates smaller or larger than 0.5 $m/year$ were defined (Luijendijk et al., 2018). A planetary-scale visualization of the eroding and accreting hot spots is shown in Figure 2.5.

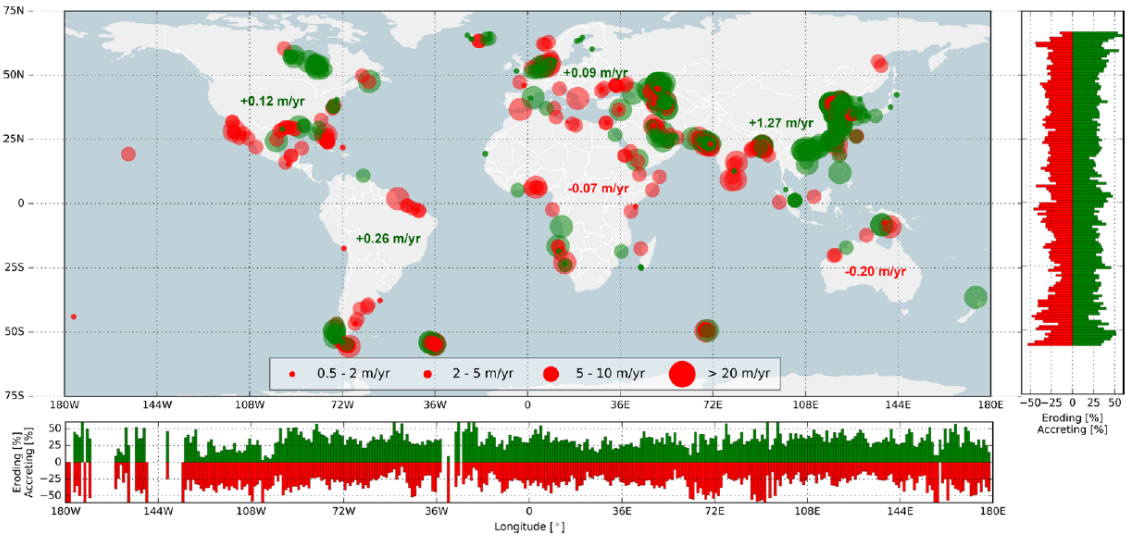


Figure 2.5: Global hot spots of beach erosion and accretion; the red (green) circles indicate erosion (accretion) for the four relevant shoreline dynamic classifications (see legend). Per degree longitude or latitude, represented by the subplot on the bottom and the right respectively, the relative occurrence of eroding or accreting sandy shorelines is shown as well. The numbers presented in the main plot represent the average change rate for all sandy shorelines per continent. (source: Luijendijk et al., 2018)

As outlined in Section 2.1.2, van Leeuwen (2018) assessed the influence of a revision of the global transect system on the Shoreline Monitor's results. van Leeuwen (2018) found that a part of the different frequently occurring inaccuracies are related to the shoreline detection method, and thus not solely to the global transect system. Different solutions to the frequently occurring inaccuracies presented in Section 2.1.2, but now related to the shoreline detection method, are listed below and were recommended by van Leeuwen (2018) to be employed in the scripting architecture behind the Shoreline Monitor.

1. Filtering on unreliable results (i.e. one-third of the maximum number of SDS positions must be present);
2. Employing different fitting methods, besides the linear regression;
3. Splitting and fitting multiple trend lines per decade.

2.2 The drivers of sandy shoreline evolution

Shoreline changes are driven by natural and human-induced spatial sediment transport gradients and / or sediment sources or sinks (Bosboom and Stive, 2015). These drivers of shoreline evolution have a wide variability in spatial and temporal scales. Wang (2018a) proved that spatiotemporal variability of multiple sandy shoreline evolution signals (features) at eleven sites, derived by the SDS detection method using biweekly composite images (outlined in Section 2.1.4), is largely in line with literature for natural and human drivers as seasonal forcing, climate variability, land subsidence and human interventions (like breakwaters and nourishments). Next to this, in the same study, it was found that the SDS detection method cannot be considered the best solution to study shoreline evolution governed by storms or SLR. Mao (2018) confirmed in another study with a greater spatial scale (the African continent) that the SDS detection method can serve to obtain (historical) coastal system understanding, for human-induced shoreline evolution driven by ports.

As stated in Section 1.3, the drivers of shoreline evolution that are considered in this research are not related to the in-depth processes behind coastal erosion or accretion, but rather to the persistent or episodic phenomena that provide the basis for the shoreline evolution trend. Hence, the terms coastal erosion / accretion are replaced by coastal retreat / progradation from this point onwards. Different drivers of shoreline evolution might be present in features in the Shoreline Monitor. Drivers within features are identified up to tertiary influences. Only drivers that are likely to be detected as a main driver at least once in this research are presented in the sections below. Drivers of shoreline evolution like subsidence, in-situ weathering and attrition, and clay exposure are not likely to be detected as main drivers. These are not discussed nor touched upon in the remainder of this research.

Information on the background (among which direct or indirect human influences) and spatiotemporal characteristics of different human and natural drivers related to shoreline evolution are outlined in Section 2.2.1 and 2.2.2. The characteristics are provided in terms of spatial scale, temporal trend (either episodic or persistent) and overall spatiotemporal variability (either small, moderate or large). Only the spatial scale consists of units and is indicated with small (1 - 10 km), moderate (10 - 50 km) or large (> 50 km) terms. The last section, Section 2.2.3, provides an overview of Sections 2.2.1 and 2.2.2.

Key points

- Sandy shoreline evolution is driven by natural and human-induced changes (with a wide variability in spatial and temporal scales) to the sediment budget or the processes that work on them;
- Various types of drivers can be related to the spatiotemporal variability in multiple sandy shoreline evolution signals (features) computed by SDS;
- A total of eight non-process related drivers of shoreline progradation or retreat are considered in this research (the first four are human-induced): nourishments, land reclamations, port / breakwaters, river sediment supply, beach sediment supply, sand bar migration, tidal inlets and storms.

2.2.1 Human drivers

Four types of human drivers related to coastal retreat / progradation are considered:

1. Nourishments;
2. Land Reclamations;
3. Port / Breakwaters;
4. River Sediment Supply.

1. Nourishments (N)

A nourishment is a 'soft solution' counteracting coastal erosion, referring to the placement of sand by artificial means (human-induced beach sediment supply) in places where the loss or lack of sand is causing problems (Linham and Nicholls, 2010 and Bosboom and Stive, 2015). A sand nourishment leaves the coast in a more natural state than 'hard solutions' and preserves its recreational values (Bosboom and Stive, 2015). The nourishment solutions to coastal erosion have a direct impact, are used frequently since the 1970's by various countries (Stive et al., 2013) and provide a progressive episodic shoreline evolution signal.

Sand nourishments can be placed at three different locations in the coastal profile: on the inner slope or the outer slope of dunes, on the (dry) beach and on the shoreface (Bosboom and Stive, 2015). Consequently, three different types of sand nourishments are recognized: dune, beach and shoreface nourishments. Recently, a fourth type of nourishment called mega nourishment was introduced Stive et al. (2013). In a pilot project called the Sand Egnine, 21 Mm^3 was placed at a relatively small stretch of the Dutch coast (Stive et al., 2013). Based on the placement location of the nourishment, different natural processes redistribute sand. Dune and beach nourishments are mostly influenced by aeolian, shoreface nourishments by marine and mega nourishments by both aeolian and marine processes (Stive et al., 2013). A conceptual diagram of the different nourishments strategies is shown in Figure 2.6.

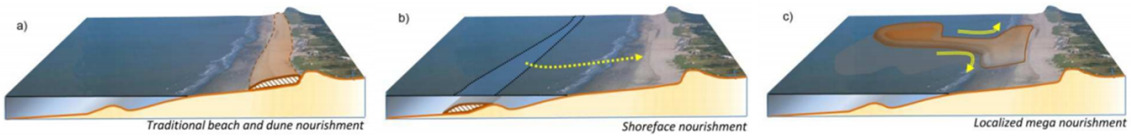


Figure 2.6: Traditional beach and dune nourishments, used frequently from the 70's onwards, place sand directly on the beach and dunes (a). Shoreface nourishments, initiated in the 90's, make use of marine processes to redistribute the sand that is placed under water in the cross-shore direction and gradually create a wider coastal defense over time (b). Mega nourishments, introduced in 2013, exploit both marine and aeolian processes to redistribute the sand both in cross and alongshore directions (c) (source: Stive et al., 2013)

Nourishments cannot halt persistent coastal erosion, but rather provide a locally placed temporary barrier against it. This means that, from time to time, a new nourishment has to be positioned (Bosboom and Stive, 2015). For this reason, the Dutch government uses schemes where design aspects like the type, place, time, location and volume of nourishments are specified. As a result of the large variation in design aspects, there is a large spatiotemporal variability present in the general nourishment term.

2. Land Reclamations (LR)

Donchyts et al. (2016) found that many countries have shaped and extended their shorelines by means of land reclamations. A land reclamation is referred to as the process where new land is created from oceans, river- and lake-beds. Land reclamations have a direct impact providing a persistent temporal shoreline evolution trend on a relatively small spatial scale. The large variation in design aspects results in a large spatiotemporal variability. Hot spots regarding land reclamations are found mainly in Asia, where China, Japan, Singapore, Hong Kong, Bahrain and the United Arab Emirates (UAE) stand out by showing large progressive shoreline signals (Luijendijk et al., 2018).

Figure 2.7: Visualization of land reclamations along the coast of Dubai. The green color indicates water converted to land, whereas the blue color indicates land converted to water. (source: aquamonitor.deltares.nl)

Figure 2.8: Visualization of land reclamations near Hangzhou, China. The green color indicates water converted to land, whereas the blue color indicates land converted to water. (source: aquamonitor.deltares.nl)

Visualizations of reclaimed land along the coast of Dubai and China are shown in Figures 2.7 and 2.8 above. The motives to reclaim land are highly diverse (Donchyts et al., 2016). The largest water-land change is found along the coast of Dubai, which provides more space for recreation (Davidson, 2009 and Donchyts et al., 2016). In Singapore, land reclamation is necessary to support economic growth.

3. Port / Breakwaters (PB)

Shore-normal breakwaters or a series of groynes are structures referred to as 'hard solutions' with a primary aim to change the longshore transport rates under both normal and extreme conditions (Bosboom and Stive, 2015). In many cases, shore-normal breakwaters are accompanied by a port. However, breakwaters can also be used as a means of river mouth stabilization or to prevent of persistent erosion at a tidal inlet (Bosboom and Stive, 2015). The usage of a breakwater is different from a series of groynes, the latter being generally used to stabilize an eroding coast (Bosboom and Stive, 2015). Breakwaters tend to influence the surrounding shorelines indirectly and locally by trapping longshore transported sediment, and provide a persistent progressive, retreating or both progressive and retreating temporal trend (Bosboom and Stive, 2015). This is dependent on various parameters like the amount of longshore transport due to wave power, the length of the breakwaters and surrounding sediment sources or sinks (Mao, 2018). Nonetheless, the overall spatiotemporal variability is small due to the indirect human influence (Garcin and Le Cozannet, 2013).

Figure 2.9: A sequence of images (1984, 1994, 2004 and 2014 from left to right) of the shoreline evolution around the Port of Nouakchott, Mauritania. (source: Mao, 2018)

A sequence of satellite images of the Port of Nouakchott, Mauritania (see Figure 2.9) shows one of the largest port breakwater influences on shoreline evolution present today (Mao, 2018). At Nouakchott, large unidirectional north-south longshore transport of sand has been blocked since 1986. This leads to shoreline retreat in the south and progression in the north of the port (Luijendijk et al., 2018).

4. River Sediment Supply (RSS)
Rivers act as a source of sediment to the coastal system and influence the moderate to large-scale sediment budget (Bosboom and Stive, 2015). The sediment supply of rivers to the coast can be reduced by means of human alterations like sand mining (extraction of sediment), the construction of dams (trapping of sediment) and changes to the land-use (stabilization of sediment) (Garcin and Le Cozannet, 2013 and Bosboom and Stive, 2015). This results in a retreating persistent shoreline evolution trend, when excluding seasonal cycles in rainfall (Rosati, 2005). The variety of indirect influences of human interventions to the river sediment supply result in an overall moderate spatiotemporal variability (Garcin and Le Cozannet, 2013).

Examples of extreme cases of decreases in river sediment supply are found in China. The Yellow River has approximately lost 90% of its sediment load over the past 60 years according to Wang et al. (2015). This is due to landscape engineering, terracing, the construction of dams and reservoirs, and vegetation restoration in China's Loess Plateau (Wang et al., 2015). Another river subject to a reduction in sediment load is the Yangtze river, which discharges to the Yellow sea adjacent to the city of Shanghai (Yang et al., 2005). The catchment area of the Yangtze river counts approximately 50.000 dams and reservoirs, where the Three Gorges Dam is the largest and most well known dam (Yang et al., 2005). Consequently, the growth rate of intertidal wetlands in the Yangtze river delta has decreased dramatically and is expected to reverse into recession if management policies are not adjusted (Yang et al., 2005).

2.2.2 Natural drivers
Also four types of natural drivers related to coastal retreat / progradation are considered:

1. Beach Sediment Supply;
2. Sand Bar Migration;
3. Tidal Inlets;
4. Storms.

1. Beach Sediment Supply (BSS)
The term beach sediment supply relates to shoreline evolution trends with both coastal progradation and coastal retreat and covers three types of phenomena in this research: the erosion of a headland or a cuspate foreland, re-orientation of the shoreline and sediment supply by eroding backshore / hinterland dunes. These phenomena are strongly related to wave or wind effects and only have a small spatial scale. Moreover, persistent temporal trends tend to flatten over time (Carter et al., 1990). This leads to a small spatiotemporal variability.

Bosboom and Stive (2015) stated that a larger part of the world's continental and island margins is lined with cliffs, consisting of rocks or cohesive clays. Cliffed coasts are commonly found on tectonically active convergent coasts, areas previously covered with Pleistocene glaciers or areas with carbonate sediments as a result of skeletal shell and coral debris (Bosboom and Stive, 2015). Due to nearshore bathymetry and refraction patterns, wave energy close to (cliffed) headlands is often higher than the surrounding shorelines (Carter et al., 1990). This ultimately leads to straightening of the coast as a result of eroding headlands and accreting adjacent bays or shorelines (Carter et al., 1990). Changes in the balance of wave energy also influence the shape of cuspate forelands as can be observed at some locations along the South-West Australian coast, where coral reefs protect the shorelines (Hollings, 2004).

Re-orientation of the shoreline results from sediment deficits at adjacent coastal stretches. This might be related to secondary effects of human interventions, as can be observed for various locations at the Nile Delta, Egypt (Frihy et al., 1998 and El Sayed et al., 2007). Beach sediment supply by eroding backshore / hinterland dunes might be reduced by stabilization due to secondary effects of human interventions like urbanization or vegetation growth as in Port Elizabeth, Africa (Bird and Lewis, 2015) as well as increased due to one prevailing wind direction as on the Caspian Sea shore of Kazakhstan.

2. Sand Bar Migration (SBM)
Both longshore and cross-shore / onshore sand bar or wave migration are considered for this driver of shoreline evolution. The spatial scale of both types of sand bar migration is small and the temporal trend shows both episodic and persistent signals of retreating and prograding shoreline evolution. The temporal trend is dependent on both human influences and natural variability like the magnitude in longshore transport Bosboom and Stive (2015). This results in a moderate overall spatiotemporal variability in the shoreline evolution signals.

Longshore sand bar migration can be found in situations where sediment bypasses shore-normal features like coastal headlands or human interventions. The latter is often found at the Nile Delta, Egypt (El Sayed et al., 2007). Seasonal variability in sediment discharges from rivers result in longshore as well as cross-shore sand bar migration and the formation of sandy spits (Bosboom and Stive, 2015). There is a strong correlation between natural cross-shore / onshore sand bar migration and the principal of shoreface nourishments. Nonetheless, natural cross-shore sand bar migration has the tendency to show stronger shoreline evolution signals as shoal migration towards the coast is also incorporated. An example of such a shoal migration is found near Warnbro, Australia (Hollings, 2004).

3. Tidal Inlets (TI)
Tidal inlets are openings in the shoreline that connect bays or lagoons to the open ocean (Bosboom and Stive, 2015). These inlets are maintained by tidal currents and are highly dynamical systems from a morphological point of view. Bosboom and Stive (2015) stated that the import or export of sand is dependent on flood- or ebb-dominance, which is related to basin geometry and sediment characteristics. Without interferences, the feedback mechanism of the basin geometry determines its own evolution (Bosboom and Stive, 2015). For long-term sediment transport through tidal inlets, human interference as well as SLR play a rol, as was shown for the Dutch Waddensea Wang (2018b). By excluding human interferences, Dissanayake et al. (2012) and Ranasinghe et al. (2013) proved that large tidal inlet / basin systems will import sediment from ebb-tidal deltas or adjacent shorelines due to relative SLR (SLR-driven basin infilling).

It was stated by Wang (2018a) that, by using the SDS detection method, the effects of SLR on open coasts are hard to capture. However, a case location at the French coast proved that the SDS method is able to capture (relative) SLR in persistently retreating shoreline evolution signals adjacent to tidal inlets (Wang, 2018a). The effects of SLR-driven basin infilling are noticeable on relatively small spatial scales, whereas there is an overall moderate spatiotemporal variability, which is dependent on various tidal inlet / basin system characteristics.

4. Storms (S)
Shorelines are subject to storm surge, large wave conditions and high-velocity winds as a result of extreme weather conditions referring to, depending on the location of occurrence, tropical cyclones, hurricanes or typhoons (Ranasinghe et al., 2012). This type of phenomenon leads to cross-shore sediment transport (Bosboom and Stive, 2015). Morton (2002) stated that storms induce beach erosion, berm migration, dune erosion, washover terrace construction, perched fan deposition, sheetwash, washover channel incision, washout formation, and forced and unforced ebb flow (listed from high frequency beach erosion to low frequency barrier island inundation). Wang (2018a) found that the SDS detection method is not the best solution to unravel shoreline evolution related to storms. This is due to the absence of satellite images induced by cloud cover. However, Wang (2018a) also stated that, at certain locations, a shoreline evolution signal does contain signatures of storm impact due to breaches and overwash. This discovery is supported by Dietz et al. (2018), who found hurricanes as a major driver of coastal erosion at the Mississipi River Delta, United States of America (USA) (see Figure 2.10).

Sediment overwash by episodic storm events leads to a loss of sediment in the coastal system (Bosboom and Stive, 2015), resulting in a rather persistent retreating shoreline evolution signal. Depending on the size of the storm, the spatial scale ranges from small to moderate. An overall moderate spatiotemporal variability is resulting from the effects of various storm characteristics like storm surge, wave conditions and wind velocity.

Figure 2.10: According to the United States Geological Survey (USGS), overwash along the Virginia barrier island shoreline changed the estuarine shoreline and buried wetland marshes and channels. The yellow arrow can be used as a reference mark in both images. (source: `https://www.usgs.gov/`)

2.2.3 Overview of natural and human-induced sandy shoreline evolution

Tables 2.1 and 2.2 provide a summary of the effects and characteristics of the main human and natural drivers of shoreline evolution, as outlined in Sections 2.2.1 and 2.2.2 respectively.

Table 2.1: A summary of the effects and characteristics of the human drivers of shoreline evolution. The column with spatial scales consists of units as indicated in the introduction of this section.

Human drivers of shoreline evolution	Prograding (+) and / or retreating (-)	Influence	Spatial scale	Temporal trend	Overall spatiotemporal variability
1. Nourishments	+	direct	small	episodic	large
2. Land Reclamations	+	direct	small	persistent	large
3. Port / Breakwaters	+ and -	indirect	small	persistent	small
4. River Sediment Supply	-	indirect	moderate / large	persistent	moderate

Table 2.2: A summary of the effects and characteristics of the natural drivers of shoreline evolution. The column with spatial scales consists of units as indicated in the introduction of this section.

Natural drivers of shoreline evolution	Prograding (+) and / or retreating (-)	Influence	Spatial scale	Temporal trend	Overall spatiotemporal variability
1. Beach Sediment Supply	+ and -	x	small	persistent	small
2. Sand Bar Migration	+ and -	x	small	episodic / persistent	moderate
3. Tidal Inlets	-	x	small	persistent	moderate
4. Storms	-	x	small / moderate	episodic / persistent	moderate

2.3 Machine Learning

ML is seen as a subset of Artificial Intelligence (AI) and is referred to as the science of getting computers to act without being explicitly programmed, but rather rely on patterns and inference instead. According to Pedregosa et al. (2011), in general, a learning problem considers a set of N samples of data. If each sample contains more than a single number, referring to a multi-dimensional entry (multivariate data), it is said to have several attributes or features. In this research, the words samples and attributes or features of the two sentences above are replaced by features and quantifications respectively, and are used throughout the remainder of this report. ML consists of two categories, supervised and unsupervised learning, which are explained in the list below (Pedregosa et al., 2011).

- With supervised learning, the data comes with additional labels that need to be predicted. Here, supervised learning considers two types of problems called classification and regression. Pedregosa et al. (2011) states that in the case of a classification problem, samples belong to two or more classes and it is opted to learn from already labeled data to predict the class of unlabeled data. In case of a regression, the desired output consists of one or more continuous variables;

- With unsupervised learning, the data comes without any corresponding labels. Here, unsupervised learning considers two types of problems called clustering and density estimation. Pedregosa et al. (2011) states that in the case of clustering, the goal is to discover groups / similarities within the data. In case of a density estimation, the goal is to determine the distribution of data within the input space.

Based on the research question and sub-questions from Section 1.4, either a supervised classification or unsupervised clustering method is appropriate for this study. Zhang and Ling (2018) stated that the predictive accuracy of the supervised classification method will be in the order of $0.67 * size^{-0.372}$, based on the size of the training dataset (i.e. the number of labeled data points). In other words, the predicting capability of classification algorithms with limited available manually labeled data points is questionable. Using the reasoning above, together with the example of using unsupervised learning in a coastal engineering application as in Burningham and French (2017), it seems more suitable to use a clustering algorithm for this research.

There is a large availability of unsupervised ML learning algorithms that can be used in case of a clustering problem. This is caused by the fact that clusters cannot be precisely defined (Estivill-Castro, 2002). Consequently, selecting the right algorithm to a clustering problem is often referred to as 'an art rather than science'. Generally, clustering algorithms attempt to minimize intra-cluster and maximize inter-cluster dissimilarity between features. By using the preprogrammed ML software of Pedregosa et al. (2011), the algorithm selection is already limited to nine distinct methods, among which K-Means and Gaussian Mixture Model (GMM). There are three distinct advantages of a GMM over K-Means, pointing to the probabilistic cluster assignment, the flexibility in cluster shapes and the absence of the requirement to pre-specify the number of clusters (Pedregosa et al., 2011). The latter two are incorporated in the algorithm hyper-parameters, which are referred to as the parameters that are manually set before the learning process begins. For this research, a GMM clustering algorithm is considered and outlined in the paragraphs below. Neither of these paragraphs go into full detail on the mathematical background, as all functions behind a GMM are well founded in literature and preprogrammed by Pedregosa et al. (2011).

Key points

- Unsupervised clustering methods are a part of ML in which the goal is to discover groups / similarities by minimizing intra-cluster and maximizing inter-cluster dissimilarity within the input data;
- A GMM, used in this research, is a probabilistic clustering algorithm that contains various options to constrain the covariance matrix and assumes the input data to be generated from a mixture of a finite number of Gaussian distributions with unknown parameters;
- To select the most appropriate number of clusters, a GMM incorporates information-theoretic criteria to overcome the generally acknowledged statistical problems of overfitting and underfitting.

Pedregosa et al. (2011) states that a GMM is a probabilistic model that assumes all features to be generated from a mixture of a finite number of Gaussian distributions with unknown parameters, $N(\mu, \sigma^2)$. The model might behave badly if the input data does not more or less look like standard normally distributed data (i.e. a Gaussian distribution with zero mean and unit variance) (Pedregosa et al., 2011). This problem can be solved by means of standardization or normalization of the input data, implying that the shape of a quantification distribution is ignored.

The main difficulty in learning a GMM from unlabeled data is that one usually doesn't know which feature came from which latent cluster. Here, a latent cluster is defined as a cluster that is not directly observed but rather inferred through a mathematical model. To get around the above-mentioned difficulty, a GMM implements the Expectation-Maximization (EM) algorithm for fitting a mixture-of-Gaussian models to iterate towards minimum intra-cluster and maximum inter-cluster dissimilarity between features. The EM algorithm uses two steps, the E- and M-step. In the E-step, random clusters are assumed and the probability of features being generated by each cluster of the model is computed. In the M-step, the cluster parameters are

adjusted by computing $(x-\mu)^2/\sigma^2$, which requires the μ and σ parameters to be estimated by maximizing the likelihood of the given feature-cluster assignments. By means of iterations, this process is guaranteed to always converge to a local optimum. To obtain a globally optimal solution, multiple random cluster initializations are used (Pedregosa et al., 2011).

The likelihood function $\mathscr{L}(\theta, y)$ describes the probability of obtaining the observed feature y within the parameters θ of the parameter space Θ. In case of a multi-dimensional parameter space, which is the case for a Gaussian distribution with parameters μ and σ, the maximum likelihood is found by computing $\hat{\theta} = argmax_{\theta\in\Theta}(\mathscr{L}(\theta, y))$. This process is called Maximum Likelihood Estimation (MLE) and is explained as computing the combination of values that are most likely to be the parameters of the joint probability distribution (covariance) underlying the observed feature y (Pedregosa et al., 2011). As the MLE function is non-linear, it is more convenient to compute the maximum likelihood through $\hat{\theta} = -\sum log(\mathscr{L}(\theta, y))$, which is referred to as the maximum log-likelihood.

According to Pedregosa et al. (2011), one can think of a GMM as generalization of K-Means clustering to incorporate information about the covariance (joint probability distribution) structure / matrix of the parameters as well as the centers of the latent Gaussians. There are different options to constrain the covariance matrix, which in fact controls the degrees of freedom of the shape and orientation of each cluster. These options are listed below, explained as outlined in Pedregosa et al. (2011) and supported by Figure 2.11.

- spherical, indicates that each cluster has its own single variance. This leads to circular clusters with equal dimensions. K-Means can be seen as a special case of a GMM with a spherical covariance matrix;
- diagonal, leads to ellipsoidal clusters constrained to align with the axes. This indicates independently variable cluster sizes along each dimension, as each cluster has its own diagonal covariance matrix;
- tied, all clusters share the same general covariance matrix. This leads to clusters that have the same size as well as orientation;
- full, each cluster has its own general covariance matrix. This allows the clusters to be represented as an ellipse with an arbitrary size and orientation.

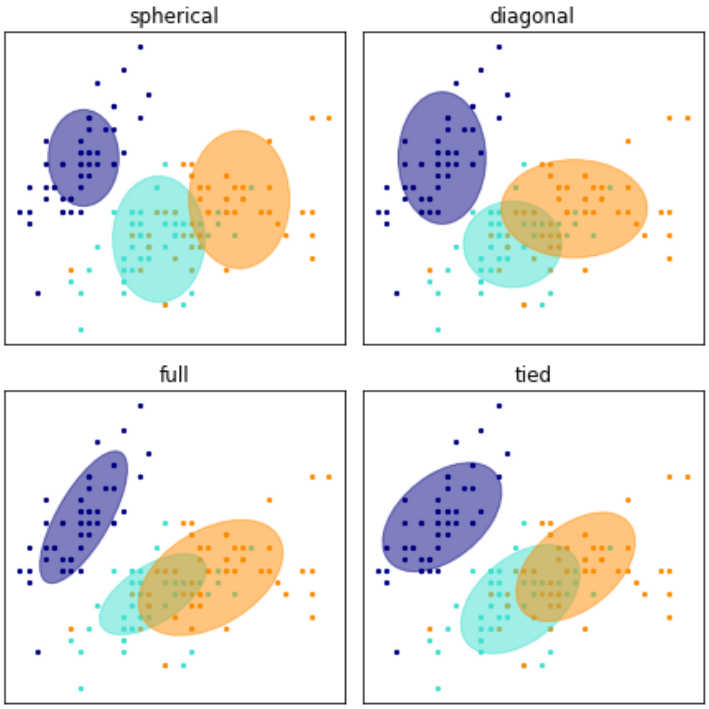


Figure 2.11: Four schematic two-dimensional visualizations of an example dataset containing labeled features (purple, blue and orange dots), after applying Gaussian Mixture Model (GMM) clustering with three clusters constrained by different types of covariance matrices. (source: adjusted from Pedregosa et al., 2011)

The GMM output assigns probabilistic or hard cluster labels to features. Cluster assignments can be explained from Figure 2.11 as well. By definition, hard cluster-feature labels refer to the list with highest probable clusters in which features are situated. These labels can be referred to as the actual color of the dots being equal to the clusters they belong to. Probable cluster-feature labels are defined as the matrix specifying the probability of a feature within all clusters. These labels can also be explained using Figure 2.11, where it is seen that some colored dots are further away from the mean or center of equally colored ellipses. The dots at a bigger distance from its ellipse have a smaller probability belonging to a cluster than the dots that are approximately at a cluster centroid. Here, cluster centroids or means indicate the differences between clusters in the dimensional space. It can be seen that some colored dots are situated in differently colored ellipses. This is explained by the possibility of a dot being separated in another dimension, not visualized in Figure 2.11 (Pedregosa et al., 2011).

Obviously, there are advantages and disadvantages of using a GMM as unsupervised ML clustering algorithm. Two advantages of a GMM are the computational speed and agnostic principle, where the latter indicates that this algorithm only maximizes the likelihood and does not bias any other parameter. Two disadvantages are related to singularities and the number of clusters (Pedregosa et al., 2011). Singularities refer to the difficulty of estimating covariance matrices whenever there are not enough features present. This leads to a diverging algorithm. The number of clusters refers to the fact that a GMM will always use all clusters it has access to.

The last disadvantage of the above-presented paragraph touches upon a subject that is generally acknowledged in statistics: overfitting and underfitting. Overfitting is related to a model that contains more variables than can be justified by the data, whereas underfitting is related to a model where some variables or terms that would appear in a correctly specified model are missing (Everitt and Skrondal, 2010). A GMM is prone to overfit if all clusters it has access to are enabled. Therefore, a GMM incorporates the computation of information-theoretic criteria, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), to select the correct number of clusters with respect to the number of variables. Pedregosa et al. (2011) states that in theory, AIC and BIC recover the true number of components only in the asymptotic regime. Apparently, BIC is better suited to identify the correct number of clusters (Pedregosa et al., 2011 and Chakrabarti and Ghosh, 2011).

3

Research Methods

As stated in the introduction, the objective of this research is to explore the possibility of exposing and classifying the drivers of shoreline evolution on a planetary scale, by employing ML on satellite imagery. Background information regarding the Shoreline Monitor (application of satellite imagery), human and natural drivers of shoreline evolution and ML is outlined in the previous chapter. This chapter contains the research methods that are developed / employed to fulfill the research objective and consists of three sections. Each of these sections represents the method to answer a single sub-question and can be visualized in a schematic overview consisting of seven steps and three domains (see Figure 3.1). All steps (except for step zero) are connected to either the local or global application domain. The principle of unsupervised ML, discovering groups / similarities within a large dataset, has a key role in the global application domain. The generation of smaller scale insight, the local application domain, functions to provide supervised vision to the unsupervised ML part to prevent it from becoming a 'black box'.

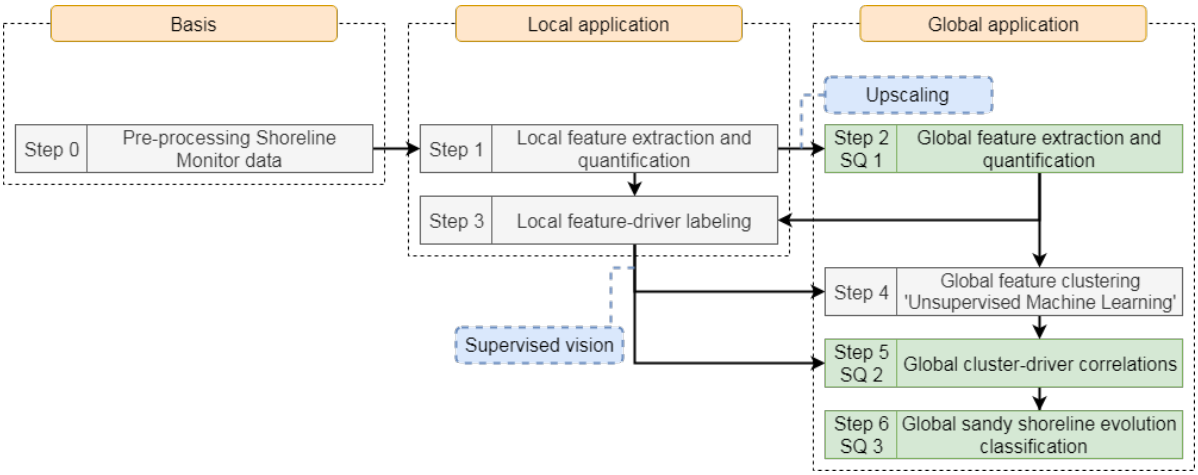


Figure 3.1: This visualization represents the different research methods in a step-wise plan. In total, there are seven steps which are divided into three domains (marked in orange): basis, local application and global application. All the sub-questions (referred to by the abbreviation SQ in the visualization) are answered in the global application domain, represented by step two, five and six (marked in green). All supporting steps to answer these questions are marked in grey. Extra information regarding the method is shown between parentheses in step four. Additional information on connected steps is shown in blue dashed boxes.

The basis of this research (step zero) considers pre-processing Shoreline Monitor data and is provided in Appendix B.1. Section 3.1, which is referred to by the underlined key words in sub-question one, contains the methods behind step one and two (see Figure 3.1). Section 3.2 represents sub-question two and contains the methods regarding step three, four and five. The last section (3.3), with the methods of sub-question three, contains only step six. All section introductions are ended with key points that consider concise statements on the most important methods.

Throughout the research, Python is used as scripting language to implement the methods described in the sections below. More information on this language is available here: `https://www.python.org/`.

3.1 Extract and quantify features from the Shoreline Monitor

The Shoreline Monitor of Luijendijk et al. (2018) contains a dataset with around 2.2 million sandy / non-sandy shore-normal transects with annual shoreline changes over the past decades. Human and natural drivers of shoreline evolution influence temporal behavior (by shaping the coastal zone) within multiple 500-m spaced transects in this dataset, considering the proof of Wang (2018a). Spatial as well as temporal influence of drivers of shoreline evolution is referred to as spatiotemporal variability and explained in Section 2.2.

The dataset behind the Shoreline Monitor is used to identify drivers by extracting and quantifying features along the world shoreline. Here, features are explained as multiple shoreline evolution signals / transects (spatial variability) with specific temporal variability. To be able to analyze these features, extraction and quantification algorithms are developed. These algorithms are based on the methodology behind the statistical procedure of Principle Component Analysis (PCA) or Empirical Orthogonal Functions (EOF), which are able to isolate dominant modes of spatial and temporal variability in a dataset with higher temporal and spatial resolution than present in this study (Miller and Dean, 2007 and Wang, 2018a).

The developed algorithms in this section consider both step one and two of Figure 3.1. These algorithms build upon step zero, provided in Appendix B.1, which allows for an 'along-the-shoreline'-analysis. In Section 3.1.1, the methods behind the local feature extraction and quantification algorithm are explained. Section 3.1.2 elaborates on the method of upscaling this algorithm to a planetary scale.

Key points

- A method is developed that is able to globally extract and quantify features, consisting of multiple shoreline evolution signals, from ordered transects on different spatial scales;
- The developed method computes twelve quantifications (four spatial and eight temporal) per feature, which are usable to explain annual spatiotemporal variability in shoreline evolution.

3.1.1 Local feature extraction and quantification

In order to capture both spatial as well as temporal variability using annual data, a feature extraction and quantification method is developed. The method to extract and quantify features in an 'along-the-shoreline'-analysis is referred to as a double-layer approach. In this approach, temporal variability is captured after spatial variability.

The development of the feature extraction and quantification method is based on two local-scale applications, considering shorelines of approximately 100 km. Hereafter, the method is tested on two moderate-scale applications with shorelines of 1000 km. The four stretches shown in Table 3.1 are selected to have a variability in two different characteristics in shorelines found in the Shoreline Monitor, which are stated below. Visualizations of the local and moderate scale application shoreline stretches presented in Table 3.1 can be found in Appendix B.2. The shoreline stretches in Table 3.1 are resulting from the the ordered continuous shoreline dataset by manually selecting transect id's in the Shoreline Monitor.

1. The uniformity of the shoreline;
2. The percentage of sandy transects.

Table 3.1: An overview of the different characteristics of the shoreline stretches that are used for step one of the research methods.

Shoreline stretch	Stretch start	Stretch end	Uniformity	Sandy transects [%]
The Netherlands	Hoek van Holland	Den Helder	high	99.5
Belgium	Bray-Dunes	Cadzand	moderate	99.6
South-West France	Biarritz	Plogoff	low	76.5
South-West Australia	Wilyabrup	Tamala	moderate	94.8

First layer: spatial variability

The first layer of the extraction and quantification method captures and characterizes features containing spatial variability in shoreline evolution on different spatial scales and uses the change rates of the linear regression. This change rate, whenever the uncertainty bandwidth is smaller than 50%, serves as an indicator of the representativeness of the linear fit and thus the shoreline evolution over time (Luijendijk et al., 2018).

Transects that were initially filtered out from the annual dataset behind the Shoreline Monitor and retrieved by pre-processing Shoreline Monitor data (see Section B.1) are excluded again. Five filters reject transects labeled 'non-sandy', labeled 'undetermined sediment composition', with a linear change rate bigger than 100 $m/year$, containing less than 5 out of 33 shoreline positions or having a temporal coverage smaller than seven years by putting the linear change rate to NaN.

Based on the recommendation of van Leeuwen (2018) (see Section 2.1.2), a new fault detection filter is built in. Ideally, due to changing transects spacings, transects are excluded on the basis of maximal change rate per m alongshore. For feasibility purposes, the sixth filter considers six different situations to exclude transects on the basis of absolute transect linear change rate differences with adjacent ones (by inserting NaN values based on a threshold called st). This is outlined in the numbered list below and schematically visualized (with the same list numbers) in Figure 3.2. Note that for explanatory purposes, only the case of a uniform shoreline is considered in Figure 3.2. Moreover, the transect colors in Figure 3.2 are representing positive and negative linear change rates, and not correct or filtered transects as in Figure B.1.

1. a positive change rate surrounded by two much bigger positive change rates;
2. a big positive change rate surrounded by two much smaller positive change rates;
3. a big positive change rate surrounded by two negative change rates;
4. a negative change rate surrounded by two much bigger negative change rates;
5. a big negative change rate surrounded by two much smaller negative change rates;
6. a big negative change rate surrounded by two positive change rates.

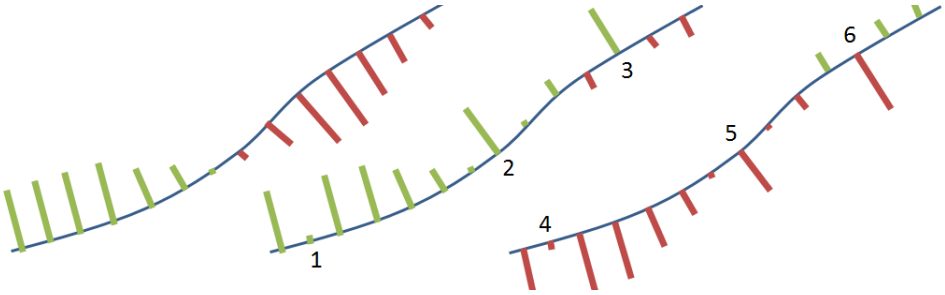


Figure 3.2: Schematic visualization of a uniform shoreline (blue) with correct linear change rates indicated by the fluent patterns on the left. Here, green (red) transects represent positive (negative) linear change rates. In the centered shoreline, the fault detection method excludes transects numbered with 1,2 and 3. On the right, the fault detection method excludes transects numbered with 4, 5 and 6.

To capture and quantify spatial variability in shoreline evolution on different scales, an along-the-shoreline moving average linear change rate is calculated. This is based on the summation of all available transect linear change rates divided by the total summed length between all transects closest to different moving windows of 2.5, 10, 20, 50 and 100 km, and is referred to as rolling mean from now on. The smallest moving window is set to 2.5 km in order to avoid localized hot spots or inaccuracies in the 500-m spaced transects. The rest of the moving windows are selected arbitrarily to incorporate drivers of shoreline evolution with a range of spatial scales. The minimum available transects should be at least a fixed percentage of the total number of transects in a moving window, again, to avoid localized hot spots or inaccuracies. The resulting computed rolling mean is appointed to the approximate center transect of the moving window, determined by the floored center position of the first moving window on the shoreline. A schematic visualization of this method on a uniform shoreline is shown for a moving window of 2.5 km in Figure 3.3.

Close to the equator, where the transect spacing is approximately 500 m, a total of six transect linear change rates is averaged for the 2.5 km window. If the minimum number of transects is set to be 70% of the total number of transects in a moving window, the algorithm starts generating rolling means at four transects (see

t_1 in Figure 3.3). The floored position is found at the second transect on this shoreline. Here, the rolling mean is appointed. The transects before the floored position of the first window do not contain a rolling mean value and are appointed NaN's. The windows only start to move and compute rolling means along the shoreline if there are six transects included from the start of the computation (this happens between step t_3 and t_4). At the end of the continuous shoreline, the windows decrease in size and stop computing rolling means whenever the minimum required number of available transects or the end of the list of transects is reached (t_{N-1} and t_N). The $t_{..}$ window shows the method to avoid a localized hot spot or inaccuracy. Here, the number of included transects is less than 70% of the total number of transects. This results in an appointed NaN value. It might happen that the spacing between two transects is so large, that there are only very little transects present in a moving window. This also induces localized hot spots in the rolling mean method. Therefore, a warning is raised if a minimum number of transects is not exceeded.

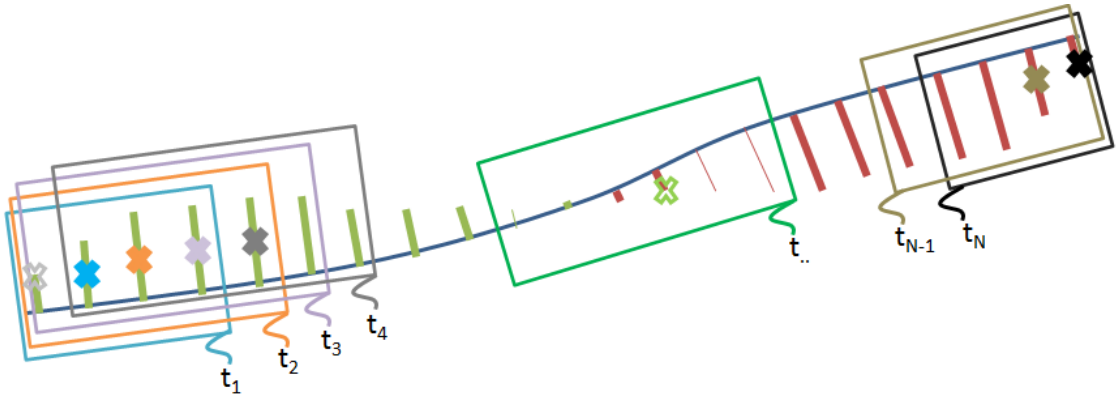


Figure 3.3: Schematic visualization of a uniform (blue) shoreline where the rolling mean method is applied. The thick green and red transects refer to positive and negative linear change rates, where the thin green and red transects refer to filtered / excluded transects with a NaN value. The colored square boxes represent the moving windows, which are labeled t_{xx}. The (empty) colored crosses represent the position and (NaN) value of the computed rolling mean.

With the rolling means for a shoreline stretch, various statistical parameters are calculated to explain the shoreline behavior at different spatial scales. These statistical parameters are listed and explained below.

- The mean, represents the average of all rolling means;
- The standard deviation (std), represents the amount of variation around the mean and is computed by $\sqrt{\sum_{i=1}^{N}(x_i-\overline{x})^2/N-1}$. Here x_i, $\overline{x}$ and N indicate the rolling mean, the average of all rolling means and the number of transects respectively;
- The max / min range, where the max (min) represents the maximum (minimum) rolling mean;
- The Probability Density Function (PDF), indicates the relative likelihood. This is explained as a point in the stretch where a random sample from the stretch has the exact similar value to that point;
- The standardized fit, provides a continuous function of the PDF. Here, the integral of the area under the function equals 1.

The automatic feature extraction method, for the different spatial scales, is based on the rolling means. This method uses a top-bottom approach, which is explained with the help of the uniform shoreline for an arbitrary spatial scale in the schematic visualization in Figure 3.4. First, all rolling means inside the stable regime (see Section 2.1.4) are excluded from the analysis. Hereafter, all the tops and bottoms are computed. This is done by comparing the change rate of a transect with its direct neighbors. A top (orange cross in Figure 3.4) is computed if the transect has a value higher than its neighbors above the stable regime or lower than its neighbors below the stable regime. For the detection of a bottom (black cross), this is exactly the other way around. If a straight line between two dots from Figure 3.4 passes through the shoreline, both these dots are recognized as bottoms as well. Up to this point, the top-bottom approach leads to a large number of tops and bottoms. These points are connected by means of a linear line, referred to as the top-bottom fit. The number of tops and bottoms is decreased by using a parameter called *cf* (see Figure 3.4). This parameter forms a threshold for the minimum required difference between consecutive tops and bottoms and removes local maxima and minima. By connecting the remaining points with a linear line, the reduced fit is obtained.

In order to move from the reduced fit to the feature fit, which in fact represents all the extracted features, *tb*, *hd* and *ia* thresholds are applied to remove remaining noise. Here, the *tb* threshold considers a minimum value between the height of a top compared to the projected linear line between two bottoms. *Hd* represents a minimum value for the height difference between a top and bottom after applying the *cf* and *tb* thresholds, while *ia* takes the maximum number of NaN valued transects compared to the total transects contained in a feature into account. In this way, on each different spatial scale, the extracted features capture spatial variability, represented by the rolling mean at the transect positions within that feature.

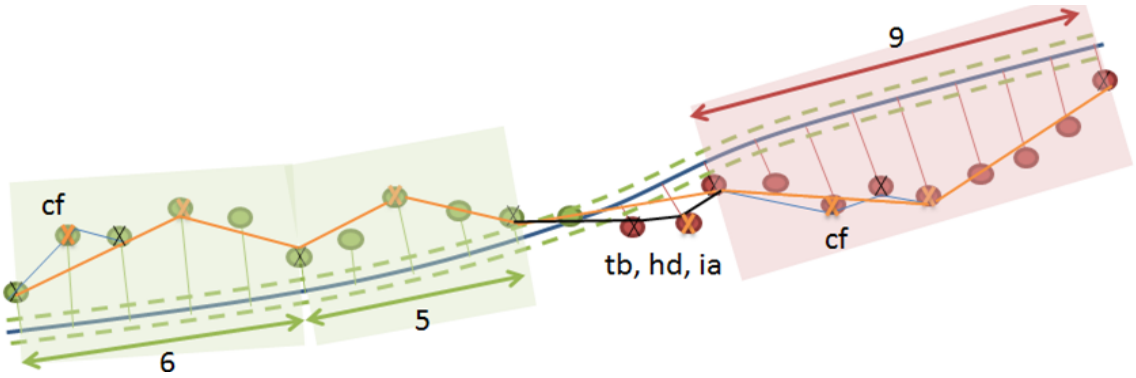


Figure 3.4: Schematic visualization of a uniform shoreline (blue) where the positive (negative) rolling mean values for transects are shown with green (red) dots and thin lines. The stable regime is indicated with dashed green lines and is characterized by the values of +/- 0.5 $m/year$. Computed tops (bottoms) are shown with orange (black) crosses. *Cf*, *tb*, *hd* and *ia* indicate thresholds which reduce the number of tops and bottoms. The orange line represents the feature fit, obtained after applying all thresholds. Based on this fit, the features are extracted. Positive (negative) features are indicated by transparent green (red) boxes and the feature-included transects with green (red) arrows with a number. The black line represents the reduced fit, which remains after applying only the *cf* threshold. The thin blue line is called the top-bottom fit, which is the fit through all tops and bottoms without any threshold reductions.

To check the reliability of the thresholds in the top-bottom approach, the top-bottom, reduced and feature fit are used to compute two different performance indicators. The first performance indicator is called Root Mean Square Error (RSME) and is calculated with the following formula: $\sqrt{(1/N)\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$. The second performance indicator is based on the Mean Absolute Error (MAE), which is calculated as follows: $(1/N)\sum_{i=1}^{N} \mid y_i - \hat{y}_i \mid$. In both formulas, N, y_i and $\hat{y}_i$ indicate the number of transects, the observed position and the predicted position respectively. For both performance indicators, lower scores are better. The difference in RMSE and MAE lies primarily in the fact that RMSE penalizes larger errors to a greater extent, which is desirable if large errors are not tolerated. Both indicators have the same unit as the dependent variable. Hence, no good or bad value can be connected to RMSE and MAE. The performance indicators are used, together with visual fit inspections, to determine the reliability of the thresholds for the development and test shoreline stretches from Table 3.1.

The quantification method in this layer considers characteristics of the spatial variability in the remaining features. Per feature in an arbitrary spatial scale, four quantifications are derived by using the rolling mean at the transect positions. The four spatial quantifications are listed and explained below, which is supported with the help of Figure 3.5.

1. Spatial width in km, represented by the alongshore distance between two black crosses within a feature (orange arrow in Figure 3.5). This quantification indicates the width of the shoreline evolution;
2. Spatial smallest skewness in $m/(year * km)$. This characteristic is represented by the smallest shore-normal difference between an orange and black cross, divided the alongshore distance between those crosses within a feature (gentlest blue arrow in Figure 3.5). Together with the next quantification, this characteristic indicates the asymmetry in the shape of the shoreline evolution;
3. Spatial largest skewness in $m/(year * km)$. This characteristic is represented by the largest shore-normal difference between an orange and black cross, divided the alongshore distance between those crosses within a feature (steepest blue arrow in Figure 3.5);
4. Spatial magnitude range in $m/year$, represented by the shore-normal distance between the orange cross and the projected line (dashed black line) between two black crosses within a feature (purple arrow in Figure 3.5). This quantification indicates the relative magnitude of the shoreline evolution compared to adjacent shorelines.

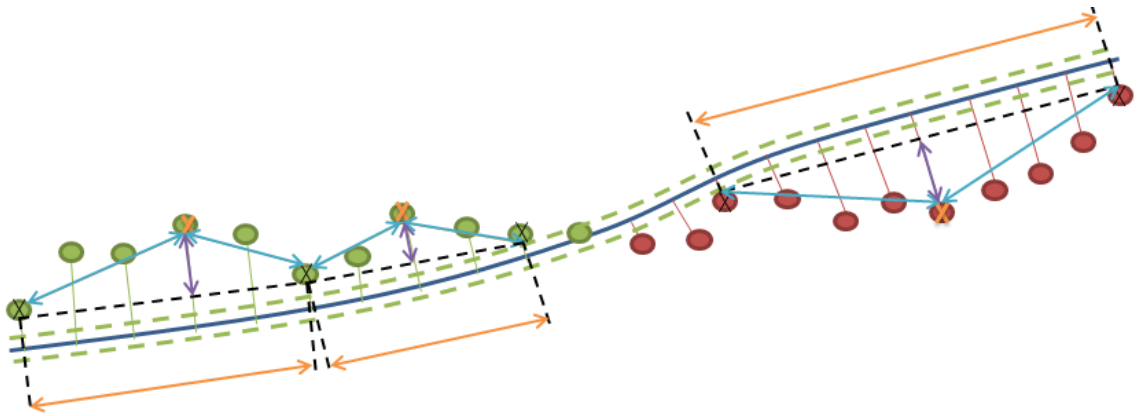


Figure 3.5: Schematic visualization of a uniform shoreline (blue) where the positive (negative) rolling mean values for transects are shown with green (red) dots and thin lines. The stable regime is indicated with dashed green lines and computed tops and bottoms with orange and black crosses respectively. The two light blue arrows indicate either the spatial smallest or largest skewness. The purple arrow provides the spatial magnitude range, whereas the orange arrow the spatial width. The black dashed lines are shown for guidance.

Second layer: temporal variability

In the second layer, the rolling means are rejected and replaced by the actual shoreline evolution signal at every feature-included transect position. In this way, temporal variability is added to the extracted features of the first layer. Figure 3.6 visualizes the aforementioned for the six transects of the first green feature of Figure 3.4. Due to the window averaging of linear change rates on different spatial scales, the edges of a positive feature may contain transects with a negative shoreline evolution trend. The other way around, it might also happen that the edges of a negative feature contains transects with positive shoreline evolution trends. These transects are excluded from the features before the temporal quantification method starts. This method requires various dataset modifications and is elaborated on in the paragraphs below.

By definition, the linear fits through shoreline evolution signals (not shown in Figure 3.6) start at 0 shoreline position. All the shoreline positions for one shoreline evolution signal are corrected by its connected linear fit. Therefore, a shoreline evolution signal does not necessarily start at a shoreline position that equals zero. Moreover, a signal might start later than 1984 and / or end earlier than 2016. Also, various shoreline evolution signals contain gaps, referring to an absent or filtered (outlier) shoreline position for a particular year (see red dots in Figure 3.6).

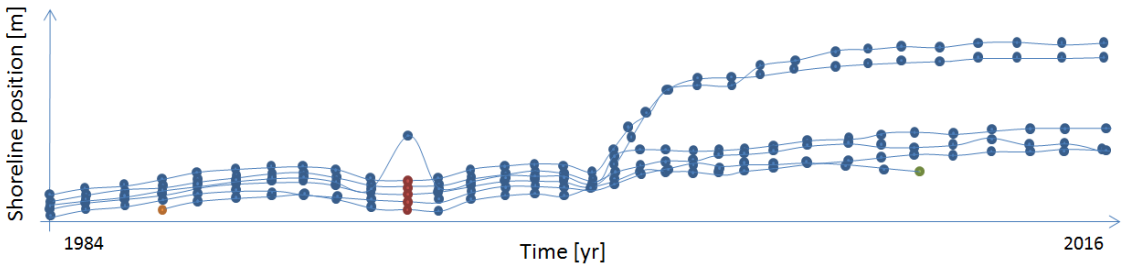


Figure 3.6: Schematic visualization of six shoreline evolution signals captured by the first green feature from Figure 3.4. The blue dots connected by the thin lines show the annual shoreline positions, at most 33 per signal. One orange (green) dot indicate the start (end) of a shoreline evolution signal that starts later than 1984 (ends earlier than 2016). The red dots represent gaps in the shoreline evolution signals.

It is rather complex to compare temporal variability between features when all transect-related temporal trends are present. For this reason, a mean feature trend is considered. This trend is obtained by averaging the summation of all shoreline positions per year. In case of the presence of gaps in shoreline evolution signals, this mean trend would be computed by only little shoreline positions. Therefore, gaps in shoreline evolution signals are linearly interpolated by using the shoreline positions at adjacent years. By computing the standard deviation of all shoreline positions compared to the mean shoreline position, information regarding the underlying transect-related shoreline evolution signals is preserved in the analysis. Both the mean feature trend (black line with crosses) and the standard deviation range are shown in Figure 3.7. A

linear regression to the mean feature trend provides the feature average change rate and uncertainty bandwidth. For explanatory purposes, the annual shoreline positions per single shoreline evolution trend (blue, red, orange and green dots in Figure 3.6) are not shown in Figure 3.7.

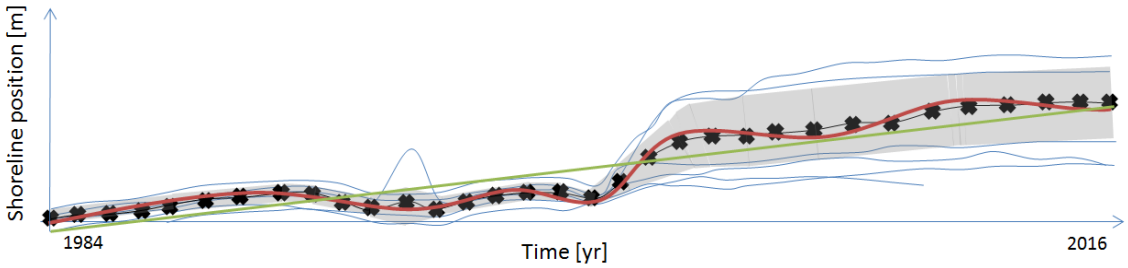


Figure 3.7: Schematic visualization of six shoreline evolution signals (blue lines) captured by the first green feature from Figure 3.4. The average shoreline positions per year are indicated with black crosses and connected with a black line. The gray band shows the standard deviation range around the mean feature trend. A thick red (green) line indicates the best (linear) curve-fitting model related to the mean feature trend.

Based on the first recommendation of van Leeuwen (2018) (see Section 2.1.4), a filter to reject unreliable results is applied. If a mean feature trend contains less than one-third of 33 shoreline positions (i.e. 10 or less), the belonging feature is disregarded. Apart from the linear regression (green line in Figure 3.7), the second recommendation of van Leeuwen (2018) considers the possibility of employing different fitting methods. By using the built-in models of Newville et al. (2014), seventeen preprogrammed curve-fitting algorithms (besides a linear regression) are applied to the mean feature trend. The application of these algorithms is consistent with the annual dataset behind the Shoreline Monitor, as it was also used to compute the initial linear regression. More information on the applied total of eighteen preprogrammed curve-fitting algorithms is included in Appendix B.2.

Different preprogrammed performance indicators are used to find the best fitting model per feature, out of the seventeen applied curve-fitting models. Newville et al. (2014) states that when comparing fits with different numbers of varying parameters, one typically selects the model with lowest reduced chi-square, AIC, and / or BIC. More information on these performance indicators is included in Appendix B.2. The performance indicators are different from the earlier introduced RMSE and MAE by incorporating the number of variable parameters of the fitting model. In this way, it is tried to balance the quality of fit with the number of variable parameters used in the fit, which is required to prevent overfitting or underfitting (Newville et al., 2014). This topic was introduced in Section 2.3.

Newville et al. (2014) stated that, generally, the BIC is considered the most conservative of the aforementioned performance indicators. Chakrabarti and Ghosh (2011) acknowledged that the BIC is the most suitable for selecting the correct model and the AIC is more appropriate in finding the best model for predicting future observations. However, due to the relatively small number of data points in a feature (at most 33), the AIC and in particular the corrected AIC (AICc) appeared to be more useful to select the correct curve-fitting model. More information on the AICc is included in Appendix B.2. The best fit to the mean feature trend, obtained by considering the minimum AICc, is used to correct all the data inside a feature. From Figure 3.7 it can be seen that all data is slightly lowered, such that the best fit starts at a shoreline position equal to zero.

The mean feature trend and the related standard deviation range, linear and best curve-fitted model (see Figure 3.7) allow to derive temporal quantifications. Both the linear and best curve-fitted model produce dependent variable parameters (values and std's) of the mathematical function, based on the independent variable shoreline position. Dependent variable parameters of the best-curve fitted model are not usable for quantification purposes due to the incomparability of these variables between features with different best fits. Moreover, neither of the annual shoreline positions in the mean feature trend or smoothed mean feature trend (best fit model) are used, as the latest start of a feature might be later than the earliest end of another feature. Per feature, there are eight quantifications regarding temporal variability that are extracted. These are based on the complete trend and are listed and explained below. The numbering in the below-presented list continues on the numbers of the spatial quantification, leading to a total of twelve quantifications for spatiotemporal variability in a shoreline evolution feature.

5. Temporal maximum annual value change in m, which is related to the mean feature trend (black line with crosses in Figure 3.7). This quantification elaborates on the largest annual step in the mean feature trend by computing the largest difference in shoreline positions between two consecutive years;
6. Temporal minimum annual value change in m, which is related to the mean feature trend (black line with crosses in Figure 3.7). This quantification elaborates on the smallest annual step in the mean feature trend by computing the smallest difference in shoreline positions between two consecutive years;
7. Temporal maximum annual std change in m, which is related to the standard deviation range (gray band in Figure 3.7). This quantification elaborates on the temporal variability of the underlying transect-related shoreline evolution signals in a feature. The largest difference in the standard deviation range between two consecutive years is computed;
8. Temporal minimum annual std change in m, which is related to the standard deviation range (gray band in Figure 3.7). This quantification elaborates on the temporal variability of the underlying transect-related shoreline evolution signals in a feature. The smallest difference in the standard deviation range between two consecutive years is computed;
9. Temporal maximum annual best fit change in m, which is related to the best fit shoreline evolution trend (red line in Figure 3.7). This quantification elaborates on the largest annual step in the best fit shoreline evolution trend by computing the largest difference in smoothed shoreline positions between two consecutive years;
10. Temporal minimum annual best fit change in m, which is related to the best fit shoreline evolution trend (red line in Figure 3.7). This quantification elaborates on the smallest annual step in the best fit shoreline evolution trend by computing the smallest difference in smoothed shoreline positions between two consecutive years;
11. Temporal change rate in $m/year$, which is related to the linear fit of the feature (green line in Figure 3.7). This quantification elaborates on the change rate of the mean feature trend and is computed by the first dependent variable value of the mathematical description of the linear fit (slope of the line);
12. Temporal change rate std in $m/year$, which is related to the (slope of the) linear fit (green line in Figure 3.7). This quantification elaborates on the correctness of the change rate of the mean feature trend and is computed by the first dependent variable std of the mathematical description of the linear fit.

3.1.2 Global feature extraction and quantification

The feature extraction and quantification method is developed and tested on shoreline stretches originating from the continuous shoreline from Appendix B.1. In this step, the local feature extraction and quantification method is upscaled to the global domain using the complete continuous shoreline. Upscaling this method allows to analyze features explaining shoreline evolution on a planetary scale and is only considered whenever the performance indicators and visual fit inspections regarding the various thresholds from Section 3.1.1 are found to be satisfactory.

Five statistical parameters, the same as in the first layer of Section 3.1.1, are computed to explain the shoreline behavior at different spatial scales for the global application. Moreover, fit performance indicators are presented to check the reliability of the thresholds in the top-bottom approach on a planetary scale. Eventually, a feature map is derived. This feature map indicates the geographical distribution of prograding and retreating features, in which a feature contains multiple sandy shoreline evolution signals that explain the spatiotemporal variability. Behind the map, there are two feature datasets (a prograding and a retreating) with four spatial and eight temporal variability quantifications. These datasets are used to compute the number of incorporated transects and influenced km of shoreline within extracted and quantified features. The total number of feature dataset-included transects and change rates is compared with the data incorporated in the total analyzed ordered continuous shoreline dataset. This forms a proxy for the representativeness of the data in extracted and quantified features compared to the analyzed global data. The variability in the extracted features is analyzed for all quantifications, using the same statistical parameters that are computed to explain global shoreline behavior.

3.2 Correlate feature similarities to drivers of sandy shoreline evolution

In Section 3.1, a method is developed that is able to extract and quantify shoreline evolution features globally. At various feature locations across the world, shoreline evolution may impose hazards. However, the drivers of shoreline evolution at these locations may not be known yet. This section describes the methods to correlate drivers of sandy shoreline evolution to the extracted and quantified features on a planetary scale, and refers to step three, four and five of Figure 3.1.

Before going directly to the global application, drivers are manually correlated to sandy shoreline evolution on a local scale by means of a labeling method presented in Section 3.2.1. The function of this step is twofold, as is it used to provide supervised vision to the global methods elaborated on in Section 3.2.2 and 3.2.3. Both the sections with global methods are founded on the hypothesis presented in Section 1.3. In Section 3.2.2, an unsupervised ML clustering method called GMM is employed to automatically group similarities in spatiotemporal variability in sandy shoreline evolution (see Section 2.3 for background information). Finally, Section 3.2.3 contains the method that correlates drivers of sandy shoreline evolution to the automatically clustered features from Section 3.2.2.

Key points

- A Machine Learning model is developed that automatically clusters similarities in spatiotemporal variability in shoreline evolution;
- A method is developed that correlates drivers of sandy shoreline evolution to automatically clustered features by using a supervised vision consisting of driver labels.

3.2.1 Local feature-driver labeling

It is seen from Figure 3.1 that step three is situated in the local application domain and obtains its input from both step one and two, while it provides supervised vision to step four and five. This section provides a method that generates local insight in the drivers of globally extracted and quantified features of shoreline evolution, before the quantifications are used to construct global correlations from the GMM output.

To generate insight, the four shoreline stretches from Section 3.1.1 as well as four new selected ones (all 1000 km shorelines) are used (see Table 3.2). The four new stretches are resulting from the continuous shorelines from Section 3.1.2 by manually selecting transect id's in the Shoreline Monitor, i.e.: Egypt, Mid-West Africa, Mid-East China and the Gulf of Mexico. Compared to Section 3.1.1, these new stretches are selected to incorporate variability in two other characteristics in shorelines found in the Shoreline Monitor. These characteristics are stated in the list below. Visualizations of the new local application shoreline stretches presented in Table 3.2 can be found in Appendix B.3.

1. Single driver influence on shoreline evolution;
2. Drivers with different spatial scales.

Table 3.2: An overview of the different characteristics of the shoreline stretches that are used for step three of the research methods.

Shoreline stretch	Stretch start	Stretch end	Single driver influence	Different spatial scales
The Netherlands	Hoek van Holland	Den Helder	no	no
Belgium	Bray-Dunes	Cadzand	no	no
South-West France	Biarritz	Plogoff	more or less	no
South-West Australia	Wilyabrup	Tamala	yes	no
Egypt	Rafah	El Salloum	yes	yes
Mid-West Africa	Accra	Akassa	more or less	yes
Mid-East China	Hutou Shanzui	Shanghai	yes	yes
South USA (Gulf of Mexico)	Matagorda	Morgans Point	no	yes

The feature-driver labeling method is explained as (subjectively) connecting various drivers of shoreline evolution (see Section 2.2) to extracted and quantified features from the eight shoreline stretches above, using the two different media listed below. The word 'subjectively' is stated between brackets as the first medium

of the below-presented list is not entirely dependent on user interpretation. Labels are added to features up to tertiary influences, which means that shoreline evolution signals within a feature might be identified to be correlated to a maximum of three drivers (no single driver influence). From the described media below, secondary as well as tertiary driver labeling is only possible when enough information is present.

- Literature, if available along the considered shoreline stretch, provides the most accurate method to correlate drivers to spatiotemporally variable features. Suitable literature is recognized as detailed studies into sedimentary budgets, forcing mechanisms or local datasets that keep track of interventions along the shoreline stretch;
- The Shoreline Monitor, the Aqua Monitor and GEE Timelapse all provide a way to determine the drivers of shoreline evolution in case literature is absent. The Shoreline Monitor is used to find the exact location of a feature along the shoreline (by using transect id's). Moreover, the high resolution background map is used to correlate drivers to features based on their spatial location. The Aqua Monitor provides quick insight into possibly erroneously extracted features. At the feature location, either green (water transferred to land) or blue (land transferred to water) patches must be present. This indicates persistent shoreline evolution over the past 33 years. GEE Timelapse makes it possible to look into annual changes to the shoreline over the past 35 years by means of a video. Only drivers of obvious changes to the shoreline are correlated to features using this method, as the resolution is lower than the aforementioned applications. More information on or visualizations of the Shoreline Monitor, the Aqua Monitor and GEE Timelapse can be found in Section 2.1, 2.2.1 and 2.1.1 respectively.

The variability in the extracted features with driver labels is analyzed on the total of twelve quantifications (see Section 3.1.1) per identified driver. Based on this analysis, it is decided whether it is needed to split the prograding and retreating feature datasets. Splitting datasets increases the ML model performance in case the variability between quantifications is high. This is elaborated on the next section.

3.2.2 Global feature clustering

This section outlines various topics regarding the employment of the unsupervised ML clustering method applied in this research. The feature quantifications from Section 3.1.2 function as input to a preprogrammed clustering algorithm called GMM (see Section 2.3). This algorithm automatically clusters similarities in spatiotemporal variability in shoreline evolution features around the world, creating various smaller prograding and retreating datasets based on feature-cluster labels. Below, different steps in the setup of the ML method are elaborated on: pre-processing model input, model setup and post-processing model output.

Pre-processing model input

From the quantification method of Section 3.1.1, it follows that all extracted features from the Shoreline Monitor contain four spatial and eight temporal characteristics (multivariate data). These feature quantifications are preserved in (split) prograding and retreating datasets (see Section 3.2.1). By means of duplication of the spatial quantifications, equal weight between spatial and temporal variability is obtained. This implies that the GMM input data has a mxn matrix structure as follows: N x 16. Here, N represents the number of extracted features in a dataset, while 16 represents the dimensional space consisting of eight spatial and eight temporal quantifications. The applicability of the quantification space is elaborated on further in Chapter 5.

To ensure a converging GMM, all feature quantification input is scaled according to the process of standardization by using $x^* = x - \overline{x}/\sigma$. Here, an individual feature quantification (x) is centered by removing the standard deviation-scaled mean value (σ / $\overline{x}$) of all feature quantifications.

Model setup

The first step in setting up a GMM considers developing a local model (LM) for every feature dataset. A LM consists of three hyper-parameters that deviate from the default settings of the preprogrammed GMM algorithm of Pedregosa et al. (2011) and is used to create initial insight. The GMM algorithm is fed with a large range of number of clusters, which is considered the first deviating hyper-parameter. Secondly, all types of covariance matrices are considered. The number of initializations to perform is a hyper-parameter that is not adjusted. By default, this hyper-parameter equals 1, which implies that a local optimum is obtained in a relatively short computational time. The third deviating hyper-parameter is referred to as the convergence threshold tolerance, which is set to a fixed number of 0.01. The EM iterations stop when the lower bound average gain is below this threshold. The tolerance is 10 times larger than the default setting to assure model

stability and a reasonable local optimum as result. By computing the minimum BIC value, a first order assessment regarding the best type of covariance matrix and number of clusters is obtained.

The second step considers developing a global model (GM) for every feature dataset. In order to obtain a reasonable global optimum, the GM consists of three hyper-parameters that deviate from the LM. The input of these parameters results from the created insight of the first order assessment from the paragraph above. The GMM algorithm is fed with a smaller range of number of clusters, more centered around the local optimum of the LM. Next to this, only the best type of covariance matrix of the LM is considered. The number of initializations to perform is the third hyper-parameter that is changed (set to a value in the order of 100). By computing the minimum BIC value, a global optimum is obtained in a computational time that is not significantly larger than the first order assessment. This optimum indicates the most suitable number of clusters to minimize intra-cluster and maximize inter-cluster dissimilarity between features in a dataset.

Post-processing model output

The GMM algorithm, with set hyper-parameters that belong to the global optimum, provides cluster means in a sixteen-dimensional space (same structure as the model input) as well as assigns probable or hard feature-cluster labels for the most suitable number of clusters in a dataset. The latter is analyzed by the methods presented in Section 3.2.3, while the cluster means are used to analyze the differences between the automatically computed groups of similarity in shoreline evolution in this section.

The cluster mean sixteen-dimensional space is reduced to a mean twelve-dimensional space for every dataset, which is possible as the four spatial quantifications were duplicated. For analysis purposes, two mean cluster quantifications are selected from the twelve-dimensional space. In this research, the spatial width and the temporal change rate are used to draw up cluster mean spatiotemporal signatures for the total of all datasets (see Figure 3.8). These signatures have three points, one top represented by the change rate and a two symmetrically positioned extents. The term 'width' is rejected as the symmetrical position induces a negative sign at the left of Figure 3.8. An arbitrary second-order polynomial connects the three points within a signature with a line. Whereas the three points have a physical meaning, the line connecting the points is only added for visualization or explanatory purposes and is stated to mimic the shape of shoreline evolution features as seen in Figures 3.4 and 3.5. Boundaries are added to the cluster mean spatiotemporal signatures (not shown in Figure 3.8). These so-called signature boundaries represent the line within which approximately 97.5% of the total number of extracted features fall. Two boundaries are already determined by the Shoreline Monitor's transect system (transect spacing or width at least 500 m) and SDS detection method (change rate at least +/- 0.5 $m/year$). Per cluster, the number of incorporated transects and influenced km of shoreline is computed as additional information.

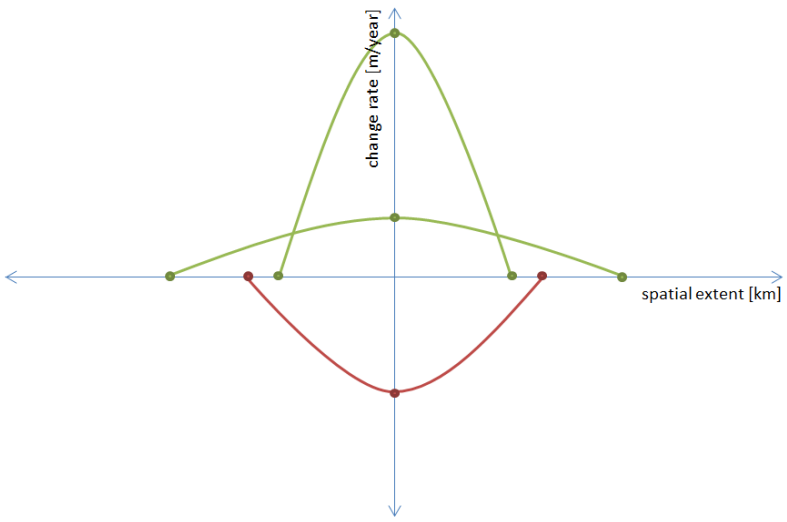


Figure 3.8: Schematic visualization of two prograding (green, upwards) and one retreating (red, downwards) cluster mean spatiotemporal signatures. These mean signatures indicate the differences between clusters based on two of the twelve quantifications. Per cluster signature, three points are present and referred to as one change rate top and two symmetrical extents (green and red dots). These three points are connected by a second order polynomial (green and red line), which is added for visualization purposes only.

3.2.3 Global cluster-driver correlations

In this section, the method to couple the local supervised vision feature-driver labels from Section 3.2.1 to the global unsupervised feature-cluster labels from Section 3.2.2 is elaborated on. Simply said, it opted to eliminate the 'feature'-term in both above-mentioned labels to find cluster-driver correlations. The method to derive cluster-driver correlations is presented in the first paragraph and the validation method regarding these correlations is outlined in the second paragraph. The third paragraph explains the method to analyze one of the outcomes of the validation method.

By definition, as outlined in Section 3.2.2, a GMM assigns hard and probable feature-cluster labels. Consequently, either hard or probable cluster-driver correlations can be computed. Per identified main driver (secondary and tertiary influences are ignored as we quantified the complete trend) resulting from Section 3.2.1, the assigned hard and probable feature-cluster labels can be traced back. With this information, the distribution of main drivers over clusters is analyzed, either be per identified driver or for summed drivers (visualized in Figure 3.9). In case of an analysis per identified driver, the assigned probable feature-cluster labels have a probability distribution over clusters, whereas the assigned hard feature-cluster labels refer to only one cluster. For the case of summed drivers, both the summed assigned probable and hard feature-cluster labels show a probability distribution over clusters. The case of summed drivers is used to determine whether hard or probable assigned feature-cluster labels are preferred. Whenever the hard and probable probability distributions are comparable, a hard cluster assignment is stated to be sufficient. For the preferred cluster assignment, cluster-driver correlations are revealed by means of bar plots. Per cluster, these plots contain computed counts and percentages (probabilities) for an identified driver.

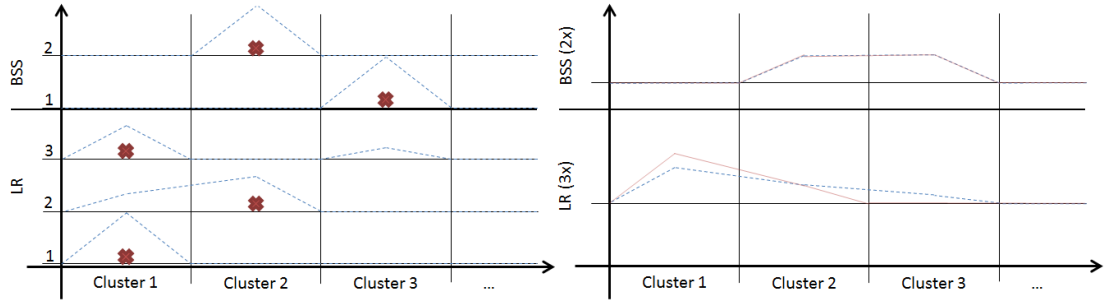


Figure 3.9: Schematic visualization of probable and hard feature-clusters labels over three clusters per identified main driver (left) or per summed main driver (right). The drivers are indicated by BSS (Beach Sediment Supply) or LR (Land Reclamation). The numbers in the left schematic refer to individual features, whereas the numbers between brackets behind the drivers in the right schematic refer to the number of summed features. The probable feature-cluster labels are probability distributions in both the left and right plot (blue dashed line). Hard feature-cluster labels are indicated with a red cross in the left and a red line (probability distribution) in the right schematic.

The validation method consists of a few steps. First of all, the extracted and quantified features of the local application shoreline stretches are excluded from the clusters. These features were used to derive the cluster-driver correlations and cannot be used again for validation purposes. Secondly, per cluster, ten features are randomly selected and put into separate datasets. Each of the randomly selected features in a cluster is manually labeled with drivers of shoreline evolution, following the same method as presented in Section 3.2.1. Subsequently, the probability of a driver being present in a cluster is computed. Three options appear when the cluster-driver correlations from the validation method are compared to the ones from the paragraph above. These options are listed below. By comparing the summed number of a manually labeled driver from Section 3.2.1 with the summed number of a validated driver, a proxy for the reliability (in terms of drivers) of the manually analyzed features compared to randomly selected features is presented.

- Not comparable. The computed probabilities (probability) of drivers (a driver) being present in a cluster are (is) rejected and only a cluster with features containing similarities in spatiotemporal variability in shoreline evolution remains. More insight is required to correctly correlate the drivers of shoreline evolution to this cluster;
- Approximately comparable, one driver of shoreline evolution present. The probability is preserved, leading to a cluster with one driver correlation which can easily be understood if the outcome is mapped;
- Approximately comparable, multiple drivers of shoreline evolution present. The probabilities are preserved, which leads to a cluster with driver correlations. More analyses are required to enhance the understanding of the above-stated outcome. This is explained below.

In case the majority of features belong to the last item in the list above, the focus is on (in)direct human and natural influences instead of drivers of shoreline evolution. Here, direct human influences are referred to as nourishments and land reclamations. Indirect human influences consider port / breakwaters and river sediment supply, where natural influences consist of beach sediment supply, sand bar migration, tidal inlets and storms. The feature-driver labels from Section 3.2.1 are transfered to feature-influence labels. The distributions of (in)direct human and natural influences are added to the cluster mean spatiotemporal signature visualization of Section 3.2.2, not taking cluster assignments into account. In this way, a first order assessment regarding obvious differences in the aforementioned influences on shoreline evolution is obtained.

3.3 Derive a planetary-scale classification

The developed and employed methods from Section 3.2 create semi-automated insight in the globally extracted and quantified features of Section 3.1 by computing cluster-driver correlations. These validated correlations are derived from a number of manually labeled features in a cluster and indicate the probable driver presence for all features in the same cluster. Drivers of shoreline evolution are transfered to (in)direct human and natural influences as outlined in Section 3.2.3. The coming paragraphs describe the method to analyze the aforementioned influences, in a more objective way than the first order assessment, to derive a planetary-scale classification of sandy shoreline evolution (step six of Figure 3.1).

Key point

- A method is developed that considers cluster maps to classify sandy shoreline evolution using spatial and temporal characteristics on a planetary scale.

Geographical driver distributions of all features in a cluster are exposed by generating cluster maps. This is only done for a selection of interesting (combined) clusters with driver correlations. In case a single cluster mean spatiotemporal signature obviously differs from all other cluster signatures, a single cluster map is derived. Whenever a prograding and retreating cluster show similar spatiotemporal signatures that differ clearly from other signatures, a combined cluster map is presented. Depending on the main message a (combined) cluster map is selected to radiate, feature locations are enlarged on one of the three different characteristics listed below. Cluster-driver correlations indicating the probable presence of drivers of shoreline evolution are shown in the legend of (combined) cluster maps, whenever these were validated to be correct, and provide support to the below-presented items.

- Temporal change rate (quantification) in $m/year$, if a cluster spatiotemporal signature has a relatively large change rate. A cluster map with enlarged prograding or retreating feature locations based on this characteristic indicates the presence of (in)direct human influences on shoreline evolution. However, in some cases, this characteristic can also indicate natural influences;
- Spatial width (quantification) or radius in km, if a cluster spatiotemporal signature has a relatively large width. Here, the radius is explained as the presence of another prograding or retreating feature in a 50 km radius. A cluster map with scaled features according to one of these characteristics outlines the presence of natural influences on shoreline evolution. Nonetheless, in some cases, this characteristic can also indicate indirect human influences;
- Pairs in km, if none of the cases above are applicable. Pairs are explained as the presence of a prograding and retreating features in a 15 km radius. A cluster map with highlighted features based on this characteristic touches upon the presence of (in)direct human as well as natural influences on shoreline evolution, also referred to as compound influences or interactions (see Chapter 1).

The selected maps with enlarged features and supporting information are used to derive various temporal and spatial thresholds for probable maximum and minimum (in)direct human and natural influences on shoreline evolution. The percentiles of manually labeled feature influences above or below these thresholds, depending on a probable maximum or minimum, are computed. This forms a proxy for the reliability (in terms of influence) of the cluster analysis derived feature influence thresholds considering the manually labeled feature influences. A statistically generated planetary-scale shoreline evolution classification, presented in terms of regimes, is based on the signature boundaries of Section 3.2.2 and the thresholds derived in this section. The thresholds and classification regimes are shown in a cluster mean spatiotemporal signature visualization like the one shown in Figure 3.8.

To indicate the relative importance of the regimes when speaking about planetary-scale shoreline evolution features, the percentages related to the number of features within the statistically generated shoreline evolution classification regimes are computed. Next to this, also the percentages related to the number of manually labeled feature influences within the regimes are computed. This forms a proxy for the representativeness (in terms of regimes) of the manually analyzed features compared to the total of global features.

As globally extracted and quantified features contain continent labels, the mean spatiotemporal signature of all shoreline evolution features within a continent can be derived. The main influence on shoreline evolution for a continent is revealed by the position of that continent signature within the statistically generated shoreline evolution classification.

4
Results

All presented results consider the revised annual dataset behind the Shoreline Monitor, where unique and small-scale inaccuracies in the global transect system were solved manually. The findings that form the basis for this research (step zero - Pre-processing Shoreline Monitor data), are provided in Appendix C.1. These findings indicate that around 54% of the Shoreline Monitor's available transects as well as the global shoreline length represented by these transects is taken into consideration for an 'along-the-shoreline'-analysis. An elaboration on the research-included data can be found in Chapter 5.

Based on the developed and employed methods presented in Chapter 3, this chapter provides the answers to the three sub-questions from Section 1.3 in three separate sections. Within each of these sections, results are split in a number of steps according to the schematic overview as seen in Figure 3.1 (similar to Chapter 3). Section 4.1 elaborates on the findings that consider extracting and quantifying shoreline evolution features with possibly hazardous effects on a planetary scale. This section contains the results behind step one and two. In Section 4.2, the results regarding step three, four and five are presented. This section considers the global generation of insight in the drivers behind extracted and quantified shoreline evolution features. In Section 4.3, the final outcome of a planetary-scale sandy shoreline evolution classification is discussed (step six). All section introductions are ended with key points that consider concise statements on the most important results.

4.1 Extract and quantify features from the Shoreline Monitor

This section consists of two steps that result in the global extraction and quantification of features from the Shoreline Monitor. In Section 4.1.1, the results of the local feature extraction and quantification method are analyzed. The outcome of upscaling this local feature extraction and quantification method to a planetary scale is elaborated on in Section 4.1.2.

Key points

- In this research, approximately 985 thousand transects (around 390 thousand km of shoreline) from the dataset behind the Shoreline Monitor are taken into consideration;
- After locally assessing the performance indicators related to various thresholds in the feature extraction and quantification method for a 2.5 km spatial scale, two global datasets with 3033 prograding and 2121 retreating features are obtained;
- Up to this point, it is known where shoreline evolution features may impose hazards to coastal communities, whereas the drivers behind these features are not brought to light yet.

4.1.1 Local feature extraction and quantification

The local feature extraction and quantification method uses linear regression change rates and various thresholds to identify shoreline evolution features, as explained in Section 3.1.1. Within these features, spatial variability is captured before temporal variability. The thresholds that are incorporated in the feature extraction and quantification method, one for the fault detection filter (st = 5.5) and four for the actual feature extraction method (cf = 1.0, tb = 0.5 , hd = 0.5 and ia = 0.5) are found iteratively by comparison of the results of two local-scale and two moderate-scale applications. The minimum number of required transects captured in a moving window (before a warning is raised) is set to 4, 8, 15, 19, 71 and 141 for spatial scales of 2.5, 5, 10, 20, 50 and 100 km respectively. Obviously, as the spatial scale increases and more transects are incorporated, this requirement weighs less. The results are explained by highlighting two (one local-scale development and

one moderate-scale test) shoreline stretches from Table 3.1 in Section 3.1.1, namely: The Netherlands and South-West France. The results of the other two shoreline stretches are presented in Appendix C.2.

The Netherlands

The uniform shoreline stretch from Hoek van Holland to Den Helder has a length of around 118 km and contains 379 shore-normal defined transects. So, on average, there is a spacing of around 312 m between two consecutive transects. This is different from the initially stated 500 m spacing between transects, as explained in Appendix B.1. After applying six filters (see Section 3.1.1), 370 of the stretch-included transects contain a linear change rate (9 transects are set to NaN).

A first step towards extracting and quantifying features with spatial variability on different scales results from setting out the rolling mean method along the flat considered shoreline stretch. This is visualized in Figure 4.1, where the location of Hoek van Holland is represented by the left and Den Helder by the right of the plot. Here, the rolling mean is computed on 2.5, 5, 10, 20, 50 and 100 km scales with the minimum number of required transects being 70% of the total number of transects in a moving window. Every single dot in Figure 4.1 represents a transect-related shoreline evolution signal with a spatial scale averaged linear change rate.

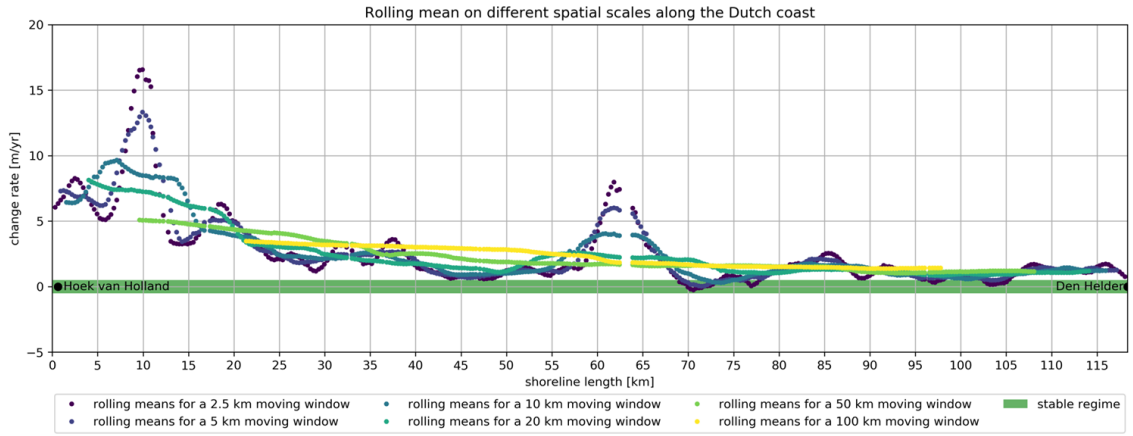


Figure 4.1: Visualization of the application of the rolling mean method (pre-work for the actual feature extraction and quantification method) to the Dutch shoreline stretch of Hoek van Holland (left) to Den Helder (right) for different spatial scales. The rolling means of each of the six different spatial scales are indicated with a different color. The stable regime is visualized with a green band.

Statistical results of the application of the rolling mean method on the 370 stretch-included transects with linear change rates are shown in Figure 4.2. This is another way of visualizing the effect of the different spatial scales on the variability along the shoreline and supports findings with numbers more easily. Figures 4.1 and 4.2 indicate that the smaller scale rolling means show a great prograding variability around the mean, which is decreasing with an increasing spatial scale due to averaging. The shoreline mean of the rolling means is prograding with approximately +2 to +3 $m/year$, with some variability between the spatial scales.

From this point onwards, the focus is on the 2.5 km scale to explain the results related to the feature extraction and quantification method for the case of the Dutch coast. In Figure 4.3, the result of the first step of the top-bottom approach in the first layer of the aforementioned method is shown. It can be seen that there are a lot of initially identified tops and bottoms connected by a line called top-bottom fit. Applying the *cf* threshold results in the exclusion of local maxima and minima. The remaining tops and bottoms are represented by a reduced fit that only slightly deviates from the top-bottom fit for the considered shoreline stretch.

In Figure 4.4, the result of the final step of the top-bottom approach in the first layer of the feature extraction and quantification method is visualized. Compared to Figure 4.3, the number of tops as well as bottoms are reduced again. This time, this results from the application of *tb*, *hd* and *ia* thresholds that remove remaining noise. The feature fit, the line that connects the remaining tops and bottoms, results in the extraction of a total of seven prograding features (without warnings) from the considered shoreline stretch. These features contain multiple shoreline evolution signals (dots in the visualizations) that are obviously different from the adjacent ones. A number of smaller 'bumps' in Figure 4.4 are not recognized as features, for instance from

approximately 24 to 28 km. This is a clear disadvantage of the application of various thresholds to come to the feature fit and is elaborated on in Section 4.1.2 for the global application and in Chapter 5.

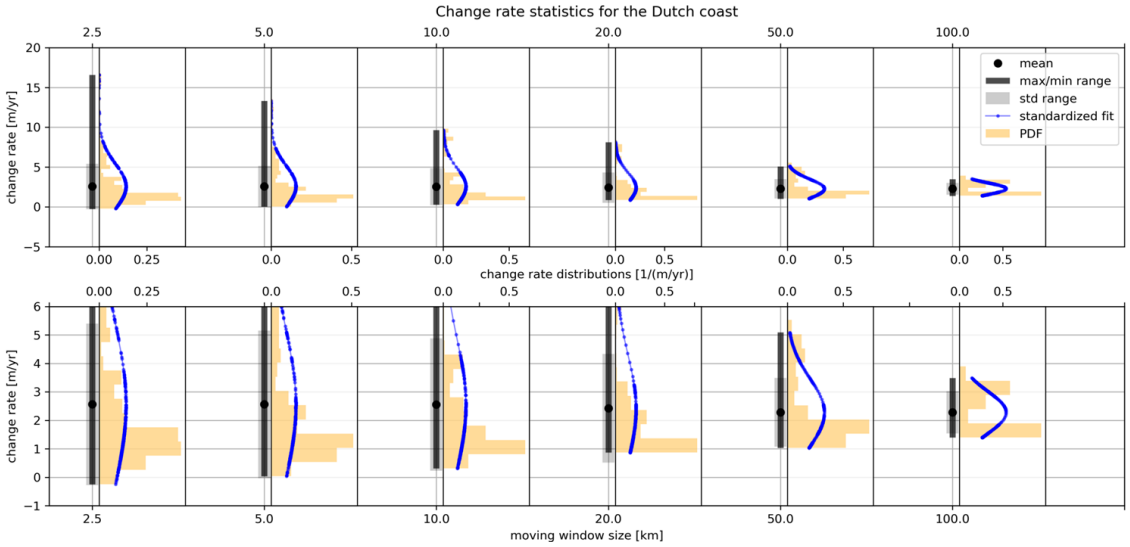


Figure 4.2: Two-panel statistical visualization of the application of the rolling mean method to the Dutch shoreline stretch for different spatial scales, represented on the top of the top panel and bottom of the bottom panel x-axes. The lower panel is a zoom-in on the standard deviation range (grey bar) of the upper panel. The mean of the rolling means is indicated with a black dot and the maximum / minimum range with a black bar. The x-axes between the two panels show the distribution of change rate values, in a PDF with orange bars and a standardized fit with a thin blue line with dots, within the shoreline stretch maximum / minimum rolling mean range.

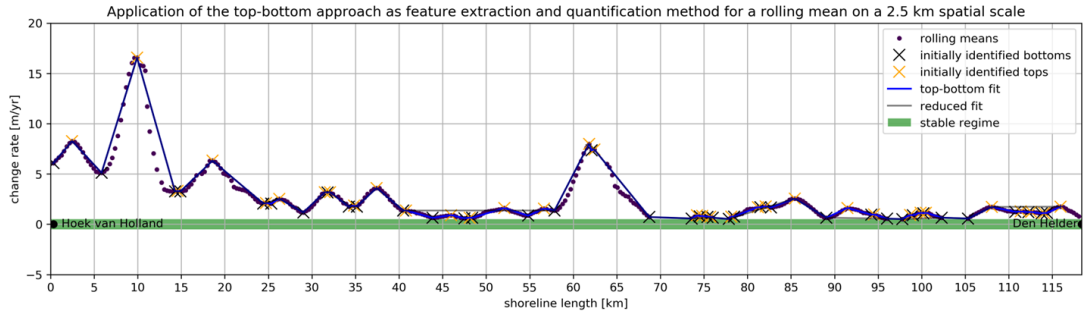


Figure 4.3: Visualization of the application of the first step in the top-bottom approach for the Dutch coast on a 2.5 km scale. The rolling means are indicated with dark purple dots and the initially identified tops and bottoms with orange and black crosses respectively. The stable regime is shown with a green band. The top-bottom fit and the reduced fit are shown with a thin blue and black line, the latter slightly deviates from the top-bottom fit due to the *cf* threshold.

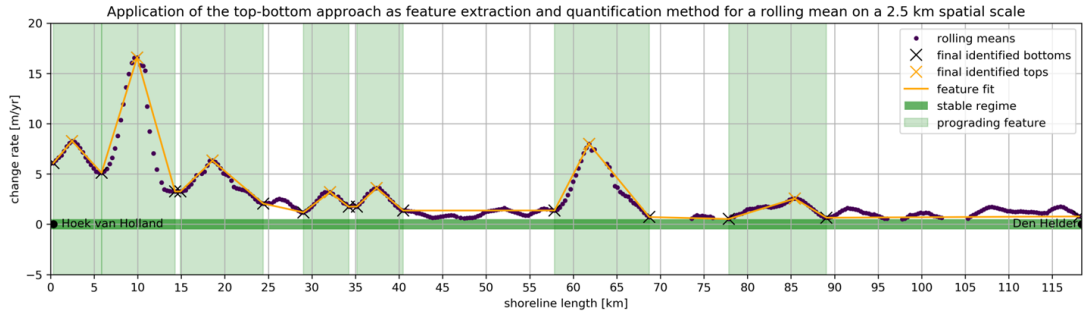


Figure 4.4: Visualization of the application of the final step in the top-bottom approach for the Dutch coast on a 2.5 km scale. The rolling means are indicated with dark purple dots and the final identified tops and bottoms (after applying the *tb*, *hd* and *ia* thresholds) with orange and black crosses respectively. The stable regime is shown with a green band and the feature fit with a thin orange line. The identified prograding features are shown with transparent vertical green bands, indicating the total feature range.

By visually inspecting Figures 4.3 and 4.4, it is stated that the top-bottom approach with its thresholds performs accordingly. The top-bottom, reduced and feature fit are seen to represent the alongshore rolling means quite well. This is supported by the absence of significant differences between fit performance indicators RMSE and MAE, which are presented in Table 4.1. From Table 4.1 it is also seen that, compared to the MAE, the RMSE penalizes larger errors to a greater extent. The errors become bigger from top-bottom to feature fit, which is caused by the application of thresholds.

Table 4.1: Performance indicators RMSE and MAE for the top-bottom, reduced and feature fit. These are related to the feature extraction and quantification method development shoreline stretch of the Dutch coast.

Type of fit	RMSE $[m/year]$	MAE $[m/year]$
Top-bottom	0.6	0.24
Reduced	0.66	0.37
Feature	0.68	0.41

The seven extracted prograding features along the Dutch coast are spatially quantified on the four characteristics presented in Section 3.1.1. These feature quantifications are shown in Table 4.2, where the features are numbered one to seven from left to right in Figure 4.4. Based on Table 4.2, it is shown that feature seven has the largest width, whereas feature two has the largest magnitude range compared to adjacent shorelines. Moreover, it is computed that feature four is approximately symmetrical, while feature six is more skewed.

Table 4.2: The four spatial quantifications of the extracted prograding features that result from the development shoreline stretch of the Dutch coast. The green-highlighted prograding feature row is considered for the temporal quantification.

Feature	Width $[km]$	Smallest skewness $[m/(year*km)]$	Largest skewness $[m/(year*km)]$	Magnitude range $[m/year]$
1	5.5	0.9	1.0	2.6
2	8.4	2.8	3.1	12.4
3	9.5	0.7	0.8	3.5
4	5.2	0.7	0.7	1.7
5	5.3	0.7	0.8	2.0
6	10.9	1.1	1.7	6.9
7	11.2	0.3	0.5	1.9

A first step towards quantifying temporal variability in the extracted features results from replacing the rolling means by actual feature-included shoreline evolution transect signals. Figure 4.5 shows an example of the outcome of the aforementioned for the second green feature of Figure 4.4. Twenty-five transects are included in this feature. All transects cover the range from 1984 to 2016 as well as miss shoreline positions from 1992 to 1997. The shoreline evolution signals in the middle of the feature (greenish transects) show the largest temporal variability in shoreline position (approximately around 800 m). This is mainly due to the rapid change around 2012 / 2013. The same transects show an erosional trend from 2013 onwards. Contrarily, the transects at the sides of the feature (purple and yellow) show a rather persistent prograding trend.

The eight temporal quantifications presented in Table 4.3 result from a total of four dataset modifications outlined in the developed methods in Section 3.1.1. The interpolated shoreline positions that fill the gaps in Figure 4.5 are solely computed to derive the modifications and are not seen as a result. The aforementioned four dataset modifications are visualized in Figure 4.6 for the same prograding feature as in the figure above and refer to the mean feature trend, the standard deviation range, the best mean curve-fitting model and the linear mean curve-fitting model. The latter allows to compute the feature change rate, referred to as change rate from now on, which is different from the single transect-related change rate used in first layer of the extraction and quantification method. The considered feature from Figure 4.6 shows a maximum annual shoreline position change of 137 m between two consecutive years and has a change rate of around 9.8 $m/year$. Next to this, the change rate std is 1.4 $m/year$, which indicates that the linear change rate is not particularly accurate.

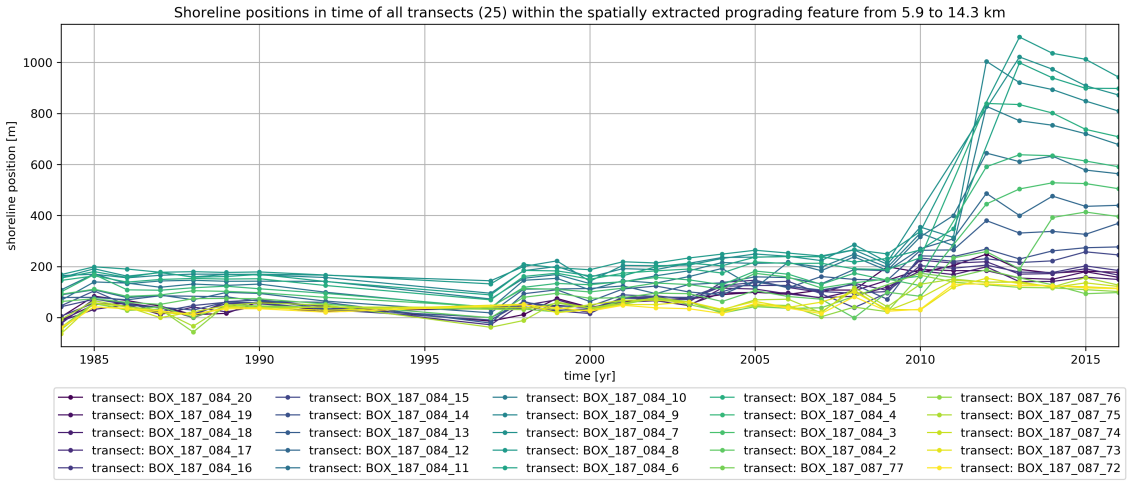


Figure 4.5: Example visualization for the second layer in the feature extraction and quantification method on a 2.5 km scale. This example considers the second prograding feature from the left (5.9 to 14.3 km in Figure 4.4) on the Dutch coast. Twenty-five transects are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. In case of the absence of an annual shoreline position for a temporal trend, no dot is visualized. The temporal trends are identified by their label (see legend).

Table 4.3: The eight temporal quantifications of the extracted second prograding feature that results from the development shoreline stretch of the Dutch coast. The labels of the temporal quantifications are split in two sets of four, represented by the first and third row.

Feature	Maximum annual value change $[m]$	Minimum annual value change $[m]$	Maximum annual std change $[m]$	Minimum annual std change $[m]$
2	137.0	-19.6	140.5	-22.4
Feature	**Maximum annual best fit change $[m]$**	**Minimum annual best fit change $[m]$**	**Change rate $[m/year]$**	**Change rate std $[m/year]$**
2	62.1	-75.1	9.8	1.4

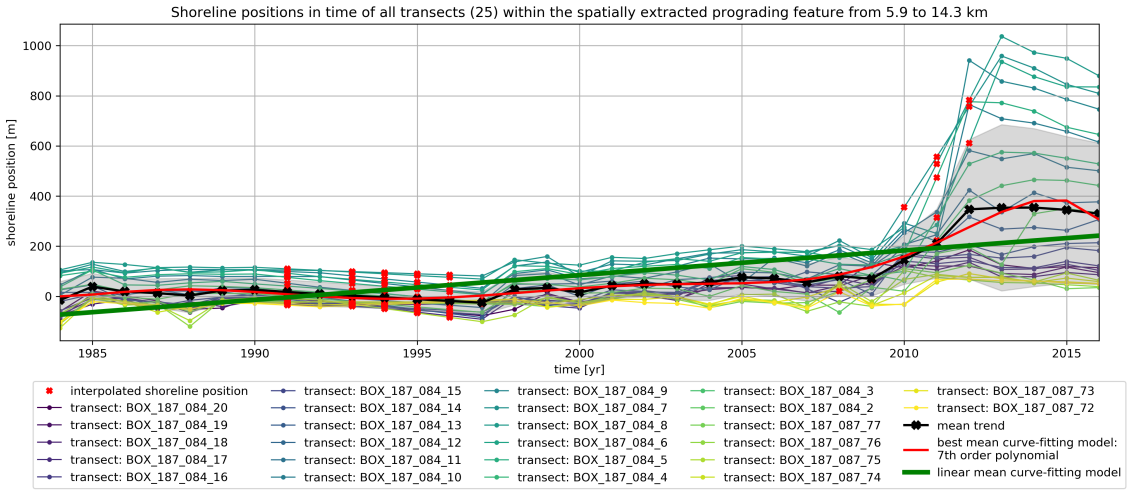


Figure 4.6: Example visualization for the quantification of temporal variability in the second layer of feature extraction method on a 2.5 km scale. This example considers the second prograding feature from the left (5.9 to 14.3 km in Figure 4.4) on the Dutch coast. Twenty-five transects (see label id's in the legend) are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. In case of the absence of an annual shoreline position for a temporal trend, a red interpolated dot is shown. The mean (feature) trend, obtained by summing all shoreline positions per year, is indicated with a black line with crosses. The linear fit and best fit to the mean trend are shown in a thick green and red line respectively. The gray band shows the standard deviation range around the mean trend.

South-West France

The non-uniform shoreline stretch from Biarritz to Gavres has a length of around 986 km and contains 2732 shore-normal defined transects. So, on average, there is a spacing of around 361 m between two consecutive transects. This is different from the above-presented case of the Dutch coast, which can be explained by the French coast being somewhat closer to the equator as well as having more frequent accidental large spacings between transects. After applying six filters, 1999 of the stretch-included transects contain a linear change rate (733 transects are set to NaN). The high number of NaN values is primarily due to the relatively low percentage of sandy transects along this shoreline stretch.

Again, the rolling mean method is set out for the spatial scales of 2.5, 5, 10, 20, 50 and 100 km along the flat considered shoreline stretch. This is visualized in Figure 4.7, where the location of Biarritz is represented by the left and Gavres by the right of the plot. The minimum number of required transects in a moving window is kept the same as for the Dutch coast.

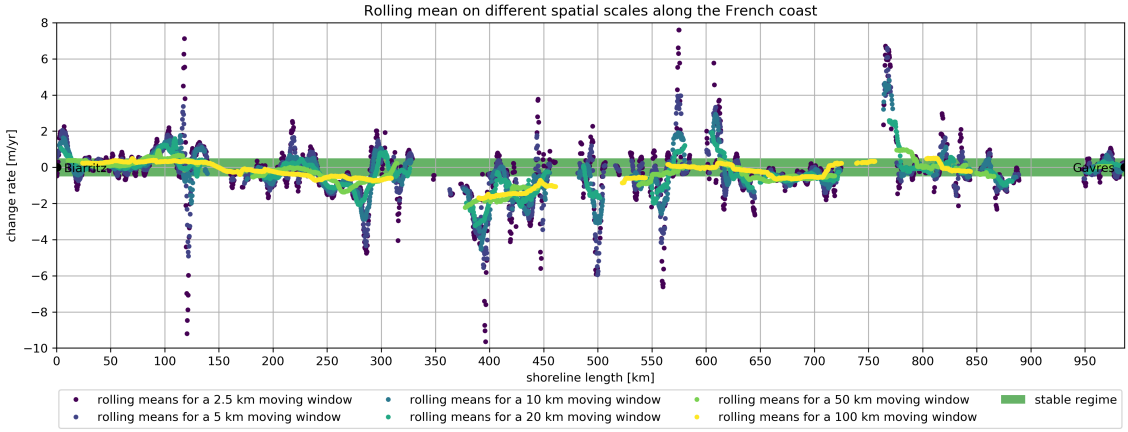


Figure 4.7: Visualization of the application of the rolling mean method (pre-work for the actual feature extraction and quantification method) to the French shoreline stretch of Biarritz (left) to Gavres (right) for different spatial scales. The rolling means of each of the six different spatial scales are indicated with a different color. The stable regime is visualized with a green band.

By comparing Figure 4.1 (Dutch coast) and Figure 4.7 (French coast), it is found that statistical visualizations become more important when the considered shoreline stretch is significantly larger than 100 km. Statistical results of the application of the rolling mean method on the 1999 stretch-included transects with linear change rates are shown in Figure 4.8. As can be seen from Figures 4.7 and 4.8, the smaller scale rolling means show a great variability around the mean. Peaks are found in the prograding as well as retreating regime. The variability around the mean is decreasing with an increasing spatial scale. The mean of the rolling means are indicating a shoreline that is retreating with approximately 0 to -0.5 $m/year$, without too much variability between the spatial scales.

From this point onwards, again, the focus is on the 2.5 km scale to explain the results related to the feature extraction and quantification method for the case of the French coast. In Figure 4.9, the result of the first step of the top-bottom approach is shown. The exclusion of local maxima and minima, by application of the *cf* threshold, results in a reduced fit that only slightly deviates from the top-bottom fit for this shoreline stretch.

In Figure 4.10, the result of the final step of the top-bottom approach is visualized. Compared to Figure 4.9, the number of tops as well as bottoms are reduced once more by the application of *tb, hd* and *ia* thresholds. The feature fit results in the extraction of a total of thirteen prograding and seventeen retreating features with one warning for the second last retreating feature from the left. When analyzing Figure 4.10, it is difficult to see whether smaller 'bumps' are not recognized as features. This is caused by the higher number of non-sandy transects, the non-uniformity of the shoreline and the significantly larger considered shoreline stretch compared to the case of the Dutch coast (see Section 4.1.2 and Chapter 5 for a more in-depth analysis).

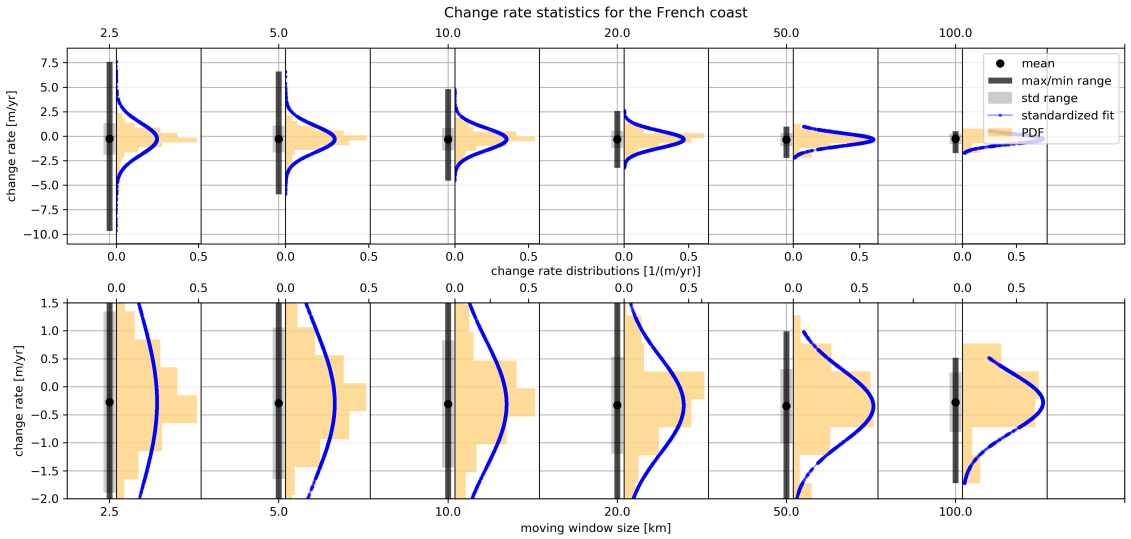


Figure 4.8: Two-panel statistical visualization of the application of the rolling mean method to the French shoreline stretch for different spatial scales. These scales are represented on the top of the top panel and bottom of the bottom panel x-axes. The lower panel is a zoom-in on the standard deviation range (transparent grey bar) of the upper panel. The mean of the rolling means is indicated with a black dot and the maximum / minimum range with a black bar. The x-axes between the two panels show the distribution of change rate values, in a PDF with orange bars and a standardized fit with a thin blue line with dots, within the shoreline stretch maximum / minimum rolling mean range.

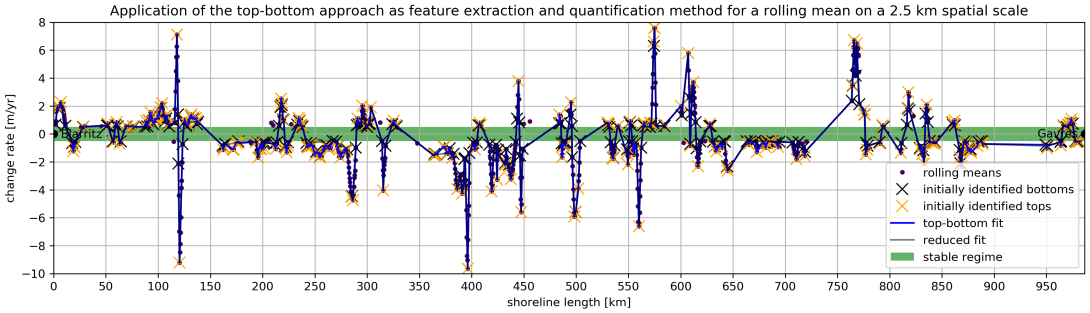


Figure 4.9: Visualization of the application of the first step in the top-bottom approach for the French coast on a 2.5 km scale. The rolling means are indicated with dark purple dots and the initially identified tops and bottoms with orange and black crosses respectively. The stable regime is shown with a green band. The top-bottom fit and the reduced fit are shown with a thin blue and black line, the latter slightly deviates from the top-bottom fit due to the *cf* threshold.

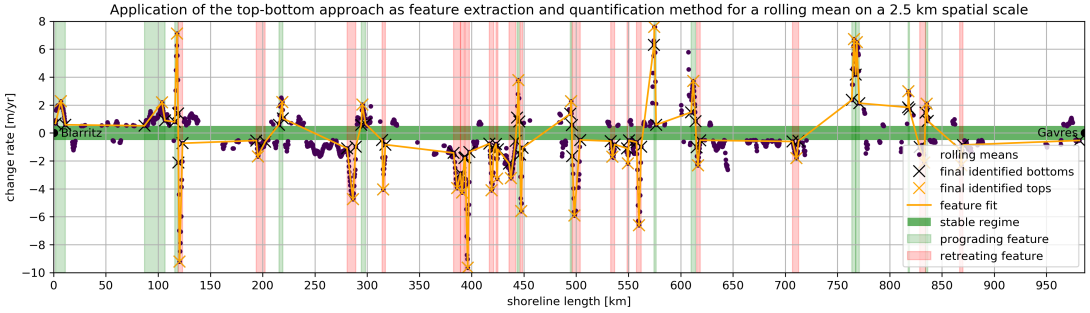


Figure 4.10: Visualization of the application of the final step in the top-bottom approach for the French coast on a 2.5 km scale. The rolling means are indicated with dark purple dots and the final identified tops and bottoms (after applying the *tb*, *hd* and *ia* thresholds) with orange and black crosses respectively. The stable regime is shown with a green band and the feature fit with a thin orange line. The identified prograding (retreating) features are shown with transparent vertical green (red) bands, indicating the total feature range.

Figures 4.9 and 4.10 imply that the top-bottom approach with its thresholds captures most of the features / peaks with rolling mean shoreline evolution signals. Where highly scattered rolling means are represented by the top-bottom and reduced fit, the feature fit does not take these points into account. This is supported by the relatively large difference between the reduced and feature fit performance indicators RMSE and MAE, which are presented in Table 4.4. Nonetheless, the thresholds are still stated to be sufficient as the number of included transects for the French coast is around 5.5 times the number of included transects for the Dutch coast.

Table 4.4: Performance indicators RMSE and MAE for the top-bottom, reduced and feature fit. These are related to the feature extraction and quantification method test shoreline stretch of the French coast.

Type of fit	RMSE $[m/year]$	MAE $[m/year]$
Top-bottom	0.27	0.11
Reduced	0.38	0.24
Feature	0.91	0.59

The total of thirty extracted features along the French coast are spatially quantified on the four characteristics presented in Section 3.1.1. These feature quantifications are shown in Table C.8 in Appendix C.2, where the prograding and retreating features are numbered one to thirteen and one to seventeen from left to right in Figure 4.10 respectively. From this table, it is seen that prograding feature two has the largest width, whereas retreating feature seven has the largest magnitude range compared to adjacent shorelines. Various features are approximately symmetrical, while some features are highly skewed.

The rolling means are replaced by actual feature-included shoreline evolution transect signals to quantify temporal variability in the extracted features. Figure 4.11 shows an example of the outcome of the aforementioned for the eight red feature (around 420 km) in Figure 4.10. Ten transects are included in this feature. All transects cover the range from 1984 to 2016 as well as miss shoreline positions from 1989 to 1999. Moreover, all transects show approximately the same temporal behavior with a variability in the angle of the trends. The transects at the left of the feature (purple) show smaller trend angles than the transects at the right (yellow).

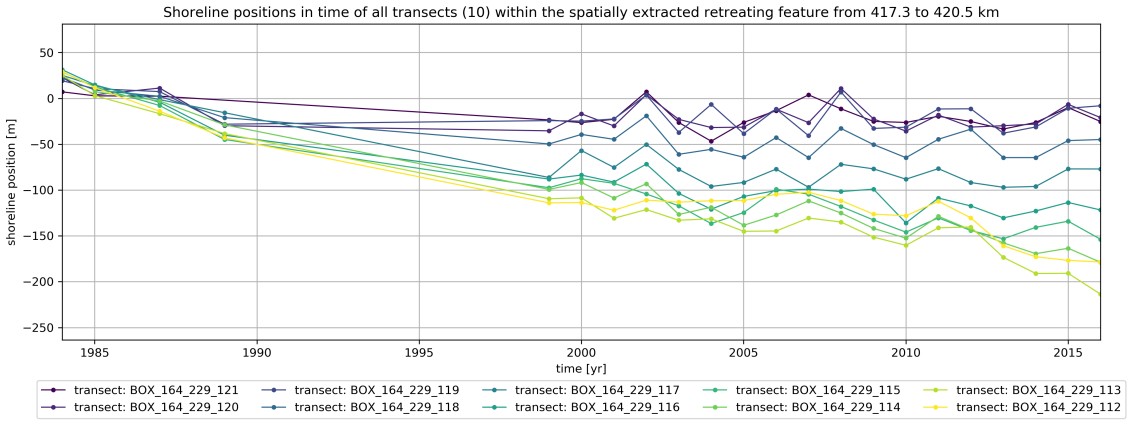


Figure 4.11: Example visualization for the second layer in the feature extraction method on a 2.5 km scale. This example considers the eight retreating feature from the left (417.3 to 420.5 km in Figure 4.10) on the French coast. Ten transects are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. In case of the absence of an annual shoreline position for a temporal trend, no dot is visualized. The temporal trends are identified by their label, presented in the legend.

The eight temporal quantifications presented in Table 4.5 result from the total of four dataset modifications. The interpolated shoreline positions that fill the gaps in Figure 4.11 are only computed to derive the aforementioned. The four dataset modifications are visualized in Figure 4.12 for the same retreating feature as in the figure above. The considered feature from Figure 4.12 shows a maximum annual shoreline position change of -26.5 m between two consecutive years and has a change rate of around -3.2 $m/year$. The change rate std is 0.2 $m/year$, which indicates that the linear change rate is quite accurate for this feature.

Table 4.5: The eight temporal quantifications of the extracted eight retreating feature that results from the test shoreline stretch of the French coast. The labels of the temporal quantifications are split in two sets of four, represented by the first and third row.

Feature	Maximum annual value change [m]	Minimum annual value change [m]	Maximum annual std change [m]	Minimum annual std change [m]
8	18.5	-26.3	13.2	-8.4
Feature	Maximum annual best fit change [m]	Minimum annual best fit change [m]	Change rate [$m/year$]	Change rate std [$m/year$]
8	-1.4	-11.3	-3.2	0.2

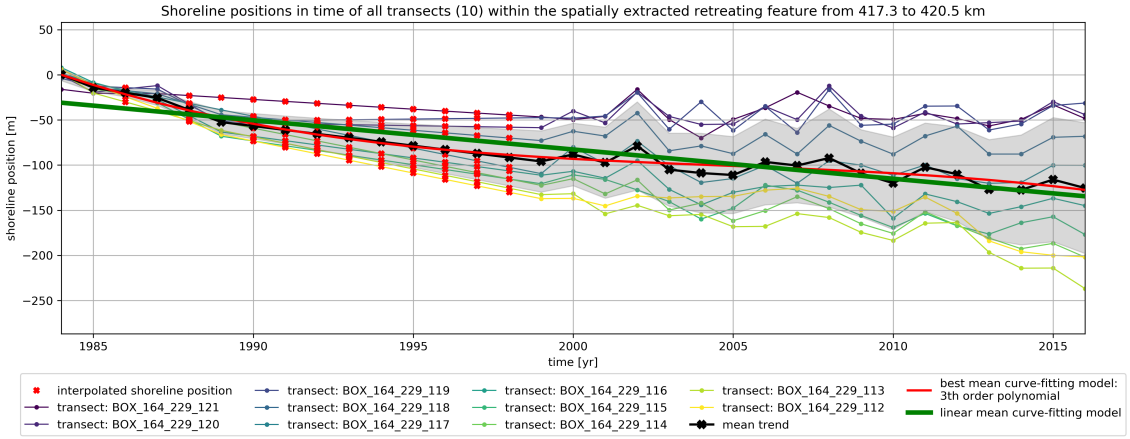


Figure 4.12: Example visualization for the quantification of temporal variability in the second layer of feature extraction method on a 2.5 km scale. This example considers the eight retreating feature from the left (417.3 to 420.5 km in Figure 4.10) on the French coast. Ten transects (see legend label id's) are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. In case of the absence of an annual shoreline position for a temporal trend, a red interpolated dot is shown. The mean (feature) trend, obtained by summing all shoreline positions per year, is indicated with a black line with crosses. The linear fit and best fit to the mean trend are shown in thick green and red lines. The gray band shows the standard deviation range around the mean trend.

4.1.2 Global feature extraction and quantification

Based on the results of the local feature extraction and quantification method in Section 4.1.1 and Appendix C.2, the iteratively determined thresholds were found to be sufficient considering the performance indicators and visual fit inspections. This section elaborates on the results regarding the upscaling of the aforementioned method to the global domain, as outlined in Section 3.1.2.

The 33 considered continuous shorelines have a combined length that amounts to approximately 390 thousand km with around 985 thousand shore-normal defined transects (see Appendix C.1). So, on average, there is a spacing of nearly 396 m between two consecutive transects. This is smaller than 500 m, which is primarily caused by the increasing offset in transect spacings further away from the equator. After applying six filters, 400 thousand of the stretch-included transects contain a linear change rate, whereas 585 thousand transects are set to NaN. The high number of NaN values is primarily due to the relatively low percentage of sandy transects (44%) along the global shoreline.

Statistical results of the application of the rolling mean method on the 400 thousand analyzed world shoreline-included transects with linear change rates are shown in Figure 4.13. The mean of the rolling means in this figure are indicating a global shoreline that is slightly prograding, with nearly no variability between the spatial scales. This result is supported by the findings in the paper of Luijendijk et al. (2018), which states that 28% of the world's sandy beaches are prograding while 24% is retreating. The smaller scale rolling means show a great variability around the mean, with higher peaks in the prograding than in the retreating regime. Some transects in the 2.5 km spatial scale show change rates of nearly +70 or -60 $m/year$ (see Figure 4.13). However, almost 90% of the transects contains a change rate between +2 and -2 $m/year$. The variability around the mean is decreasing with an increasing spatial scale.

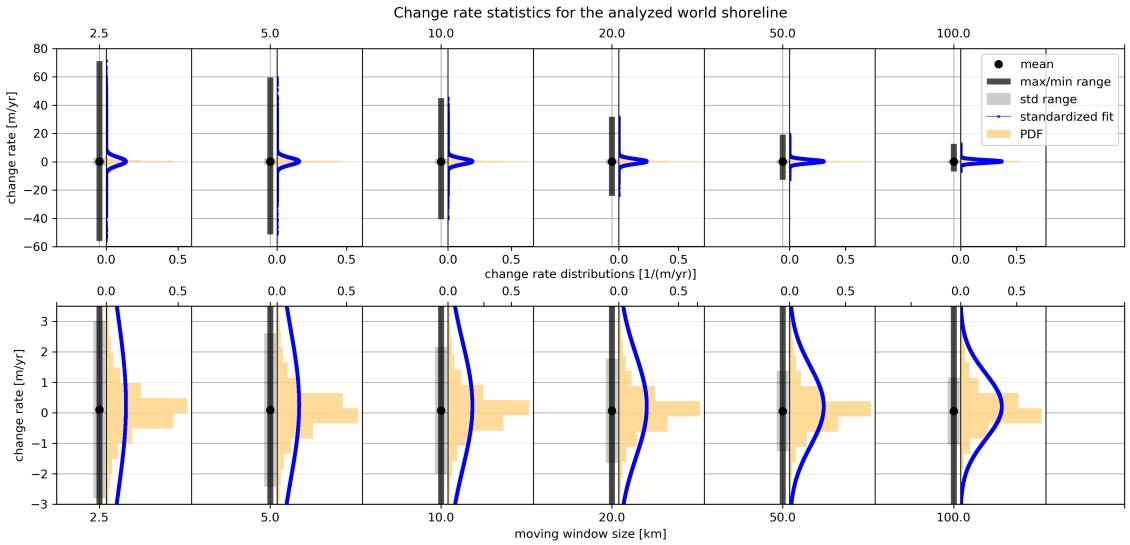


Figure 4.13: Two-panel statistical visualization of the application of the rolling mean method to the analyzed world shoreline for different spatial scales, represented on the top of the top panel and bottom of the bottom panel x-axes. The lower panel is a zoom-in on the standard deviation range (grey bar) of the upper panel. The mean of the rolling means is indicated with a black dot and the maximum / minimum range with a black bar. The x-axes between the two panels show the distribution of change rate values, in a PDF with orange bars and a standardized fit with a thin blue line with dots, within the shoreline stretch maximum / minimum rolling mean range.

From this point onwards, once more, the focus is on the 2.5 km scale to explain the results related to the feature extraction and quantification method for the analyzed world shoreline. Computing the top-bottom, the reduced and finally the feature fit along the analyzed shorelines results in the extraction and quantification of a total of 5154 (3033 prograding and 2121 retreating) features on a planetary scale. The performance indicators regarding the aforementioned fits are presented in Table 4.6. There is a relatively large difference between the reduced and feature fit performance indicators RMSE and MAE. This is comparable to the discussed findings for the French coast. Again, it stated that the thresholds are sufficient as the number of included transects increased significantly for the global application.

Table 4.6: Performance indicators RMSE and MAE for the top-bottom, reduced and feature fit. These are related to the feature extraction and quantification method for the analyzed world shoreline.

Type of fit	RMSE $[m/year]$	MAE $[m/year]$
Top-bottom	0.91	0.27
Reduced	0.97	0.37
Feature	2.70	1.49

The total of 5154 globally extracted features are spatially and temporally quantified on the four and eight characteristics presented in Section 3.1.1, the latter after four dataset modifications. All extracted features can be visualized geographically by means of a feature map (see Figure 4.14). This map indicates the position of prograding or retreating shoreline evolution features that may impose hazards. Features are caused by certain drivers of shoreline evolution that are unknown up to this point.

Whenever shoreline evolution features are absent along a stretch in the globally analyzed shorelines in Figure 4.14, almost the entire considered shoreline stretch-included transects are labeled 'non-sandy' or consist of change rates that fall in the stable regime. The former can be supported by looking at the percentage of sandy transects per continuous shoreline in Table C.1 in Appendix C.1. For instance, New Guinea, Sulawesi, Mindanao and New Britain all contain less than 4% sandy transects.

From Figure 4.14, it can be seen that the number of prograding features for the continent of Asia is almost twice the number of retreating features. A large amount of prograding features is seen at the coast of China, Japan and UAE, which is possibly highly related to land reclamations. The number of prograding features

compared to the number of retreating features is also way larger for the continents of Europe and South-America. In Europe, this might be related to a high number of human interventions. For South-America, many prograding features seem to be caused by natural drivers that are mainly affecting the central west part of the continent. Africa, Oceania and North-America contain approximately the same number of prograding and retreating features.

The total of prograding and retreating features incorporate approximately 27 and 22 thousand transects that affect around 12 and 10 thousand km of shoreline respectively. These features are formed from almost 224 thousand transects (of the total of approximately 400 thousand considered transects with a change rate) with change rates smaller or larger than 0.5 $m/year$, as nearly 176 thousand transects are located in the stable regime. Hence, somewhat less than 22% of the available transects formed a feature and are present in the prograding and retreating feature datasets. It can be stated that almost a quarter of the considered global shoreline contains long-term shoreline changes. As this is a relatively low number, it can be questioned whether the feature datasets are representative compared to the considered global dataset. This is highly dependent on the definition of a feature and thus inherently connected to the employed thresholds (st, cf, tb, hd, ia) in the extraction algorithm (see Section 3.1.1). By splitting the total feature dataset-incorporated transect change rates according to the classification presented in Luijendijk et al. (2018), an enhanced understanding of the effect and impact of the thresholds in the feature extraction algorithm as well as the representativeness of the feature datasets is obtained. This is presented in Table 4.7.

Table 4.7: Overview of the percentages of transect change rates for the total analyzed ordered continuous shoreline dataset as well as for the feature datasets in each of the classification regimes from Luijendijk et al. (2018). The second, third, fourth and fifth column refer to the absolute change rate regimes and have a unit presented in $m/year$. The total number of dataset-included transects with change rates larger or smaller than 0.5 $m/year$ are shown between brackets in the first column.

Dataset(s)	0.5 - 1	1 - 3	3 - 5	> 5
Global (223968x)	39.2	35.8	8.5	16.5
Feature (48607x)	13.3	42.1	17.4	27.2

It can be stated that the feature extraction method thresholds generate an offset towards larger shoreline evolution, which is supported by the percentages in Table 4.7. In this Table, it is seen that the total of feature dataset-included transects incorporate one-third of the global analyzed transects within the 0.5 - 1 $m/year$ regime. The difference of 26% between the global and feature datasets for the aforementioned regime is distributed over all other regimes. From all these other regimes, the increase of almost 11% in the regime that contains change rates larger than 5 $m/year$ is the largest. The impact of the above-stated is elaborated on in Section 4.2.2 and Chapter 5, whereas effect of the different thresholds is outlined in the two paragraphs below.

Four thresholds (cf, tb, hd, ia) in the feature extraction method are found to be the cause of the above-mentioned offset. The st threshold, contrarily to the other four thresholds, is found to decrease the offset by removing a large portion of transects in the regime with change rates larger than 5 $m/year$. The effect of the cf, tb and hd thresholds is dependent on the spatial variation between multiple temporal shoreline evolution signals as well as the behavior of adjacent shorelines. These thresholds are required to remove excessively identified tops, bottoms and remaining noise, as seen in the local application shoreline stretches for The Netherlands, Belgium and South-West Australia. Here, the cf, tb and hd thresholds are found to be constricted by the spatial and temporal resolution of the annual dataset behind Shoreline Monitor and primarily remove transects with lower change rates. This is elaborated on in Chapter 5. The effect of the ia threshold becomes evident when considering shorelines with more NaN-valued transects, like in the case of the French coast. This threshold removes transects with change rates in all regimes.

Apart from the fact that the feature extraction method generates an offset towards larger shoreline evolution, it is stated that the total of extracted and quantified features are representative on the global scale as all possible drivers of long-term shoreline changes are incorporated. This is supported by Table 4.7, where it is seen that all regimes contain at least 6500 transects from the total feature-included transects in the two (one prograding and one retreating) datasets.

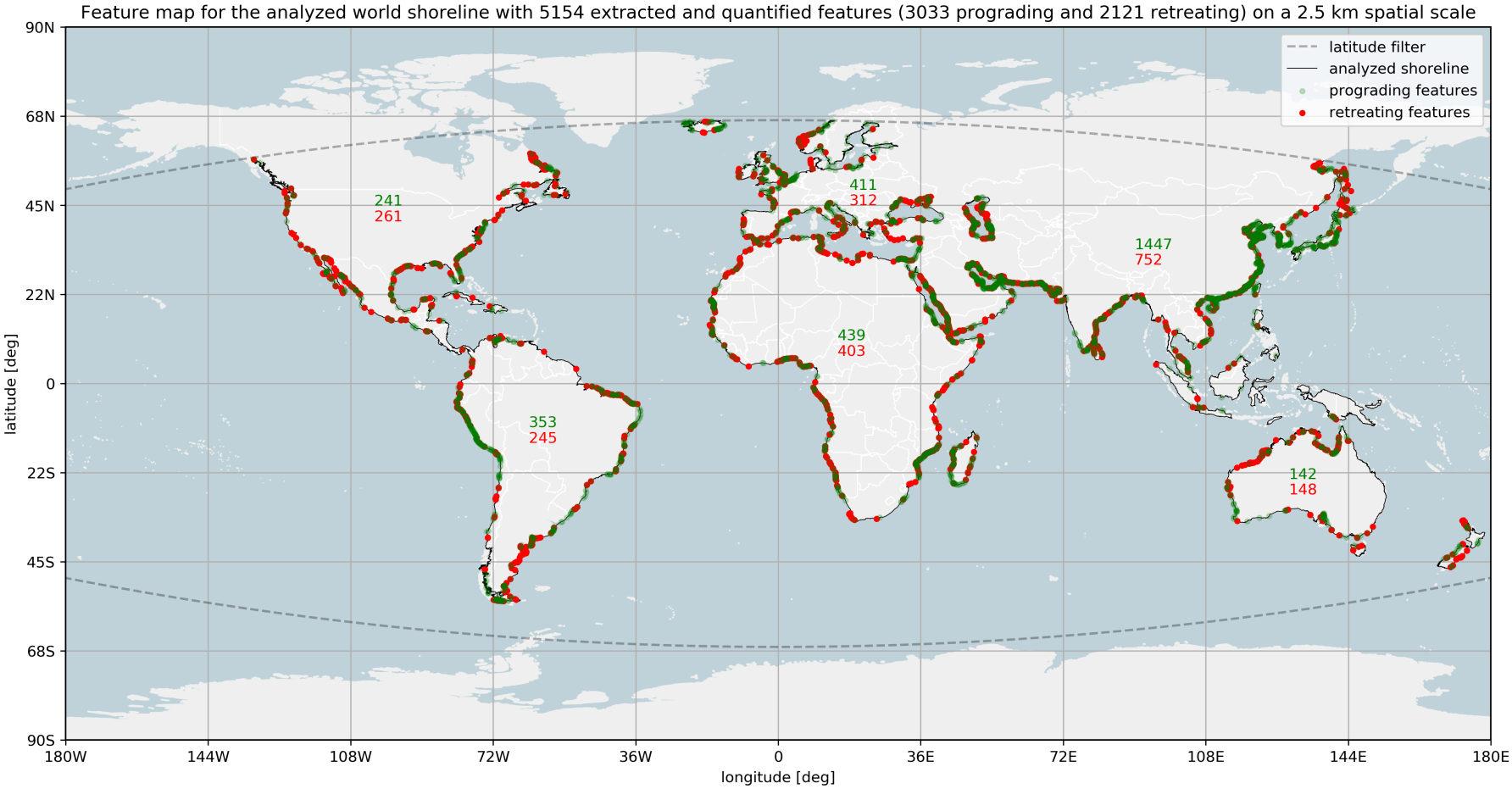


Figure 4.14: Global visualization of the analyzed 33 continuous shorelines (black lines) together with 3033 prograding (green dots) and 2121 retreating (red dots) extracted and quantified features. This visualization is referred to as a feature map, where each feature contains multiple shoreline evolution signals that explain the spatiotemporal variability. The green / red figures in the map indicate the number of extracted prograding / retreating features per continent. Latitude filters are shown with dashed gray lines.

Behind the feature map, the two datasets with a total of twelve quantifications explain the spatiotemporal variability in shoreline evolution features. The variability in the twelve-quantification space is shown in Figure 4.15 for the same statistical parameters as in Figure 4.13. There is a very large variability in temporal quantifications five to ten, while spatial quantifications one to four and temporal quantification eleven and twelve show considerably less variability. Visualizations of the mean feature trends for both datasets, which are used to derive the eight temporal quantification, are presented in Appendix C.3. These visualizations support the discussed outcome related to the high variability in six of the eight temporal quantifications.

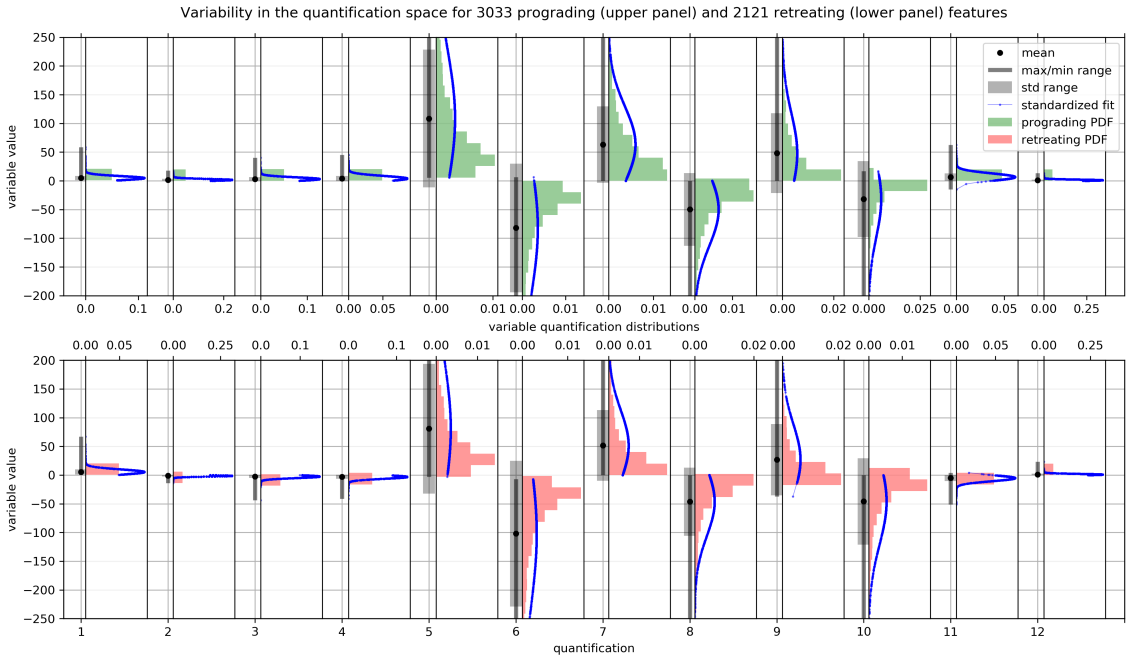


Figure 4.15: Two-panel statistical visualization of the variability in the twelve quantifications (see Section 3.1.1) for the 3033 prograding (upper panel) and 2121 retreating (lower panel) features. Both panels are a zoom-in on the absolute minimum and maximum standard deviation range (transparent grey bar) of all twelve quantifications. The mean of a single quantification is indicated with a black dot and the maximum / minimum range with a black bar. The x-axes between the two panels show the distribution of quantification values, in a PDF with green (prograding) and red (retreating) bars and a standardized fit with a thin blue line with dots, within the maximum / minimum range.

By comparing retreating and prograding variability from Figure 4.15, it can be seen that retreating features have a wider spread in the spatial width (quantification one), while prograding features have a larger temporal change rate (quantification eleven). Quantification eleven indicates that there are a few prograding as well as retreating features that have change rates with an opposite sign. These features are labeled 'wrong' as this is caused by dataset modifications. Local generation of insight in the drivers behind extracted and quantified shoreline evolution features enhances the understanding of the quantification variability in Figure 4.15 and is elaborated on in Section 4.2.1.

4.2 Correlate feature similarities to drivers of sandy shoreline evolution

The three considered steps in this section result in the discussion of planetary-scale correlations between feature similarities and drivers of sandy shoreline evolution. Section 4.2.1 contains findings regarding local feature-driver labels, which provides more insight in the variability in the globally extracted and quantified features of the previous section. In Section 4.2.2, the results of the globally employed ML model that clusters feature similarities are presented and elaborated on. The outcomes of both aforementioned sections are merged in Section 4.2.3.

Key points

- On a 2.5 km spatial scale that contains all types of drivers, a total of 212 features are manually labeled with drivers of shoreline evolution. This resulted in four split prograding and retreating feature datasets that enhance the ML algorithm performance and global correlations;
- For each of the four split datasets, a GMM computed the minimum intra-cluster and maximum inter-cluster dissimilarity between feature quantifications. This resulted in a total of seventeen clusters containing similar shoreline evolution shapes in space and time;
- Up to this point, it is known what the global driver correlations regarding 80% of the shoreline evolution features in thirteen clusters are, whereas an objective enhanced understanding on these correlations is not obtained yet.

4.2.1 Local feature-driver labeling

As stated in Section 4.1.2, the drivers behind planetary-scale shoreline evolution features are unknown up to this point. This section considers the coupling of drivers to extracted and quantified features (called labeling) on eight local application shoreline stretches.

Using the labeling method presented in Section 3.2.1, features are 'subjectively' coupled to drivers up to tertiary influences for the 2.5 km spatial scale. Only the main driver behind shoreline evolution features is considered in the remainder of this research, as spatial and temporal quantifications are based on the complete trends. Table 4.8 provides an overview of the results of the labeling method for the eight shoreline stretches separately and for the total of all of these shorelines.

Table 4.8: Overview of the results of the application of the driver labeling method on eight local shoreline stretches. The drivers are referred to by their ID (see Section 2.2). The prograding and retreating features are considered in the top and bottom panel respectively. 'Total' refers the summed number of either prograding or retreating feature-driver labels, whereas 'combined total' is the sum of all labels. A '-'-sign means that the driver cannot be identified in the considered panel and an empty cell refers to a driver being identified zero times. 'X' indicates erroneously extracted features labeled 'wrong' and '?' relates to 'unknown' drivers.

Shoreline stretch	N	LR	PB	RSS	BSS	SBM	TI	S	X	?	Total
The Netherlands	6		1	-			-	-			7
Belgium	7		1	-			-	-			8
South-West France		2	1	-		8	-	-	2		13
South-West Australia	1	5	1	-	6	1	-	-	1		15
Egypt		2	13	-		6	-	-	3		24
Mid-West Africa			2	-		7	-	-	9		18
Mid-East China		33		-			-	-	2		35
South USA (Gulf of Mexico)	1			-		2	-	-	1	2	6
Total	15	42	19	-	6	24	-	-	18	2	126
The Netherlands	-	-									0
Belgium	-	-	1								1
South-West France	-	-	2		3	2	2		8		17
South-West Australia	-	-			1						1
Egypt	-	-	7	7					2		16
Mid-West Africa	-	-	10	6	5				9		30
Mid-East China	-	-		3					4		7
South USA (Gulf of Mexico)	-	-	4	1		1	3	5			14
Total	-	-	24	17	9	3	5	5	23		86
Combined total	15	42	43	17	15	27	5	5	41	2	212

In total, around seven thousand km of shoreline with 212 extracted and quantified features is analyzed manually (this is elaborated on further in Chapter 5). 41 of the 212 shoreline evolution features are labeled 'wrong' as these are erroneously extracted and quantified from the Shoreline Monitor. This is caused by frequently occurring inaccuracies, like estuaries and bays, rivers and deltas, and shallow shorelines, but also due to minor errors in dataset modifications. Drivers of shoreline evolution could not be identified for two of the 212

features. These are labeled 'unknown', resulting in a total of 210 (4% of the world, 124 prograding and 86 retreating) known feature-driver labels for approximately 2% of the world shoreline.

As seen in Table 4.8, a lot of prograding features are related to the direct human influence of land reclamations along the shoreline stretch in Mid-East China. This supports the findings in Section 4.1.2. The Dutch and Belgium coasts are protected by many direct 'soft' human influences referred to as nourishments. For the case of the Dutch coast, the second prograding feature in Figure 4.4 (with belonging temporal trends in Figure 4.6) is found to be primarily related to the mega nourishment project called the Sand Engine. For the case of the Belgian coast, the considered prograding feature in Figure C.8 is found to be driven by beach nourishments. Egypt as well as Mid-West Africa contain numerous shoreline evolution features that are traced back to indirect human influences as port / breakwaters and river sediment supply. In South-West France, indirect human influence related to port / breakwaters is only identified three times as main driver. One of these is seen in the eight retreating feature considered in the case of the French coast in Figures 4.10 and 4.12. The retreating feature that is considered for the case of the South-West Australian coast is found to be related to the natural influence of beach sediment supply. For the same stretch, nearly half of the prograding features of shoreline evolution are seen to be related to beach sediment supply as well. From Table 4.8, it follows that the Gulf of Mexico is subject to shoreline retreat caused by storms. Erroneously extracted features are primarily identified in South-West France (shallow shorelines) and Mid-West Africa (rivers running parallel to the coast), and are discussed further in Chapter 5.

In Section 2.2, drivers of shoreline evolution like river sediment supply and storms were stated to have moderate or large spatial scales. However, these drivers are also present in the relatively small 2.5 km spatial scale. This is explained by the change rate averaging of the rolling mean method. Apparently, shoreline evolution features with a small spatial width and a large change rate are smeared out. Features with a large spatial width and a small change rate are averaged out when the spatial scale is increased from 2.5 to 100 km. For this reason, from this point onwards, it is solely focused on the 2.5 km spatial scale. The 5, 10, 20, 50 and 100 km spatial scales were of use to explain shoreline behavior at the start of the analysis in Sections 4.1.1 and 4.1.2 and can be used for another application elaborated on in Chapter 5.

By visualizing the twelve-dimensional quantification space for the identified 126 prograding (including 'unknown') and 86 retreating drivers of shoreline evolution, the understanding of driver-related quantification variability is enhanced. This is shown in Figures C.17 and C.20 in Appendix C.4, where each driver is assigned a unique color. The maximum and minimum annual value change (quantification five and six) show the largest variability for the identified drivers, primarily caused by the labels 'wrong' for the retreating and land reclamations as well as 'wrong' for the prograding feature dataset. It is decided to split the prograding and retreating feature datasets to decrease the variability between feature quantifications. This enhances the ML model performance in the next section. By using limits of 140 and -140 for both prograding and retreating features in quantification number five and six of the aforementioned appendix figures, both datasets are split into regular and irregular ones. Hence, a total of four feature datasets are obtained. The limits indicate that natural trends do not have an annual change in shoreline position bigger or smaller than 140 $m/year$ and are based on the locally generated insight in 212 feature-driver labels. For this reason, the irregular datasets contain extracted and quantified features showing primarily unnatural trends.

The two prograding datasets contain 1865 regular and 1165 irregular quantified features with 85 and 41 (together 126) identified drivers respectively. The temporal trends and the identified drivers (with unique colors) in both these datasets are shown in Figures C.18 and C.19 in Appendix C.4. The retreating regular dataset contains 1414 quantified features with 65 identified drivers, whereas the retreating irregular dataset incorporates 707 quantified features with 21 identified drivers. The temporal trends and identified drivers for these datasets are also shown in Appendix C.4.

4.2.2 Global feature clustering

The four split datasets with feature quantifications (see previous section) function as input for the unsupervised ML clustering method called GMM. Before the input is employed, the datasets are standardized to ensure a well-performing ML model. The first two paragraphs below discus the findings regarding the ML model setup. The remaining paragraphs elaborate on the post-processing of results.

Four GMMs are created for the four split datasets with a combined total of 5154 features with sixteen quantifications (the four spatial quantifications are duplicated). Each of the GMMs consists of a LM and a GM (see Section 3.2.2). The LM incorporates a large range of values for various hyper-parameters and functions to derive a local minimum for a first order assessment. The GM uses a smaller range of values than the LM and results in a global optimum, also stated to be the final outcome of a GMM. The results of the two model steps for the four split datasets are summarized in Table 4.9. Visualizations to support the findings presented in Table 4.9 can be found in Appendix C.5. As can be seen from these visualizations, the clustering algorithm does not perform well with spherical and diagonal constrained covariance matrices (GMM type). These types of matrices constrain clusters in the shape and / or orientation and are unable to minimize intra-cluster and maximize inter-cluster dissimilarity within the input data. The tied as well as full covariance matrices are better capable of deriving an optimal number of clusters. The best type of covariance matrix is not included in Table 4.9 below, as it is found that a full covariance matrix results in the lowest BIC for every split dataset. In case of a full covariance matrix, each cluster is represented by an ellipse with an arbitrary size and orientation.

Table 4.9: Summary of the outcomes of the LM and GM steps for the four GMMs with split datasets. The number of extracted and quantified features that function as input for the ML models are stated between brackets in the first column. If applicable, the last row provides summed figures for the total of four models.

Feature dataset		LM minimum BIC	LM optimal number of clusters	GM minimum BIC	GM optimal number of clusters
Regular prograding	(1865x)	-44082.0	6	-44479.0	5
Irregular prograding	(1165x)	-24704.0	4	-24713.0	4
Regular retreating	(1414x)	-34107.0	6	-34534.0	5
Irregular retreating	(707x)	-14566.0	3	-14596.0	3
Combined total	(5154x)		19		17

From Table 4.9, it is seen that the LM minimum BIC is approximately similar to the GM minimum BIC. Apparently, the first order assessment with one random initialization is quite capable of iterating towards minimum intra-cluster and maximum inter-cluster dissimilarity. However, compared to the LM, the GM derived minimum BIC results in a lower number of optimal clusters for two of the four datasets. This means that feature similarities are captured in a total of seventeen clusters, resulting from the GM step, instead of the first order assessment computed nineteen clusters. The minimum BIC values for the irregular datasets are higher (smaller negative value) than for the regular datasets, this is due to the larger variability between features in the irregular datasets (containing mainly unnatural features). Therefore, the GMMs with irregular datasets are not capable of creating clusters with features that have as much similarity as in the clustered regular datasets. The applicability of the usage of the minimum BIC to derive the optimal number of clusters is questioned and elaborated on in Chapter 5.

The differences between automatically grouped feature similarities are analyzed by computing the cluster means. Here, a cluster mean is referred to as the twelve-dimensional average and can be seen as an overall representative shoreline evolution feature of all cluster-included feature quantifications. These cluster mean quantifications are provided in Table C.9 in Appendix C.5. As outlined in Section 3.2.2, the spatial width and the temporal change rate (quantification one and eleven, see Section 3.1.1) are selected to draw up cluster mean spatiotemporal signatures. It is found that this is the easiest and most understandable way of visualizing the differences between clusters in the twelve-dimensional quantification space for this research. However, one has to keep in mind that clusters contain differences in ten other quantifications. These are not shown in cluster mean spatiotemporal signature visualizations. Second-order polynomials connect the change rates (shown on the y-axis) with symmetrically positioned extents (width to the left and right on the x-axis) to mimic the shape of shoreline evolution features as seen in for instance Figures 4.4 and C.6.

In Figure 4.16, the power of unsupervised ML clustering is visualized by showing all 5154 feature input signatures as well as the automatically derived representative seventeen cluster mean spatiotemporal signatures (nine prograding and eight retreating). It can be seen that the maximum spatial width of an extracted feature is somewhat less than 70 km. Next to this, the maximum prograding change rate is nearly 65 $m/year$, whereas the maximum retreating change rate is around 50 $m/year$. This is supported by the analysis of Figure 4.15 in Section 4.1.2. Figure 4.17 shows a zoom-in on the cluster mean spatiotemporal signatures of

Figure 4.16, making it easier to outline differences between clusters. The change rate maxima are more or less 15 $m/year$ and the width approximately 13 km for both prograding and retreating clusters. Based on the change rate, the largest two and the smallest three prograding and retreating clusters are seen to be resulting from the irregular and regular databases respectively. The former indicates mainly unnatural cluster-included features. Next to this, many clusters are have a spatial width around the 5 km. In Table C.10 in Appendix C.5, the cluster-included feature-related number of transects and affected km of shoreline is presented. From this table, it can be seen that regular prograding and retreating cluster five contain less than 200 features with a relatively large number of included transects and affected km of shoreline. This substantiates the low and wide mean spatiotemporal signatures, as seen in Figure 4.17. Contrarily, the two highest and small mean prograding and retreating spatiotemporal signatures contain a relatively small number of included transects as well as affected km of shoreline.

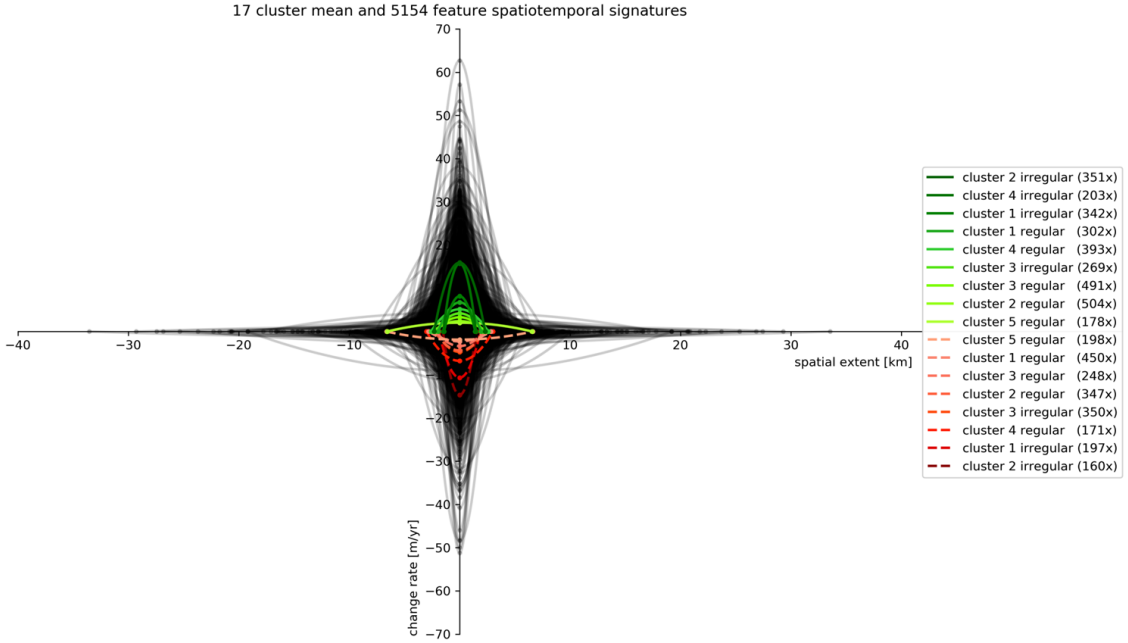


Figure 4.16: Visualization of seventeen automatically derived and representative cluster mean (greenish colored lines for prograding and dashed reddish colored lines for retreating) and 5154 feature input (grey lines) spatiotemporal signatures for two of the twelve quantifications. The spatial extent on the x-axis shows the symmetrically positioned width and the y-axis the temporal change rate of a feature or cluster. The legend indicates the cluster labels (number of included features between brackets), which are ordered from high to low positive change rates on the positive y-axis and low to high negative change rates on the negative y-axis.

To be able to analyze similarities between prograding and retreating clusters from Figure 4.17, the latter are mirrored on an absolute change rate y-axis. Next to this, as the x-axis is symmetrical, it is decided to show only one quadrant. These two alterations are shown in Figure 4.18. As can be seen, based on the two visualized quantifications, various prograding and retreating clusters have similar mean spatiotemporal signatures. Table C.9 allows to compare clusters on the other ten quantifications not shown in Figure 4.18.

The feature extraction method thresholds that were stated to generate an offset towards larger shoreline evolution are found to have a similar impact on the cluster mean spatiotemporal signatures. This outcome is supported by the cluster mean spatiotemporal signatures in Figure 4.18, which indicate a minimum cluster average absolute change rate of around 2 $m/year$ (see also Table C.9 in Appendix C.5). These smaller clusters are located in the middle of the second change rate regime presented in Table 4.7 in Section 4.1.2. None of the cluster mean signatures is found to be located in the regime with change rates between the 0.5 - 1 $m/year$. Therefore, it is stated that the impact of the 13.3% feature-included transects with change rates in the aforementioned regime is insignificant compared to the number of transects in other regimes. This is also elaborated on in Chapter 5.

Signature boundaries are shown in Figure 4.18 as well. These boundaries indicate the area in which, for 97.5% of the time, feature input signatures fall. Signature boundaries are derived from scatter plots (see Appendix C.5), which are only a different way of visualizing the message that Figure 4.18 is transferring. Note that

individual feature scatter data is clustered in a twelve-dimensional space and never shown in cluster mean plots as this complicates the aforementioned message for two of the twelve quantifications. The signature boundaries are only applicable to a spatiotemporal visualization with width and change rate quantifications.

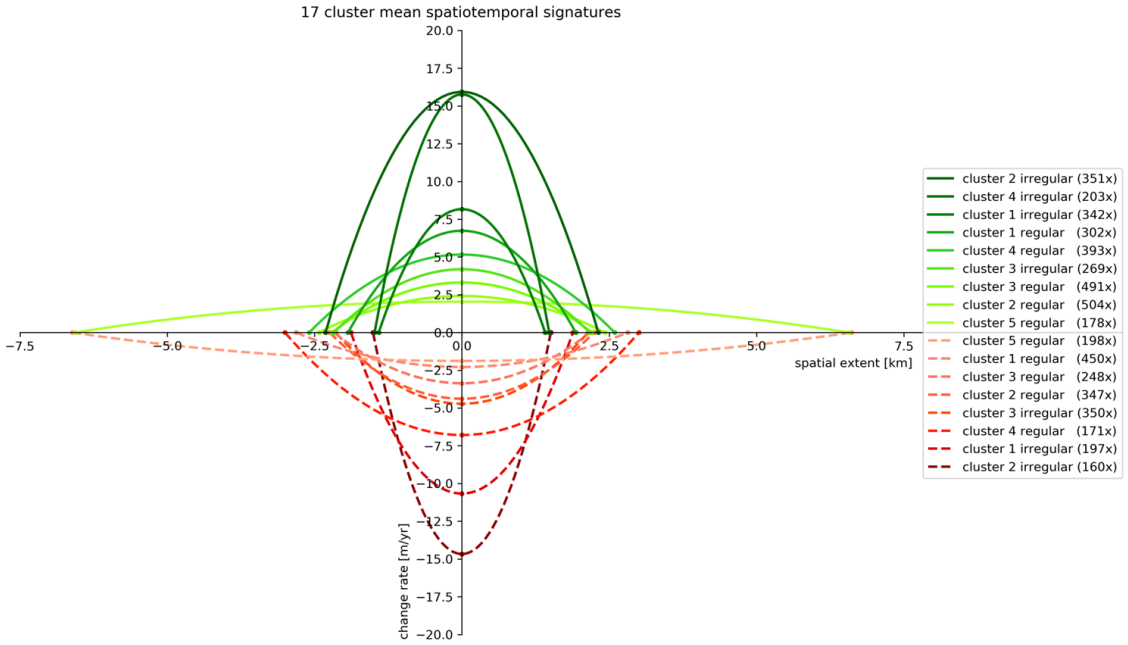


Figure 4.17: Visualization of seventeen automatically derived and representative cluster mean (greenish colored lines for prograding and dashed reddish colored lines for retreating) spatiotemporal signatures for two of the twelve quantifications (zoom-in of Figure 4.16 without the 5154 feature input signatures). The spatial extent on the x-axis shows the symmetrically positioned width and the y-axis the temporal change rate of a cluster. The legend indicates the cluster labels (number of included features between brackets), which are ordered from high to low positive change rates on the positive y-axis and low to high negative change rates on the negative y-axis.

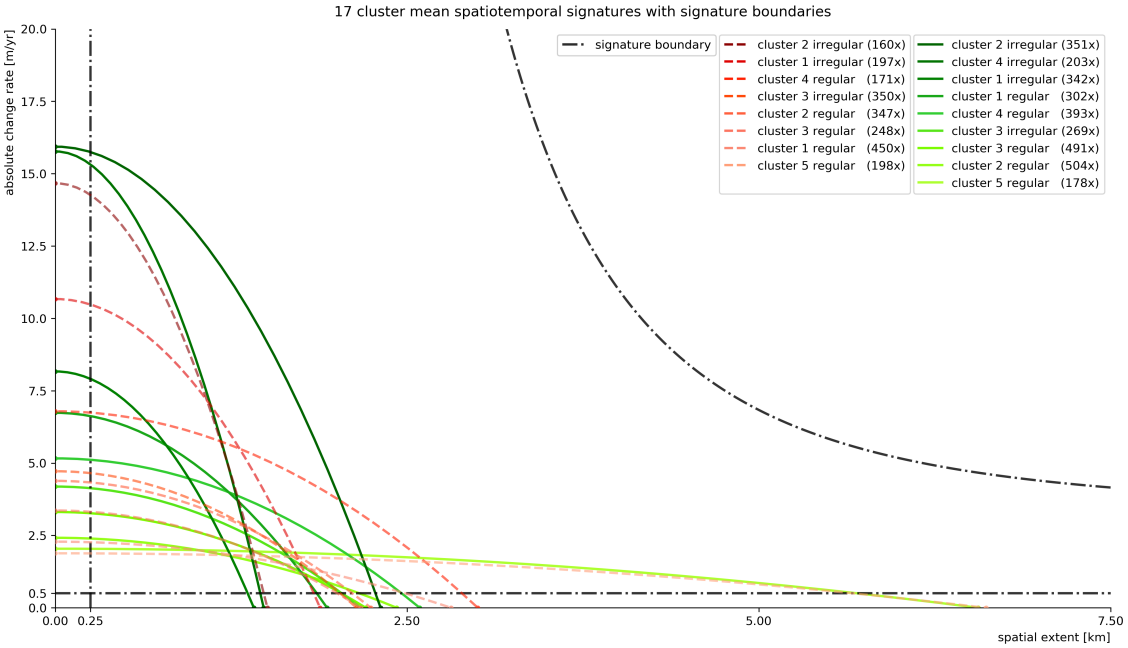


Figure 4.18: Visualization of seventeen automatically derived and representative cluster mean (greenish colored lines for prograding and dashed reddish colored lines for retreating) spatiotemporal signatures for two of the twelve quantifications (mirrored and one-quadrant alteration to Figure 4.17). The signature boundaries are shown by means of dashed-dotted black lines. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster. The legend indicates the cluster labels (number of included features between brackets), which are ordered from high to low absolute change rates.

Up to this point, on a global scale, shoreline evolution features are grouped in seventeen clusters according to twelve quantifications. This means that one is able to compare shoreline evolution features with similar ones for any place on earth. Moreover, it is known what the differences between clusters (and thus feature similarities) are and what their mean spatiotemporal shoreline evolution signature looks like. However, it is not known yet what these clusters represent. This is elaborated on in the next section.

4.2.3 Global cluster-driver correlations

In the last paragraph of Section 4.2.2, it is outlined that it is not known yet what the the automatically derived feature-cluster correlations represent. By merging this outcome with the locally generated insight in feature-driver labels from Section 4.2.1, cluster-driver correlations are obtained. The first paragraph below discusses the findings regarding summed as well as single identified hard or probable driver labels, where probable labels are an additional outcome of the GMMs. The remaining paragraphs present the findings related to the aforementioned correlations, the validation method and the final outcome of this step.

By looking at the summed main driver distributions over clusters for all split datasets in Figures C.33 to C.36 in Appendix C.6, it can be concluded that hard driver labels are sufficiently accurate. These are approximately similar to the probable driver labels. The decision to use hard driver labels in the remainder of this research, which is based on locally generated insight, is discussed further in Chapter 5. Visualizations regarding hard as well as probable driver labels per identified driver are shown in Figures C.37 to C.42 in Appendix C.6. These visualizations can be used to elaborate on the importance of taking multiple driver influences into account, instead of using only the main driver, and are discussed in Chapter 5.

For all locally labeled prograding drivers (124 as unknown labels are rejected), the hard cluster assignments and thus the cluster-driver correlations are made visible by means of a bar plot in Figure 4.19. This is also done for the locally labeled retreating drivers (86) as shown in Figure 4.20. The cluster-driver correlations are based on locally generated insight in eight shoreline stretches that are not equally distributed geographically nor over clusters. Therefore, using the method described in Section 3.2.3, the cluster-driver correlations are validated by selecting ten random features per cluster. This results in an additional 90 prograding and 80 retreating (170 in total for seventeen clusters) driver labels. Consequently, drivers of shoreline evolution are known for 380 of the total of 5154 extracted and quantified features. 96 of the 380 driver labels are identified 'wrong' as these features are recognized to be erroneously extracted and quantified. Thus, 284 features are labeled with actual drivers. Generally stated, insight in automatically derived groups from a ML algorithm is obtained from around 33% known labels. The low number of known labels, around 7%, is discussed further in Chapter 5.

As can be seen in the lower panels from Figures 4.19 and 4.20, clusters with approximately similar patterns between locally identified drivers generated from local application shoreline stretches (bars) and validated drivers (black lines) are marked in green. The similarity in patterns is considered together with a proxy for the reliability of the locally and manually analyzed features compared to the randomly selected validation features, as seen in Table 4.10. For the prograding dataset, the locally analyzed features are labeled nourishments for 12% of the time. Nourishments are not identified in a single randomly selected validation feature. Consequently, the locally analyzed shoreline is stated to be completely unrepresentative for other features in relation to nourishments. So, these are excluded from the aforementioned pattern similarity comparison. For the retreating dataset, a considerable smaller amount of port / breakwaters and larger amount of 'wrongs' are identified when comparing the local generation and validation feature sets. These are not excluded from the pattern similarity comparison but were analyzed with caution when marking a cluster green or red, as the locally analyzed shoreline is stated to be somewhat unrepresentative for other features. Clusters marked red are generally seen to have little locally identified drivers (upper panels in the figures). It is stated that more insight is required to correctly correlate drivers to these clusters, as for now, only clusters with features containing similarities in spatiotemporal variability in shoreline evolution remain.

From all green-marked clusters in Figures 4.19 and 4.20, only irregular retreating cluster 2 contains one identified driver ('wrongs') that can easily be understood by mapping the geographical distribution of the features. This is shown in Figure C.43 in Appendix C.6. The rest of the clusters contain multiple identified drivers and are stated to have validated cluster-driver correlations. This is the case for around 70% of the total number of clusters and 77% of the cluster-included features. It is foreseen that more insight in the drivers of shoreline

evolution for the red-marked clusters will also result in validated cluster-driver correlations. This is based on the fact that there are already multiple drivers present in the relatively low number of identified ones. Apparently, up to this point, assumptions, decisions and restrictions have a large influence on the separability of different drivers among clusters. This is elaborated on further in Chapter 5.

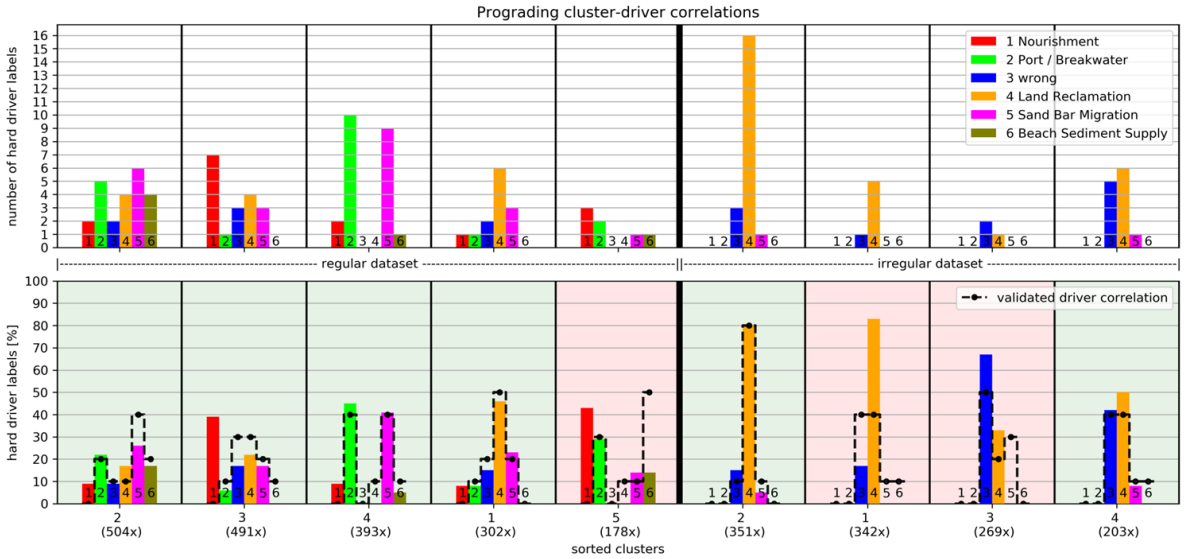


Figure 4.19: Bar plot visualization of nine prograding hard cluster-driver correlations, where the upper panel represents the counted number and the lower panel the percentage of driver labels within a cluster (total of 124 labels for the dataset with 3033 features). Here, each driver is assigned a unique color and number that refers to its belonging bar. Between the two panels, the extents of the regular and irregular datasets are given. Per dataset, the clusters on the lower x-axis (automatically numbered by the GMM algorithm) are ordered on the number of included features (stated between brackets). The lower panel contains validated cluster-driver correlations (with a total of 90 labels) that are based on ten manually labeled features per cluster (dashed-dotted black line). Whenever the validated correlations are more or less similar to the locally generated insight correlations (colored bars), the cluster is marked in transparent green in the lower panel. If not, the cluster is marked in transparent red.

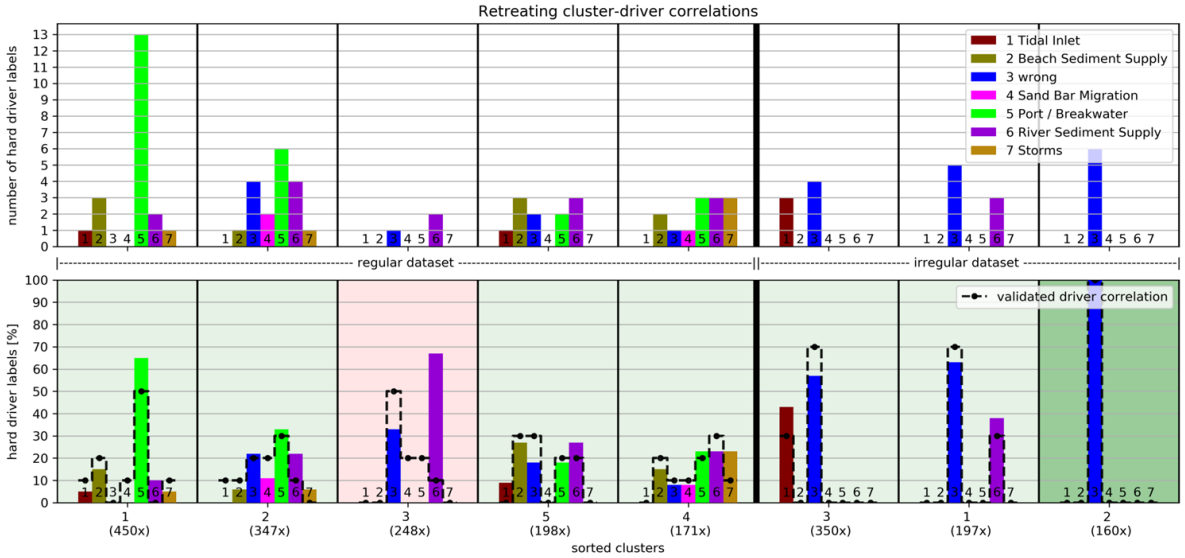


Figure 4.20: Bar plot visualization of eight retreating hard cluster-driver correlations, where the upper panel represents the counted number and the lower panel the percentage of driver labels within a cluster (total of 86 labels for the dataset with 2121 features). Here, each driver is assigned a unique color and number that refers to its belonging bar. Between the two panels, the extents of the regular and irregular datasets are given. Per dataset, the clusters on the lower x-axis (automatically numbered by the GMM algorithm) are ordered on the number of included features (stated between brackets). The lower panel contains validated cluster-driver correlations (with a total of 80 labels) that are based on ten manually labeled features per cluster (dashed-dotted black line). Whenever the validated correlations are more or less similar to the locally generated insight correlations (colored bars), the cluster is marked in transparent green in the lower panel. If not, the cluster is marked in transparent red. In case of a 100% driver correlation, the cluster is marked in brighter green.

Table 4.10: Overview of the relative percentages of driver labels for the locally generated driver insight (126 prograding and 86 retreating) as well as randomly selected and validated feature-driver labels (90 prograding and 80 retreating). The top panel represents the prograding labels and the lower panel the retreating labels. The drivers are referred to by their ID (see Section 2.2). 'X' indicates erroneously extracted features labeled 'wrong' and '?' relates to 'unknown' drivers. A '-'-sign means that the driver cannot be identified in the considered panel. Cells are marked grey if the relative difference between rows in a panel is larger than 10%.

Labels	N	LR	PB	RSS	BSS	SBM	TI	S	X	?
Local generation	12	33	15	-	5	19	-	-	14	2
Validation	0	32	12	-	12	21	-	-	22	0
Local generation	-	-	28	20	10	3	6	6	27	0
Validation	-	-	18	13	10	8	6	3	44	0

As outlined in the paragraph above, the majority of the clusters contain multiple correlated drivers. Therefore, the remainder of this research will focus on (in)direct human and natural influences on shoreline evolution (see Section 3.2.3). By rejecting the 96 erroneously extracted and quantified features, this results in 86, 95 and 103 identified direct human, indirect human and natural influences respectively. Without taking cluster assignments into account, the width and change rate quantifications of these influences are analyzed by adding the distributions to the visualization of Figure 4.18, resulting in Figure 4.21.

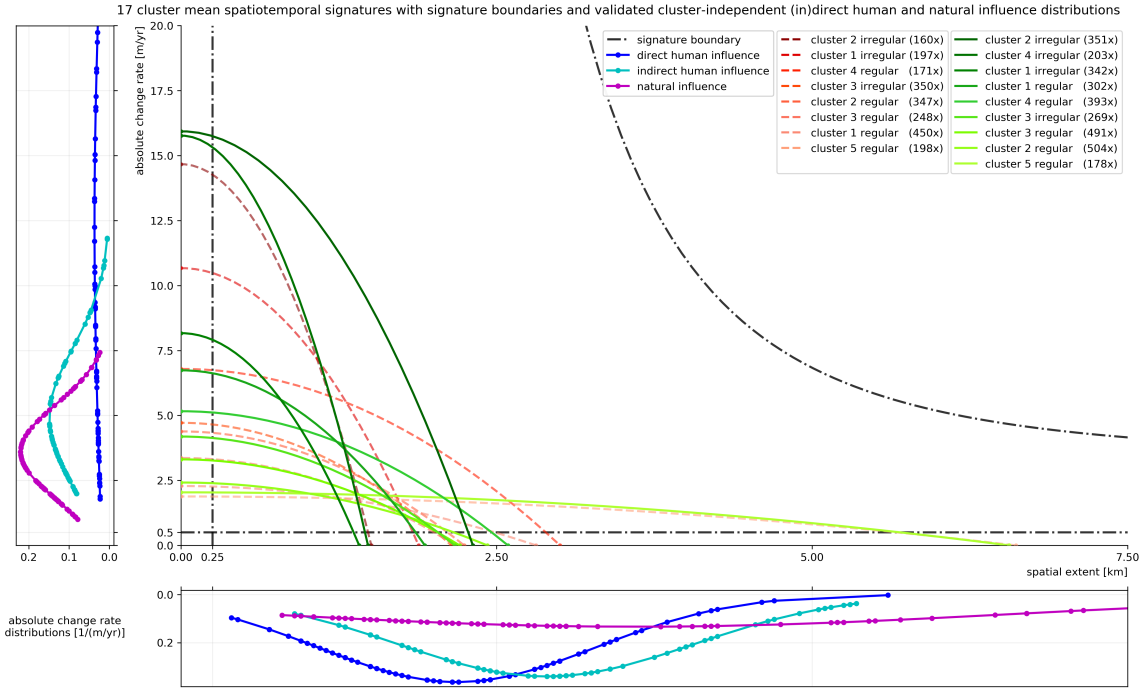


Figure 4.21: Visualization of seventeen automatically derived and representative cluster mean (greenish colored lines for prograding and dashed reddish colored lines for retreating) spatiotemporal signatures for two of the twelve quantifications. The signature boundaries are shown by means of dashed-dotted black lines. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster. The legend indicates the cluster labels (number of included features between brackets), which are ordered from high to low absolute change rates. The left and bottom panel show the validated direct human (blue line with dots), indirect human (light blue line with dots) and natural (purple line with dots) influence distributions (independent from cluster assignments) for the absolute change rate and the extent respectively.

Some interesting patterns can be seen in the cluster-independent influence distributions. Direct human and natural influences seem to have a large distribution in the change rate and one-sided width respectively, not entirely shown in Figure 4.21 (the complete distributions are visualized in Figure C.44). Besides this, (in)direct human influences seem to halt at approximately 10 km width. Furthermore, direct human influences appear to have a smaller mean extent (peak of the distribution) than indirect human as well as natural influences. Finally, the change rate distribution regarding natural influences has its peak and halts lower than (in)direct human influences. It is rather subjective to base final results on the total of 284 manually labeled feature

influences for only two quantifications (width and change rate). Therefore, the next section considers incorporating cluster dependency to elaborate on a more objective way of analyzing influence distributions to come to conclusions.

4.3 Derive a planetary-scale classification

In the section above, it was found that most of the automatically derived clusters contain multiple correlated drivers that explain shoreline evolution. Adjusting the focus from drivers to (in)direct human and natural influences led to the discovery of interesting subjective cluster-independent patterns. In this section, geographical driver distributions of all features (not only the manually labeled ones) within interesting (combined) clusters are mapped to analyze the aforementioned patterns more objectively (see Section 3.3). Here, while the results are shown for two quantifications (width and change rate), the word 'objective' also refers to considering the total of twelve quantifications as cluster-dependent patterns are analyzed. Eventually, findings result in a statistically derived planetary-scale classification of natural and human-induced sandy shoreline evolution. In the first two paragraphs below, an explanation on the selected interesting clusters to come to the classification is presented. The remaining paragraphs provide a discussion on the findings in the visualized (combined) cluster maps and provide an overview of the final outcome with an application.

Key points

- Natural and human-induced shoreline evolution accounts for approximately 16 and 25% of the total of globally extracted and quantified features respectively, whereas 57% is present in the compound regime with complex and combined influences;
- A continental analysis revealed that Asia's features are primarily due to (in)direct human influences, whereas the other five continents are located in the compound regime.

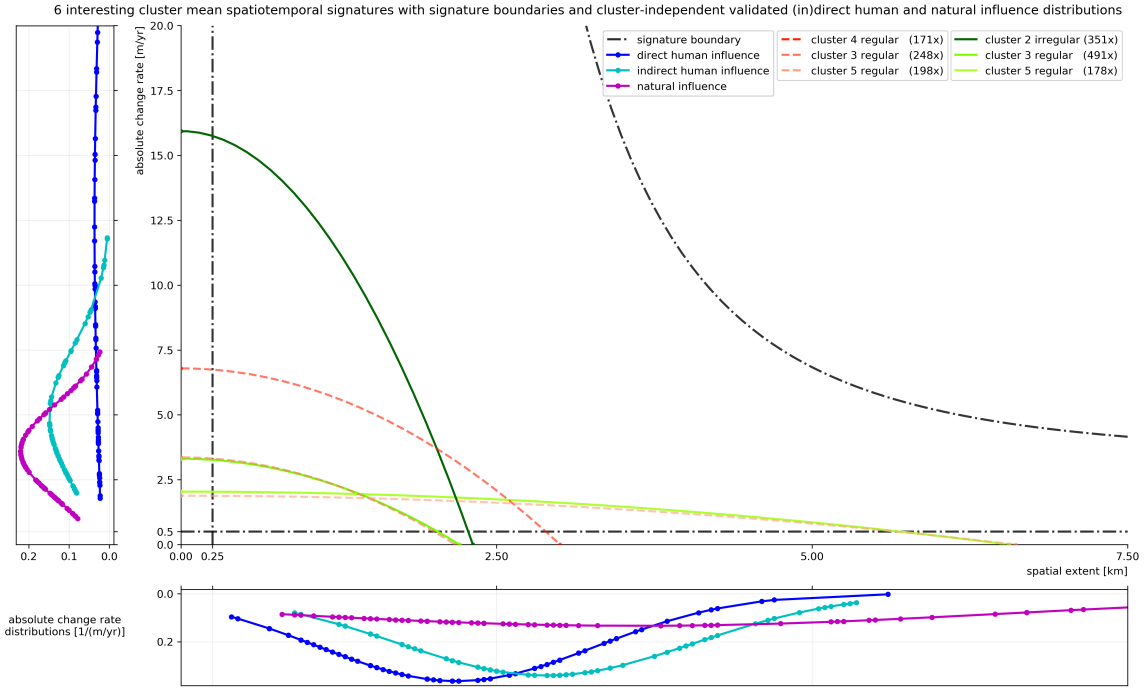


Figure 4.22: Visualization of six interesting automatically derived and representative cluster mean (greenish colored lines for prograding and dashed reddish colored lines for retreating) spatiotemporal signatures for two of the twelve quantifications. The signature boundaries are shown by means of dashed-dotted black lines. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster. The legend indicates the cluster labels (number of included features between brackets), which are ordered from high to low absolute change rates. The left and bottom panel show the validated direct human (blue line with dots), indirect human (light blue line with dots) and natural (purple line with dots) influence distributions (independent from cluster assignments) for the absolute change rate and the extent respectively.

As can be seen from Figure 4.21, the highest three prograding and two retreating clusters are resulting from the irregular dataset with mainly unnatural trends. It is decided to analyze the cluster on the extreme of the change rate axis, prograding cluster 2 irregular. This cluster also has a relatively large width that is different from other irregular clusters. Next to this, based on the change rate, retreating cluster 4 regular and prograding cluster 1 regular are seen to be the highest clusters with more natural trends. As retreating cluster 4 regular has a deviating width compared to other clusters, this cluster is selected to be analyzed as well. Also the highly similar clusters on the extreme of the extent axis are considered for the analysis, prograding and retreating clusters 5 regular. Multiple clusters have a change rate of around 2.5 to 5.0 $m/year$ and an extent around 2.5 km. Prograding and retreating clusters 3 regular are very similar and are approximately in the center of the clusters in the aforementioned change rate and extent ranges. These clusters are the last ones that are selected to be analyzed. All the clusters that are referred to in the text above are highlighted with thick lines in Figure C.45 in Appendix C.7. Excluding all the non-highlighted clusters results in the creation of an overview for the coming analysis, which is is shown in Figure 4.22.

Some clusters will form single maps, where others will become combined maps. In this research, two single maps and two combined maps are created. These four maps radiate a particular message that is based on one of the three analysis characteristics to enlarge feature locations, as outlined in Section 3.3. Table 4.11 provides an overview of the maps with its cluster(s), the selected analysis characteristic and the main message a map is transferring.

Table 4.11: Overview of the selected interesting clusters to derive a planetary-scale classification of natural and human-induced sandy shoreline evolution. In case the words prograding or retreating are not mentioned in the second column, it is referred to both types of clusters.

Cluster map	Cluster(s)	Map type	Analysis characteristic	Main message (influences)
1	Prograding cluster 2 irregular	Single	Change rate	Direct human
2	Retreating cluster 4 regular	Single	Change rate	Indirect human and natural
3	Clusters 5 regular	Combined	Width / radius	Natural and indirect human
4	Clusters 3 regular	Combined	Pairs	Compound

1. Single map with direct human influences

The single map regarding prograding cluster 2 irregular is shown in Figure 4.23, with enlarged shoreline evolution features based on the change rate. The areas in the Middle-East (UAE) and China show a considerable amount of enlarged features with change rates bigger than 15 $m/year$. This is supported by the number of prograding features (265) for the continent of Asia, which is significantly larger than all other continents.

The driver correlations in the legend show that the shoreline evolution features in this map are primarily related to the direct human influence of land reclamations (80%), whereas only a small part (15%) belongs to erroneously extracted features. The former is substantiated by coupling the World Port Index (WPI) dataset, which contains locations of around 3700 ports worldwide (source: `https://www.nga.mil/`). In total, 78 features in this map are possibly related to WPI ports (yellow '+'-signs). These are primarily found at the enlarged features in the aforementioned areas. For this reason, it is concluded that at these 78 locations, possible WPI ports refer to large land reclamations.

It is also seen that there are some features with smaller change rates that can be related to land reclamations, following the same reasoning in the paragraph above. However, there are nearly no features with change rates in the range of 0.5 - 2.0 $m/year$. Apparently, direct human influences have a large variability in the change rate quantification with an offset towards bigger values. Consequently, it is proposed to set a temporal threshold for the minimum direct human influence at a change rate of around 2.0 $m/year$.

2. Single map with indirect human and natural influences

Figure 4.24 shows a single map related to retreating cluster 4 regular with enlarged features based on the change rate. Other areas than the ones in the first single map are highlighted. The Gulf of Mexico, the African continent and the Asian continent stand out with numerous features having change rates between the 7 and 11 $m/year$. The latter two are supported by the relatively large summed number of retreating features. However, not once, a feature has a change rate above the 11 $m/year$ in this cluster. Enlarged features in the map

have an equal probability (46%) of being influenced indirectly by humans or by nature when considering the correlations from the legend. Direct human influences are absent. This touches upon the smaller variability in indirect human and natural influences, with an offset towards smaller values.

For the case of the Gulf of Mexico, it is known that enlarged features are due to natural influences of storms with feature averaged change rates a little larger than 7 $m/year$. Besides this, features with change rates between the 0.5 and 2.0 $m/year$ are rarely seen and are all due to natural influences. One of the fourteen WPI port-related retreating shoreline evolution features is known to be driven by the Port of Nouakchott, Mauritania. This indirect human influence is one of the largest of its kind and has a change rate of approximately 10 $m/year$ and a width of around 10 km. Moreover, features with enlarged change rates between 7 and 11 $m/year$ at coastal stretches in Africa and India are found to be related to the indirect human influence of river sediment supply. Apparently, in general, the indirect human influences have a tendency to incorporate larger change rates than natural influences.

Based on the reasoning above, three different temporal thresholds are proposed. One for the minimum indirect human influence at a change rate of around 2 $m/year$ and another for the maximum indirect human influence at approximately 10.7 $m/year$. The last threshold considers the maximum natural influence, at a change rate of more or less 6.8 $m/year$.

3. Combined map with natural and indirect human influences

The combined map that takes prograding and retreating clusters 5 regular into account is shown in Figure 4.25. This map consists of shoreline evolution features with an enlarged width, where the African continent stands out. Enlarging multiple prograding or retreating features within a 50 km radius is also considered for the aforementioned clusters and shown in Figure C.46 in Appendix C.7. This map indicates large-scale influences and highlights, besides the continent of Africa, also the South-American continent.

There is an approximate equal probability of a feature belonging to an indirect human or natural influence for retreating features, where direct human influences are completely absent. Driver correlations were rejected for prograding cluster 5 regular in Section 4.2.3. However, it is known that the enlarged prograding features in Madagascar and Brazil in Figure C.46 are related to the natural influence of beach sediment supply. It is found that nearly all highlighted features resulting from the radius analysis are related to natural influences, as only the enlarged prograding features at North-East America and Japan are driven by large-scale indirect human interventions. The 35 possible WPI port '+'-signs are generally connected to features with a width smaller than 10 km, as seen in Figure 4.25. Apparently, features wider than approximately 10 km (proposed spatial threshold) are mainly related to natural influences that have a larger spatial scale.

4. Combined map with compound influences

Figure 4.26 visualizes combined prograding and retreating clusters 3 regular. Pairs (enlarged orange dots) are more or less randomly distributed over the world. These pairs indicate a prograding and retreating shoreline evolution feature within a 15 km radius. The map contains indirect human influences (10%), direct human influences (27%) and natural influences (27%). Note that these percentages differ from the driver correlations in the legend as it was stated that nourishments are not representative for the rest of the world features (see Section 4.2.3).

Possible WPI ports are identified 111 times, primarily in prograding features for Europe and Asia. This is supported by the large difference between the summed number of prograding and retreating features for these continents. For Asia, the above-mentioned is mainly due to direct human influences of land reclamations. In Europe, indirect human influences related to port / breakwaters are more evident. Driver correlations were rejected for retreating cluster 5 regular in Section 4.2.3. Nonetheless, it can be stated that for this cluster map, retreating features are not likely to be human-induced as this would have resulted in more pairs. The large difference in prograding and retreating features for the continent of South-America is found to be primarily related to natural influences due to the absence of identified WPI ports.

Compared to the previously analyzed maps, for this map, it is rather difficult to say something general about natural and human-induced shoreline evolution features without looking at a more local scale. This is due to large differences between the primary influences on features between continents. Consequently, this map is stated to indicate compound influences.

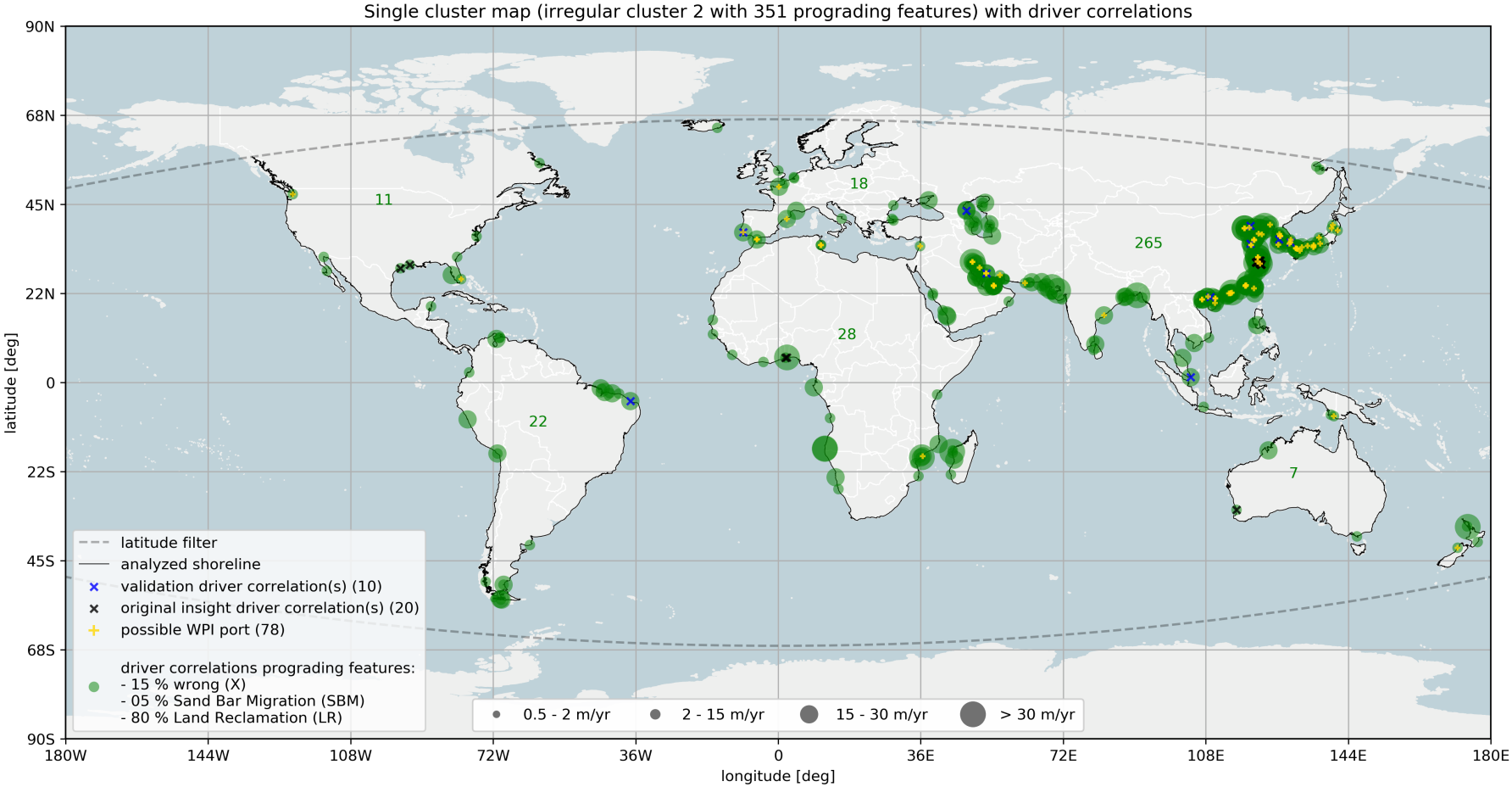


Figure 4.23: Single cluster map with driver correlations for 351 extracted prograding shoreline evolution features (green dots) on the analyzed 33 continuous shorelines (black lines). These features result from irregular cluster 2, which is part of the split dataset with 1165 features. The bottom-centered legend in the visualization indicates the change rate value in relation to the size of enlarged green dots. The driver correlations for this cluster (see bottom-left legend) are based on locally generated insight (black crosses), which is validated on randomly selected features (blue crosses). Features that are possibly related to a WPI port are marked with yellow '+'-signs. The green figures in the map indicate the number of extracted prograding features per continent. Latitude filters are shown with dashed gray lines.

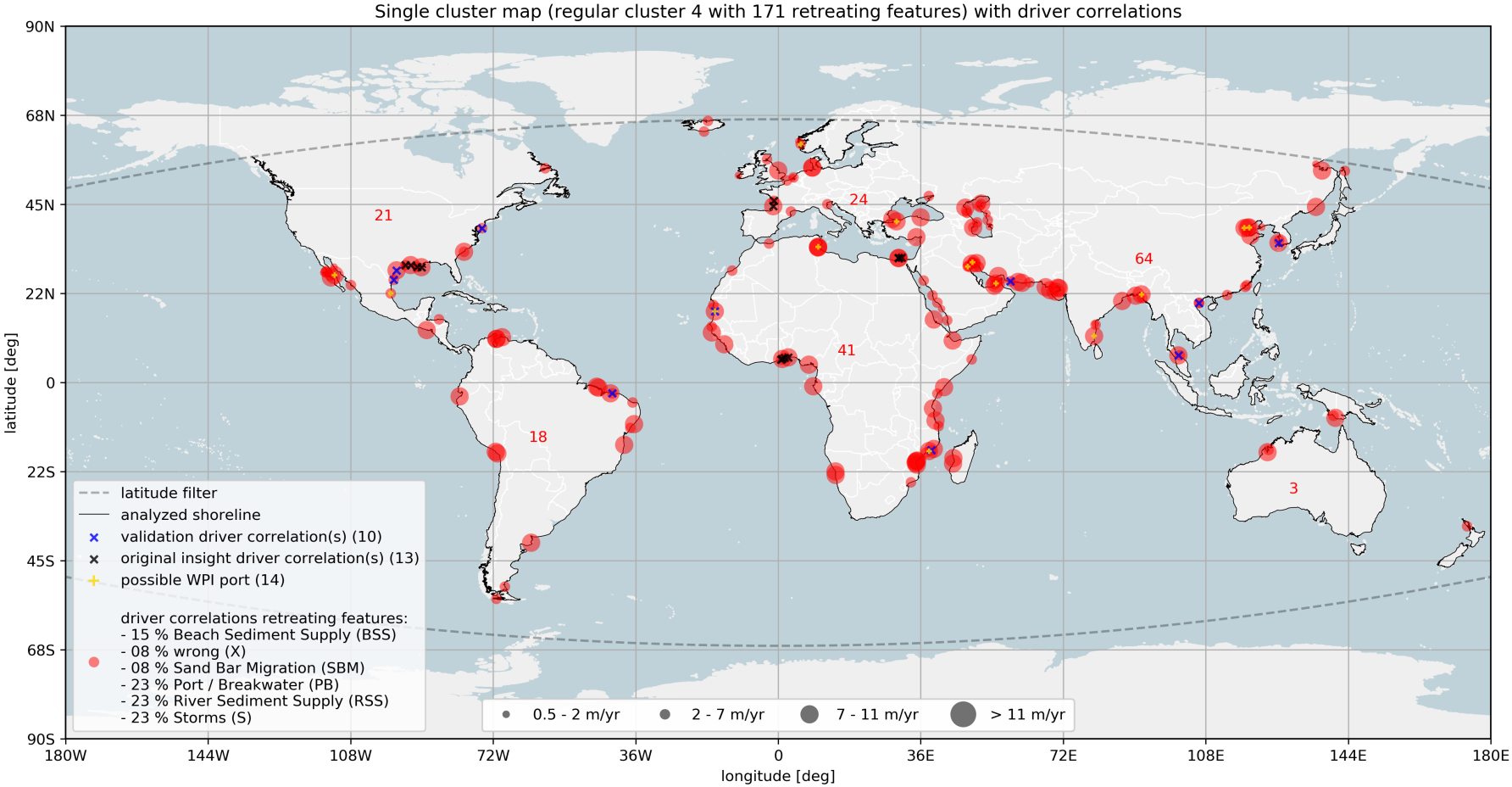


Figure 4.24: Single cluster map with driver correlations for 171 extracted retreating shoreline evolution features (red dots) on the analyzed 33 continuous shorelines (black lines). These features result from regular cluster 4, which is part of the split dataset with 1414 features. The bottom-centered legend in the visualization indicates the change rate value in relation to the size of enlarged red dots. The driver correlations for this cluster (see bottom-left legend) are based on locally generated insight (black crosses), which is validated on randomly selected features (blue crosses). Features that are possibly related to a WPI port are marked with yellow '+'-signs. The red figures in the map indicate the number of extracted retreating features per continent. Latitude filters are shown with dashed gray lines.

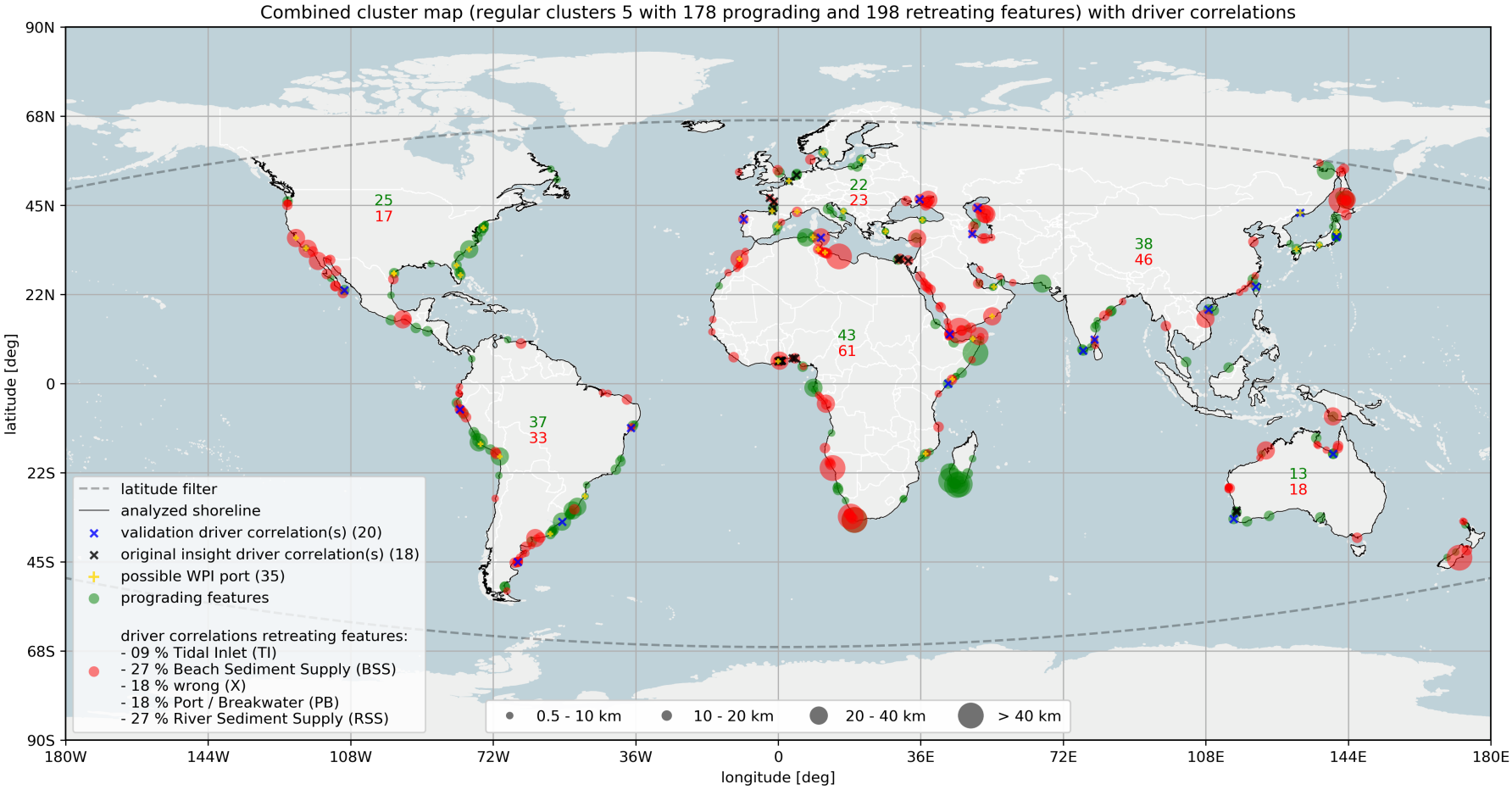


Figure 4.25: Combined cluster map with 178 extracted prograding shoreline evolution features (green dots) and 198 extracted retreating features (red dots) with driver correlations on the analyzed 33 continuous shorelines (black lines). These features result from regular clusters 5, which are part of the split datasets with 1865 and 1414 features. The bottom-centered legend in the visualization indicates the width value in relation to the size of enlarged green or red dots. The driver correlations for this cluster (see bottom-left legend) are based on locally generated insight (black crosses), which is validated on randomly selected features (blue crosses). There was an inconsistency in the locally generated insight and randomly selected validation features for the prograding cluster, which resulted in the rejection of driver correlations. Features that are possibly related to a WPI port are marked with yellow '+'-signs. The green and red figures in the map indicate the number of extracted prograding and retreating features per continent. Latitude filters are shown with dashed gray lines.

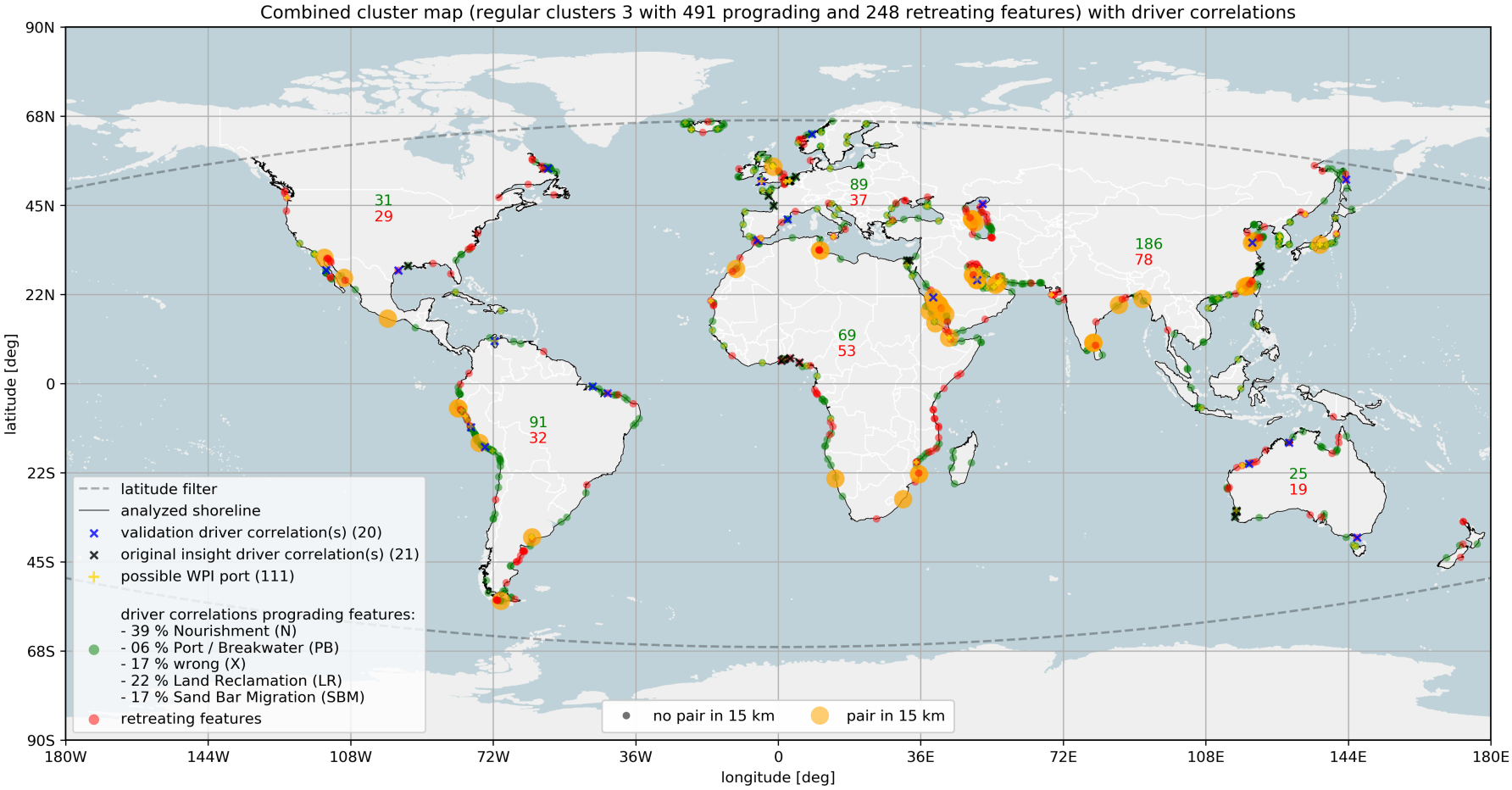


Figure 4.26: Combined cluster map with 491 extracted prograding shoreline evolution features (green dots) with driver correlations and 248 extracted retreating features (red dots) on the analyzed 33 continuous shorelines (black lines). These features result from regular clusters 3, which are part of the split datasets with 1865 and 1414 features. The bottom-centered legend in the visualization indicates the presence of a prograding and retreating feature in a 15 *km* radius in relation to the size of enlarged orange dots. The driver correlations for this cluster (see bottom-left legend) are based on locally generated insight (black crosses), which is validated on randomly selected features (blue crosses). There was an inconsistency in the locally generated insight and randomly selected validation features for the retreating cluster, which resulted in the rejection of driver correlations. Features that are possibly related to a WPI port are marked with yellow '+'-signs. The green and red figures in the map indicate the number of extracted prograding and retreating features per continent. Latitude filters are shown with dashed gray lines.

Overview

An overview of the cluster map analysis derived thresholds is shown in Table 4.12. Merging the derived thresholds into the cluster mean spatiotemporal signature visualization of Figure 4.21 results in Figure 4.27. This figure represents the final result of this study and is stated to show a statistically derived classification of natural and human-induced sandy shoreline evolution, where various regimes refer to different classes. The classification is based on the total of twelve quantifications, as thresholds were determined by analyzing cluster maps. However, it is only visualized for two quantifications. The subjective (cluster-independent) pattern analysis from Section 4.2.3, regarding manually labeled feature influences on the left and bottom panels in Figure 4.27, are approximately consistent with the derived thresholds. This is supported by the high percentiles presented in Table 4.12, which form a proxy for the reliability of the more objective cluster analysis derived thresholds compared to the subjective patterns from the manually labeled influences.

Table 4.12: Overview of the cluster map analysis derived thresholds. Besides, the percentiles of the split 284 manually labeled influences are shown in the last column. These percentiles refer to the number of data points above or below a threshold, depending on whether it is a minimum or maximum.

Influence threshold	Cluster map	Value [$m/year$ or km]	Manually labeled influence percentile
Temporal minimum direct human	1	2.0	97.5
Temporal maximum indirect human	2	10.7	95.0
Temporal maximum natural	2	6.8	92.5
Temporal minimum indirect human	2	2.0	97.5
Spatial direct human	3	5.0	99.0
Spatial indirect human	3	5.0	92.5

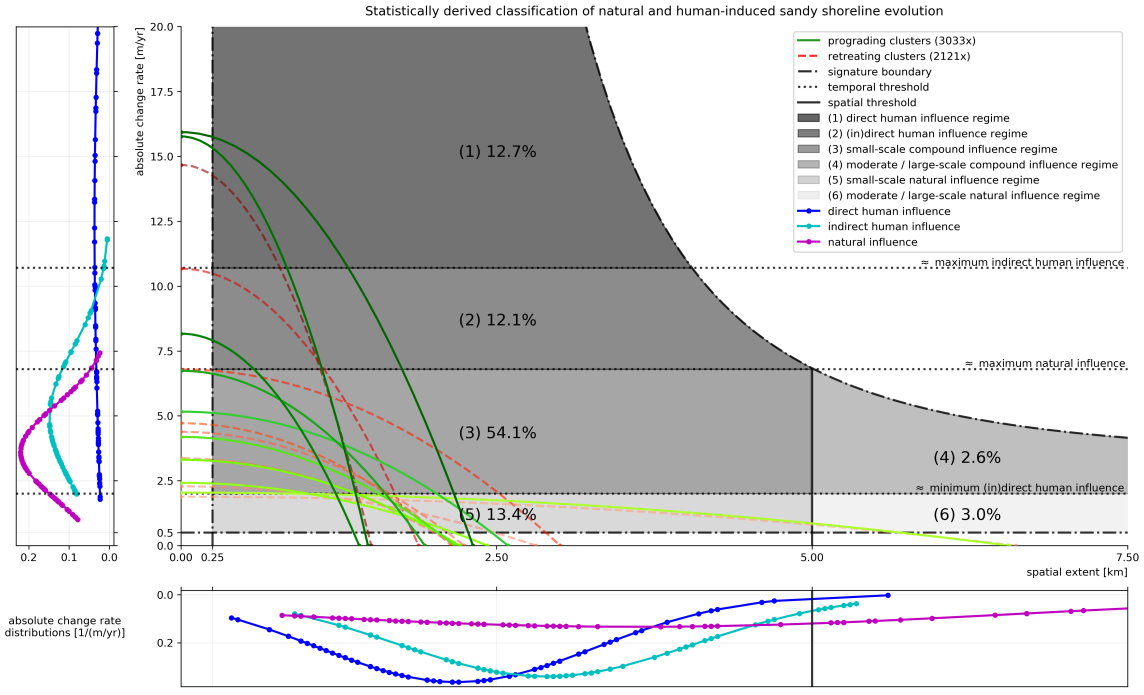


Figure 4.27: Visualization of the statistically derived classification of natural and human-induced sandy shoreline evolution for two of the twelve quantifications of the seventeen automatically derived and representative cluster mean spatiotemporal signatures. The greenish and dashed reddish colored lines are related to prograding clusters (nine times) and retreating clusters (eight times) respectively. The number of included features for the total of prograding and retreating clusters are shown between brackets in the legend. The signature boundaries are shown by means of dashed-dotted black lines and temporal thresholds with horizontal dotted black lines with indicative text on the far right of these lines. A vertical black line indicates the spatial threshold. All regimes (see legend) within the signature boundaries are indicated with different grayish colors, dependent on their temporal and spatial threshold splits. Percentages regarding the number of features (from the total of 5154 extracted features) within each of these regimes are also shown. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster. The left and bottom panel show the validated direct human (blue line with dots), indirect human (light blue line with dots) and natural (purple line with dots) influence distributions (independent from cluster assignments) for the absolute change rate and the extent respectively.

Per regime, the number of extracted and quantifed shoreline evolution features compared to the total of 5154 features are computed and shown in percentages in Figure 4.27. Considering planetary-scale shoreline evolution, these percentages indicate the relative importance of each of the regimes. Based on the classification, it is identified that around 25% of the world's shoreline evolution is due to (in)direct human influences (regime 1 and 2). Moreover, somewhat more than 16% of the extracted features is found to be related to natural influences (regime 5 and 6). The compound regimes (3 and 4) take natural and human-induced shore evolution into account and are stated to incorporate approximately 57% of all features. The remaining 2% of the features lay outside of this classification, which is due to the definition of the signature boundaries.

Table 4.13 outlines the percentage of manually labeled features as well as the global percentages within the classification regimes of Figure 4.27. Comparing these percentages results in a proxy for the representativeness of the manually analyzed features compared to the total of features (see Section 3.3). From Table 4.13, it follows that the manually analyzed shoreline quite accurately represents the global features in regime 1, 2 and 3. This is different for the other three regimes and the features outside the signature boundaries. It is seen that regime 5 contains a considerable lower percentage of shoreline evolution features compared to the world, which is added to regime 4, 6 and the area outside the signature boundaries. This relates to the fact that the eight locally analyzed shorelines were selected to incorporate obvious relatively large (-scale) shoreline evolution features.

Table 4.13: Overview of the percentages of global features as well as manually labeled features in each of the classification regimes (number of features between brackets). The column numbers refer to the regimes from Figure 4.27 and the column 'outside' to the white areas outside the boundaries. Cells are marked grey if the differences in a column are relatively large.

Feature dataset	1	2	3	4	5	6	outside
Global (5154x)	12.7	12.1	54.1	2.6	13.4	3.0	2.5
Local (284x)	12.7	11.8	54.7	5.3	6.1	4.9	4.4

By using the derived classification together with feature-continent labels (see Section 3.3), a continental mean spatiotemporal shoreline evolution signature analysis is enabled. This is shown in Figure 4.28.

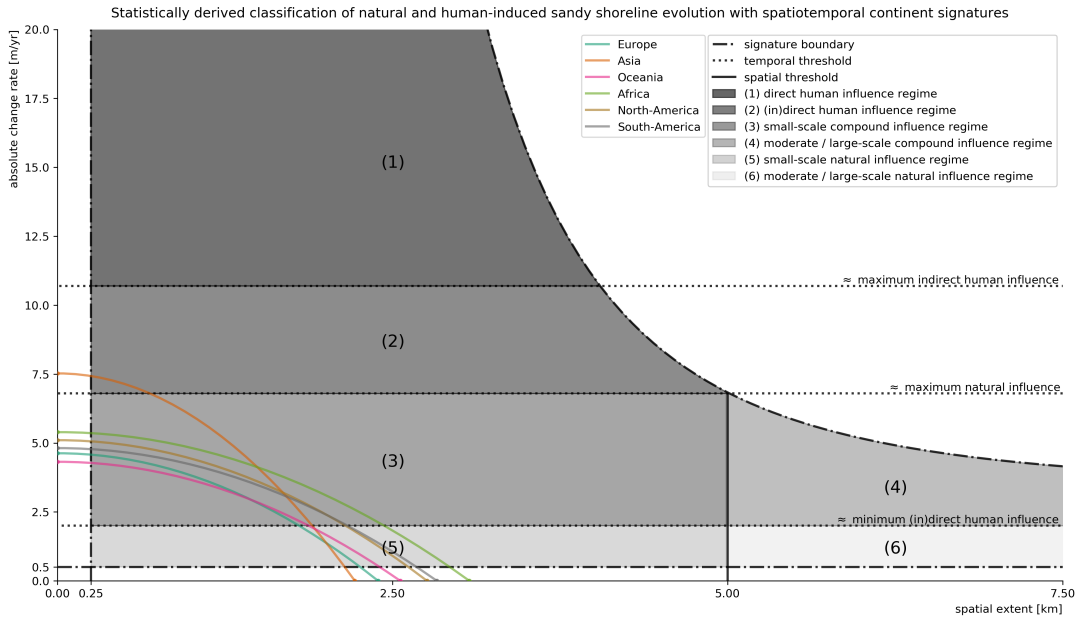


Figure 4.28: Visualization of the statistically derived classification of natural and human-induced sandy shoreline evolution for two of the twelve quantifications. Continental mean spatiotemporal signatures are shown and indicated with different colored lines as seen in the legend. The signature boundaries are shown by means of dashed-dotted black lines and temporal thresholds with horizontal dotted black lines with indicative text on the far right of these lines. A vertical black line indicates the spatial threshold. All regimes (see legend) within the signature boundaries are indicated with different grayish colors, dependent on their temporal and spatial threshold splits. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate.

As seen from Figure 4.28, the mean shoreline evolution signature of the continent of Asia is located in the (in)direct human influence regime. This is supported by various analyses in previous sections. The continent of Africa has the largest mean spatial extent, the second largest average change rate and is located in the compound regime. It can be stated that this continent consists of more natural influences but, whenever human interfere, mean long-term shoreline changes are generally larger due to the lack of knowledge or economic resources. In contrast to the continent of Africa, Europe and Oceania have approximately similar average shoreline evolution signatures with more human influences that have smaller mean effects. Note that the above-stated analysis on the differences in mean shoreline evolution signatures for the three continents located in the compound regime is found to be relatively insignificant. When standard deviation ranges are included, quite some overlap is present.

Implications for practice regarding the usability of the final outcome of this research (the actual classification), which was elaborated on in the paragraphs above, are presented in Chapter 5. For instance, features recognized to be erroneously extracted were excluded from the analysis when the focus was changed to influences instead to drivers (in the last part of Section 4.2.3). The statistical classification of Figure 4.27 can be used to identify shoreline evolution features that have a high possibility of containing the 'wrong' label. Potential outcomes are presented in Chapter 5 as well. These are, amongst others, related to the opportunities an enhanced insight in the compound regime (containing nearly 60% of all features) have to offer.

5 Discussion

This chapter contains a reflection on the limitations in developed or employed methods and obtained results presented in previous chapters (see Section 5.1). These limitations are mainly focused on assumptions, decisions and restrictions. Apart from limitations, implications for practice outlined in Section 5.2 involve discussing what the research outcomes might mean for individuals who work in this field of study. Considering the ability to solve some limitations, the potentials of this research explained in Section 5.3 indicate rather subjective 'out-of-the-box' thoughts regarding various possible future research outcomes.

5.1 Limitations

Three major assumptions, decisions and restrictions incorporated in the research methods and results of Chapters 3 and 4 are connected to literature and listed below.

1. Natural processes and location dependency

As outlined in the introduction, the world's shorelines are shaped by complex natural processes. These processes are location dependent as regional and global climate variability defines the rate at which local shorelines change. The relatively low temporal and spatial resolution of the Shoreline Monitor strongly complicates the inclusion of the aforementioned. Therefore, it is assumed to neglect the natural processes and the location dependency. This is a crude assumption when considering assessing the effect of different drivers on the spatiotemporal variability in shoreline evolution on a global scale and is the main reason the hypothesis introduced in the first chapter is found to be not entirely valid. Increasing the spatiotemporal resolution of the Shoreline Monitor allows to incorporate various higher resolution datasets like wave climates in the analyis. This can function to extract location dependent natural processes from long-term shoreline changes (see Wang, 2018a). It is foreseen that the above-stated will influence presented results and conclusions.

2. Multiple driver influences

Using the labeling method, features are correlated to drivers up to tertairy influences. However, only the main driver behind a shoreline evolution feature is considered. This decision is based on the fact that all twelve quantifications (four spatial and eight temporal) are restricted to the complete feature-included shoreline evolution trends due to the spatiotemporal resolution of the Shoreline Monitor. Consequently, intra-feature complex and combined (compound) influences are absent in the analysis (Stive et al., 2002). An increase in the resolution of the Shoreline Monitor and the application of the recommendation of van Leeuwen (2018) to split and fit multiple trend lines per decade might help to incorporate the aforementioned. This is elaborated on in Section 5.3 and will influence obtained findings and conclusions within this research.

3. Oscillatory drivers of shoreline evolution

This limitation is related to the developed feature extraction and quantification algorithm that uses linear change rates to detect features, making it impossible to identify oscillatory drivers. However, temporary coastal accretion or erosion might also lead to problems as outlined in Bird and Lewis (2015). Next to this, as outlined in the first point above, various shoreline evolution signals across the world are influenced by climate variability on a seasonal and inter-annual scale. The latter incorporates the phenomena of North Atlantic Oscillation (NAO) and El Niño - Southern Oscillation (ENSO). Again, the assumption is based on the relatively low temporal and spatial resolution of the Shoreline Monitor. Adjusting the feature extraction and quantification algorithm and increasing the spatiotemporal resolution of the Shoreline Monitor will inherently change presented results and conclusions.

Seven minor limitations incorporated in the research methods and results are connected to literature and listed below as well.

1. Feature extraction thresholds
To exclude excessive tops, bottoms and remaining noise in the feature extraction and quantification method, various thresholds are applied. These are iteratively determined to represent the most optimal configuration regarding performance indicators for four local application shoreline stretches. However, as seen in Figure 4.4, the small bump at around 25 km is not recognized as a shoreline evolution feature. From the nourishment scheme of `https://www.rijkswaterstaat.nl`, it is found that this bump is related to a shoreface nourishment. Generally, it can be stated that due to the thresholds, smaller shoreline evolution features (approximate difference between the largest top and largest bottom smaller than 0.5 $m/year$) are excluded from the analysis. This results in an offset towards larger extracted shoreline evolution features, which has an impact on all the findings presented thereafter. For instance, it is found that this offset is the reason behind the generation of larger cluster mean spatiotemporal signatures. Once more, the spatiotemporal resolution of the Shoreline Monitor as well as the accuracy of the transect system and SDS method are appointed to be the main cause of the offset towards the extraction of larger features. An increase in resolution or accuracy of the three aforementioned causes will allow to adjust thresholds and consequently also allows to decrease the offset. This has an effect on results and might even influence conclusions as well.

2. The quantification space
The GMM input for every split dataset consists of a sixteen dimensional space, as it is decided to duplicate the four spatial quantifications to weigh spatial and temporal variability in shoreline evolution features equally. Although it is found that reference runs with various other input dataset configurations do not lead to better cluster-driver correlations, it is not investigated what the correlations between quantifications are. Next to this, an increase in the resolution of the Shoreline Monitor is expected to allow to derive more relevant quantifications related to for instance natural processes (or physics, as in Wang, 2018a) or shoreline orientation. Changes to the quantification space inherently induce different results and conclusions than the ones presented here.

3. Erroneously extracted features
96 of the 380 manually labeled shoreline evolution features in this research are found to be erroneously extracted. This means that, after local investigations, it is concluded that the considered feature is stable instead of showing shoreline evolution. Most of the time, this is caused by frequent inaccuracy locations in the Shoreline Monitor transect system and SDS method (estuaries and bays, deltas and river mouths, cliffed coasts, and shallow shorelines). van Leeuwen (2018) recommended six solutions to reduce the number of inaccuracy locations. Three of these solutions are implemented, leaving the possibility to increase the accuracy of the Shoreline Monitor by plotting transects on SDS instead of on OSM shoreline, adjusting the transect length dynamically or statically, and splitting and fitting multiple trend lines per decade untouched. The latter is considered in Section 5.3, while a different solution to erroneously extracted features is provided in Section 5.2. A reduced number of erroneously extracted features might change various results.

4. Manual insight
Due to time restrictions, a relatively low percentage of the total of extracted features (7%, 380 features) are manually labeled using either literature, the GEE, the Shoreline Monitor or the Aqua Monitor. Besides, the latter three are rather subjective and most of the time only allow to correlate one main driver. This influences cluster-driver correlations, dataset splits, the decision to go for hard cluster label assignments and the first order assessment regarding classification thresholds, which are all obtained after the ML models are set-up. The inclusion of more manual insight may change outcomes regarding the aforementioned topics. Generally stated, ML problems consider cross-validation to assess the estimator performance using around 30 to 50% of known labels (Pedregosa et al., 2011). Cross-validation is not necessary as the unsupervised ML clustering method is used as classifier in this research. However, this does give a direction how much manual insight is required to form strongly substantiated conclusions.

5. Cluster assignments
Based on insight in the manually labeled features, it is decided to use hard cluster assignments while a GMM also provides the possibility to use 'soft' or probable cluster assignments (Pedregosa et al., 2011). From the

single feature label visualizations in Figure C.37 to C.42 in Appendix C.6, it can be seen that the usage of hard cluster assignments are a crude assumption for some of the features. Whenever more manual insight is generated, this decision might be adjusted and consequently findings and conclusions will be affected. The deviating probable and hard cluster assignments in the aforementioned figures might be due to multiple driver influences, as outlined in the second major limitation. Therefore, in case it is possible to split multiple driver influences in a feature, it may be worthwhile to investigate the effect of using probable cluster labels. This will, of course, also have an effect on presented research findings.

6. Planetary-scale assessment
This research is referred to as a planetary-scale assessment as it considers a large dataset with a wide geographical distributions of shorelines. However, 'only' 985 thousand of the 1.8 million available transects in the Shoreline Monitor dataset (54%) are taken into account. This is the result of a decision to only include continuous shorelines that surround (combined) continents or islands with an area larger than 25.000 km^2. Various results and conclusions might change whenever more data is included. Moreover, only 40% of the research-included transects contain actual change rates after the application of six filters. Therefore, 400 thousand transects are incorporated in the global analysis. This is mainly due to the limitations of the Shoreline Monitor employed SDS method, which was found to be inaccurate for 'non-sandy' labeled transects (Luijendijk et al., 2018). When the accuracy of the SDS method is increased, for instance when considering another shoreline detection approach as in Mentaschi et al. (2018), results and conclusions may also change.

7. Machine Learning algorithms and decision criteria
It is decided to consider only one ML algorithm (called GMM) for this research. The reason for this is related to time restrictions and the fact that this research is intended for the field of Coastal Engineering rather than Data Science. Besides, the usage of the BIC criterion to select the optimal number of cluster is questionable (see `https://jakevdp.github.io/PythonDataScienceHandbook/`). Contrarily, it is found that a smaller or larger number of clusters (than the BIC appointed number) does not necessarily result in better cluster-driver correlations. The application of a different ML approach or algorithm and other decision criteria might influence various research outcomes and conclusions.

5.2 Implications for practice
Taking into consideration all the limitations outlined the section above, this section elaborates on three main research outcomes (outlined in the list below) and examples of their applicabilities.

1. Generation of insight in large planetary-scale datasets.
   For any shoreline stretch on earth, we are able to extract spatial and temporally variable shoreline evolution features that may impose hazards to coastal communities;

2. Semi-automated derivation of plantary-scale patterns.
   We can show statistical similarities in characteristics between an extracted feature and any other feature in the world, which tells something about the state of the shoreline evolution, allows to couple literature from other locations and reveals global distributions that were not known before;

3. Planetary-scale classification of natural and human-induced sandy shoreline evolution.
   We can support detailed local-scale investigations by outlining the most probable or relevant long-term driver of shoreline change for the extracted features as well as indicating whether the features are located in the (in)direct human, natural or compound influence regime.

1. Shoreline evolution on different spatial scales
As outlined in Section 3.1.1, the feature extraction method is developed to incorporate all drivers of shoreline evolution by capturing and quantifying spatiotemporal variability on different spatial scales or moving windows (2.5, 10, 20, 50 and 100 km). This is based on the summation of all available transect linear change rates divided by the total summed length between all transects closest to a window. It is stated that the 5, 10, 20, 50 and 100 km spatial scales are only of use to explain shoreline behavior at the start of the analysis in Sections 4.1.1 and 4.1.2, as all types of drivers are found to be contained in the 2.5 km window. Here, it is outlined that the 5, 10, 20, 50 and 100 km windows can be used for another application.

Instead of using moving windows to extract and quantify features from continuous shorelines, hard sections can be used to visualize and explain the alongshore mean shoreline evolution on different spatial scales. This is shown for hard 100 km sections in Figure D.1 in Appendix D. This figure indicates that there are 1409 hard 100 km sections, implying that around 140 thousand km of shoreline is either prograding or retreating (around 35% of the total analyzed shoreline). 65% of the total analyzed shoreline is either stable (within +/- 0.5 $m/year$) or contains too many 'non-sandy' labels (primarily around the equator). Hard section shoreline evolution maps can also be created for the other spatial scales, which allows to compare long-term shoreline changes on different levels of littoral or sedimentary cells.

2. Identification of erroneously extracted features
In the third minor limitation of the section above, it is stated that frequent inaccuracy locations in the Shoreline Monitor induce erroneously extracted features. These features were first brought to light in Section 4.2.1. Rather than implementing the three remaining solutions of van Leeuwen (2018) to improve the accuracy of the Shoreline Monitor, a different solution (introduced in the last paragraph of Section 4.3) to identify erroneously extracted features is elaborated on below.

For the 96 manually labeled erroneously extracted features (see Section 4.2.3), the temporal change rate and one-sided spatial width (referred to as extent) distributions are added to Figure 4.27. This results in the creation of Figure D.2 in Appendix D, where interesting subjective patterns can be seen again. Based on the change rate, it can be stated that erroneously extracted features are approximately indicating a similar distribution as direct human influences. From the spatial extent quantification, it becomes clear that erroneously extracted features are mainly present up to approximately 5 km width. Consequently, Figure D.2 allows to assess whether a certain extracted feature somewhere on the earth has a probability of being labeled 'wrong'.

3. Arbitrary shoreline evolution classification
This application is approximately similar to the continental analysis presented in the final discussed findings in Section 4.3. As the global feature datasets also contain other informative labels, figures like Figure 4.28 can be created on arbitrary levels like regions, countries, or even single features.

5.3 Potentials

This section elaborates on three rather subjective potential outcomes, provided that some limitations outlined in Section 5.1 are resolved. In other words, the topics presented here establish a link with the recommendations provided in Chapter 7 and thus function as possible research directions.

1. Enhanced understanding of compound influences
At the beginning of the research, the following hypothesis is formulated: *'a persistent or episodic driver of shoreline evolution shows approximately similar characteristics in spatiotemporal variability everywhere on earth'*. This hypothesis stands at the basis of the research methods up to approximately Section 3.2.3 and is based on various assumptions, decisions and restrictions.

In Section 4.2.3, it is stated that the majority of the clusters contain multiple correlated drivers. It is found that all major limitations and the first five of seven minor limitations of Section 5.1 have an influence on the separability of different drivers among clusters. This is substantiated in Figures 4.27 and 4.28, which indicate that almost 60% of the total of all extracted features and five of the six continents contain (mean) spatiotemporal shoreline evolution signatures that are located in the compound regime. For this reason, the question is raised whether it is really possible to verify the above-mentioned hypothesis. To answer this question, one might start investigating ways to resolve the eight aforementioned limitations. However, the three major limitations are stated to be the primary reason of the incapability of finding similar characteristics for a persistent or episodic driver of shoreline evolution anywhere on earth. In the two paragraphs below, possible solutions to the first two major limitations are elaborated on.

The considered Shoreline Monitor dataset-included 500-m spaced transects contain yearly shoreline positions that construct the shoreline evolution trend within these transects. There is a possibility of increasing the temporal resolution to biweekly shoreline positions, as stated in Section 2.1.4. As the generation of 1.8 million transects containing biweekly data is very computationally intensive, it might be interesting to replace feature-included annual transects (approximately 49 thousand, see Section 4.1.2) by biweekly ones to

begin with (also referred to as spatiotemporal subsets in Wang, 2018a). Next to this, one might decide to increase the spatial resolution within these features by adding more transects. With an increased resolution, well founded methods like PCA or EOF can be used to isolate the dominant modes of spatial or temporal variability within the shoreline evolution features. This allows to couple more natural processes or physics, rather than solely depending on spatial and temporal characteristics.

Besides incorporating more natural or physical aspects, which might allow to identify, separate and isolate natural influences within features, intra-feature compound human influences require to be separated as well (if possible). So far, rather large differences in annual shoreline positions have restricted the number of quantifications that could be derived. This resulted in the decision to consider only the main driver. However, from Figure 5.1 it is seen that the usage of annual data and the application of a trend break algorithm (as recommended by van Leeuwen, 2018) with a manually set threshold of 90 $m/year$ already provides a possibility to split different human interventions. The trend breaks at the years 2011 and 2012 in the above-mentioned figure are identified to belong to the construction of the Sand Engine. Therefore, from 2011 onwards, the main influence on shoreline evolution is seen to be related to the complex processes around a mega nourishment. Before 2011, smaller effects on shoreline evolution are due to groynes and shoreface nourishments. Increasing the temporal resolution of the Shoreline Monitor inherently leads to having a larger number of available characteristics. Furthermore, it also allows to decrease the aforementioned threshold to create a possibility of considering drivers of shoreline evolution up to tertiary influences.

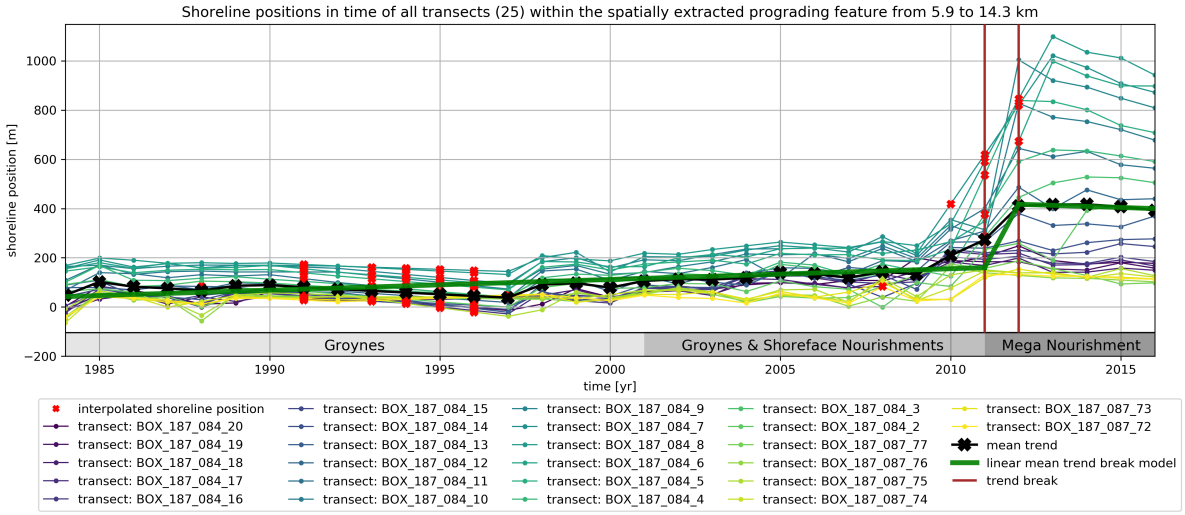


Figure 5.1: Example visualization of intra-feature compound human interactions in shoreline evolution on a 2.5 km scale. This example considers the second identified prograding feature from the left (5.9 to 14.3 km in Figure 4.4) on Dutch coast. Twenty-five transects (see label id's in the legend) are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. The mean trend of all shoreline positions per year is shown by means of a black line with crosses, while the interpolated shoreline positions are indicated by red crosses. The mean trend is split by an automatic trend break algorithm, indicating three different linear trends (green lines) split on the years indicated by a vertical brown line.

2. Impact assessments

This potential outcome is touching upon an extension of the developed framework or methods presented in this research and is thus not related to any of the limitations from Section 5.1. As the geographical locations as well as the characteristics of features are known and incorporated in planetary-scale datasets, it might be interesting to consider impact assessments regarding rather persistent shoreline evolution hazards to coastal communities or terrestrial and marine protected areas.

An analysis on the above-mentioned may be molded in two consecutive steps. First, together with the appointed (in)direct human or natural influences, the annual (or even biweekly) incorporated shoreline positions within features might be used to derive accurate predictions for future shoreline behavior. Secondly, different geospatial databases might be coupled to the above-mentioned. For instance, assessments regarding impacts on nature might consider coupling the World Database on Protected Areas (WDPA), as in Luijendijk et al. (2018). In relation to human impact assessments, one might think of coupling databases like

Gross Domestic Product (GDP), Global Roads Inventory Project (GRIP), Human Built-up And Settlement Extent (HBASE) or Dynamic Interactive Vulnerability Analysis (DIVA) as in Cowan (2017). The latter already takes into account the first step related to future projections for coastal progradation or retreat.

3. Comprehensive classification of coastal systems

Similar to the section above, this section does not relate to any of the limitations from Section 5.1. As stated in Chapter 1, a coastal classification can be used as a means to inventory the large variability in coastal system / zones around the world. Finkl (2004) proposed to use a unified system for integrated and systematic approaches to coastal classifications, as the availability of information is ever-increasing. Earlier general classifications were broad in scope but lacked specificity or narrowly focused, providing uneven coverage of shorelines around the world. This research resulted in a statistically derived classification of natural and human-induced sandy shoreline evolution by using spatial and temporal characteristics on a planetary scale. This may be of use as an addition to the proposed comprehensive classification model of Finkl (2004) to provide an enhance understanding in overlapping and interrelated problems in beach and surf zone environments.

6

Conclusion

So far, detailed local-scale studies are able to expose human and natural drivers of long-term shoreline changes and provide a possibility to make a step towards intentional rather than accidental coastal engineering. Recent developments in satellite imagery allow to generate large geospatial datasets with an increased spatial and temporal scale. The annual dataset behind the Shoreline Monitor of Luijendijk et al. (2018) is an application of satellite imagery. This dataset incorporates an assessment of the occurrence of sandy beaches and a change detection method called Satellite Derived Shorelines (SDS) to represent shoreline evolution on a planetary scale, every 500 m (transect) along the coast for the past 33 years. Nonetheless, the opportunity to develop a model that exposes the drivers of shoreline evolution on a planetary scale remains unexplored. This is due to the required computational effort and the large variability in coastal systems around the world.

With an ever-increasing data availability, data-driven models incorporating Machine Learning (ML) are recognized as an efficient alternative approach to heavy computing classical process-driven models in civil engineering practice. Next to this, a coastal classification can be used as a means to inventory the aforementioned variability. Therefore, the research objective in this study is to explore the possibility of exposing and classifying the drivers of shoreline evolution on a planetary scale, by employing ML on the annual dataset behind the Shoreline Monitor. The objective is translated into a research question, stated below.

To what extent can the annual dataset behind the Shoreline Monitor be used to derive a planetary-scale classification of sandy shoreline evolution when employing Machine Learning?

To answer this research question, we developed, employed and molded various methods into a multi-step framework that is outlined in the next paragraphs. Here, the following hypothesis stands at the basis: *'a persistent or episodic driver of shoreline evolution shows approximately similar characteristics in spatiotemporal variability everywhere on earth'*.

Worldwide, we first extracted and quantified 3033 prograding and 2121 retreating features (5154 in total) with an offset towards larger long-term shoreline changes. Here, features are defined as multiple sandy shoreline evolution transects with change rates larger or smaller than 0.5 $m/year$ and quantifications are related to twelve characteristics (four spatial and eight temporal) that explain the spatiotemporal variability in long-term shoreline changes. So, for any place, country or continent on earth, we are able to extract spatial and temporally variable shoreline evolution features that may impose hazards to coastal communities.
Hereafter, on a global scale, we semi-automatically derived seventeen groups / clusters with driver-related similarities in spatiotemporal variability by using an unsupervised ML method (Gaussian Mixture Model, GMM) and local manual insight. This means that we can show statistical similarities in characteristics between an extracted feature and any other feature in the world. Hence, we can tell something about the state of the shoreline evolution, couple literature from other similar locations and reveal global distributions that were not known before. Besides, we found that nearly all of the clusters contain multiple drivers of shoreline evolution. Therefore, it is concluded that the aforementioned hypothesis is not entirely valid considering various assumptions, restrictions and decisions. Nonetheless, we indicated that redirecting the research focus from single drivers to the classes of (in)direct human and natural influences on shoreline evolution shows interesting patterns.
Finally, considering the redirected research focus, we analyzed geographical driver distributions of all features in six (combined) clusters to statistically derive a planetary-scale classification of natural and human-induced sandy shoreline evolution. We found that natural and human-induced shoreline evolution accounts

for approximately 16 and 25% of the total of globally extracted and quantified features respectively. By looking at mean shoreline evolution on continent level, we derived that Asia's features are mainly due to (in)direct human influences. 57% of all features is present in the compound regime with complex and combined influences. We also know what this compound regime consists of in terms of potential drivers. For instance, we found that five continents are located in the aforementioned regime. On average, the continent of Africa consists of more natural influences. However, whenever humans interfere, mean long-term shoreline changes are generally larger due to the lack of knowledge or economic resources. In contrast to the continent of Africa, Europe and Oceania show more human influences that have smaller mean effects. Consequently, we can indicate the most probable or relevant driver of the extracted shoreline evolution features as well as whether the features are located in the (in)direct human, natural or compound influence regime.

All the above-stated outcomes can support detailed local-scale investigations and therefore provide an enhanced opportunity to make a step towards intentional rather than accidental coastal engineering. So, it is concluded that the developed and applied methods that employ ML on the annual dataset behind the Shoreline Monitor can be used to expose and classify (in)direct human and natural influences on sandy shoreline evolution using spatial and temporal characteristics on a planetary scale. Nevertheless, almost 60% of the shoreline evolution features are present in the compound regime, which still requires (instead of supports) local-scale investigations to determine the correct influence or driver. Different causes for this outcome were elaborated on in the discussion in the previous chapter and are primarily due the exclusion of variability in natural processes and location dependency, multiple driver influences, and oscillatory drivers of shoreline evolution. More research is necessary to elaborate on the opportunities that can enhance insight in the compound regime, improve the obtained results and increase the applicability of this study.

7

Recommendations

Primarily based on the limitations and the potential outcomes outlined in the discussion of Chapter 5, but also on the conclusion of Chapter 6, this chapter provides an overview of recommendations for a future agenda regarding possible research directions.

A total of six concrete and feasible recommendations are provided in the list below, which is prioritized from top to bottom. Here, the first four items consider the recommendations to achieve the first potential outcome (incorporating a total of eight limitations), while item five and six take potential outcomes two and three into account respectively. Note that item five and six are not linked to any limitation.

1. Increase the temporal and possibly also the spatial resolution of the dataset behind the Shoreline Monitor that is incorporated in the research methods. From all limitations listed in Section 5.1, it is evident that 50% (three major and the first two minor) is due to the research-included annual and 500 m spatiotemporal resolution of the Shoreline Monitor dataset. Increasing the resolution requires a significantly larger and possibly even unworkable computational effort, as the limits of computational power are already tested for the planetary-scale assessment in this research. Therefore, it is recommended to assess whether it is possible to enhance the understanding of compound influences with an increase in resolution on either continent or country level, while only adjusting the data within extracted feature ranges (as outlined in Section 5.3);

2. Include various adjustments to the Shoreline Monitor transect system and SDS method or global shoreline evolution feature datasets to reduce the number of erroneously extracted features. One adjustment can be related to the application outlined in Section 5.2, where the number of erroneously extracted features can be reduced by iteratively assessing the most probable driver considering the width quantification. Another adjustment relates to the possibility to filter noise regarding excessively computed SDS, that induce erroneously extracted features, by filtering on the actual transect-included sand polygon (see Section 2.1.3) instead of the complete transect. However, it is recommended to investigate whether it is possible to apply adjustments regarding the remaining three recommendations of van Leeuwen (2018) first. This refers to plotting transects on SDS instead of on OSM shoreline, adjusting the transect length dynamically or statically, and splitting and fitting multiple trend lines per decade. The latter already indicated promising outcomes for annual data, as outlined in Section 5.3;

3. Verify the research findings and conclusions by including more manually labeled extracted and quantified features. Due to time restrictions, only 7% of the 5154 extracted features is used to generate insight in this research. It is recommended to increase the 7% to at least 30% (see Section 5.1). Making the extracted and quantified prograding and retreating feature datasets publicly available and accessible through the frontend application of the Shoreline Monitor (`shorelinemonitor.deltares.nl`) might be an interesting option to achieve the above-mentioned recommended percentage more easily and objectively;

4. Verify the assumption that hard cluster assignments are sufficiently accurate after consideration of the three items above. In case this is not sufficient, it is recommended to adjust the research methods from Section 3.2.3 onwards. It is also recommended to use figures like the ones provided in Appendix C.6 to assess whether it is possible to enhance the understanding of compound influences even more;

5. Extend the research methods framework to incorporate impact assessments considering episodic or persistent shoreline evolution features. It is recommended to investigate this opportunity by using methods like multivariate linear regression or autoregressive regression models before coupling various other geospatial datasets. These kinds of statistical methods can be trained with data in the feature-included period and can be used to extrapolate non-linear shoreline evolution behavior in the (near) future. Note that an increased temporal resolution (item one) will inherently lead to better predictions;

6. Assess whether it is possible to merge the final outcome regarding a statistically derived classification of natural and human-induced sandy shoreline evolution into the unified (comprehensive) open ended system for coastal classifications of Finkl (2004). It is recommended to investigate this opportunity by determining whether the research outcome is to be classified, through establishing links or relations between shoreline evolution features and coastal landforms.

When facing difficulties or having trouble understanding the followed procedures, assumptions or decisions considering the developed and employed methods, results or recommendations for research directions, please do not hesitate to get in touch: etiennekras@gmail.com.

Bibliography

Aarninkhof, S. G., Turner, I. L., Dronkers, T. D., Caljouw, M., and Nipius, L. (2003). A video-based technique for mapping intertidal beach bathymetry. *Coastal Engineering*, 49(4):275 – 289.

Addo, K. A., Walkden, M., and Mills, J. (2008). Detection, measurement and prediction of shoreline recession in accra, ghana. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(5):543 – 558. Theme Issue: Remote Sensing of the Coastal Ecosystems.

Ahsan, R., Kellett, J., and Karuppannan, S. (2016). 19 - climate migration and urban changes in bangladesh. In Shaw, R., ur Rahman, A., Surjan, A., and Parvin, G. A., editors, *Urban Disasters and Resilience in Asia*, pages 293 – 316. Butterworth-Heinemann.

Airoldi, L. and Bulleri, F. (2011). Anthropogenic disturbance can determine the magnitude of opportunistic species responses on marine urban infrastructures. *PLOS ONE*, 6(8):1–9.

Al-Tahir, R. and Ali, A. (2004). Assessing land cover changes in the coastal zone using aerial photography. *Surveying and Land Information Science*, 64:107–112.

Almonacid-Caballer, J., Sánchez-García, E., Pardo-Pascual, J. E., Balaguer-Beser, A. A., and Palomar-Vázquez, J. (2016). Evaluation of annual mean shoreline position deduced from landsat imagery as a mid-term coastal evolution indicator. *Marine Geology*, 372:79 – 88.

Anthoff, D., Nicholls, R., Tol, R., and Vafeidis, A. (2006). *Global and regional exposure to large rises in sea-level: A sensitivity analysis*. Number 96 in Working paper. Tyndall Centre for Climate Change Research.

Bird, E. and Lewis, N. (2015). *Beach Renourishment*. Springer International Publishing AG, Cham.

Boak, E. H. and Turner, I. L. (2005). Shoreline Definition and Detection: A Review. *Journal of Coastal Research*, 214:688–703.

Bosboom, J. and Stive, M. (2015). *Coastal Dynamics I: lecture notes CIE4305*. Delft Academic Press (DAP).

Burningham, H. and French, J. (2017). Understanding coastal change using shoreline trend analysis supported by cluster-based segmentation. *Geomorphology*, 282:131 – 149.

Bush, D., Pilkey, O., and Neal, W. (2001). Coastal topography, human impact on*. In Steele, J. H., editor, *Encyclopedia of Ocean Sciences (Second Edition)*, pages 581 – 590. Academic Press, Oxford, second edition edition.

Carter, R. W. G., Jennings, S. C., and Orford, J. D. (1990). Headland erosion by waves. *Journal of Coastal Research*, 6(3):517–529.

Chakrabarti, A. and Ghosh, J. K. (2011). Aic, bic and recent advances in model selection. In Bandyopadhyay, P. S. and Forster, M. R., editors, *Philosophy of Statistics*, volume 7 of *Handbook of the Philosophy of Science*, pages 583 – 605. North-Holland, Amsterdam.

Church, J. A. and White, N. J. (2011). Sea-level rise from the late 19th to the early 21st century. *Surveys in Geophysics*, 32(4):585–602.

Colwell, R. N., Ulaby, F. T., Simonett, D. S., Estes, J. E., and Thorley, G. A. (1983). *Manual of remote sensing (Theory, instruments and techniques, Vol. 1)*. Falls Church: American Society of Photogrammetry.

Cowan, T. (2017). New developments in global beach erosion assessment - improving projections using the diva model. Master's thesis, Delft University of Technology.

Davidson, C. M. (2009). Dubai: foreclosure of a dream. *Middle East report*, 251 (Summer):8–13.

de Bruijn, L. (2018). Maintenance dredging in the port of rotterdam. Master's thesis, Delft University of Technology.

Dietz, M. E., Liu, K., and Bianchette, T. A. (2018). Hurricanes as a major driver of coastal erosion in the mississippi river delta: A multi-decadal analysis of shoreline retreat rates at bay champagne, louisiana (usa). *Water*, 10(10).

Dissanayake, P., Ranasinghe, R., and Roelvink, D. J. (2012). The morphological response of large tidal inlet/basin systems to relative sea level rise. *Climatic Change*, 113:253–276.

Dolan, R., Hayden, B., and Heywood, J. (1978). Analysis of coastal erosion and storm surge hazards. *Coastal Engineering*, 2:41 – 53.

Donchyts, G., Baart, F., Winsemius, H., Gorelick, N., Kwadijk, J., and Van De Giesen, N. (2016). Earth's surface water change over the past 30 years.

El Sayed, W., Ali, M., Iskander, M., and Fanos, A. (2007). Evolution of rosetta promontory on nile delta coast during the period from 1500 to 2005, egypt. *Proceedings of the 8th International Conference on the Mediterranean Coastal Environment, MEDCOAST 2007*, 2:1003–1016.

Estivill-Castro, V. (2002). Why so many clustering algorithms: A position paper. *SIGKDD Explor. Newsl.*, 4(1):65–75.

Everitt, B. and Skrondal, A. (2010). *The Cambridge dictionary of statistics*. Cambridge University Press Cambridge, UK ; New York, 4th ed. edition.

Finkl, C. (2004). Coastal classification: Systematic approaches to consider in the development of a comprehensive scheme. *Journal of Coastal Research*, 20:166–213.

Frihy, O. E., Dewidar, K. M., Nasr, S. M., and Raey, M. M. E. (1998). Change detection of the northeastern nile delta of egypt: Shoreline changes, spit evolution, margin changes of manzala lagoon and its islands. *International Journal of Remote Sensing*, 19(10):1901–1912.

Garcin, M. and Le Cozannet, G. (2013). The driving factors of coastal evolution: toward a systemic approach.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18 – 27. Big Remotely Sensed Data: tools, applications and experiences.

Hagenaars, G. (2017). Accuracy assessment of coastline dynamics based on satellite images; application to the holland coast. Master's thesis, Delft University of Technology.

Hagenaars, G., de Vries, S., Luijendijk, A. P., de Boer, W. P., and Reniers, A. J. (2018). On the accuracy of automated shoreline detection derived from satellite imagery: A case study of the sand motor mega-scale nourishment. *Coastal Engineering*, 133:113 – 125.

Hayes, M. O. (1967). Relationship between coastal climate and bottom sediment type on the inner continental shelf. *Mar. Geol. 5: 111-132*.

Hinkel, J., Nicholls, R. J., Tol, R. S., Wang, Z. B., Hamilton, J. M., Boot, G., Vafeidis, A. T., McFadden, L., Ganopolski, A., and Klein, R. J. (2013). A global analysis of erosion of sandy beaches and sea-level rise: An application of diva. *Global and Planetary Change*, 111:150 – 158.

Hoepffner, N. and Zibordi, G. (2009). Remote sensing of coastal waters. In Steele, J. H., editor, *Encyclopedia of Ocean Sciences (Second Edition)*, pages 732 – 741. Academic Press, Oxford, second edition edition.

Hollings, B. (2004). Sediment dynamics of warnbro sound, western australia. Master's thesis, The University of Western Australia.

LaGro, J. (2005). Land-use classification. In Hillel, D., editor, *Encyclopedia of Soils in the Environment*, pages 321 – 328. Elsevier, Oxford.

Linham, M. M. and Nicholls, R. J. (2010). *Technologies for Climate Change Adaptation: Coastal Erosion and Flooding*.

Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G., and Aarninkhof, S. (2018). The State of the World's Beaches. *Scientific Reports*, 8(1).

Mao, Y. (2018). Coastline evolution around african seaports - an evidence database from space. Master's thesis, Delft University of Technology.

McGranahan, G., Balk, D., and Anderson, B. (2007). The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization*, 19(1):17–37.

Mentaschi, L., Vousdoukas, M. I., Pekel, J.-F., Voukouvalas, E., and Feyen, L. (2018). Global long-term observations of coastal erosion and accretion. *Scientific Reports*, 8(1):12876.

Miller, J. K. and Dean, R. G. (2007). Shoreline variability via empirical orthogonal function analysis: Part i temporal and spatial characteristics. *Coastal Engineering*, 54(2):111 – 131.

Morton, R. (2002). Factors controlling storm impacts on coastal barriers and beaches - a preliminary basis for near real-time forecasting. *Journal of Coastal Research*, 18:486–501.

Neumann, B., Vafeidis, A. T., Zimmermann, J., and Nicholls, R. J. (2015). Future coastal population growth and exposure to sea-level rise and coastal flooding - a global assessment. *PLOS ONE*, 10(3):1–34.

Newville, M., Stensitzki, T., Allen, D. B., and Ingargiola, A. (2014). Lmfit: Non-linear least-square minimization and curve-fitting for python.

OpenStreetMap (2019). Coastlines.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

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

Pontee, N. I. (2011). Reappraising coastal squeeze: a case study from north-west england. *Proceedings of the Institution of Civil Engineers - Maritime Engineering*, 164(3):127–138.

Ranasinghe, R. (2016). Assessing climate change impacts on open sandy coasts: A review. *Earth-Science Reviews*, 160:320 – 332.

Ranasinghe, R., Callaghan, D., and Stive, M. J. F. (2012). Estimating coastal recession due to sea level rise: beyond the bruun rule. *Climatic Change*, 110(3):561–574.

Ranasinghe, R., Duong, T., Uhlenbrook, S., Roelvink, D. J., and Stive, M. (2013). Climate-change impact assessment for inlet-interrupted coastlines. *Nature Climate Change*, 3:83–87.

Reich, Y. (1996). Machine learning techniques for civil engineering problems. *Microcomputers in Civil Engineering*, 12.

Richardson, L. (1961). The problem of contiguity: An appendix to statistics of deadly quarrels. *General System Yearbook*, 6:139–187.

Rosati, J. D. (2005). Concepts in sediment budgets. *Journal of Coastal Research*, pages 307–322.

Seto, K. C. and Shepherd, J. M. (2009). Global urban land-use trends and climate impacts. *Current Opinion in Environmental Sustainability*, 1(1):89 – 95.

Singh, A. (1988). Digital change detection techniques using remotely sensed data. *International Journal of Remote Sensing*, 10.

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

Stive, M., De Schipper, M., Luijendijk, A., Ranasinghe, R., Thiel de Vries, J., Aarminkhof, S., van Gelder-Maas, C., de Vries, S., Henriquez, M., and Marx, S. (2013). The sand engine: A solution for vulnerable deltas in the 21st century?

Stive, M. J., Aarninkhof, S. G., Hamm, L., Hanson, H., Larson, M., Wijnberg, K. M., Nicholls, R. J., and Capobianco, M. (2002). Variability of shore and shoreline evolution. *Coastal Engineering*, 47(2):211 – 235. Shore Nourishment in Europe.

Vaidya, A., Kori, S. K., and Kudale, M. (2015). Shoreline response to coastal structures. *Aquatic Procedia*, 4:333 – 340. INTERNATIONAL CONFERENCE ON WATER RESOURCES, COASTAL AND OCEAN ENGINEERING (ICWRCOE'15).

van Leeuwen, D. (2018). Global shoreline dynamics; a revision of the shoreline monitor.

Wang, N. (2018a). Unravelling the sandy shoreline dynamics derived from satellite images. Master's thesis, Delft University of Technology.

Wang, S., Fu, B., Piao, S., Lu, Y., Ciais, P., Feng, X., and Wang, Y. (2015). Reduced sediment transport in the yellow river due to anthropogenic changes. *Nature Geoscience*, 9.

Wang, Z. B. (2018b). Long term morphological development of the tidal inlet systems in the dutch wadden sea.

Yang, S. L., Zhang, J., Zhu, J., Smith, J. P., Dai, S. B., Gao, A., and Li, P. (2005). Impact of dams on yangtze river sediment supply to the sea and delta intertidal wetland response. *Journal of Geophysical Research: Earth Surface*, 110(F3).

Zhang, Y. and Ling, C. (2018). A strategy to apply machine learning to small datasets in materials science. *npj Computational Materials*.

A

Background Information

This appendix provides additional background information that is related to the id label structure of the Shoreline Monitor transect system, shortly touched upon in Chapter 2.

The Shoreline Monitor of Luijendijk et al. (2018) contains around 2.2 million 500-m spaced transects that are located on the global shoreline (represented by all green boxes in Figure A.1). Visualizing, adjusting and performing calculations with all transects in for instance Quantum Geographic Information System (QGIS) or Python would considerably slow down any computer. For this reason, transects are split and labeled with id's according to their position, as shown in Figure A.1 and elaborated on in the next paragraph. 130 of the 330 boxes consist of only land or water and are excluded from the dataset. For another 74 of the green-marked boxes, shoreline data is unavailable or covered in ice (box position completely above or below the latitude filters of 60°N and 50°S). Consequently, 126 of the boxes shown in Figure A.1 are considered in the Shoreline Monitor dataset and contain the aforementioned 2.2 million transects.

Every transect in the Shoreline Monitor has a label id with a structure as follows: BOX_xxx_xxx_xxx. Here, the first set of x's refers to the box number found in Figure A.1. The second set of x's is represented by the number of a smaller sub-box within a box (this is not shown in Figure A.1). The last set of x's considers the alongshore transect number within a sub-box.

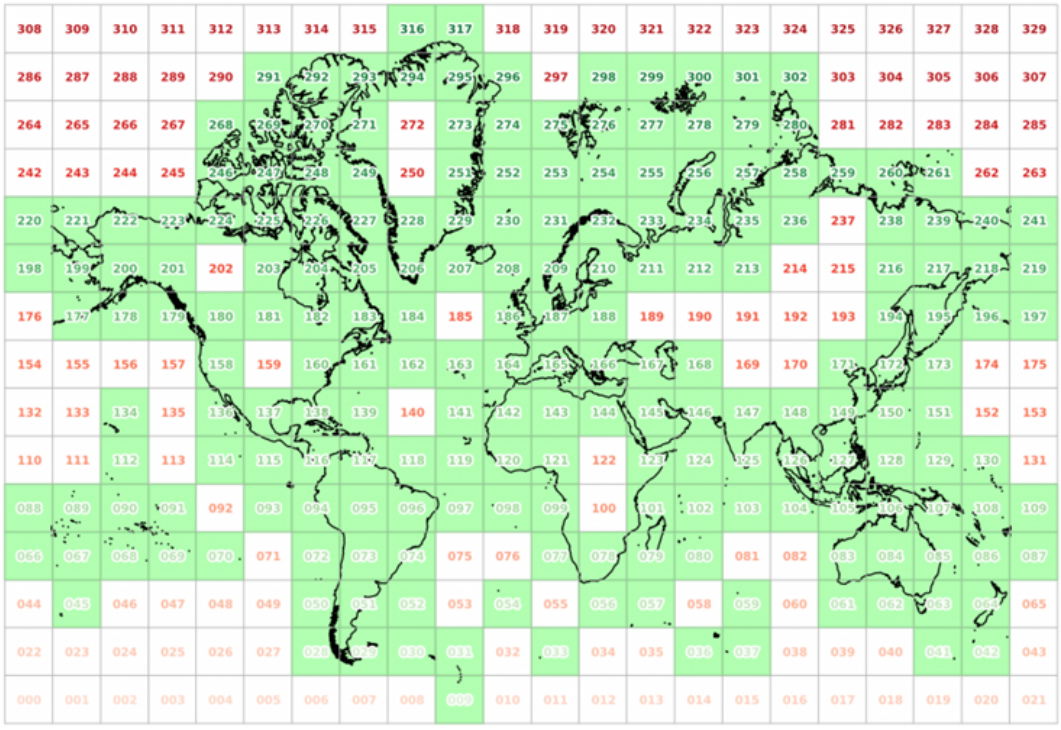


Figure A.1: Gridded world visualization related to the id label structure of the Shoreline Monitor transect system. Every green box represents a shoreline, whereas white boxes with red numbers are either water or land. The latter are excluded from the research.

B

Research Methods

This appendix provides additional information on the methods presented in Chapter 3. Section B.1 contains an explanation on the developed methods to pre-process Shoreline Monitor data (step zero from Figure 3.1). In Section B.2, four local application shoreline stretches are visualized and the different curve-fitting models with performance indicators are elaborated on. Four more local application shoreline stretches are shown in Section B.3.

B.1 Pre-processing Shoreline Monitor data

The 2.2 million sandy / non-sandy 500-m spaced annual transects of the Shoreline Monitor are originally ordered according to the transect id label system, as outlined in Appendix A. As a result of this ordering, a planetary-scale 'along-the-shoreline'-analysis is impossible. Therefore, the dataset requires to be reordered (pre-processed), by means of least distances between transects. An algorithm is developed that is able to order transects, related to shoreline evolution signals from the annual dataset behind the Shoreline Monitor, along global continuous shorelines. Filter adjustments, transect-shoreline intersections and continuous shorelines have to be computed or constructed before reordering the transects. The methods to do this are elaborated on in the paragraphs below and supported by Figure B.1.

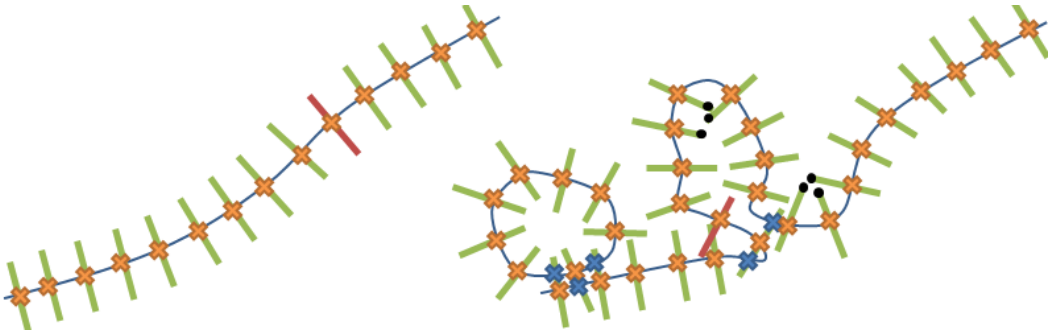


Figure B.1: Schematic visualization of a uniform (blue, left) and non-uniform (blue, right) shoreline with correct (green) and filtered (red, only once) shore-normal transects. Correct transect intersections are shown with orange crosses. In the case of a non-uniform shoreline, a single transect might intersect multiple times with either the same or another shoreline (shown with blue crosses). Transect origins and ends are shown in black dots for the peninsula and bay respectively, in the non-uniform shoreline.

Filter adjustments

As explained in Sections 2.1.2, 2.1.3 and 2.1.4, various filters have been applied to the Shoreline Monitor to increase the accuracy of results. These filters result in the exclusion of transects in the annual dataset behind the Shoreline Monitor. However, no transect in the 'along-the-shoreline'-analysis is allowed to be rejected as it can compromise the reliability of results in this research. Whenever a transect is excluded in the case of a uniform shoreline, this will not form a reliability problem as the straight line between the transect intersections to the left and right of the red-marked filtered transect will follow the shoreline approximately correctly (see Figure B.1). However, in the case of a non-uniform shoreline, exclusion of the red-marked transect will form a reliability problem. Figure B.1 shows that reordering, using least distances between transects, will result in the exclusion of the entire peninsula. Consequently, the filtered transects in the annual dataset of the Shoreline Monitor must be retrieved. By labeling these transects, they can be excluded again in the actual analysis.

Transect-shoreline intersections
The annual dataset behind the Shoreline Monitor does only contain transect origins and ends, indicated by the black dots (at the peninsula and bay respectively) in Figure B.1. When computing the least distances of the transect origins at the peninsula, one transect will be left out of the 'along-the-shoreline'-analysis. The same happens for the transect ends at the bay. For this reason, it is required to take the least distances between true transect-shoreline intersections into account.

As can be seen from the blue crosses in Figure B.1, one transect might intersect with the actual shoreline as well as with a different shoreline or multiple times with the same shoreline. Only the 'true' transect-shoreline intersection needs to be obtained. In order to achieve this, transect-shoreline intersections are computed in two different projections (WGS3857 and WGS4326, see Section 2.1.2). By using the fact that transects have a length of 1 km (500 m land- and shoreward), the distances of intersections with the center of a transect are calculated in WGS3857. If a single transect contains multiple intersections, only the coordinates of an intersection closest to the center are kept in WGS4326. The intersection that is kept also contains an id that relates to one of the OSM shoreline ids (the highest resolution OSM shorelines contains 30374 individually labeled pieces), which will be used in the next sub-step.

Continuous shorelines
The shore-normal transects in the Shoreline Monitor are defined on OSM shorelines as explained in Section 2.1.2. There are 30374 small pieces of OSM shoreline that represent the total global shoreline. For an 'along-the-shoreline'-analysis, these small individual pieces of OSM shoreline are manually welded together to form a continuous shoreline / OSM id-string for (combined) continents or islands with an area larger than 25.000 km^2. Within the boundaries of the latitude filters, 33 continuous shorelines are created for this research. Transects are grouped according to the OSM id-string, which is possible as the previous sub-step computed the correct shoreline id for transect intersections. This automatically means that small offshore islands, as seen for the right non-uniform shoreline of Figure B.1, are excluded from further computations.

The combined continent of Africa, Europe and Asia (Afro-Eurasia) is considered the largest area surrounded by a continuous shoreline (source: `https://en.wikipedia.org/wiki/List_of_islands_by_area`). The combined continent of North- and South-America (Americas) and the continent of Australia follow in the second and third place respectively. The island of Silicy (Italy) is the smallest area surrounded by a continuous shoreline taken into account.

The 500-m spaced global transect system behind the Shoreline Monitor was created on a flat projection. Flat projections of the earth (for both WGS4326 and WGS3857) generate an increasing offset in computed distance away from the equator. This means that transects are spaced with 500 m intervals at the equator, while the spacing becomes smaller further away from the equator. The total considered continuous shoreline in this research can still be computed if the offset in transect spacings is corrected. This is done using the Universal Transverse Mercator (UTM) coordinate system, where the correct system is found based on the coordinates of a transect intersection. When considering the coastline paradox, introduced by Richardson (1961), the 'true' shoreline length cannot be determined as it can be analyzed on different spatial scales. However, statements on the shoreline length accompanied by the spatial resolution of the measurements or maps are justified.

Transect reordering
The transect reordering algorithm starts from the most upper left point of every created continuous shoreline. The grouped sets of transects are ordered based on the smallest linear distance between two transect intersections in the WGS3857 projection. The linear distance is calculated using the pythagoras theorem ($c = \sqrt{a^2 + b^2}$) with a correction according to the appropriate UTM coordinate system. The original transect system contains quite some gaps bigger than the flat projection spacing of 500 m, as a result of the id label structure outlined in Appendix A. Hence, after reordering of a transect, the coordinates are replaced by Not a Number (NaN). This prevents a transect from appearing twice, if the distance to the previous transect is smaller than the smallest distance to the next transect. In case of a highly non-uniform coast, it might happen that a few transects are passed. This results in very large and incorrect connections to a few last transects at the end of some continuous shorelines. These connections are excluded manually.

B.2 Local feature extraction and quantification

The first section below shows the four local application shoreline stretches. In the second section, background information regarding the applied curve-fitting methods and performance indicators in the temporal layer of the feature extraction and quantification method are outlined.

Local application shoreline stretches

Visualizations of the local and moderate-scale application shoreline stretches that consider The Netherlands, Belgium, South-West France and South-West Australia are presented below. These stretches are used for developing and testing the local feature extraction and quantification method / algorithm outlined in Section 3.1.1 and applied in Section 4.1.1.

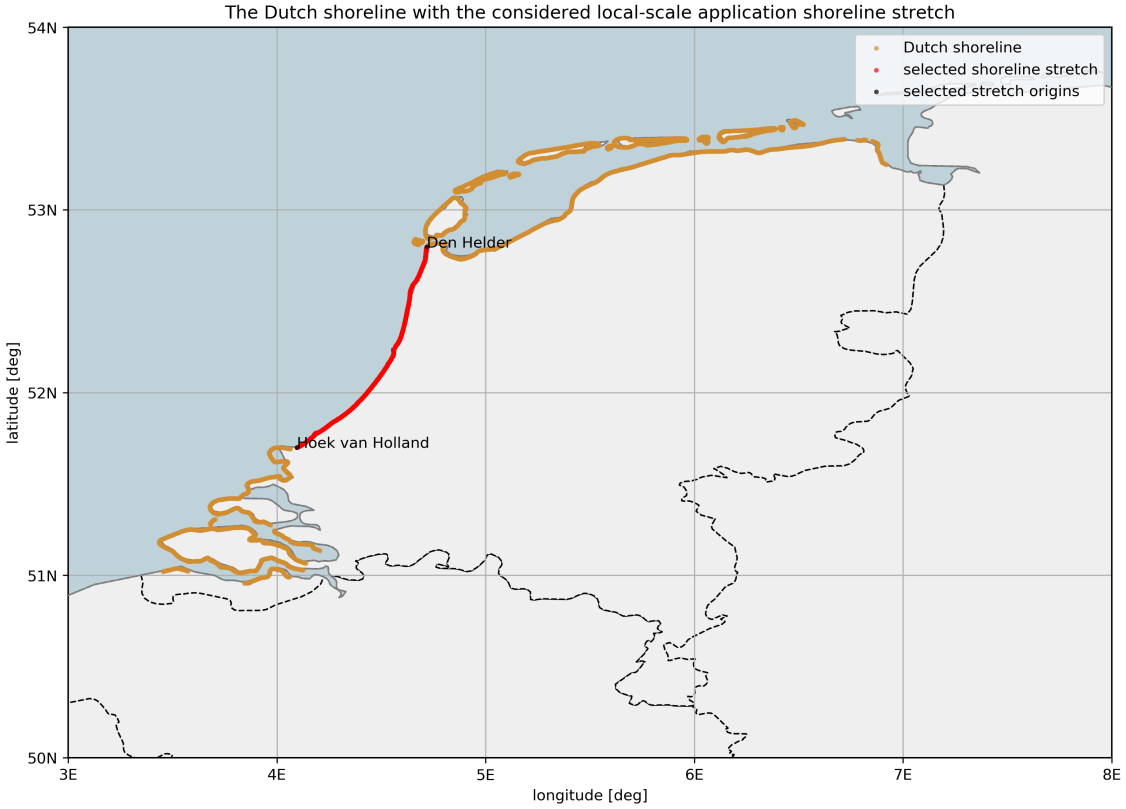


Figure B.2: Visualization of the selected local-scale application shoreline stretch (red-marked dots with black dots and indicative text at the origins) related to surrounding shorelines (brownish dots) of The Netherlands. In this case, the surrounding shorelines represent all Shoreline Monitor included transects referred to by the aforementioned country.

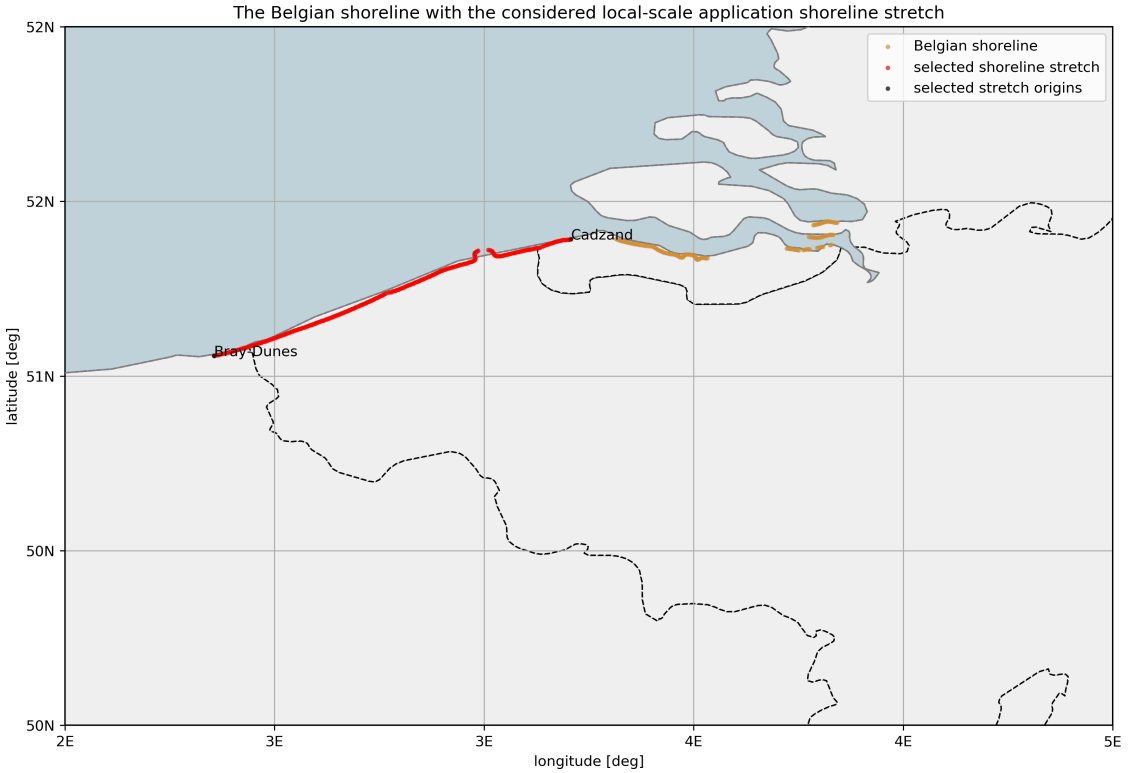


Figure B.3: Visualization of the selected local-scale application shoreline stretch (red-marked dots with black dots and indicative text at the origins) related to surrounding shorelines (brownish dots) of Belgium. In this case, the surrounding shorelines represent all Shoreline Monitor included transects referred to by the aforementioned country.

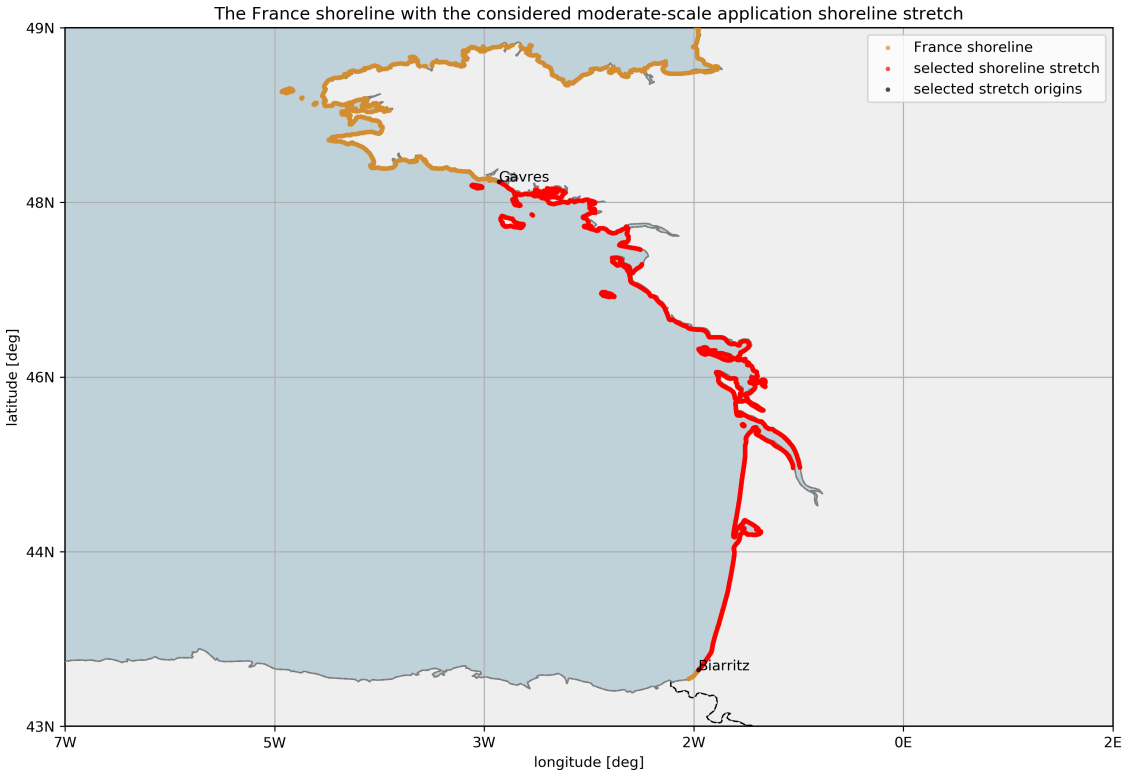


Figure B.5: Visualization of the selected moderate-scale application shoreline stretch (red-marked dots with black dots and indicative text at the origins) related to surrounding shorelines (brownish dots) of France. In this case, the surrounding shorelines represent a part of the Shoreline Monitor included transects referred to by the aforementioned country.

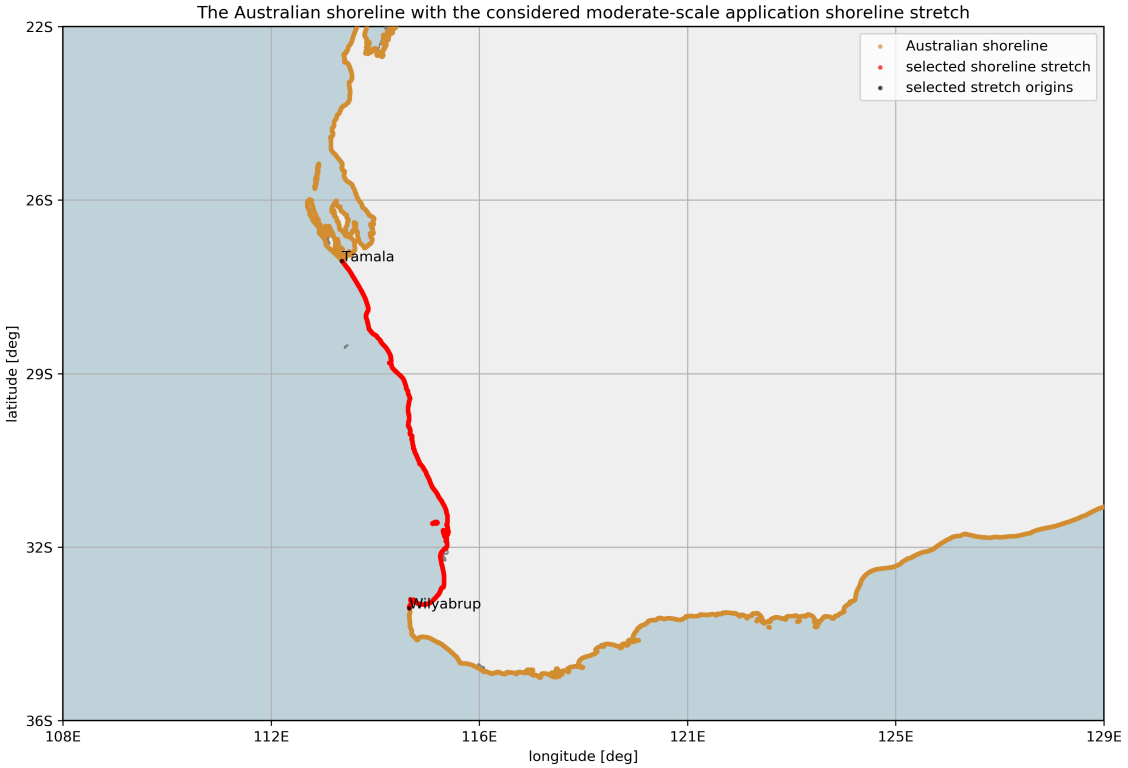


Figure B.4: Visualization of the selected moderate-scale application shoreline stretch (red-marked dots with black dots and indicative text at the origins) related to surrounding shorelines (brownish dots) of Australia. In this case, the surrounding shorelines represent a part of the Shoreline Monitor included transects referred to by the aforementioned country.

Curve-fitting models and performance indicators

Newville et al. (2014) provides a package with built-in models for non-linear least-squares problems. Here, least-squares refers to the standard approach in a regression analysis to approximate the solution of a set of equations in which there are more equations than unknowns (overdetermined system). In a formula form, least-squares look like $minimize_{x \in R^n} \sum_i f_i(x)^2$, where f_i are real-valued functions with parameters x. It is attempted to iteratively minimize the total sum of the residual squares in the results of all equations. If all f_i in the equation above are affine functions (linear function + a constant), the problem is stated to be linear. If not, the problem is referred to as a non-linear form of the least-squares analysis.

In total, eighteen different built-in curve-fitting models are applied in the temporal layer of the feature extraction and quantification method, i.e.: eight times a polynomial model (0th to 7th order), eight times a step-like model (step and rectangle with linear, arc-tan, error or logistic functions), once an exponential model and once a power law model. Here, a 0th, 1st and 2nd order polynomial refer to a constant, linear and quadratic model respectively. The mathematical background behind all these models is not elaborated on here, as this is clearly outlined in the online user manual of Newville et al. (2014). This manual can be accessed through the following link: `https://lmfit.github.io/lmfit-py/builtin_models.html`.

The performance indicators to assess the goodness of the fit, considering the earlier introduced terms of overfitting and underfitting, are referred to as the reduced chi-square, BIC, AIC or AICc. The applicability of these indicators differs per problem, while for all indicators a lower value refers to a better fit. The mathematical descriptions of the four performance indicators are given below, in the same order as stated above.

$$\chi^2_{red} = \sum_i^N r_i^2 / (N - N_{var}) \tag{B.1}$$

$$BIC = N ln(\sum_i^N r_i^2 / N) + ln(N) N_{var} \tag{B.2}$$

$$AIC = Nln(\sum_{i}^{N} r_i^2/N) + 2N_{var} \tag{B.3}$$

$$AICc = Nln(\sum_{i}^{N} r_i^2/N) + 2N_{var} + \frac{2N_{var}^2 + 2N_{var}}{N - N_{var} - 1} \tag{B.4}$$

In all formulae above, r_i refers to the residual array that is returned by the non-linear least-squares optimization. Next to this, N refers to the number of data points and N_{var} to the number of variable parameters. As AIC is prone to overfit in case N is relatively small, it may be required to use AICc. AICc corrects for small sample sizes, by means of the last term in Equation B.4. Note that when N is very large, the correction term converges to zero and AICc reduces to AIC again.

B.3 Local feature-driver labeling

Visualizations of the moderate-scale application shoreline stretches that consider Egypt, Mid-West Africa (Ghana, Togo, Benin and Nigeria), Mid-East China and South USA (Gulf of Mexico) are presented below. These stretches are used for the local feature-driver labeling method outlined in Section 3.2.1 and applied in Section 4.2.1.

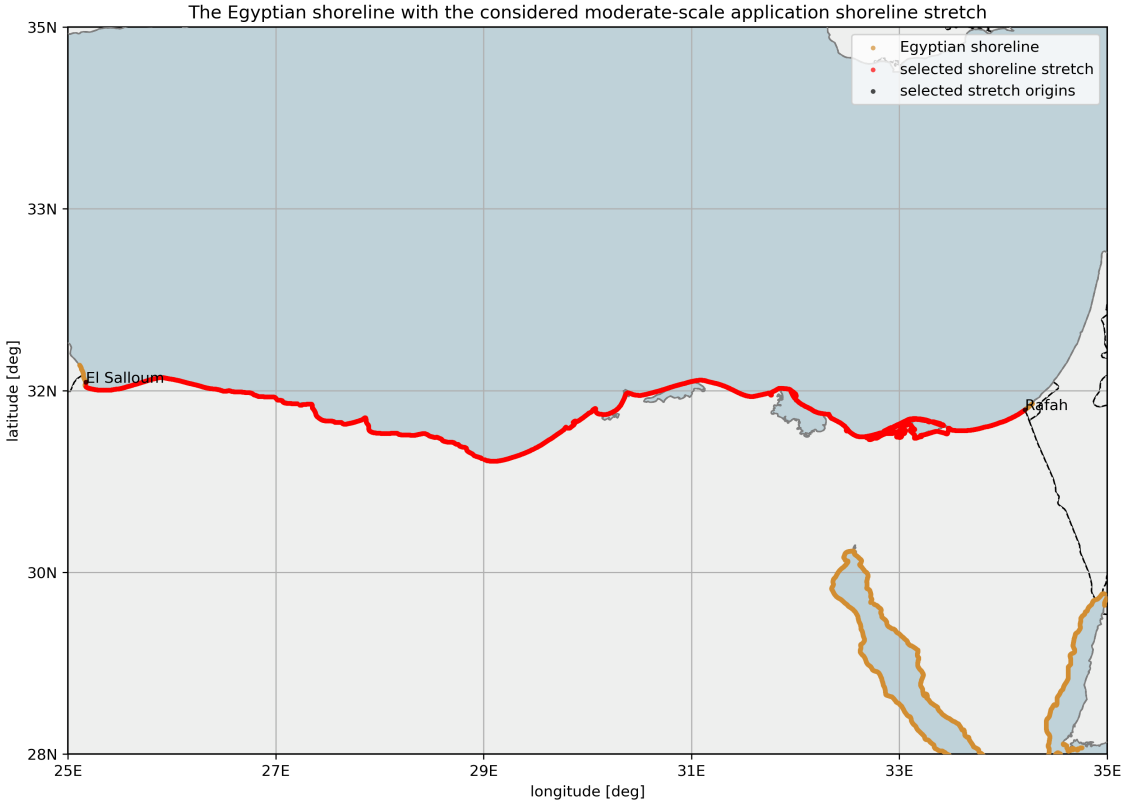


Figure B.6: Visualization of the selected moderate-scale application shoreline stretch (red-marked dots with black dots and indicative text at the origins) related to surrounding shorelines (brownish dots) of Egypt. In this case, the surrounding shorelines represent a part of the Shoreline Monitor included transects referred to by the aforementioned country.

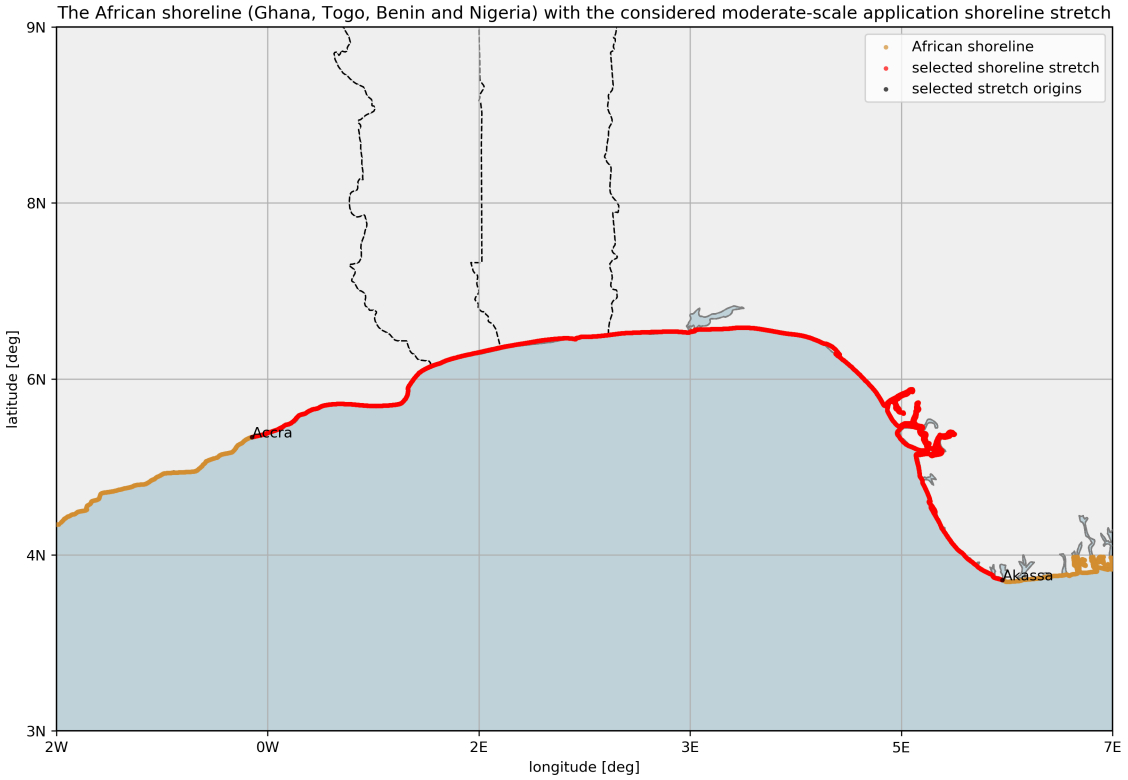


Figure B.7: Visualization of the selected moderate-scale application shoreline stretch (red-marked dots with black dots and indicative text at the origins) related to surrounding shorelines (brownish dots) of Africa's Ghana, Togo, Benin and Nigeria. In this case, the surrounding shorelines represent a part of the Shoreline Monitor included transects referred to by the aforementioned countries.

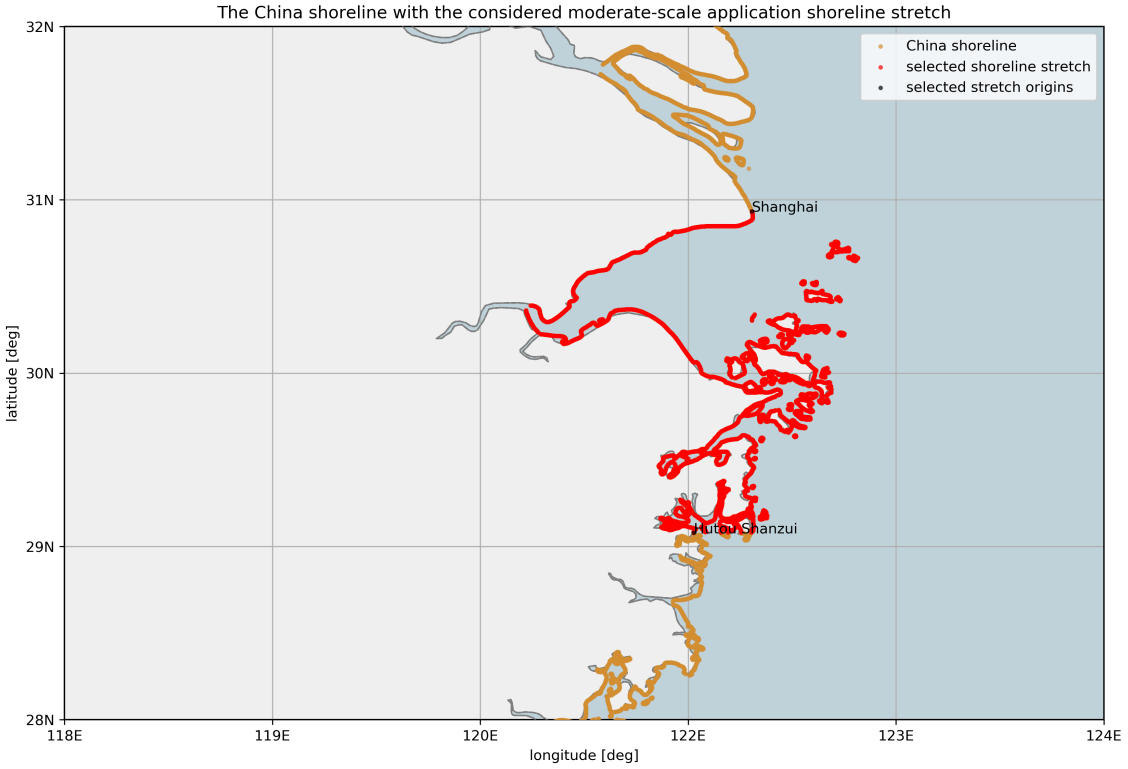


Figure B.8: Visualization of the selected moderate-scale application shoreline stretch (red-marked dots with black dots and indicative text at the origins) related to surrounding shorelines (brownish dots) of China. In this case, the surrounding shorelines represent a part of the Shoreline Monitor included transects referred to by the aforementioned country.

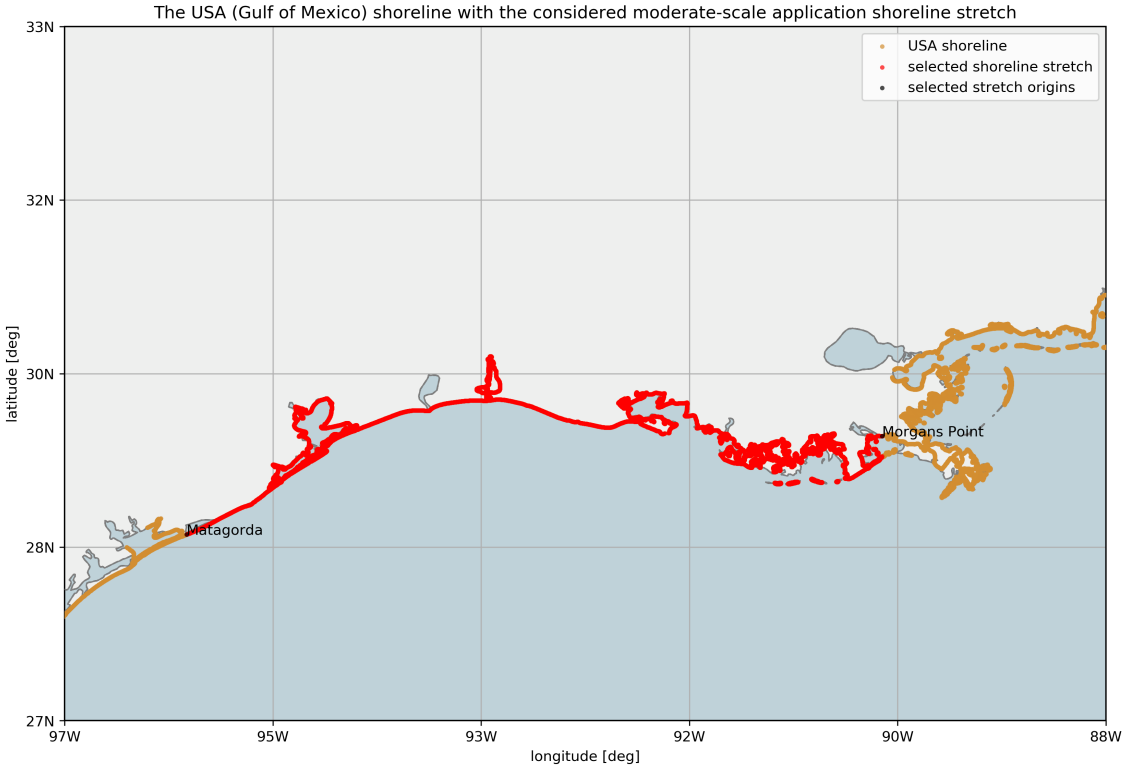


Figure B.9: Visualization of the selected moderate-scale application shoreline stretch (red-marked dots with black dots and indicative text at the origins) related to surrounding shorelines (brownish dots) of the USA. In this case, the surrounding shorelines represent a part of the Shoreline Monitor included transects referred to by the aforementioned country.

C
Results

This appendix contains additional outcomes that are related to the results presented in Chapter 4. Section C.1 provides the findings that consider pre-processing Shoreline Monitor data. This is based on the method presented in Appendix B.1. In Section C.2, the results of the local feature extraction and quantification method of the two remaining shoreline stretches as well as some additional results of the considered shorelines in Section 4.1.1 are outlined. Visualizations of the globally extracted mean feature trends in the 2.5 km scale prograding and retreating datasets are shown in Section C.3. Section C.4 shows visualizations related to the variability in the quantification space for the identified prograding and retreating drivers. Next to this, also the mean feature trends in the split datasets are shown. The results presented in Section 4.2.2 are supported by figures outlining the LM and GM steps of the GMM in Section C.5. Besides, information on cluster means and the cluster-included number of transects and km of affected shoreline is added in tables. Furthermore, scatter plots indicating cluster means, all extracted features and the signature boundaries are shown. In Section C.6, hard and probable labels regarding summed and single main identified driver distributions are visualized. Moreover, an actual driver map is presented and the zoom-out of the cluster mean visualization with complete influence distributions is shown. Finally, in Section C.7, the analysis-included clusters for Section 4.3 are highlighted in the cluster mean visualization. Also, a map with enlarged locations based on the presence of multiple features in a 50 km radius is presented.

C.1 Pre-processing Shoreline Monitor data

The original transect id labeling system contains 126 out of 330 boxes with shore-normal defined transects, which is considerably less than the originally 200 green-marked boxes. These 126 boxes are stated to incorporate the global available-for-analysis ice-free shoreline consisting of nearly 2.2 million transects. After applying latitude filters (60°N and 50°S), around 1.8 million transects remain available. Here, approximately 36% of the transects is marked 'sandy', whereas 64% is marked 'non-sandy' (including 'undetermined sediment composition'). By using a position-based UTM conversion on straight summed distances between transect intersections with the highest resolution 2016 OSM shorelines, these 1.8 million transects are found to represent a global shoreline length of nearly 720 thousand km.

Approximately 1 million transects, intersecting with one of the 33 generated continuous shorelines using the method described in Section B.1, are initially maintained and unordered as shown in Figure C.1 (indicated by a red line). These transects are found to represent more than 450 thousand km shoreline length, of which around 44% is labeled 'sandy' and 56% 'non-sandy'. Figure C.1 highlights the importance of pre-processing the annual dataset behind the Shoreline Monitor to enable an 'along-the-shoreline'-analysis.

After applying the transect reordering method, which is outlined in Appendix B.1 as well, Figure C.2 is obtained. As can be seen from this figure, all connected transects (indicated by a green line) follow the continuous shoreline. Some remaining large connections are manually excluded, resulting in nearly 985 thousand reduced and ordered transects. The shoreline length represented by these transects amounts to approximately 390 thousand km. The sandy / non-sandy distribution remains unchanged compared to the initially maintained and unordered transects. So, for this research, around 54% of the Shoreline Monitor's available transects as well as the global shoreline length represented by these transects is taken into consideration for an 'along-the-shoreline'-analysis. An overview of the 33 analyzed continuous shorelines, with the computed shoreline lengths and percentage of sandy transects, is presented in Table C.1.

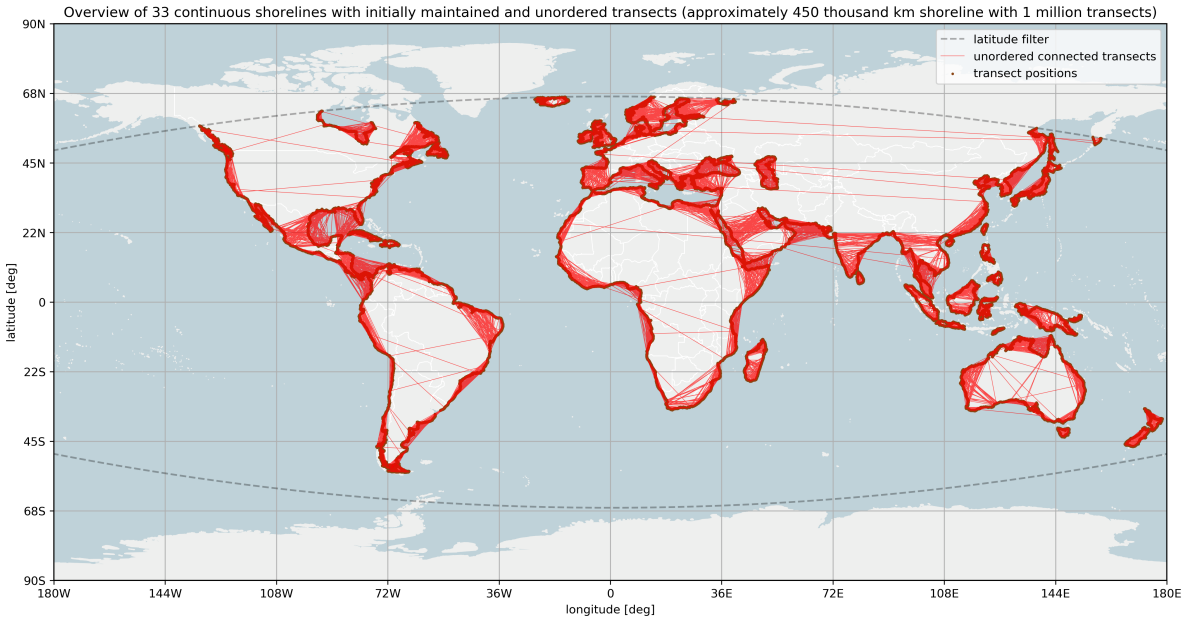


Figure C.1: Global visualization of the considered 33 continuous shoreline transect positions (dots) together with initial unordered connected transects (red lines) on these shorelines. Latitude filters are shown in dashed gray lines.

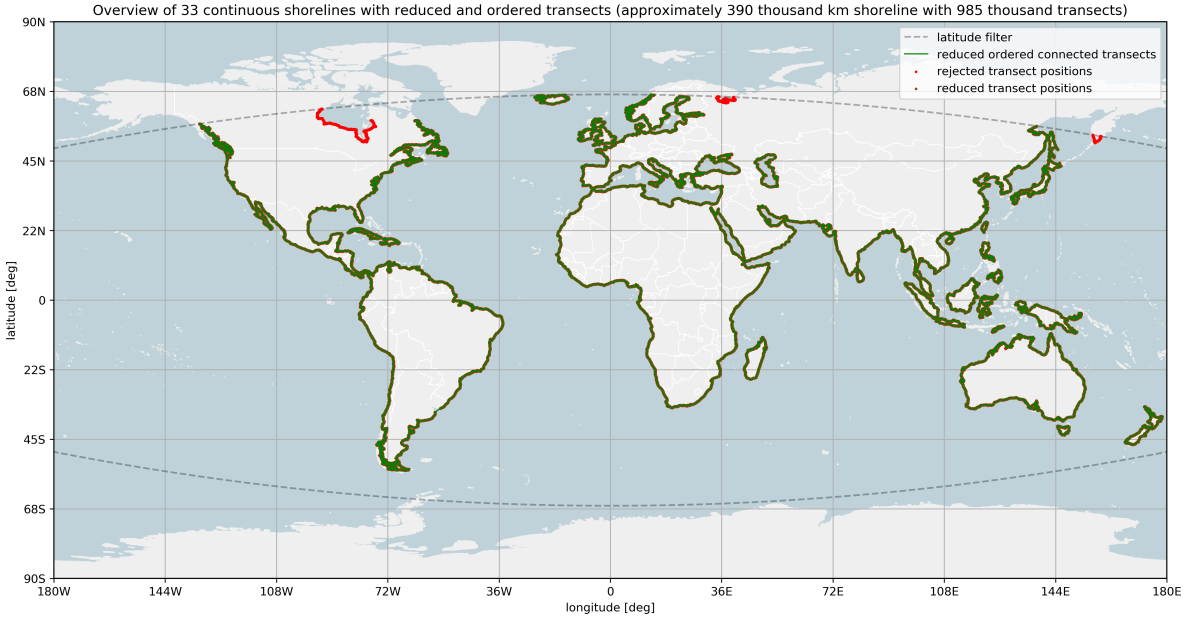


Figure C.2: Global visualization of the considered 33 continuous shoreline transect positions (dots) together with reduced ordered connected transects (green lines). Manually rejected transects (red dots) and latitude filters (dashed gray lines) are also shown.

Table C.1: The research-included 33 continuous shorelines together with various characteristics, ordered from largest to smallest area. The minimum required shoreline surrounded area for research inclusion is > 25.000 km^2. Countries related to considered islands are presented between brackets in the second column. The shoreline length is computed after manual reduction and corrected according to a position-based UTM conversion. The last column gives the percentage of sandy transects compared to the total number of continuous shoreline included transects.

Number	(Combined) Continent or island	ID	Area [km^2]	shoreline length [km]	Sandy transects [%]
1	Afro-Eurasia	AEA	84980532	143215.9	55.3
2	Americas	AME	42549000	102882.0	40.2
3	Australia	AUS	7595342	27476.9	48.0
4	New Guinea (Indonesia and Papua New Guinea)	NGU	785753	11104.6	2.5
5	Borneo (Indonesia, Malaysia and Brunei)	BNO	748168	6669.1	7.6
6	Madagascar	MDG	587041	6163.8	52.9
7	Sumatra (Indonesia)	SUM	443066	5022.1	7.2
8	Kaspian Sea	KSE	371000	5337.8	71.2
9	Honshu (Japan)	HSU	225800	6166.3	44.0
10	Great Britain	GBR	209331	6841.4	34.5
11	Sulawesi (Indonesia)	SUL	180681	5683.5	3.0
12	South Island (New Zealand)	NZS	145836	4014.8	29.2
13	Java (Indonesia)	JAV	138794	2938.4	8.5
14	North Island (New Zealand)	NSN	111583	4701.4	23.1
15	Luzon (Philippines)	LUZ	109965	3703.5	6.5
16	Newfoundland (Canada)	NFL	108860	7385.5	61.6
17	Cuba	CUB	105806	4197.2	14.2
18	Iceland	ICE	101826	4337.6	52.7
19	Mindanao (Philippines)	MIN	97530	3105.3	2.6
20	Ireland	IRE	84421	3433.6	24.8
21	Hokkaido	HOK	78719	2185.0	41.6
22	Hispaniola (Dominican Republic and Haiti)	HIS	73929	2428.5	26.4
23	Sakhalin (Russia)	SAK	72493	3312.5	73.9
24	Sri Lanka	SRL	65268	1458.7	59.5
25	Tasmania (Australia)	TAS	65022	2173.1	32.9
26	Isla Grande de Tierra del Fuego (Argentina and Chile)	TDF	47401	3282.2	54.5
27	Kyushu (Japan)	KYU	37437	2171.8	25.8
28	Taiwan	TAI	35833	1037.4	35.1
29	New Britain (Papua New Guinea)	NBR	35145	1668.4	0.7
30	Hainan (China)	HAI	33210	1059.5	70.0
31	Vancouver Island (Canada)	VAI	31285	2466.7	4.9
32	Timor (Indonesia and East Timor)	TIM	28418	1150.8	16.7
33	Silicy (Italy)	SIL	25662	933.0	68.2

C.2 Local feature extraction and quantification

This section contains the results of the two remaining (one local-scale development and one moderate-scale test) shoreline stretches related to the local feature extraction and quantification method outlined in Section 3.1.1, namely: Belgium and South-West Australia. Next to this, additional results for the case of South-West France are presented. No additional results are shown for the case of the Dutch coast, as all results are elaborated on in Section 4.1.1.

Belgium

The shoreline stretch from Bray-Dunes to Cadzand has a length of around 85 km and contains 262 shore-normal defined transects. So, on average, there is a spacing of around 323 m between two consecutive transects. After applying filers, 252 of the stretch-included transects contain a linear change rate.

The rolling mean method is set out for the spatial scales of 2.5, 5, 10, 20, 50 and 100 km along the flat considered shoreline stretch. This is visualized in Figure C.3, where the location of Bray-Dunes is represented by the left and Cadzand by the right of the plot. The minimum number of required transects in a moving window is kept the same as for the Dutch coast. From the figure below, it is seen that the spacing between two transects (at around 65 km) can be way larger than the average 323 m. If this would result in an extracted and quantified feature, due to a small number of averaged transect-related change rates, a warning is raised.

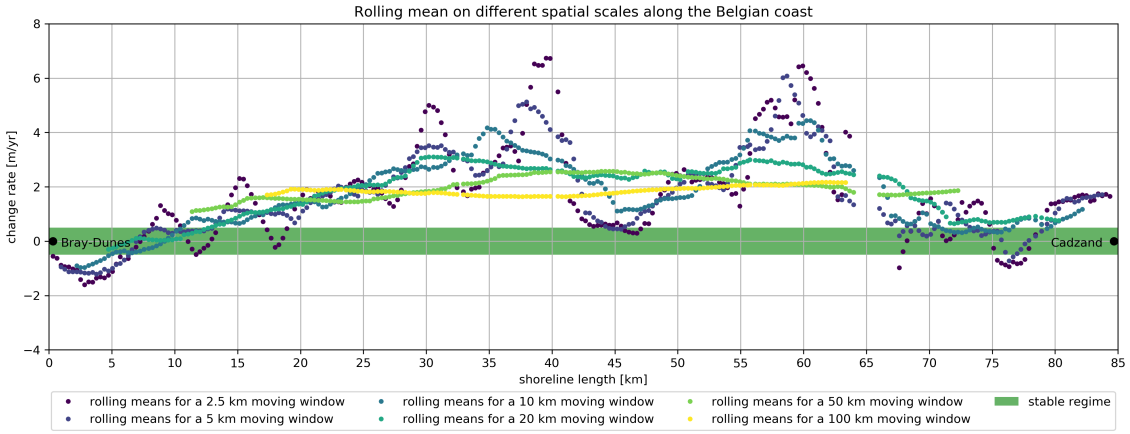


Figure C.3: Visualization of the application of the rolling mean method (pre-work for the actual feature extraction and quantification method) to the Belgian shoreline stretch of Bray-Dunes (left) to Cadzand (right) for different spatial scales. The rolling means of each of the six different spatial scales are indicated with a different color. The stable regime is visualized with a green band.

Statistical results of the application of the rolling mean method on the 252 stretch-included transects with linear change rates are shown in Figure C.4. Figures C.3 and C.4 indicate that the smaller scale rolling means show a great variability around the mean (with a prograding offset), which is decreasing with an increasing spatial scale due to averaging. The shoreline mean of the rolling means is prograding with approximately +1.5 to +2.0 $m/year$, without too much variability between the spatial scales.

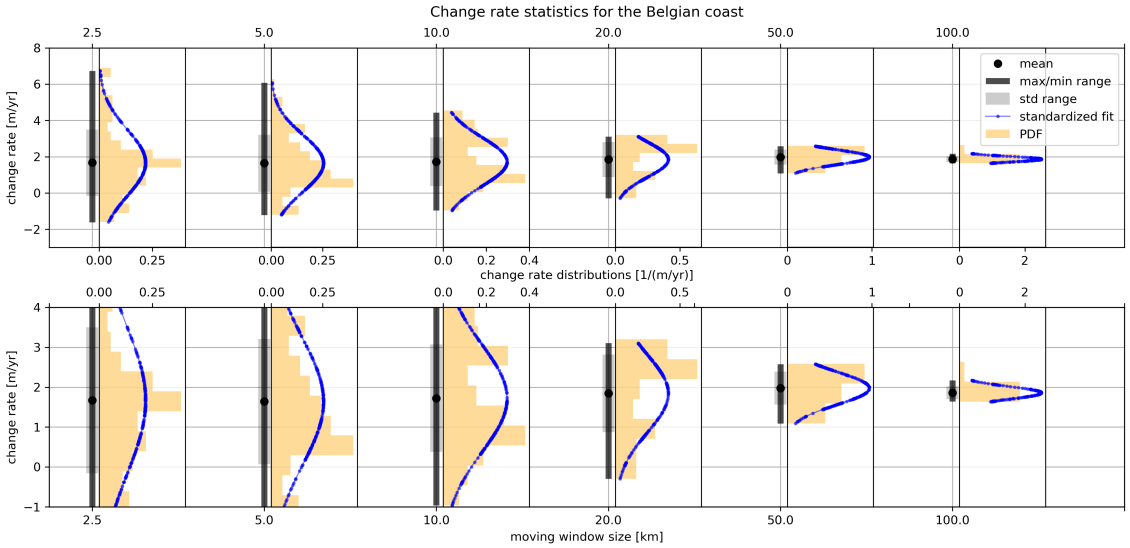


Figure C.4: Two-panel statistical visualization of the application of the rolling mean method to the Belgian shoreline stretch for different spatial scales, represented on the top of the top panel and bottom of the bottom panel x-axes. The lower panel is a zoom-in on the standard deviation range (transparent grey bar) of the upper panel. The mean of the rolling means is indicated with a black dot and the maximum / minimum range with a black bar. The x-axes between the two panels show the distribution of change rate values, in a PDF with orange bars and a standardized fit with a thin blue line with dots, within the shoreline stretch maximum / minimum rolling mean range.

From this point onwards, the focus is on the 2.5 km scale to explain the results related to the feature extraction and quantification method for the case of the Belgian coast. In Figure C.5, the result of the first step of the top-bottom approach in the first layer of the aforementioned method is shown. It can be seen that there are a lot of initially identified tops and bottoms connected by a line called top-bottom fit. Applying the *cf* threshold results in the exclusion of local maxima and minima. The remaining tops and bottoms are represented by a reduced fit that slightly deviates from the top-bottom fit for the considered shoreline stretch.

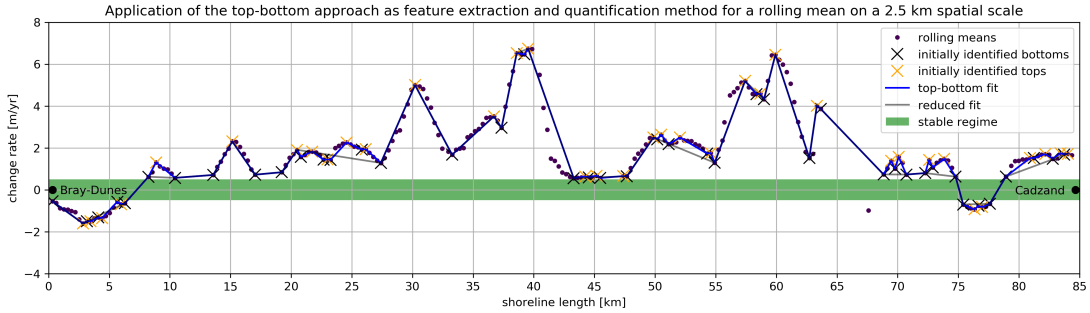


Figure C.5: Visualization of the application of the first step in the top-bottom approach for the Belgian coast on a 2.5 km scale. The rolling means are indicated with dark purple dots and the initially identified tops and bottoms with orange and black crosses respectively. The stable regime is shown with a green band. The top-bottom fit and the reduced fit are shown with a thin blue and black line, the latter slightly deviates from the top-bottom fit due to the *cf* threshold.

In Figure C.6, the result of the final step of the top-bottom approach in the first layer of the feature extraction and quantification method is visualized. Compared to Figure C.5, the number of tops as well as bottoms are reduced again. This time, this results from the application of *tb*, *hd* and *ia* thresholds that remove remaining noise. The feature fit, the line that connects the remaining tops and bottoms, results in the extraction of a total of eight prograding features and one retreating feature (without warnings) from the considered shoreline stretch. A number of smaller 'bumps' in Figure C.6 are not recognized as features, for instance around 10 km. This is a clear disadvantage of the application of various thresholds to come to the feature fit and is elaborated on in Section 4.1.2 for the global application and in Chapter 5.

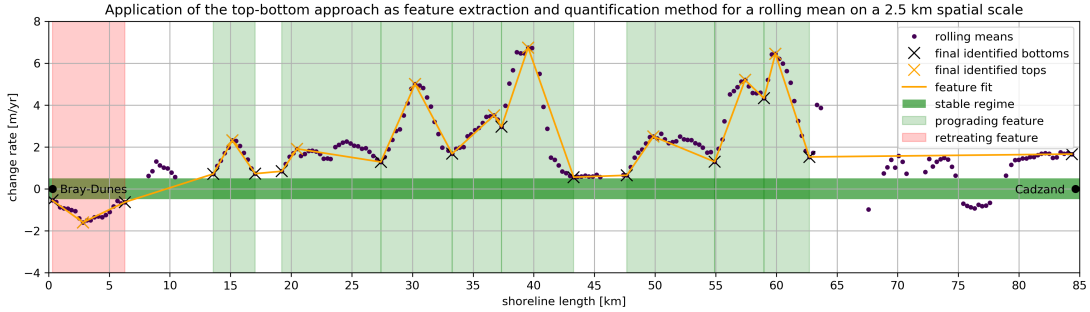


Figure C.6: Visualization of the application of the final step in the top-bottom approach for the Belgian coast on a 2.5 km scale. The rolling means are indicated with dark purple dots and the final identified tops and bottoms (after applying the *tb*, *hd* and *ia* thresholds) with orange and black crosses respectively. The stable regime is shown with a green band and the feature fit with a thin orange line. The identified prograding (retreating) features are shown with transparent vertical green (red) bands, indicating the total feature range.

By visually inspecting Figures C.5 and C.6, it is stated that the top-bottom approach with its thresholds performs accordingly. The top-bottom, reduced and feature fit are seen to represent the alongshore rolling means quite well. This is supported by the absence of significant differences between fit performance indicators RMSE and MAE, which are presented in Table C.2. The errors become bigger from top-bottom to feature fit, which is caused by the application of thresholds. By comparing the table below with Table C.2 for the Dutch coast, it is seen that the top-bottom and reduced fit perform better for the Belgian coast. Nonetheless, the feature fit shows a larger RMSE and MAE for the Belgian coast, which is supported by the exclusion of more smaller 'bumps' and scattered rolling means.

Table C.2: Performance indicators RMSE and MAE for the top-bottom, reduced and feature fit. These are related to the feature extraction and quantification method development shoreline stretch of the Belgian coast.

Type of fit	RMSE $[m/year]$	MAE $[m/year]$
Top-bottom	0.38	0.18
Reduced	0.46	0.29
Feature	0.72	0.42

The total of nine extracted features along the Belgian coast are spatially quantified on the four characteristics presented in Section 3.1.1. These feature quantifications are shown in Table C.3, where the prograding features are numbered one to eight from left to right in Figure C.6.

Table C.3: The four spatial quantifications of the extracted prograding (upper panel) and retreating feature(s) (lower panel) that result from the development shoreline stretch of the Belgian coast. The green-highlighted prograding feature row is considered for the temporal quantification.

Feature	Width $[km]$	Smallest skewness $[m/(year*km)]$	Largest skewness $[m/(year*km)]$	Magnitude range $[m/year]$
1	3.5	0.8	1	1.6
2	8.2	0.1	0.9	1.0
3	5.9	1.1	1.3	3.5
4	4.1	0.5	0.9	0.7
5	5.9	1.6	1.7	4.7
6	7.2	0.2	0.8	1.7
7	4.1	0.6	1.6	2.0
8	3.7	1.8	2.3	2.8
1	6.0	-0.3	-0.4	-1.0

The rolling means are replaced by actual feature-included shoreline evolution transect signals to quantify temporal variability in the extracted features. Figure C.7 shows an example of the outcome of the aforementioned for the first green feature (around 15 km) in Figure C.6. Eleven transects are included in this feature. One transect has a range from 1984 to 2016, whereas others range from 1985 to 2016. It also seen that all transects miss shoreline positions from 1993 to 1996 and indicate approximately the same temporal behavior from 2002 onwards. Before 2002, there is a large irregularity in some transect trends. For instance, BOX_187_044_047 shows highly deviating peaks compared to the others.

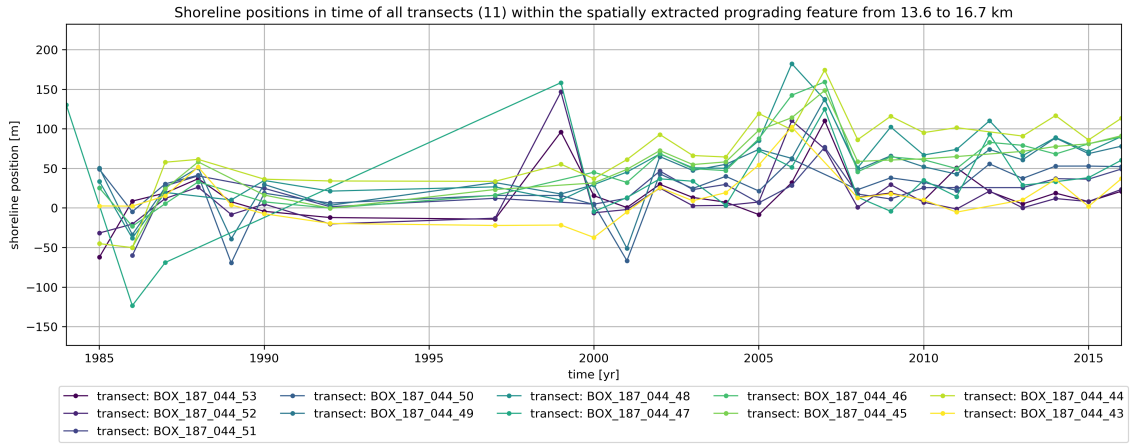


Figure C.7: Example visualization for the second layer in the feature extraction method on a 2.5 km scale. This example considers the first prograding feature from the left (13.6 to 16.7 km in Figure C.6) on the Belgian coast. Eleven transects are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. In case of the absence of an annual shoreline position for a temporal trend, no dot is visualized. The temporal trends are identified by their label, presented in the legend.

The eight temporal quantifications presented in Table C.4 result from the total of four dataset modifications. The interpolated shoreline positions that fill the gaps in Figure C.7 are only computed to derive the aforementioned. The four dataset modifications are visualized in Figure C.8 for the same prograding feature as in the figure above. The considered feature from Figure C.8 shows a maximum annual shoreline position change of 51.0 m between two consecutive years and has a change rate of around 1.8 $m/year$. The change rate std is 0.55 $m/year$, which indicates that the linear change rate is on the edge of being called accurate or inaccurate (see Luijendijk et al., 2018). After applying dataset modifications, it is also seen that all temporal trends are shifted to the retreating regime. This is caused by the single transect with a shoreline position in the year 1984. As shoreline positions are not extrapolated, this single transect has a big influence on the temporal quantifications in this feature.

Table C.4: The eight temporal quantifications of the extracted first prograding feature that results from the development shoreline stretch of the Belgian coast. The labels of the temporal quantifications are split in two sets of four, represented by the first and third row.

Feature	Maximum annual value change [m]	Minimum annual value change [m]	Maximum annual std change [m]	Minimum annual std change [m]
1	51.0	-127.4	38.8	-33.1
Feature	Maximum annual best fit change [m]	Minimum annual best fit change [m]	Change rate [$m/year$]	Change rate std [$m/year$]
1	9.6	-80.3	1.8	0.6

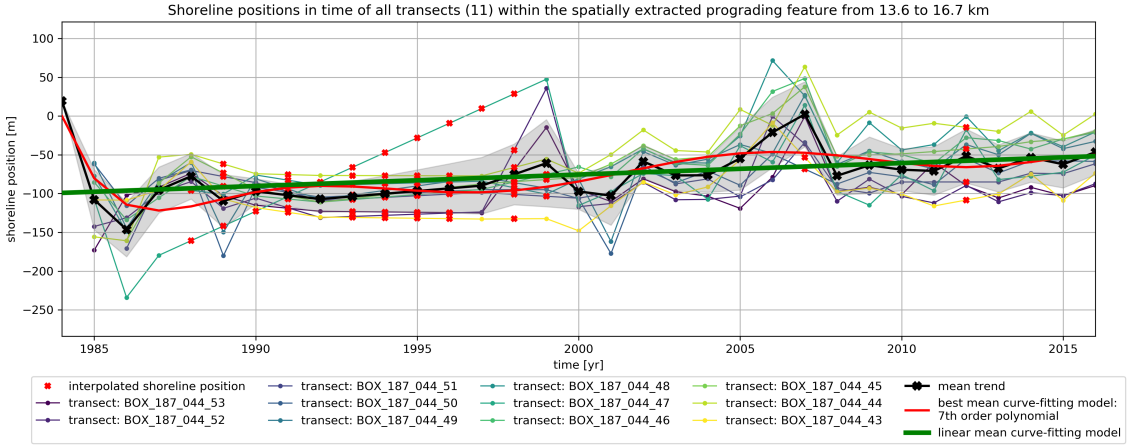


Figure C.8: Example visualization for the quantification of temporal variability in the second layer of feature extraction method on a 2.5 km scale. This example considers the first prograding feature from the left (13.6 to 16.7 km in Figure C.6) on the Belgian coast. Eleven transects (see legend label id's) are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. In case of the absence of an annual shoreline position for a temporal trend, a red interpolated dot is shown. The mean (feature) trend, obtained by summing all shoreline positions per year, is indicated with a black line with crosses. The linear fit and best fit to the mean trend are shown in thick green and red lines. The gray band shows the standard deviation range around the mean trend.

South-West Australia

The shoreline stretch from Wilyabrup to Tamala has a length of around 924 km and contains 2112 shore-normal defined transects. So, on average, there is a spacing of around 437 m between two consecutive transects. This is different from the case of the French coast, which can be explained by having less frequent accidental large spacings between transects. After applying six filters, 1822 of the stretch-included transects contain a linear change rate (290 transects are set to NaN).

Once more, the rolling mean method is set out for the spatial scales of 2.5, 5, 10, 20, 50 and 100 km along the flat considered shoreline stretch. This is visualized in Figure C.9, where the location of Wilyabrup is represented by the left and Tamala by the right of the plot. The minimum number of required transects in a moving window is again kept the same as for the Dutch coast.

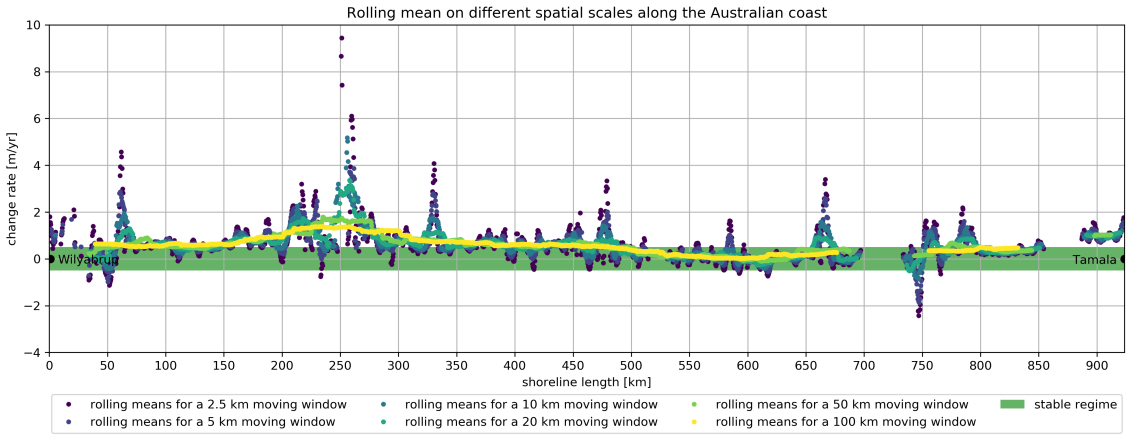


Figure C.9: Visualization of the application of the rolling mean method (pre-work for the actual feature extraction and quantification method) to the Australian shoreline stretch of Wilyabrup (left) to Tamala (right) for different spatial scales. The rolling means of each of the six different spatial scales are indicated with a different color. The stable regime is visualized with a green band.

Statistical results of the application of the rolling mean method on the 1822 stretch-included transects with linear change rates are shown in Figure C.10. As can be seen from Figures C.9 and C.10, the smaller scale rolling means show a great variability around the mean with peaks in the prograding as well as retreating regime. These peaks have an offset towards the prograding regime. The variability around the mean is decreasing with an increasing spatial scale and the mean of the rolling means are indicating a shoreline that is prograding with approximately +0.5 $m/year$, without too much variability between the spatial scales.

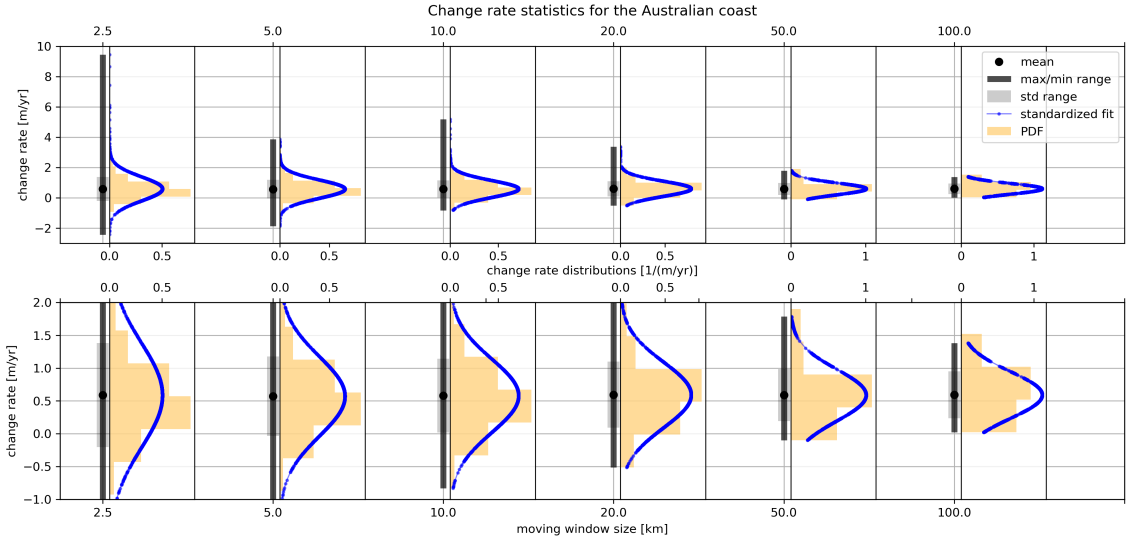


Figure C.10: Two-panel statistical visualization of the application of the rolling mean method to the Australian shoreline stretch for different spatial scales. These scales are represented on the top of the top panel and bottom of the bottom panel x-axes. The lower panel is a zoom-in on the standard deviation range (transparent grey bar) of the upper panel. The mean of the rolling means is indicated with a black dot and the maximum / minimum range with a black bar. The x-axes between the two panels show the distribution of change rate values, in a PDF with orange bars and a standardized fit with a thin blue line with dots, within the shoreline stretch maximum / minimum rolling mean range.

From this point onwards, again, the focus is on the 2.5 km scale to explain the results related to the feature extraction and quantification method for the case of the Australian coast. In Figure C.11, the result of the first step of the top-bottom approach is shown. The exclusion of local maxima and minima, by application of the cf threshold, results in a reduced fit that only slightly deviates from the top-bottom fit for this shoreline stretch.

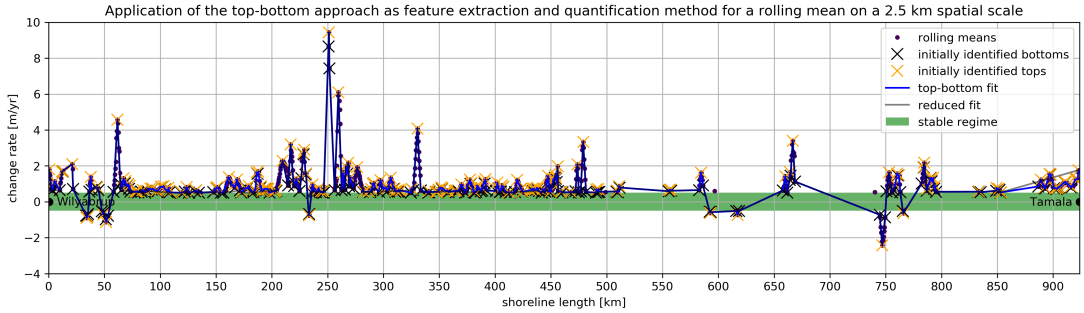


Figure C.11: Visualization of the application of the first step in the top-bottom approach for the Australian coast on a 2.5 km scale. The rolling means are indicated with dark purple dots and the initially identified tops and bottoms with orange and black crosses respectively. The stable regime is shown with a green band. The top-bottom fit and the reduced fit are shown with a thin blue and black line, the latter slightly deviates from the top-bottom fit due to the *cf* threshold.

In Figure C.12, the result of the final step of the top-bottom approach is visualized. Compared to Figure C.11, the number of tops as well as bottoms are reduced once more by the application of *tb, hd* and *ia* thresholds. The feature fit results in the extraction of a total of fifteen prograding features and one retreating feature with one warning for the fifth prograding feature from the left. When analyzing Figure C.12, it is seen that smaller 'bumps' are not recognized as features. This is elaborated on in Section 4.1.2 for the global application and in Chapter 5.

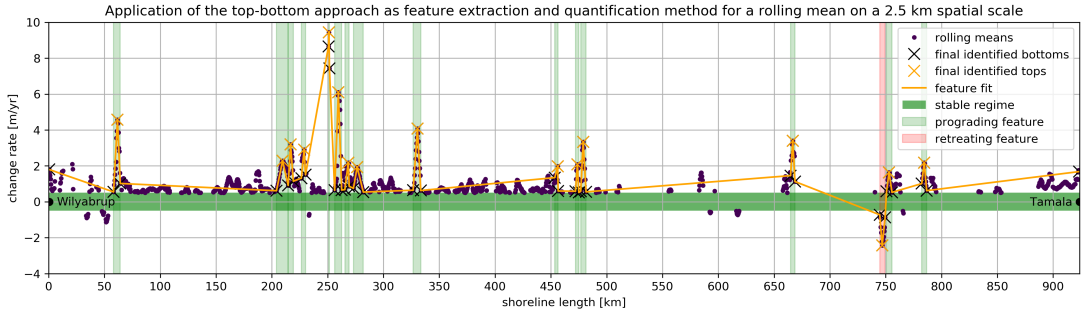


Figure C.12: Visualization of the application of the final step in the top-bottom approach for the Australian coast on a 2.5 km scale. The rolling means are indicated with dark purple dots and the final identified tops and bottoms (after applying the *tb, hd* and *ia* thresholds) with orange and black crosses respectively. The stable regime is shown with a green band and the feature fit with a thin orange line. The identified prograding (retreating) features are shown with transparent vertical green (red) bands, indicating the total feature range.

Figures C.11 and C.12 imply that the top-bottom approach with its thresholds captures most of the features / peaks with rolling mean shoreline evolution signals. Where highly scattered rolling means are represented by the top-bottom and reduced fit, the feature fit does not take these points into account. This is supported by the relatively large difference between the reduced and feature fit performance indicators RMSE and MAE, which are presented in Table C.5. Nonetheless, following the same reasoning as in the case of the French coast, the thresholds are still stated to be sufficient.

Table C.5: Performance indicators RMSE and MAE for the top-bottom, reduced and feature fit. These are related to the feature extraction and quantification method test shoreline stretch of the Australian coast.

Type of fit	RMSE $[m/year]$	MAE $[m/year]$
Top-bottom	0.19	0.07
Reduced	0.35	0.23
Feature	0.9	0.44

The total of sixteen extracted features along the Australian coast are spatially quantified on the four characteristics presented in Section 3.1.1. These feature quantifications are shown in Table C.6, where the prograding features are numbered one to fifteen from left to right in Figure C.12.

Table C.6: The four spatial quantifications of the extracted prograding (upper panel) and retreating feature(s) (lower panel) that result from the test shoreline stretch of the Australian coast. The red-highlighted retreating feature row is considered for the temporal quantification.

Feature	Width [km]	Smallest skewness [$m/(year*km)$]	Largest skewness [$m/(year*km)$]	Magnitude range [$m/year$]
1	6.0	1.1	1.4	3.7
2	10.5	0.3	0.3	1.5
3	4.6	0.8	1.1	2.2
4	3.8	0.6	1.1	1.5
5	0.8	1.9	4.8	1.4
6	6.3	1.6	1.8	5.5
7	3.9	0.5	1.7	1.4
8	8.9	0.3	0.3	1.3
9	6.8	0.9	1.2	3.5
10	3.0	0.2	3.2	1.3
11	3.0	0.9	1.2	1.5
12	4.7	1.1	1.3	2.8
13	3.9	1.0	1.1	2.1
14	4.8	0.4	0.5	1.1
15	4.4	0.5	0.7	1.4
1	4.8	-0.6	-0.8	-1.6

The rolling means are replaced by actual feature-included shoreline evolution transect signals to quantify temporal variability in the extracted features. Figure C.13 shows an example of the outcome of the aforementioned for the only red feature (around 750 km) in Figure C.12. Ten transects are included in this feature. All transects cover the range from 1986 to 2016 and show approximately the same temporal behavior with a variability in the angle of the trends.

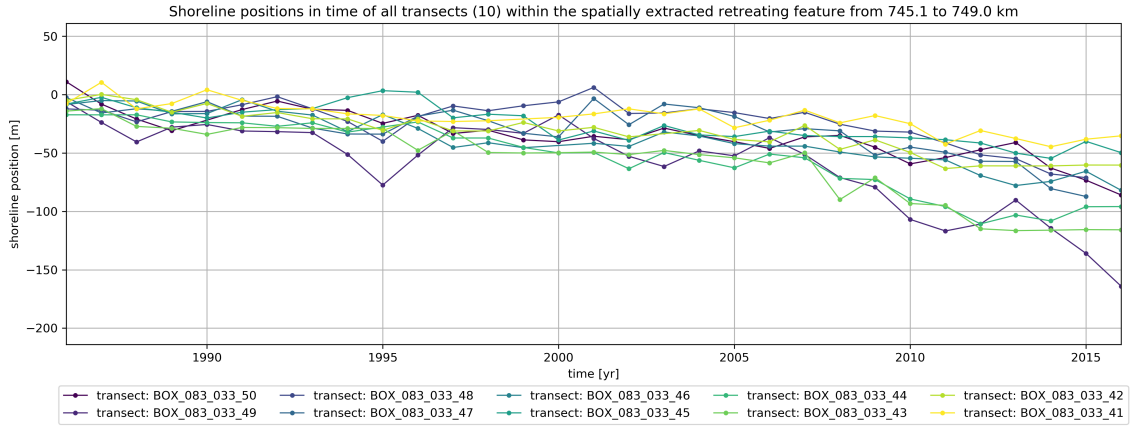


Figure C.13: Example visualization for the second layer in the feature extraction method on a 2.5 km scale. This example considers the only retreating feature (745.1 to 749 km in Figure C.12) on the Australian coast. Ten transects are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. In case of the absence of an annual shoreline position for a temporal trend, no dot is visualized. The temporal trends are identified by their label, presented in the legend.

The eight temporal quantifications presented in Table C.7 result from the total of four dataset modifications. The interpolated shoreline positions that fill the gaps in Figure C.13 are only computed to derive the aforementioned. Note that for this feature, the are only very few missing shoreline positions. The four dataset modifications are visualized in Figure C.14 for the same retreating feature as in the figure above. The considered feature from Figure C.14 shows a maximum annual shoreline position change of -12.6 m between two consecutive years and has a change rate of around -2.2 $m/year$. The change rate std is 0.1 $m/year$, which indicates that the linear change rate is quite accurate for this feature.

Table C.7: The eight temporal quantifications of the single extracted retreating feature that results from the test shoreline stretch of the Australian coast. The labels of the temporal quantifications are split in two sets of four, represented by the first and third row.

Feature	Maximum annual value change $[m]$	Minimum annual value change $[m]$	Maximum annual std change $[m]$	Minimum annual std change $[m]$
1	6.1	-12.6	8.6	-4.9
Feature	Maximum annual best fit change $[m]$	Minimum annual best fit change $[m]$	Change rate $[m/year]$	Change rate std $[m/year]$
1	-0.2	-7.6	-2.2	0.1

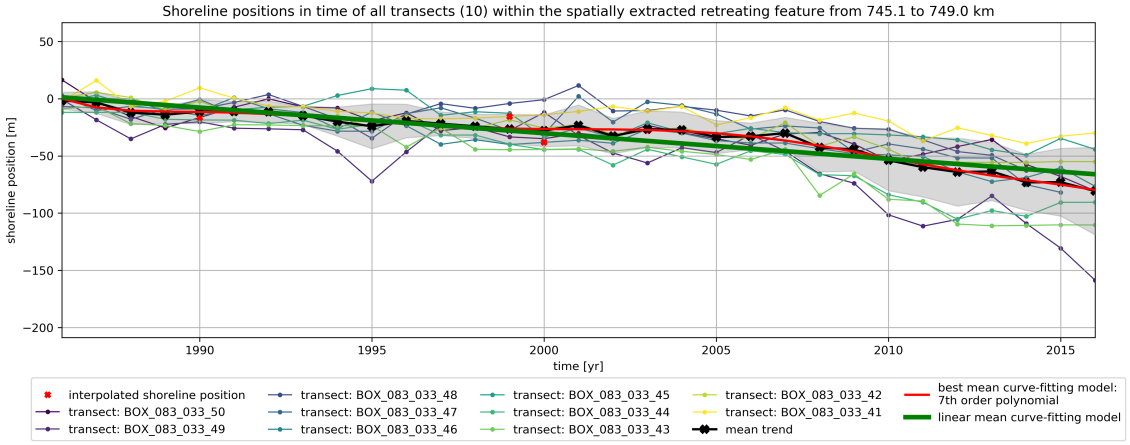


Figure C.14: Example visualization for the quantification of temporal variability in the second layer of feature extraction method on a 2.5 km scale. This example considers the only retreating feature (745.1 to 749 km in Figure C.12) on the Australian coast. Ten transects (see legend label id's) are included in this feature and the temporal trends of these transects are shown in different colored lines with dots. In case of the absence of an annual shoreline position for a temporal trend, a red interpolated dot is shown. The mean (feature) trend, obtained by summing all shoreline positions per year, is indicated with a black line with crosses. The linear fit and best fit to the mean trend are shown in thick green and red lines. The gray band shows the standard deviation range around the mean trend.

South-West France

In Table C.8 on the next page, four spatial quantifications for thirteen prograding and seventeen retreating features are presented. This table is elaborated on in Section 4.1.1 and is the only result that is moved to the appendices.

Table C.8: The four spatial quantifications of the extracted prograding (upper panel) and retreating features (lower panel) that result from the test shoreline stretch of the French coast. The red-highlighted retreating feature row is considered for the temporal quantification.

Feature	Width [km]	Smallest skewness [$m/(year*km)$]	Largest skewness [$m/(year*km)$]	Magnitude range [$m/year$]
1	9.4	0.3	0.4	1.6
2	19.6	0.1	0.5	1.4
3	3.2	3.0	5.4	5.9
4	3.9	0.5	1.7	1.3
5	4.1	0.4	1.4	1.4
6	2.1	1.9	4.2	2.9
7	1.4	1.3	2.5	1.3
8	2.1	3.8	4.1	2.3
9	4.1	1.1	1.4	2.6
10	3.7	1.5	2.1	3.4
11	3.4	1.8	2.3	2.9
12	1.4	1.3	3.4	1.2
13	2.0	0.6	1.2	0.9
1	3.9	-3.0	-6.7	-7.5
2	8.5	-0.1	-0.8	-1.1
3	8.4	-0.7	-1.3	-3.7
4	3.2	-1.5	-3.3	-3.4
5	6.5	-0.4	-0.7	-1.7
6	3.9	-0.8	-1.2	-1.9
7	4.2	-3.2	-4.7	-8.1
8	4.2	-1.4	-1.6	-3.1
9	1.7	-2.4	-3.0	-2.3
10	6.3	-0.6	-0.6	-1.5
11	2.4	-3.2	-3.9	-4.2
12	7.6	-1.0	-2.0	-4.6
13	3.8	-0.4	-0.7	-1.0
14	1.7	-1.6	-1.6	-1.3
15	4.4	-2.5	-2.7	-5.8
16	3.5	-0.9	-0.9	-1.5
17	5.8	-0.3	-0.6	-1.2

C.3 Global feature extraction and quantification

In Figures C.15 and C.16 below, the mean feature trends for the prograding and retreating datasets are shown. The prograding trends indicate that the maximum shoreline evolution over 33 years is approximately 2000 m, whereas the retreating trends have a maximum shoreline evolution reaching nearly -1500 m. Besides, very large annual steps in shoreline positions (sometimes up to 1000 $m/year$) are seen at the edges of the prograding and retreating trend clouds. One might expect that the prograding dataset contains only trends in the prograding regime and the retreating dataset only trends in the retreating regime. However, both figures show trends in both regimes. This is explained by temporal dataset modifications outlined in Section 3.1.1. The above-presented findings are also discussed in Section 4.1.2.

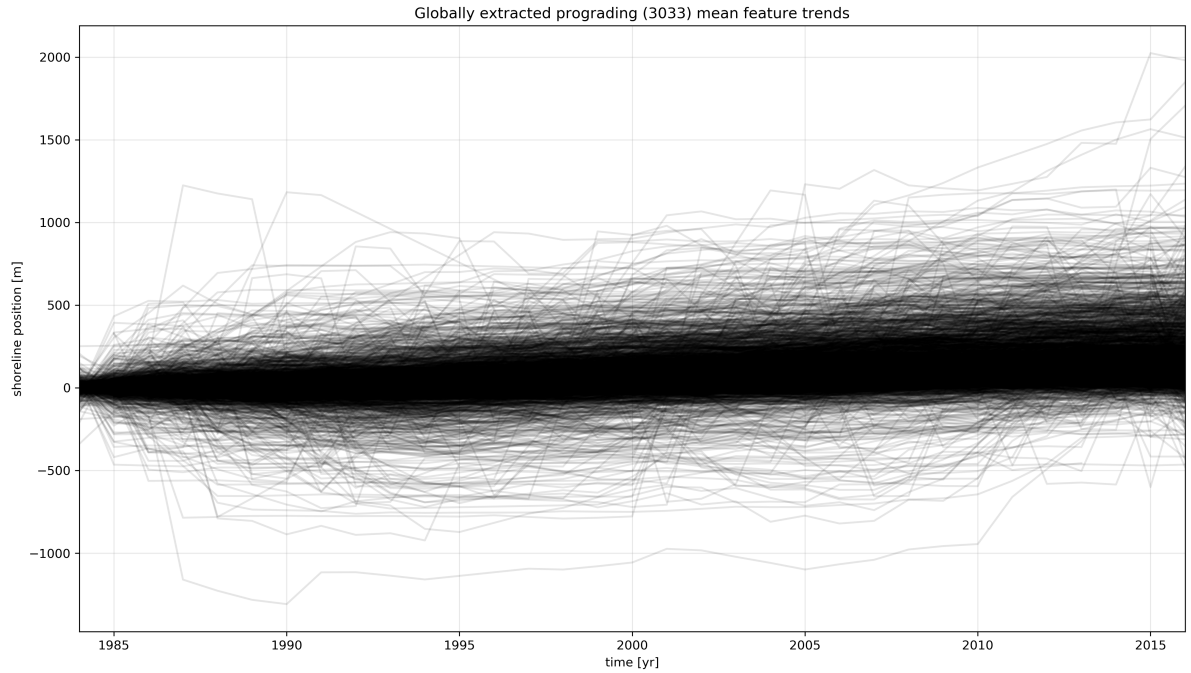


Figure C.15: Visualization of 3033 prograding mean feature trends (grey lines) resulting from globally extracted and quantified features.

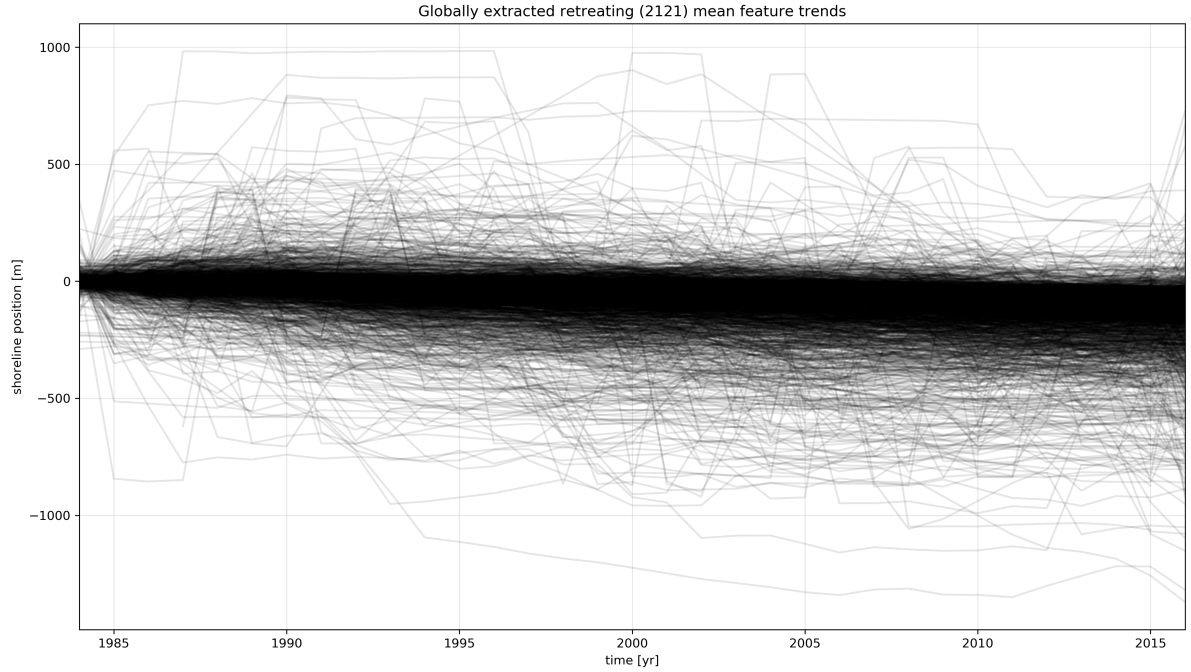


Figure C.16: Visualization of 2121 retreating mean feature trends (grey lines) resulting from globally extracted and quantified features.

C.4 Local feature-driver labeling

This section contains visualizations of the outcomes of the local feature-driver labeling method from Section 3.2.1, which are elaborated on in Section 4.2.1. Both Figures C.17 and C.20 show the identified driver variability in the quantification space, where the black lines in the fifth and sixth quantification indicate the dataset splits. Figures C.18 and C.19 and Figures C.21 and C.22 indicate the total of mean feature trends with the identified driver mean trends for the two split prograding and two split retreating datasets respectively. Both Figures C.19 and C.22 contain very large irregular trends that are primarily due to unnatural drivers.

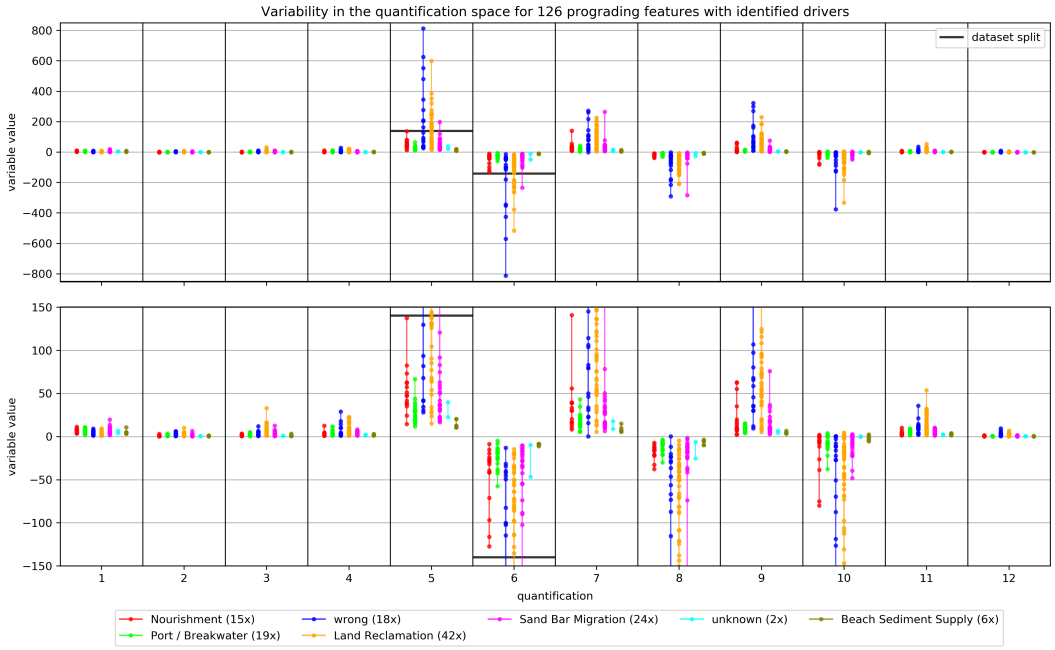


Figure C.17: Two-panel visualization of the variability in twelve quantifications (see Section 3.1.1), per driver for a total of 126 prograding shoreline evolution features. The lower panel is a zoom-in on the -150 + 150 range of the upper panel. Each of the drivers (with the summed number of labels between brackets) has a unique color that comes back in every quantification of the two panels. The black lines in quantification five and six indicate the value of 140 and -140 respectively, which are used to split the prograding dataset into two.

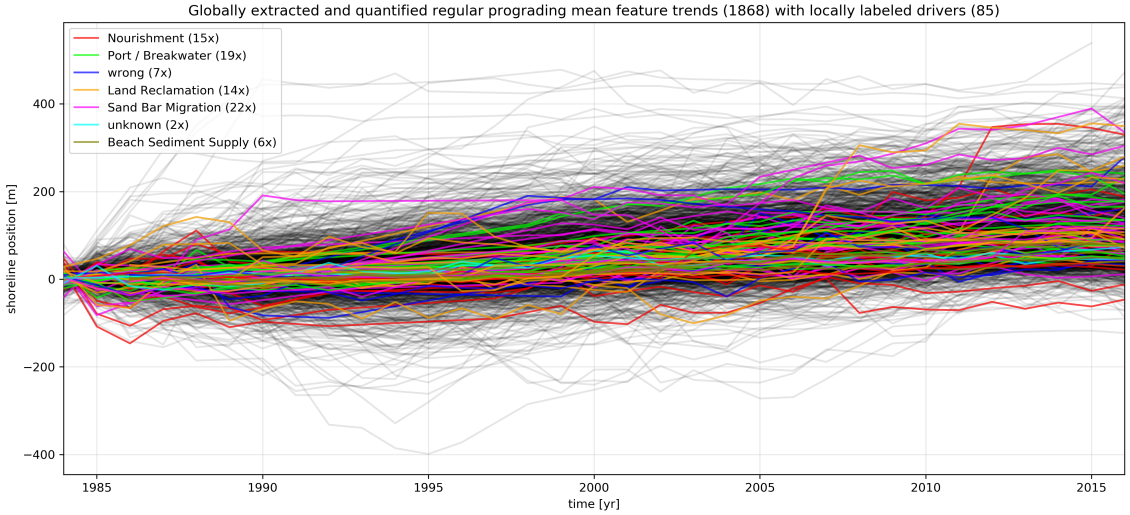


Figure C.18: Visualization of the globally split 1868 regular prograding mean feature trends (grey lines). Drivers of shoreline evolution are identified locally for 85 trends, highlighted by unique colors. The number of times a driver is detected is shown in the legend between brackets.

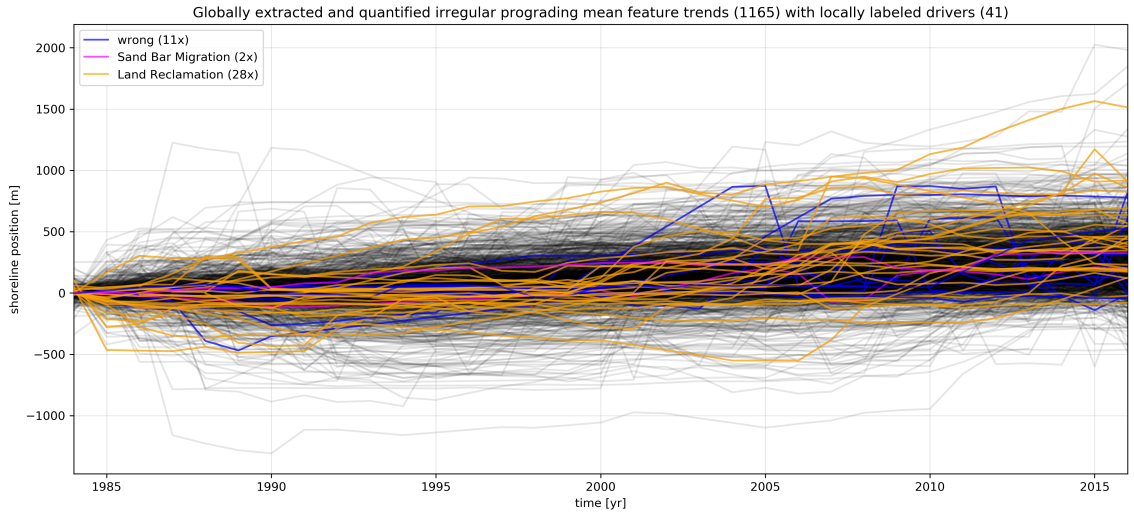


Figure C.19: Visualization of the globally split 1165 irregular prograding mean feature trends (grey lines). Drivers of shoreline evolution are identified locally for 41 trends, highlighted by unique colors. The number of times a driver is detected is shown in the legend between brackets.

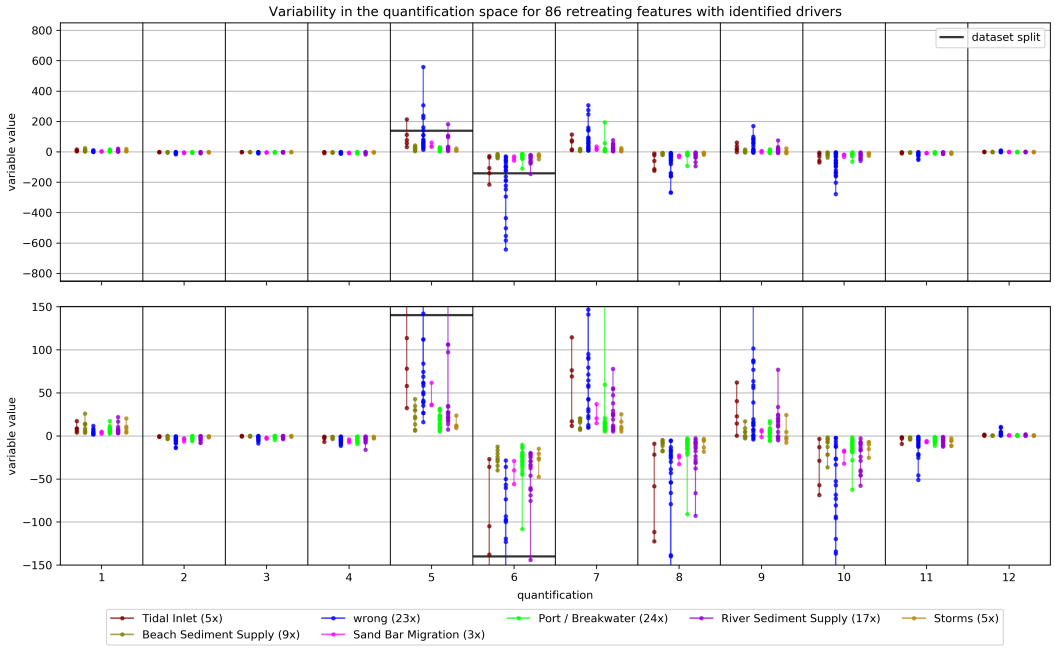


Figure C.20: Two-panel visualization of the variability in twelve quantifications (see Section 3.1.1), per driver for a total of 86 retreating shoreline evolution features. The lower panel is a zoom-in on the -150 + 150 range of the upper panel. Each of the drivers (with the summed number of labels between brackets) has a unique color that comes back in every quantification of the two panels. The black lines in quantification five and six indicate the value of 140 and -140 respectively, which are used to split the retreating dataset into two.

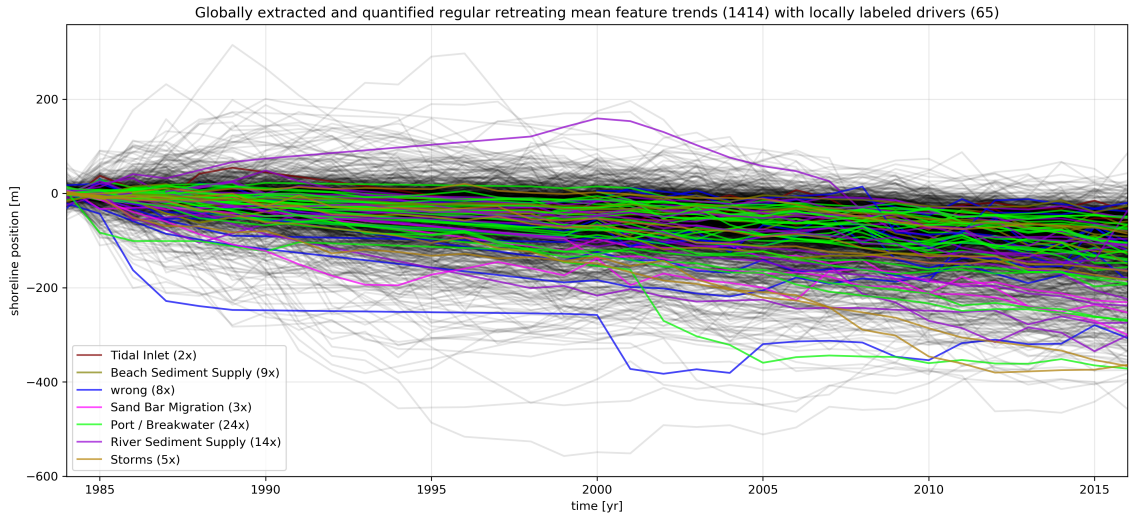


Figure C.21: Visualization of the globally split 1414 regular retreating mean feature trends (grey lines). Drivers of shoreline evolution are identified locally for 65 trends, highlighted by unique colors. The number of times a driver is detected is shown in the legend between brackets.

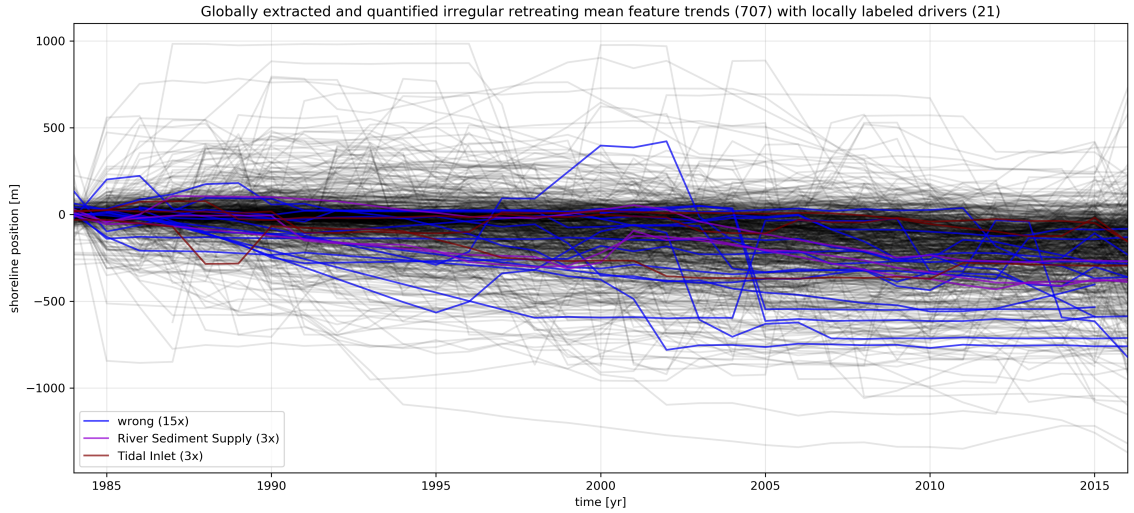


Figure C.22: Visualization of the globally split 707 irregular retreating mean feature trends (grey lines). Drivers of shoreline evolution are identified locally for 21 trends, highlighted by unique colors. The number of times a driver is detected is shown in the legend between brackets.

C.5 Global feature clustering

All eight figures below are related to the LM (first four two panel visualizations) and the GM (next four single figures) steps in the set-up of the GMMs for the four split datasets. These figures were used to derive Table 4.9 in Section 4.2.2. As seen in the LM step figures, both the model type (spherical, tied, diagonal or full configurations) and the number of clusters are variable. The wiggles in the lines are due to single random initializations for each of the considered number of clusters. For the GM step, only the full covariance matrix is considered.

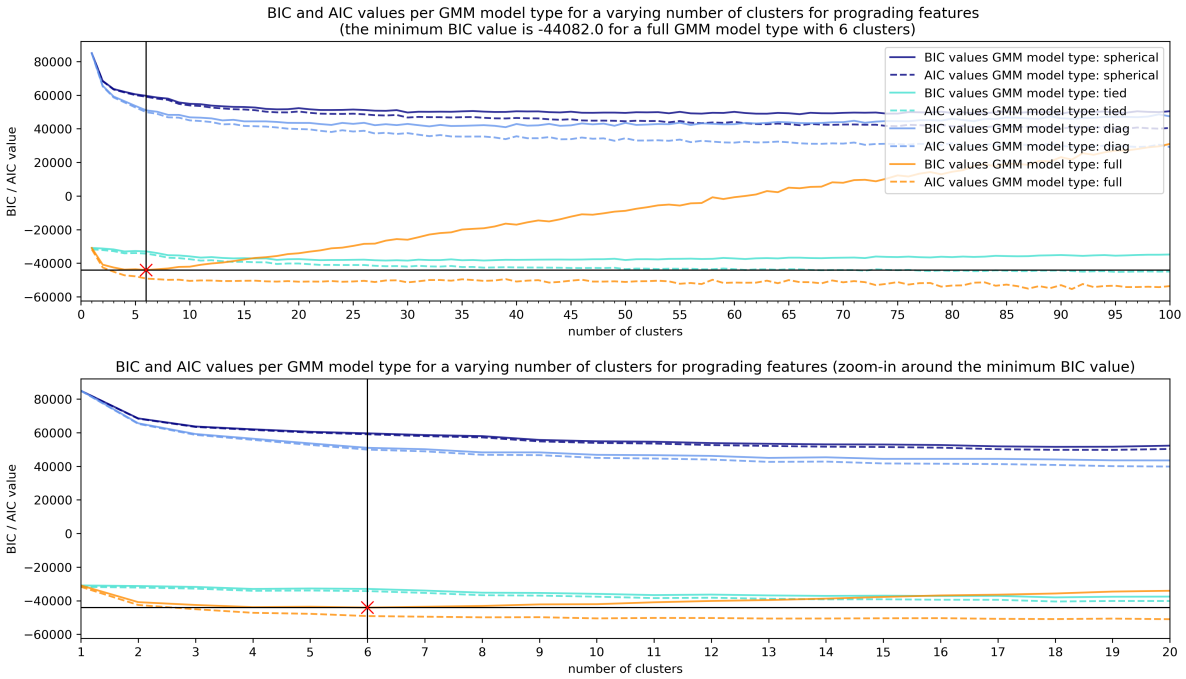


Figure C.23: Visualization of the local model computed information-theoretic criteria Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) with a varying number of clusters and Gaussian Mixture Model (GMM) (covariance) type for the regular dataset with 1868 extracted prograding features. The upper panel considers a large range of cluster variability, whereas the lower panel is a zoom-in on the minimum BIC value (indicated with black lines and a red cross). BIC (AIC) values are shown in solid (dashed) lines, colors indicate the covariance type.

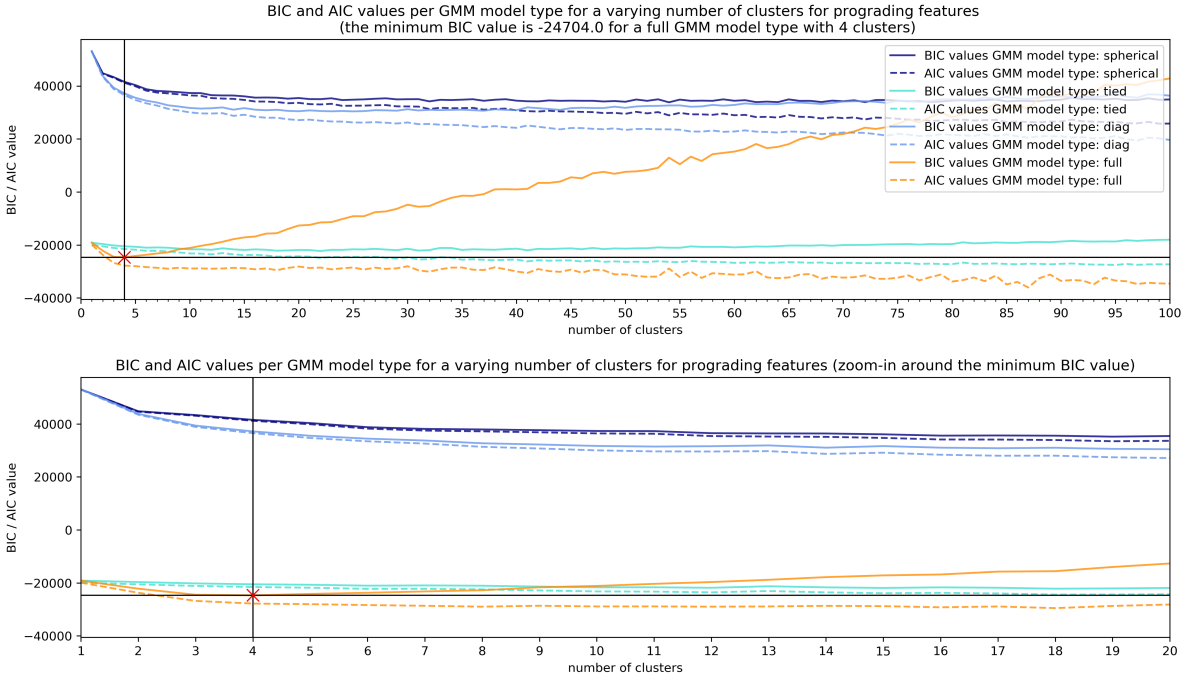


Figure C.24: Visualization of the local model computed information-theoretic criteria AIC (dashed lines) and BIC (solid lines) with a varying number of clusters and GMM (covariance) type for the irregular dataset with 1165 extracted prograding features. The upper panel considers a large range of cluster variability, whereas the lower panel is a zoom-in on the minimum BIC value (indicated with black lines and a red cross). Colors indicate the covariance type.

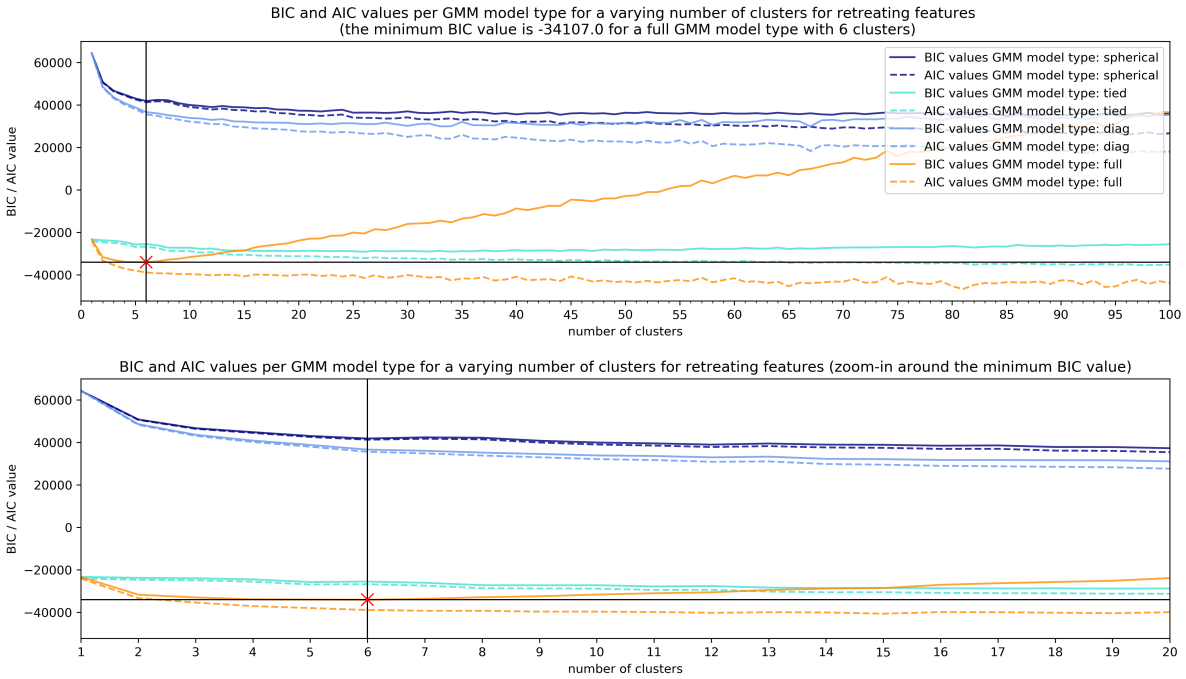


Figure C.25: Visualization of the local model computed information-theoretic criteria AIC (dashed lines) and BIC (solid lines) with a varying number of clusters and GMM (covariance) type for the regular dataset with 1414 extracted retreating features. The upper panel considers a large range of cluster variability, whereas the lower panel is a zoom-in on the minimum BIC value (indicated with black lines and a red cross). Colors indicate the covariance type.

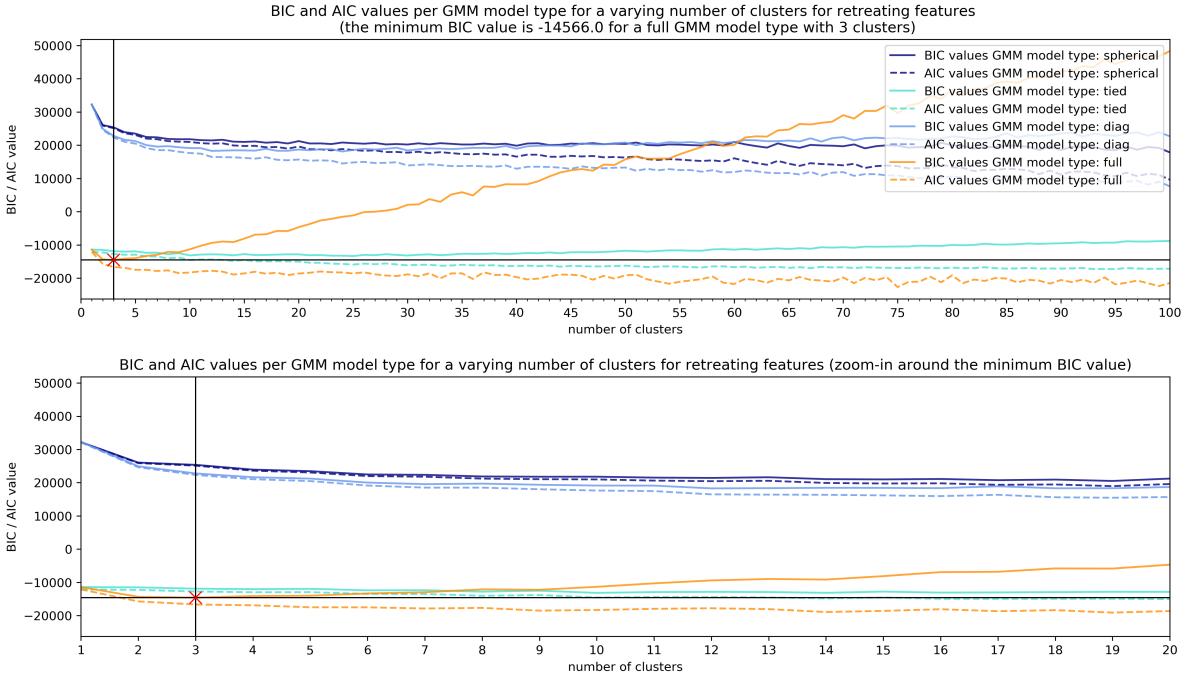


Figure C.26: Visualization of the local model computed information-theoretic criteria AIC (dashed lines) and BIC (solid lines) with a varying number of clusters and GMM (covariance) type for the irregular dataset with 707 extracted retreating features. The upper panel considers a large range of cluster variability, whereas the lower panel is a zoom-in on the minimum BIC value (indicated with black lines and a red cross). Colors indicate the covariance type.

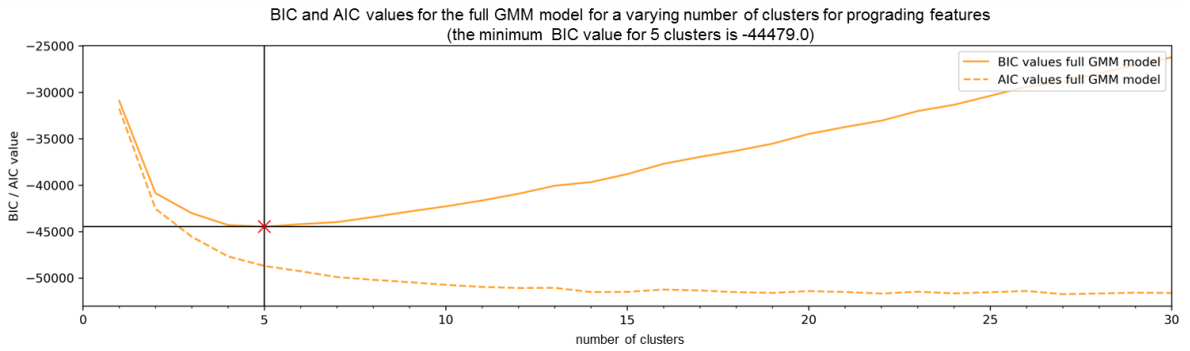


Figure C.27: Visualization of the global model (full GMM) computed information-theoretic criteria AIC (dashed lines) and BIC (solid lines) with a varying number of clusters for the regular dataset with 1868 extracted prograding features.

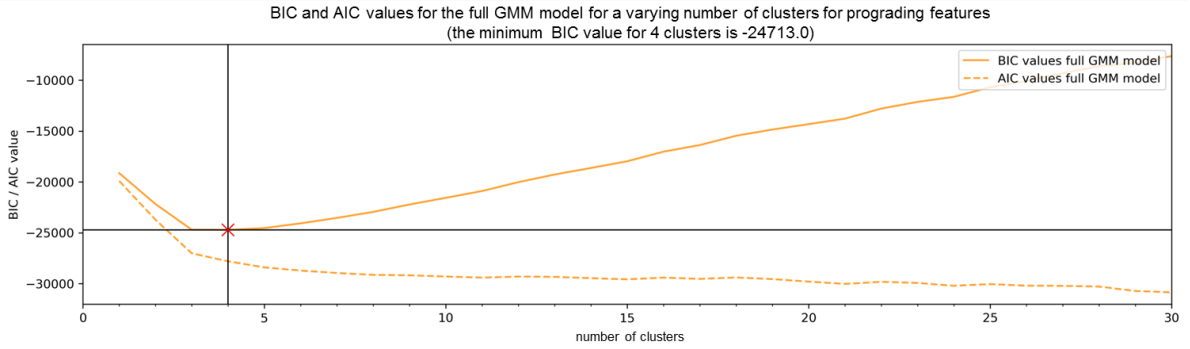


Figure C.28: Visualization of the global model (full GMM) computed information-theoretic criteria AIC (dashed lines) and BIC (solid lines) with a varying number of clusters for the irregular dataset with 1165 extracted prograding features.

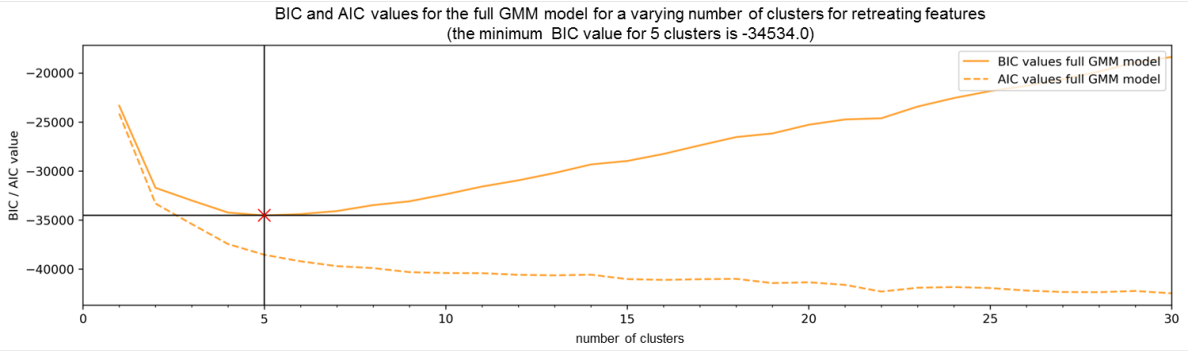


Figure C.29: Visualization of the global model (full GMM) computed information-theoretic criteria AIC (dashed lines) and BIC (solid lines) with a varying number of clusters for the regular dataset with 1414 extracted retreating features.

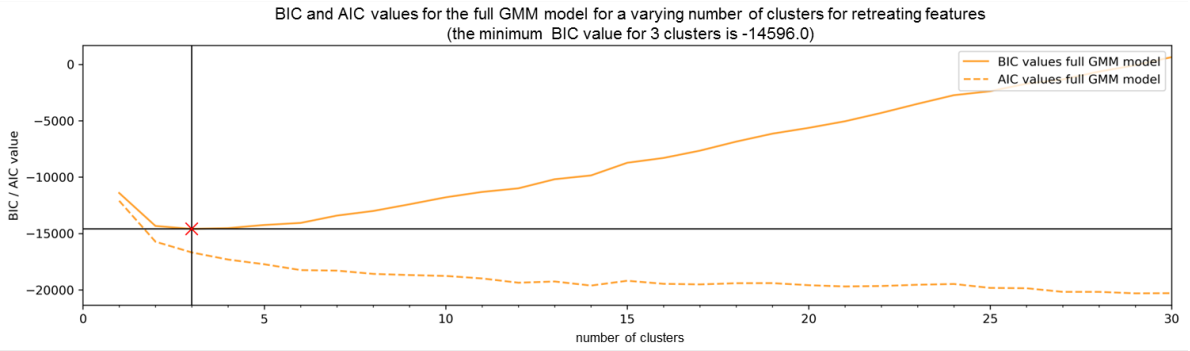


Figure C.30: Visualization of the global model (full GMM) computed information-theoretic criteria AIC (dashed lines) and BIC (solid lines) with a varying number of clusters for the irregular dataset with 707 extracted retreating features.

Both tables below show results of the GMMs for the split datasets. In Table C.9, the cluster-related mean output for every quantification is indicated. As can be seen, prograding and retreating clusters 5 regular have a relatively large width (quantification one) compared to all other clusters. Next to this, they show a relatively small change rate (quantification eleven). Prograding cluster 2 regular and retreating cluster 1 regular have the most cluster-included features and are stated to contain approximately the smallest quantifications compared to every other cluster. In other words, these cluster are likely to capture the features that are just large enough to be extracted (i.e. not threshold rejected). By highlighting quantification five and six, the ones that are used for the dataset splits, it can be seen that the irregular datasets contain high positive and negative numbers (i.e. highly irregular or fluctuating temporal trends). These quantifications indicate the temporal maximum and minimum annual value change and confirm the presence of mainly unnatural features. The irregularity in the aforementioned datasets is also seen in quantifications seven to ten.

Table C.10 presents the cluster-included number of transects as well as the affected km of shoreline. This table is used to confirm the above-presented explanation on cluster-included features and is elaborated on in Section 4.2.2. By summing the table information for the prograding and retreating datasets, it is found that features incorporate approximately 27 and 22 thousand transects that affect around 12 and 10 thousand km of shoreline respectively (see Section 4.1.2).

Table C.9: Overview of the twelve cluster mean quantifications for a total of seventeen clusters for all split datasets. The cluster labels (numbers) in the second column are determined by the GMM algorithms and ordered on a high to low number of included features, which are stated between brackets. The numbers one to twelve in the first row refer to the spatial and temporal quantifications. The unit for quantification one is km, quantification two and three $m/(year * km)$, quantification four, eleven and twelve $m/year$ and quantification five to ten m. The columns with quantification one and eleven are surrounded by vertical lines as these are selected to draw up the cluster mean spatiotemporal signature.

Feature dataset	Cluster	1	2	3	4	5	6	7	8	9	10	11	12
	2 (504x)	5	1	1	2	28	-23	17	-14	7	-3	2	0
	3 (491x)	4	1	1	2	56	-50	40	-36	26	-17	3	1
Regular prograding (1865x)	4 (393x)	5	1	3	4	45	-29	31	-20	14	-6	5	0
	1 (302x)	4	2	4	5	79	-47	66	-42	41	-21	7	1
	5 (178x)	13	0	1	2	31	-26	14	-13	10	-6	2	0
	2 (351x)	5	4	7	11	166	-71	122	-56	69	-29	16	2
Irregular prograding (1165x)	1 (342x)	3	2	3	3	212	-169	105	-96	105	-71	8	2
	3 (269x)	4	1	2	3	120	-113	68	-64	48	-37	4	1
	4 (203x)	3	3	7	6	378	-344	156	-169	180	-158	16	4
	1 (450x)	6	-1	-1	-2	21	-26	13	-11	1	-7	-2	4
	2 (347x)	4	-1	-2	-3	32	-41	23	-20	5	-12	-4	7
Regular retreating (1414x)	3 (248x)	4	-1	-1	-2	51	-65	46	-43	22	-32	-3	14
	5 (198x)	13	0	-1	-1	32	-38	15	-15	8	-17	-2	6
	4 (171x)	6	-2	-4	-5	44	-65	50	-34	13	-37	-7	13
	3 (350x)	4	-1	-2	-2	136	-173	87	-82	37	-83	-5	1
Irregular retreating (707x)	1 (197x)	4	-3	-5	-6	104	-142	90	-82	47	-68	-11	2
	2 (160x)	3	-3	-7	-6	343	-416	154	-138	140	-183	-15	4

Table C.10: Overview of the number of included transects and affected km of shoreline for a total of seventeen clusters for all split datasets. The cluster labels (numbers) in the second column are determined by the GMM algorithms and ordered on a high to low number of included features, which are stated between brackets.

Feature dataset	Cluster	Number of transects [-]	Affected shoreline [km]
Regular prograding (1865x)	2 (504x)	4879	2086.8
	3 (491x)	4113	1775.9
	4 (393x)	3767	1690.9
	1 (302x)	2207	957.0
	5 (178x)	4491	2184.5
Irregular prograding (1165x)	2 (351x)	2757	1322.1
	1 (342x)	1669	687.5
	3 (269x)	2111	931.9
	4 (203x)	962	398.1
Regular retreating (1414x)	1 (450x)	5226	2236.4
	2 (347x)	3013	1314.9
	3 (248x)	2159	921.7
	5 (198x)	5143	2497.2
	4 (171x)	1722	835.7
Irregular retreating (707x)	3 (350x)	2531	1127.9
	1 (197x)	1244	568.5
	2 (160x)	613	256.1

The two figures below show the applicability of the 97.5 percentile signature boundaries. These figures are referred to as scatter plots and are, compared to the visualizations in the main report, a different way of visualizing the outcomes of the GMMs for the split datasets. Figure C.31 visualizes the cluster means and standard deviations for the spatial extent (one-sided positioned width) and temporal change rate quantifications. As can be seen, only the edges of two cluster standard deviations extend past the signature boundaries. It is stated that this is another reason to analyze these clusters in Section 4.3. Individual feature scatter data is never shown in cluster mean plots like Figure C.31, as only two of the twelve quantifications are visualized. A feature is clustered in a twelve-dimensional space and might be separated from clusters in another dimension than the ones that are shown. Figure C.32 shows all 5154 extracted features for the width and change rate quantifications and indicates how the signature boundaries are related to the 97.5 percentile.

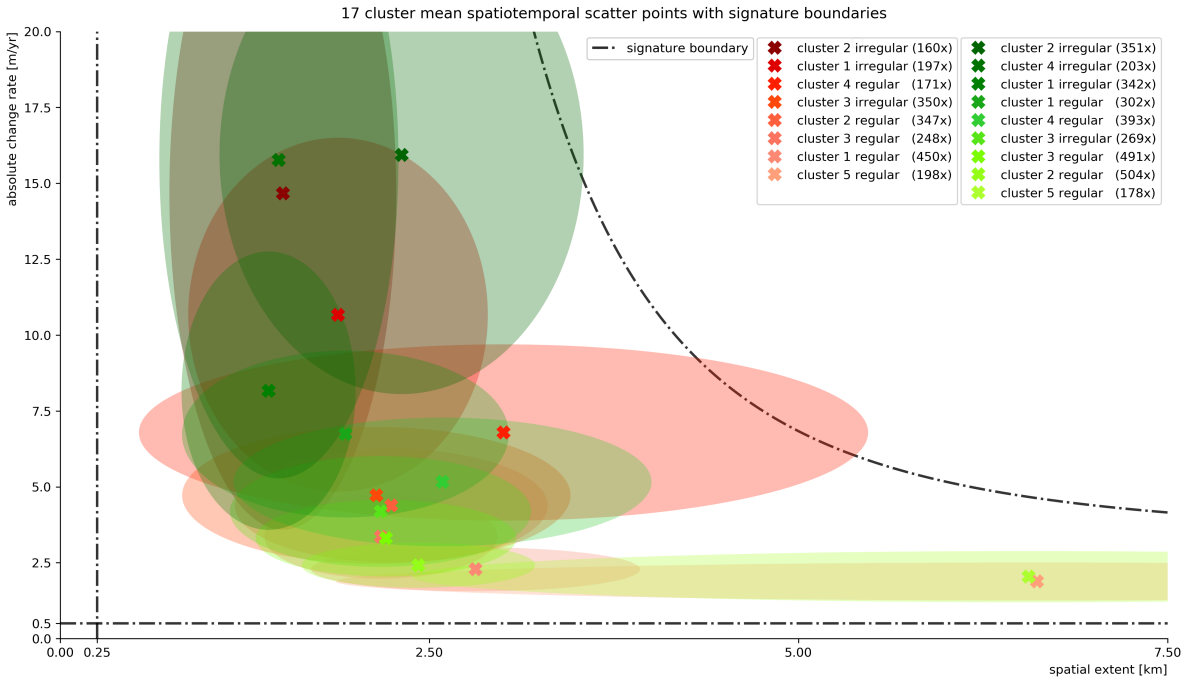


Figure C.31: Visualization of seventeen automatically derived and representative cluster mean (greenish colored crosses for prograding and reddish colored crosses for retreating) spatiotemporal scatter points for two of the twelve quantifications. The standard deviation ranges around the cluster mean scatter points are indicated by equally colored ellipses. The signature boundaries are shown by means of dashed-dotted black lines. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster. The legend indicates the cluster labels (number of included features between brackets), which are ordered from high to low absolute change rates.

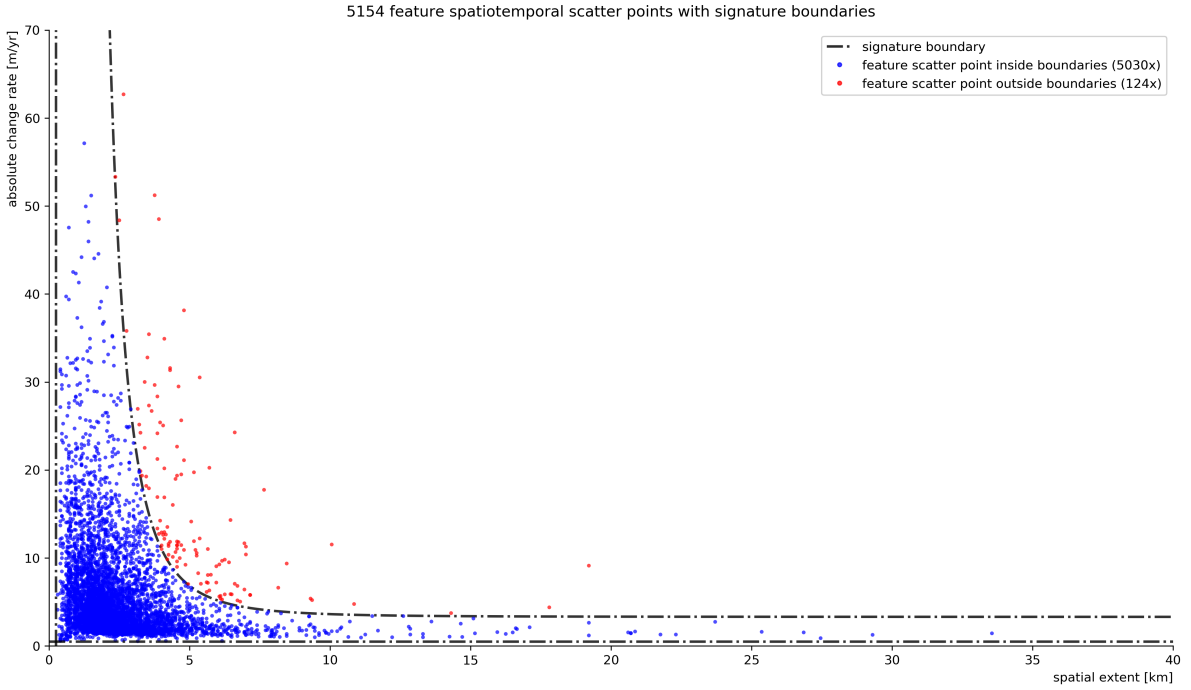


Figure C.32: Visualization of 5154 feature spatiotemporal scatter points (small dots) for two of the twelve quantifications. The signature boundaries are shown by means of dashed-dotted black lines. All feature points within the boundaries are indicated by a blue color, while points outside the boundaries are shown in red. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster.

C.6 Global cluster-driver correlations

The first four visualizations below are related to the summed hard and probable cluster-driver labels, which are elaborated on in Section 4.2.3. The next six figures are indicating single hard and probable cluster-driver labels. The applicability of these were shortly touched upon in Section 4.2.3 and discussed to a greater extent in Chapter 5.

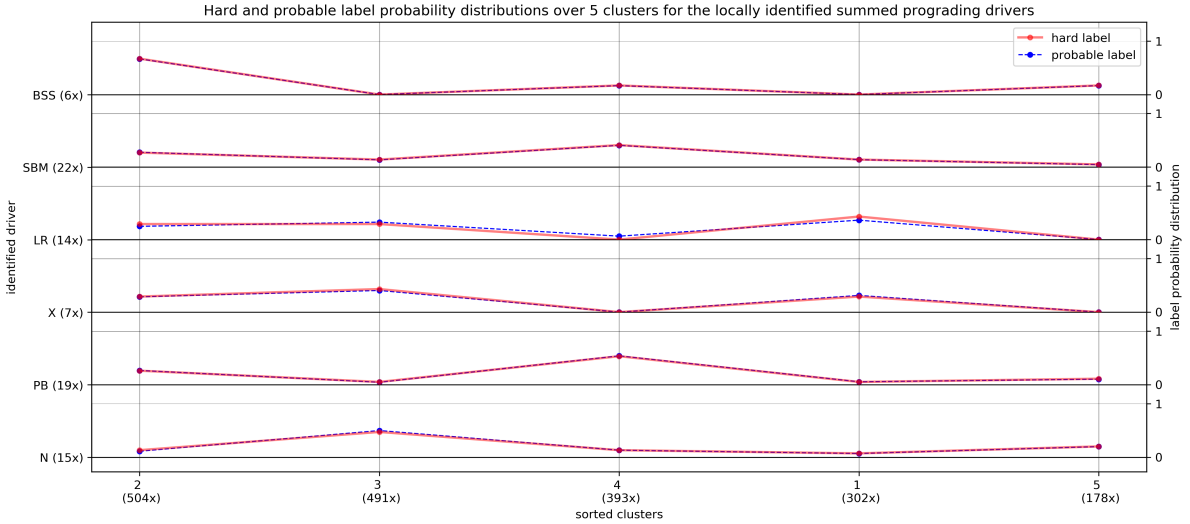


Figure C.33: Visualization of the summed hard (solid red line with dot) and probable (dashed blue line with dot) cluster label probability distributions related to 83 (unknown labels are rejected) locally identified drivers in the regular dataset with 1868 extracted and quantified prograding features. The y-axis on the left shows the identified driver id, with the summed number of labels between brackets. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets).

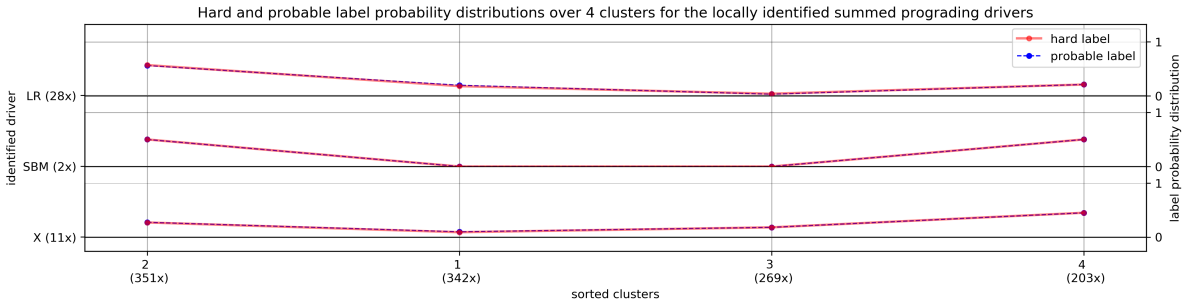


Figure C.34: Visualization of the summed hard (solid red line with dot) and probable (dashed blue line with dot) cluster label probability distributions related to 41 locally identified drivers in the irregular dataset with 1165 extracted and quantified prograding features. The y-axis on the left shows the identified driver id, with the summed number of labels between brackets. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets).

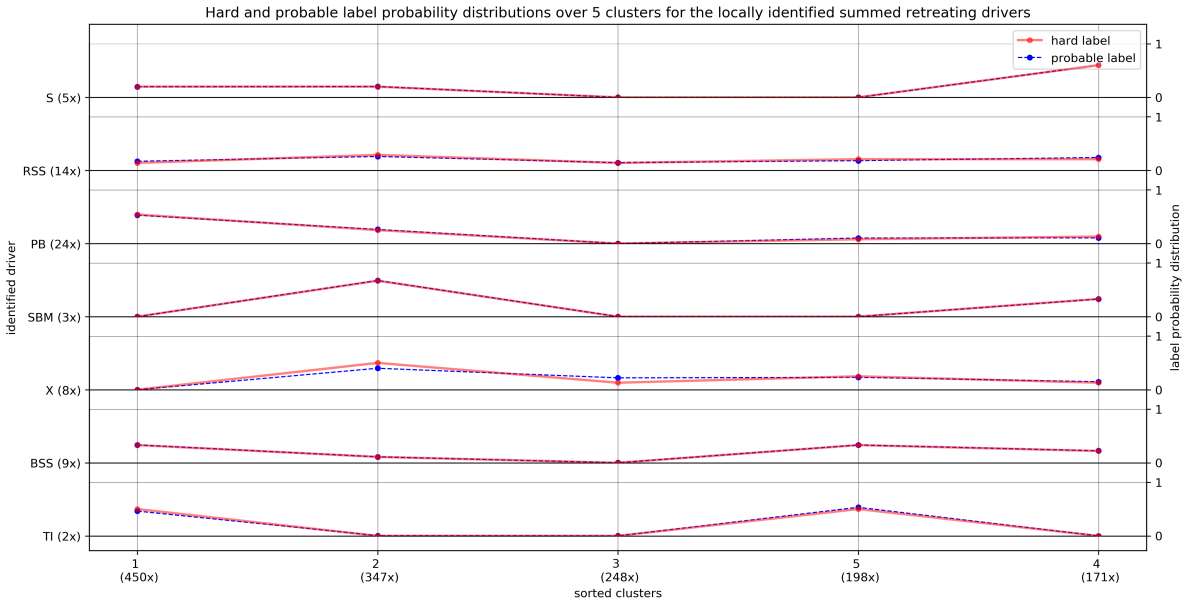


Figure C.35: Visualization of the summed hard (solid red line with dot) and probable (dashed blue line with dot) cluster label probability distributions related to 65 locally identified drivers in the regular dataset with 1414 extracted and quantified retreating features. The y-axis on the left shows the identified driver id, with the summed number of labels between brackets. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets).

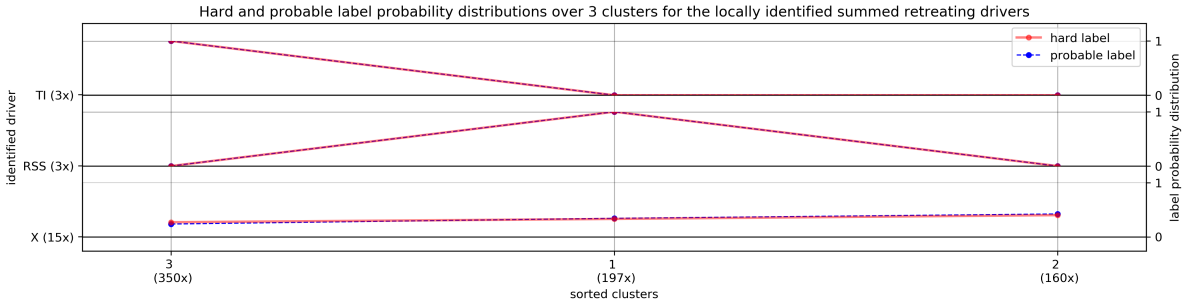


Figure C.36: Visualization of the summed hard (solid red line with dot) and probable (dashed blue line with dot) cluster label probability distributions related to 21 locally identified drivers in the irregular dataset with 707 extracted and quantified retreating features. The y-axis on the left shows the identified driver id, with the summed number of labels between brackets. The y-axis on the right indicates the label probability distribution and the x-axis shows the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets).

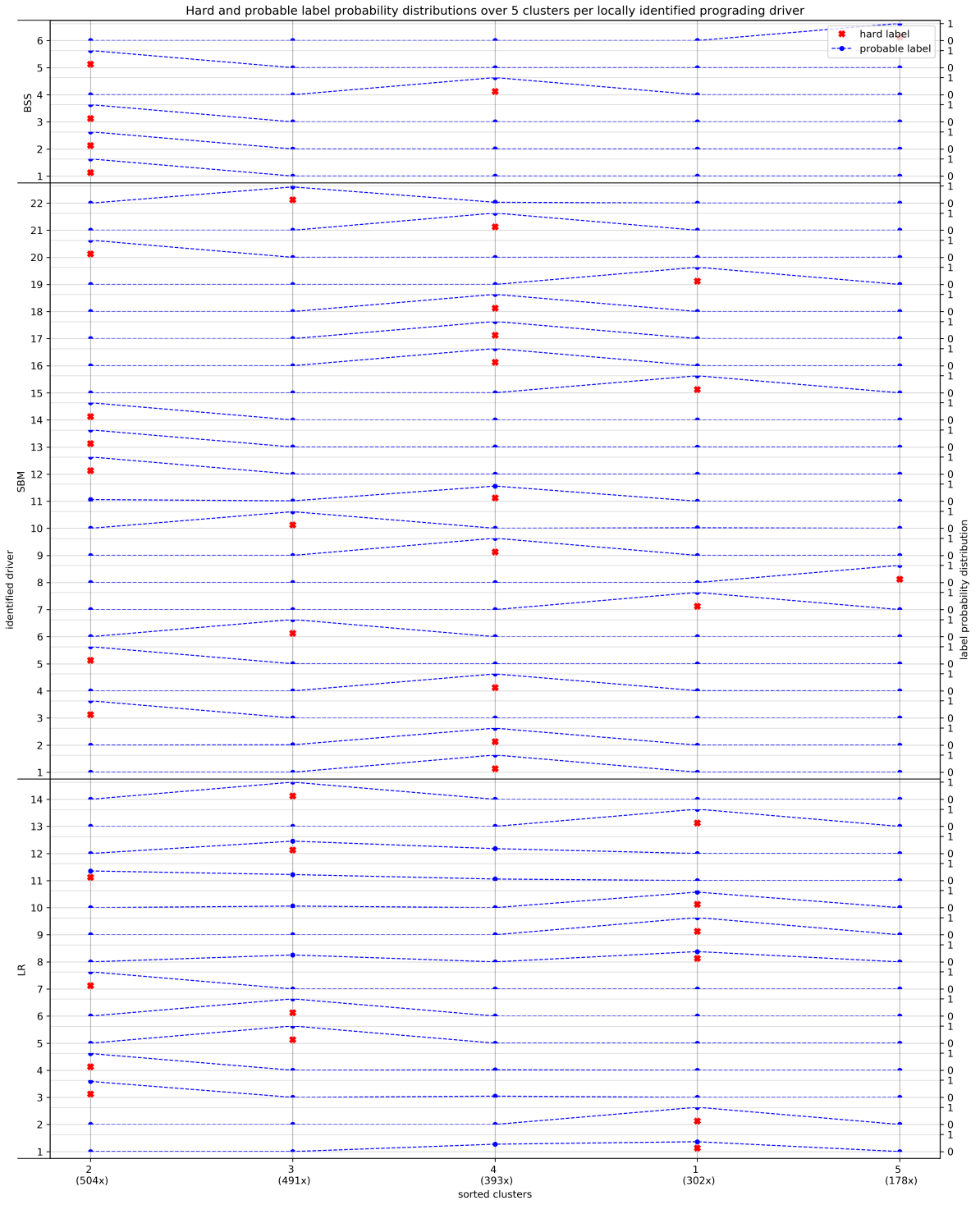


Figure C.37: Visualization of the single hard (red cross) cluster labels and probable (dashed blue line with dot) cluster label probability distributions related to 83 (unknown labels are rejected) locally identified drivers in the regular dataset with 1868 extracted and quantified prograding features. The y-axis on the left shows the identified driver id with its numbers. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets). This figure is split into two to fit in this report, the other half is shown on the next page.

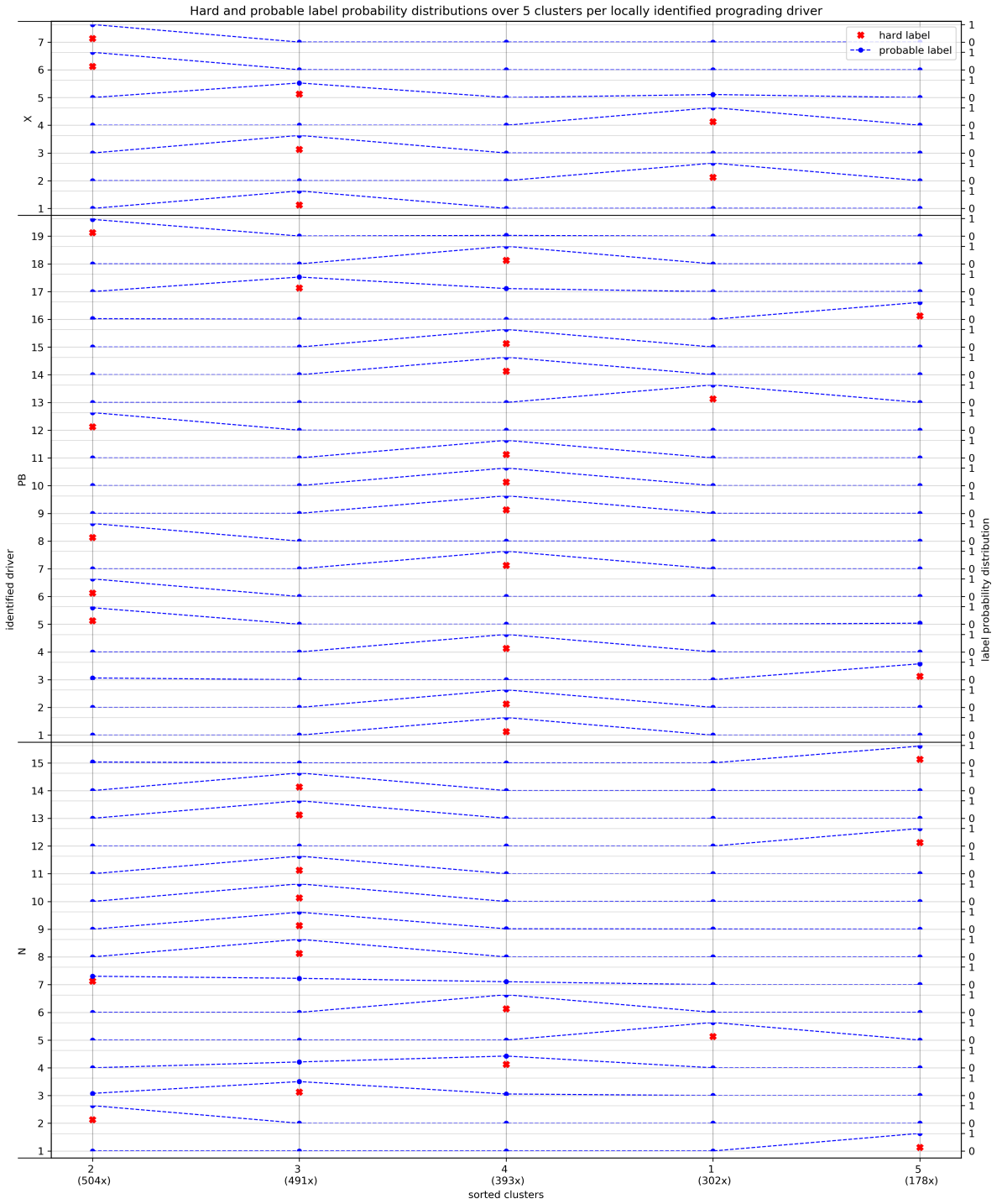


Figure C.38: Visualization of the single hard (red cross) cluster labels and probable (dashed blue line with dot) cluster label probability distributions related to 83 (unknown labels are rejected) locally identified drivers in the regular dataset with 1868 extracted and quantified prograding features. The y-axis on the left shows the identified driver id with its numbers. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets). This figure is split into two to fit in this report, the other half is shown on the previous page.

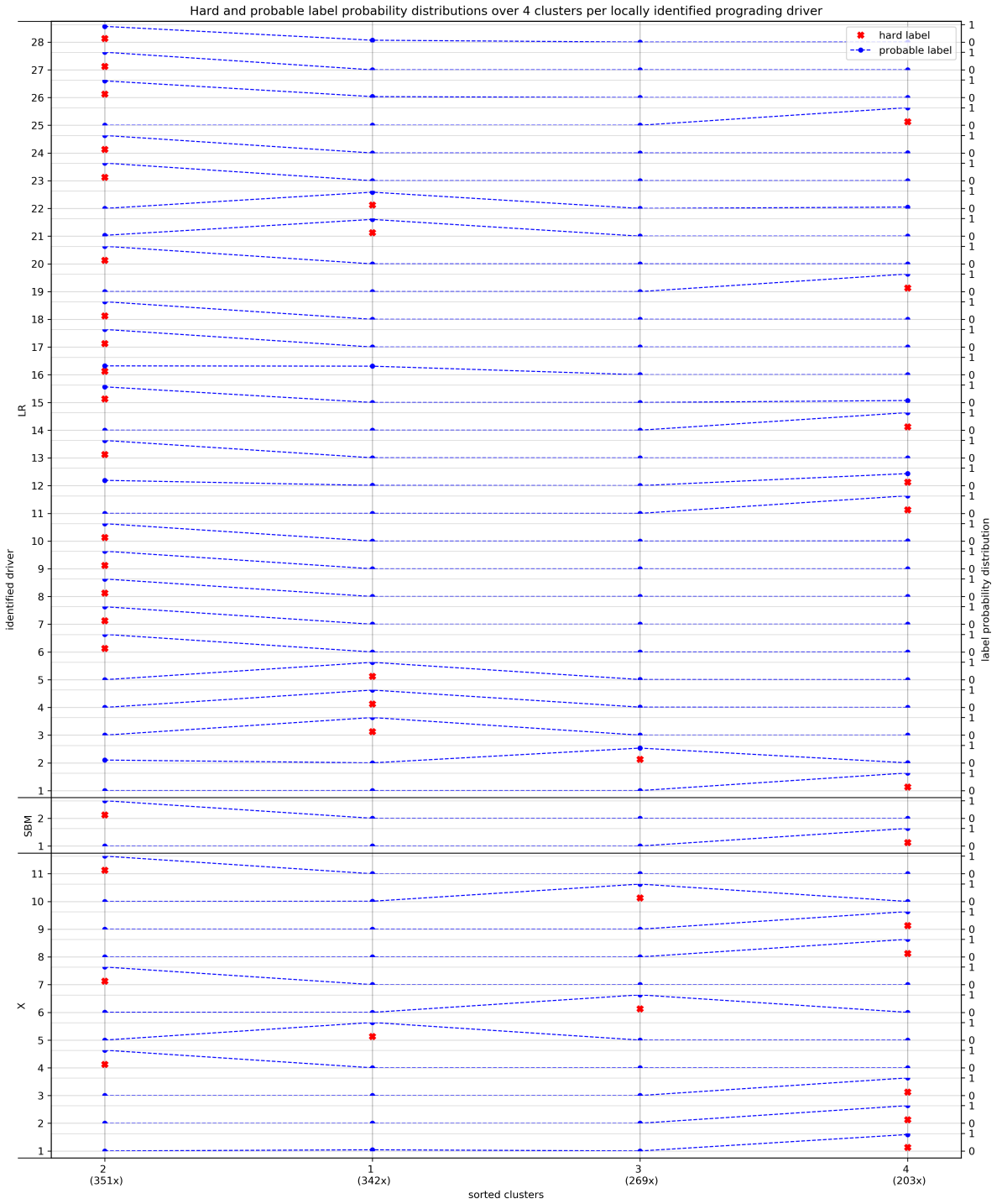


Figure C.39: Visualization of the single hard (red cross) cluster labels and probable (dashed blue line with dot) cluster label probability distributions related to 41 locally identified drivers in the irregular dataset with 1165 extracted and quantified prograding features. The y-axis on the left shows the identified driver id with its numbers. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets).

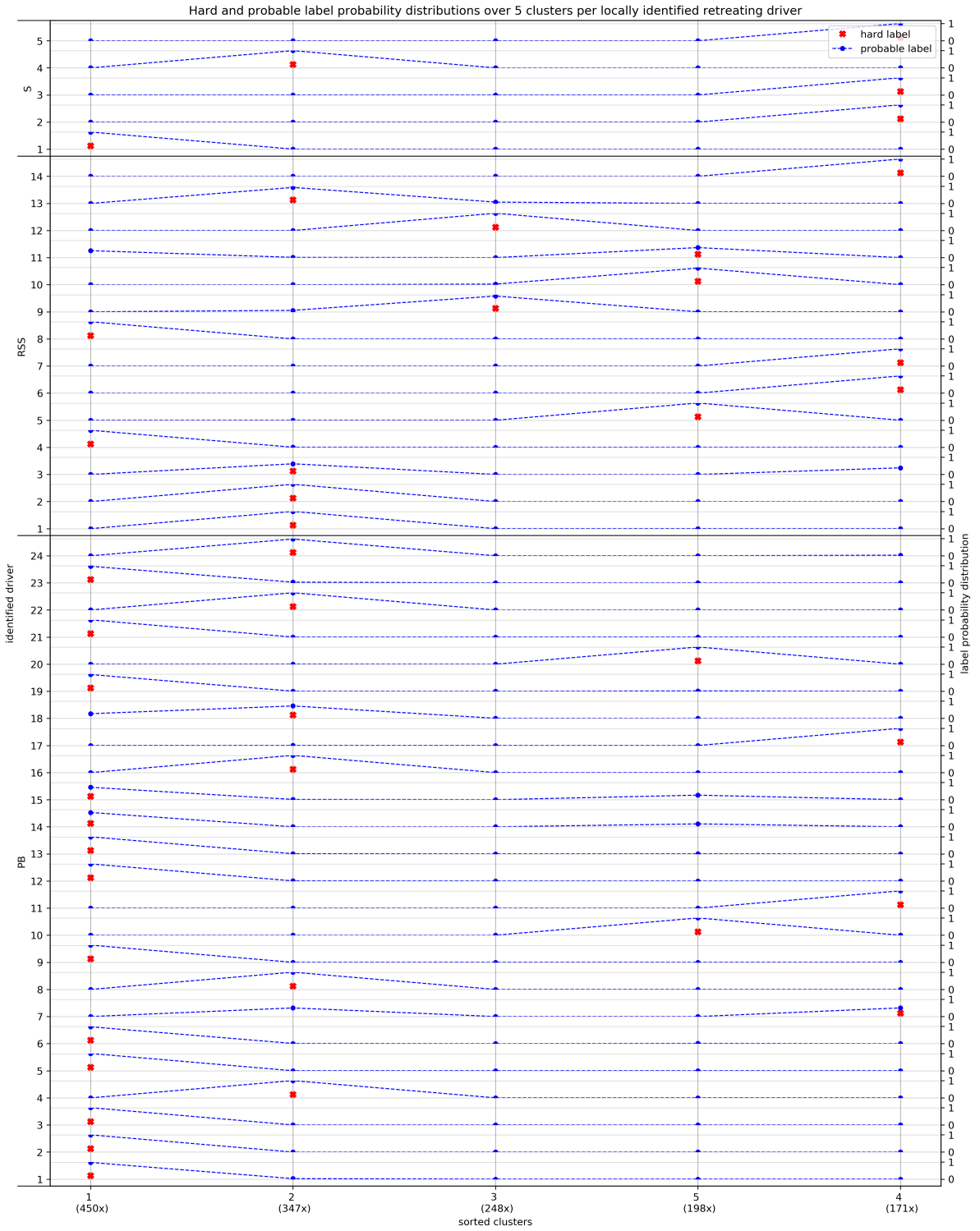


Figure C.40: Visualization of the single hard (red cross) cluster labels and probable (dashed blue line with dot) cluster label probability distributions related to 65 locally identified drivers in the regular dataset with 1414 extracted and quantified retreating features. The y-axis on the left shows the identified driver id with its numbers. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets). This figure is split into two to fit in this report, the other half is shown on the next page.

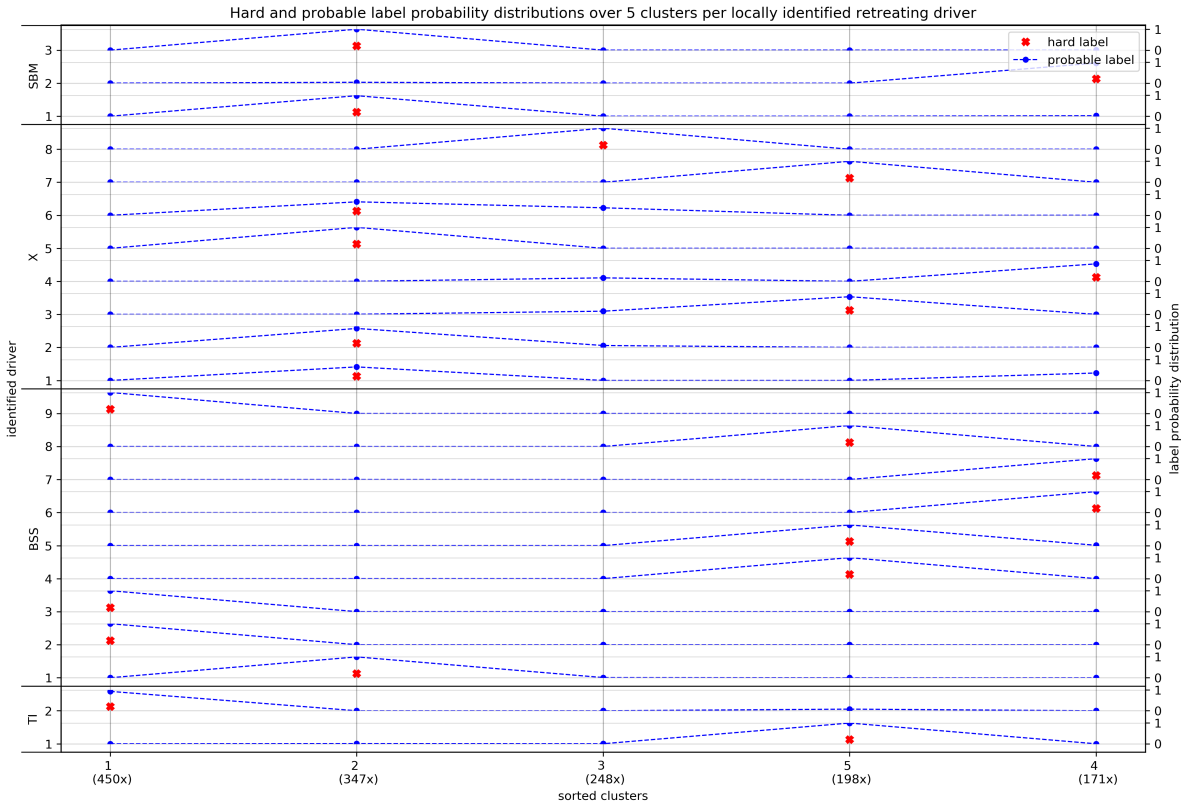


Figure C.41: Visualization of the single hard (red cross) cluster labels and probable (dashed blue line with dot) cluster label probability distributions related to 65 locally identified drivers in the regular dataset with 1414 extracted and quantified retreating features. The y-axis on the left shows the identified driver id with its numbers. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets). This figure is split into two to fit in this report, the other half is shown on the previous page.

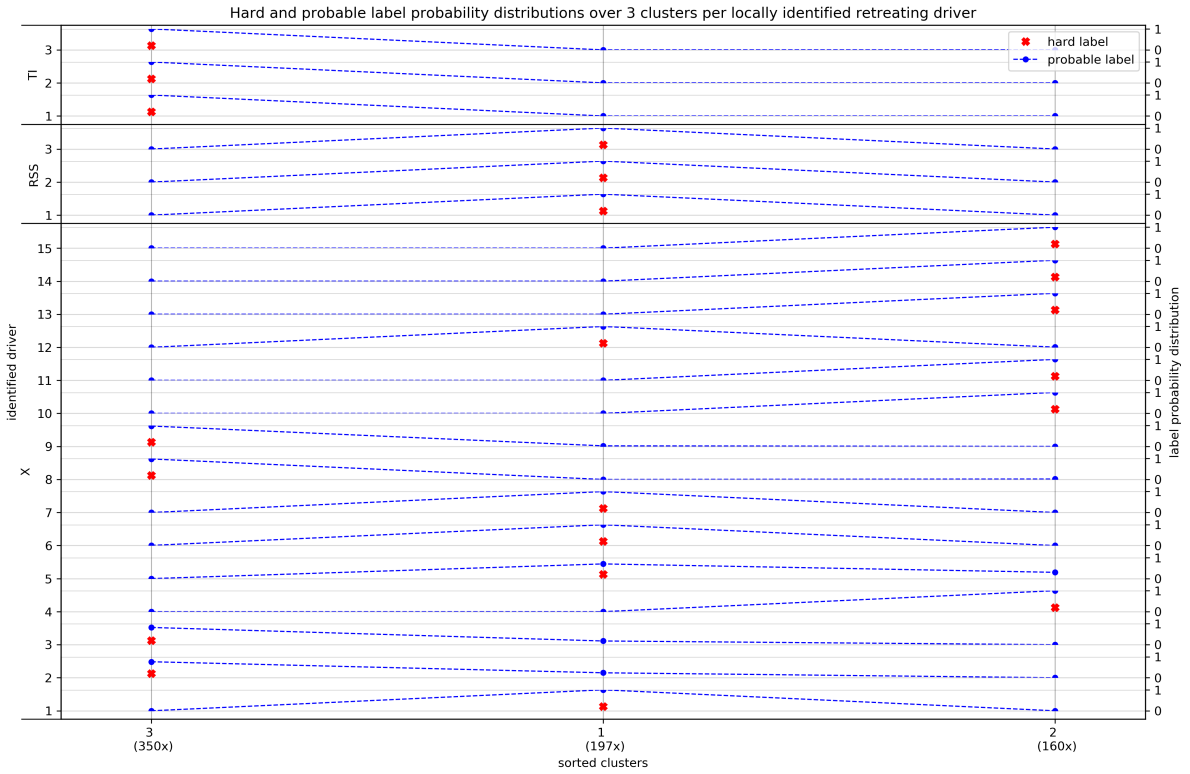


Figure C.42: Visualization of the single hard (red cross) cluster labels and probable (dashed blue line with dot) cluster label probability distributions related to 21 locally identified drivers in the irregular dataset with 707 extracted and quantified retreating features. The y-axis on the left shows the identified driver id with its numbers. The y-axis on the right indicates the label probability distribution and the x-axis the cluster labels, sorted from high to low (left to right) on the number of included features (between brackets).

As outlined in Section 4.2.3, only one cluster is validated to be correlated to only one driver. This cluster is referred to as retreating cluster 2 irregular and contains erroneously extracted features, which can easily be understood by mapping the geographical distribution of the cluster-included features as shown in Figure C.43. The locations in this map are seen to be close to frequent error locations like estuaries and bays, deltas and river mouths, and shallow shorelines.

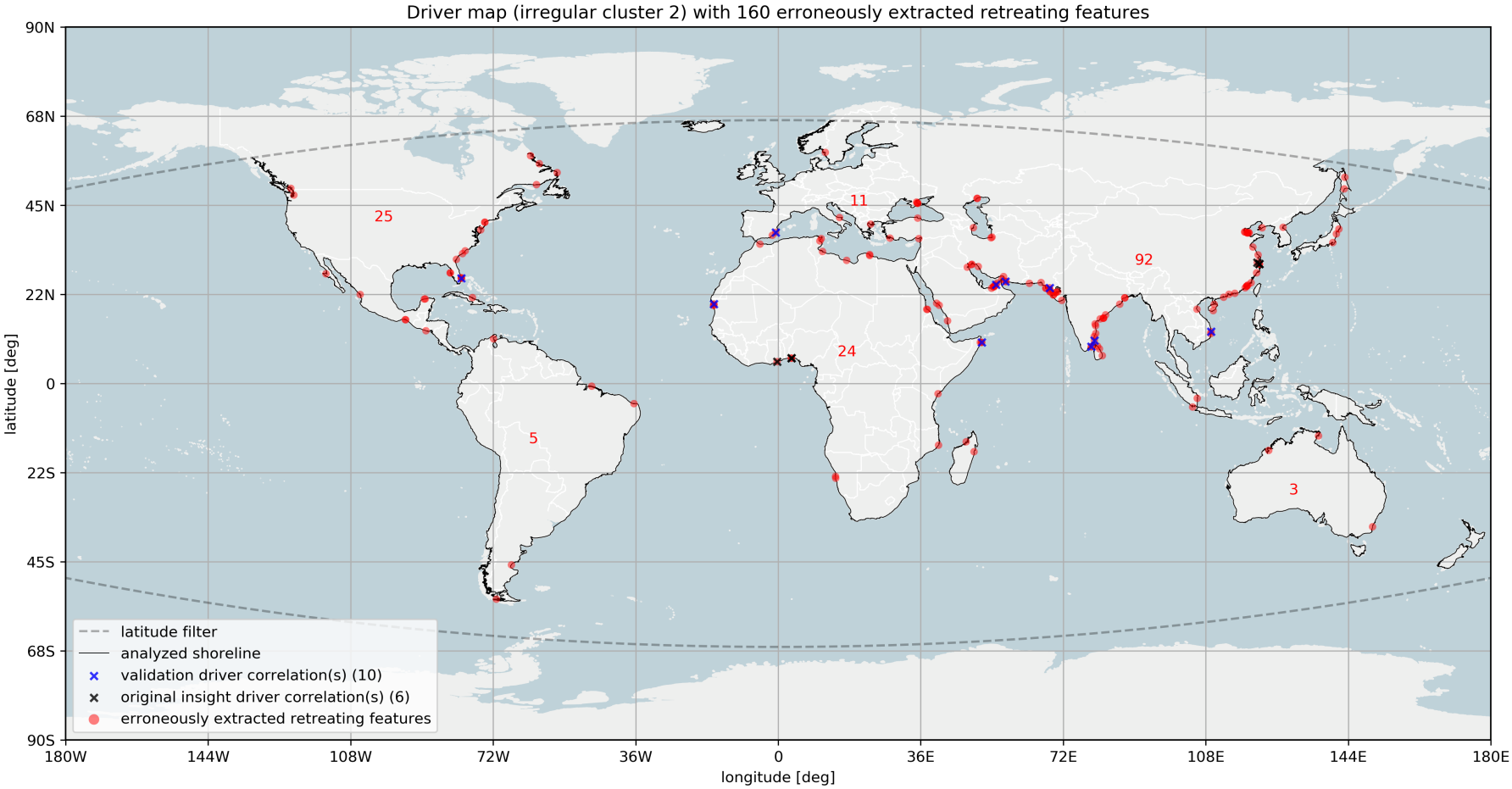


Figure C.43: Driver map for 160 erroneously extracted retreating features (red dots) on the analyzed 33 continuous shorelines (black lines). These features result from irregular cluster 2, which is part of the split dataset with 707 features. The 100% driver correlation for this cluster is based on local insight (black crosses) and randomly selected validation features (blue crosses). The red figures in the map indicate the number of erroneously extracted retreating features per continent. Latitude filters are shown with dashed gray lines.

Figure C.44 indicates the complete cluster-independent (in)direct human and natural influence distributions for the spatial width and temporal change rate quantifications. A zoom-in of this figure is visualized and elaborated on in Section 4.2.3.

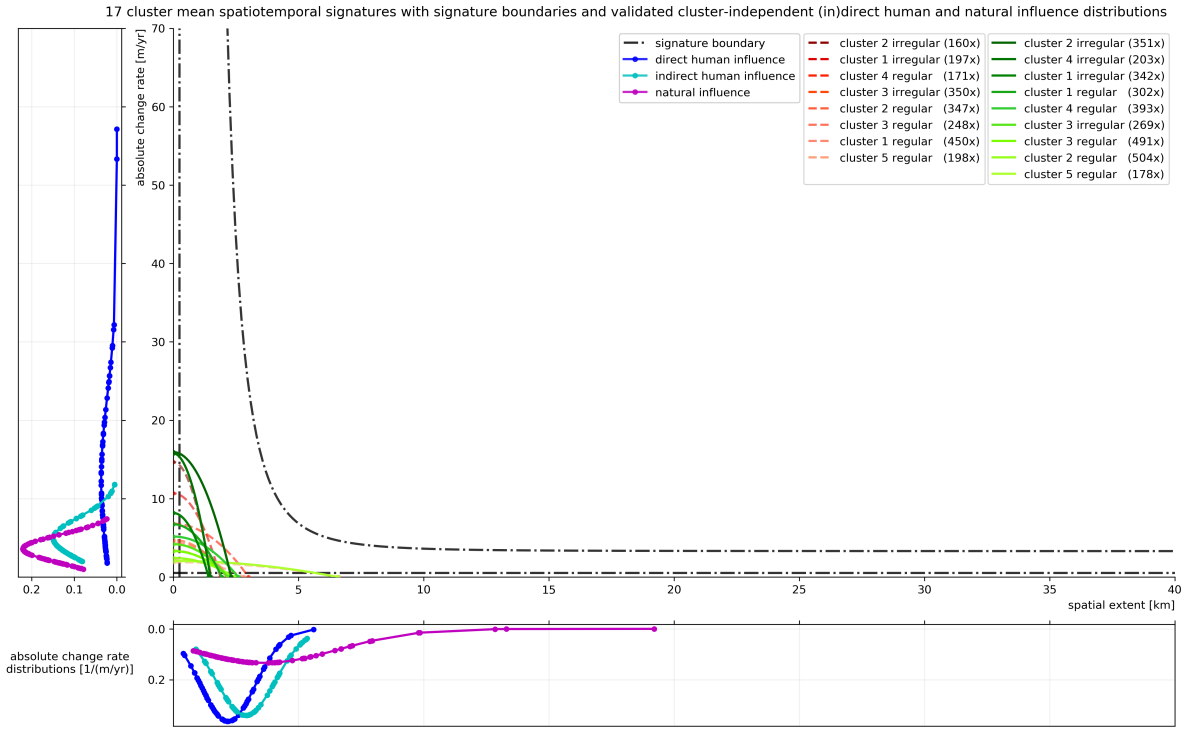


Figure C.44: Visualization of seventeen automatically derived and representative cluster mean (greenish colored lines for prograding and dashed reddish colored lines for retreating) spatiotemporal signatures for two of the twelve quantifications (zoom-out of Figure 4.21). The signature boundaries are shown by means of dashed-dotted black lines. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster. The legend indicates the cluster labels (number of included features between brackets), which are ordered from high to low absolute change rates. The left and bottom panel show the validated direct human (blue line with dots), indirect human (light blue line with dots) and natural (purple line with dots) influence distributions (independent from cluster assignments) for the absolute change rate and the extent respectively.

C.7 Global sandy shoreline evolution classification

Both figures below are related to the outcomes presented in Section 4.3, which are based on the method outlined in Section 3.3. In Figure C.45, the analysis-included clusters are highlighted by means of thicker spatiotemporal signature lines. By deleting the non-highlighted lines, Figure 4.22 is obtained. Figure C.46 shows the cluster-included features for prograding and retreating clusters 5 regular. This combined map is related to the analysis that considers natural and indirect human influences and shows enlarged locations whenever there are multiple features (either prograding or retreating) in a 50 km radius with each other. Enlarged locations might indicate the presence of large-scale drivers with relatively small change rates.

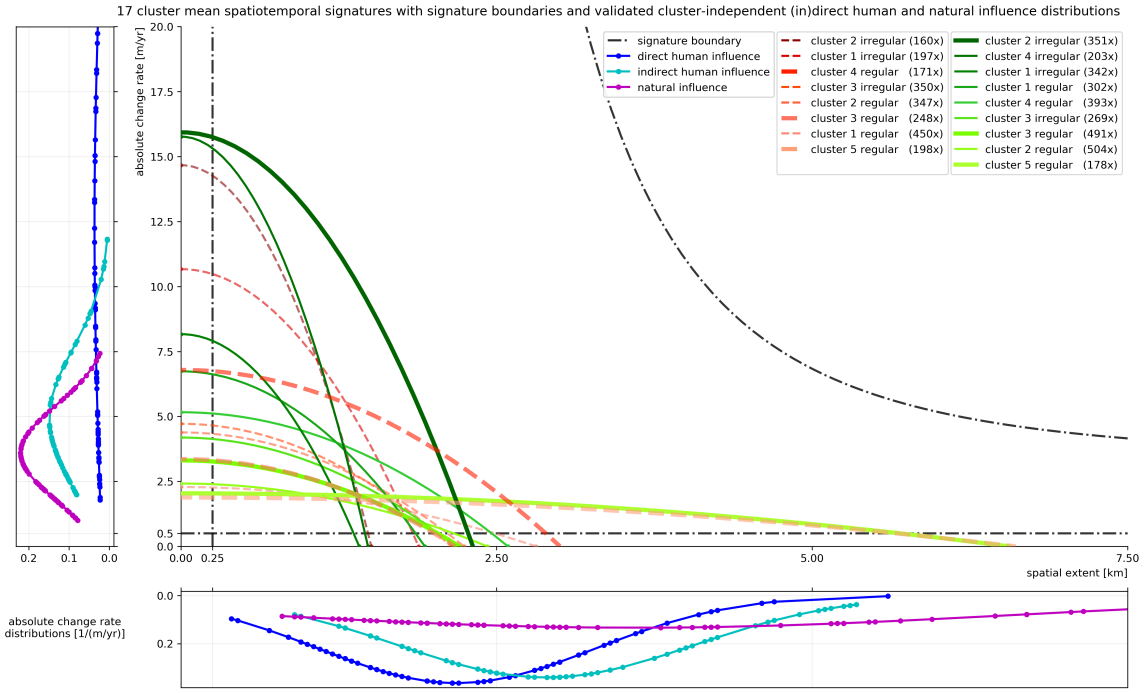


Figure C.45: Visualization of seventeen automatically derived and representative cluster mean (greenish colored lines for prograding and dashed reddish colored lines for retreating) spatiotemporal signatures for two of the twelve quantifications. The six interesting clusters that are seen in Figure 4.22 are highlighted by means of thicker lines. The signature boundaries are shown by means of dashed-dotted black lines. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster. The legend indicates the cluster labels (number of included features between brackets), which are ordered from high to low absolute change rates. The left and bottom panel show the validated direct human (blue line with dots), indirect human (light blue line with dots) and natural (purple line with dots) influence distributions (independent from cluster assignments) for the absolute change rate and the extent respectively.

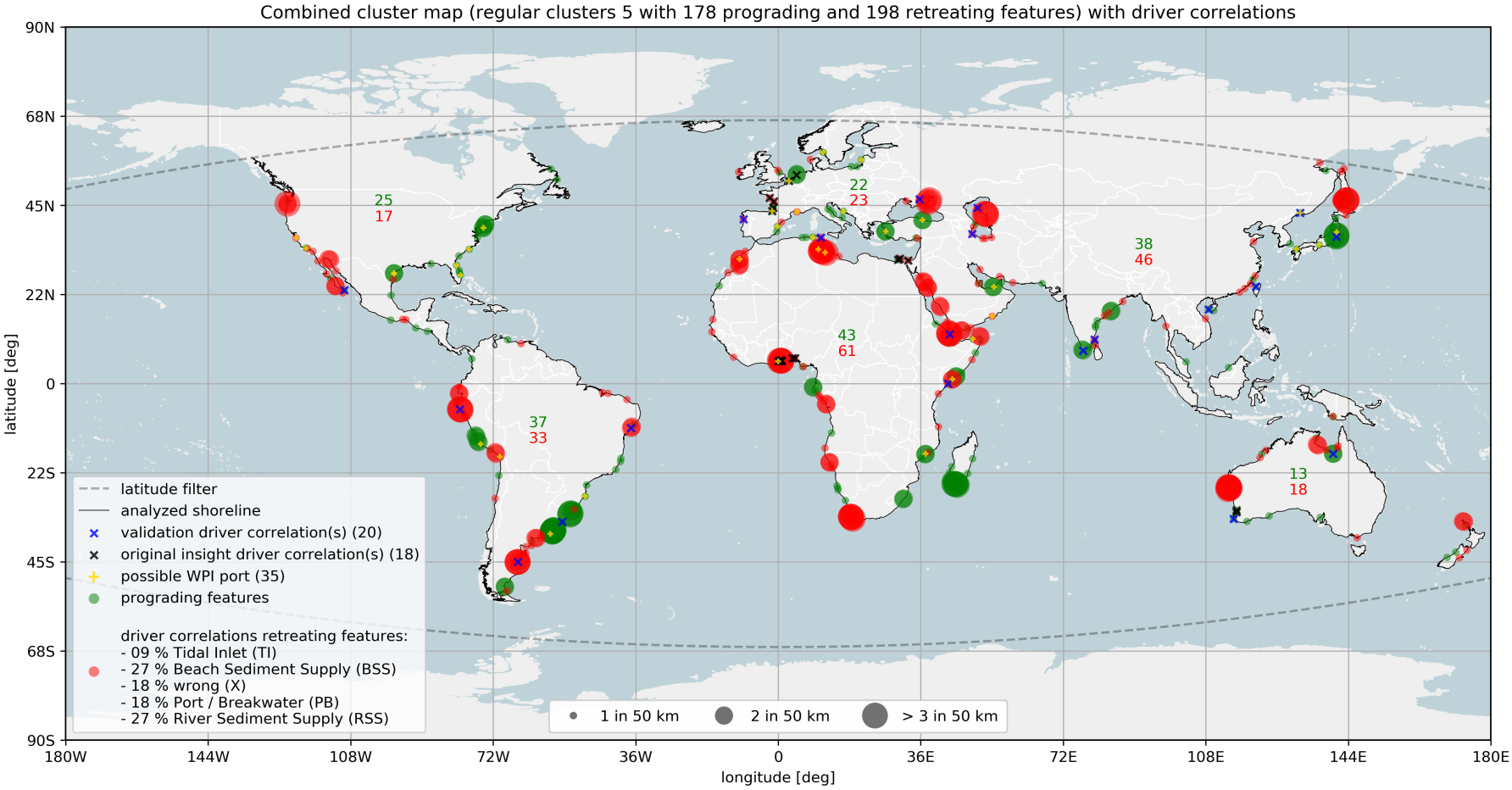


Figure C.46: Combined cluster map with 178 extracted prograding shoreline evolution features (green dots) and 198 extracted retreating features (red dots) with driver correlations on the analyzed 33 continuous shorelines (black lines). These features result from regular clusters 5, which are part of the split datasets with 1865 and 1414 features. The bottom-centered legend in the visualization indicates the presence of multiple features in a 50 km radius in relation to the size of enlarged green and red dots. The driver correlations for this cluster (see bottom-left legend) are based on locally generated insight (black crosses), which is validated on randomly selected features (blue crosses). There was an inconsistency in the locally generated insight and randomly selected validation features, which resulted in the rejection of driver correlations for the prograding features. Features that are possibly related to a WPI port are marked with a yellow '+'-sign. The green and red figures in the map indicate the number of extracted prograding and retreating features per continent. Latitude filters are shown with dashed gray lines.

D

Discussion

This appendix provides visualizations for two of the three discussed implications for practice in Chapter 5. The first figure below is related to the first topic that considers shoreline evolution on different spatial scales and shows the example case of hard 100 km sections. The second figure, on the next page, relates to the application of the identification of erroneously extracted features.

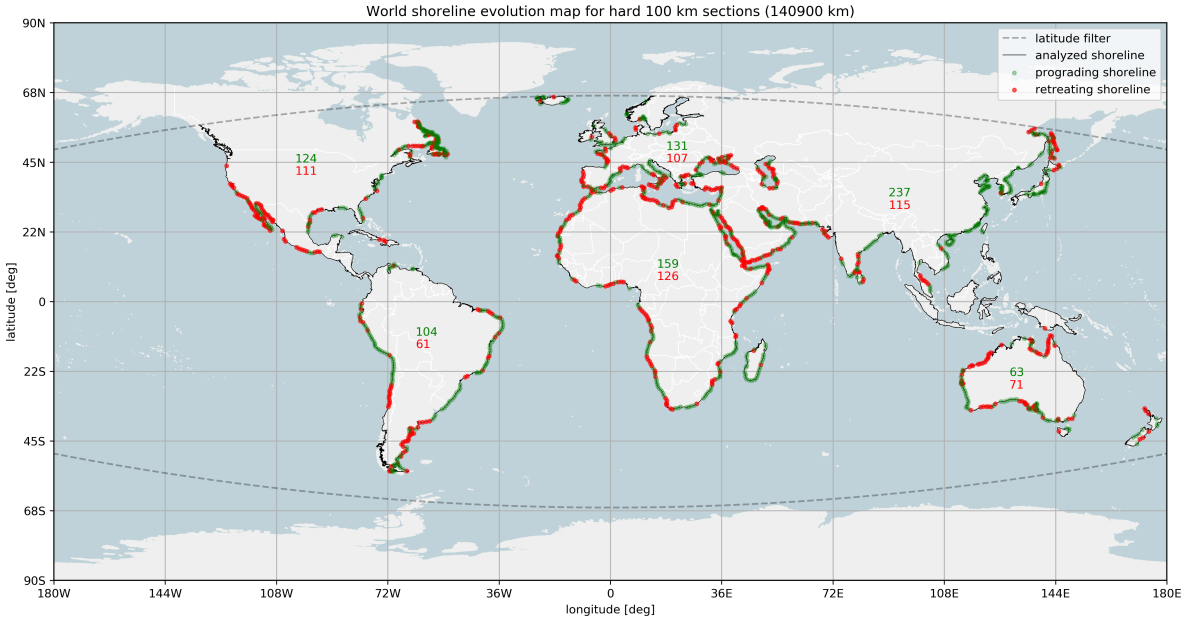


Figure D.1: Global visualization of the analyzed 33 continuous shorelines (black lines) together with 818 prograding (green dots) and 591 retreating (red dots) hard 100 km sections (1409 in total). This visualization is referred to as a shoreline evolution map, where each dot represents the average long-term change rate over the total of all section-included transects. The green / red figures in the map indicate the number of hard prograding / retreating sections per continent. Latitude filters are shown with dashed gray lines.

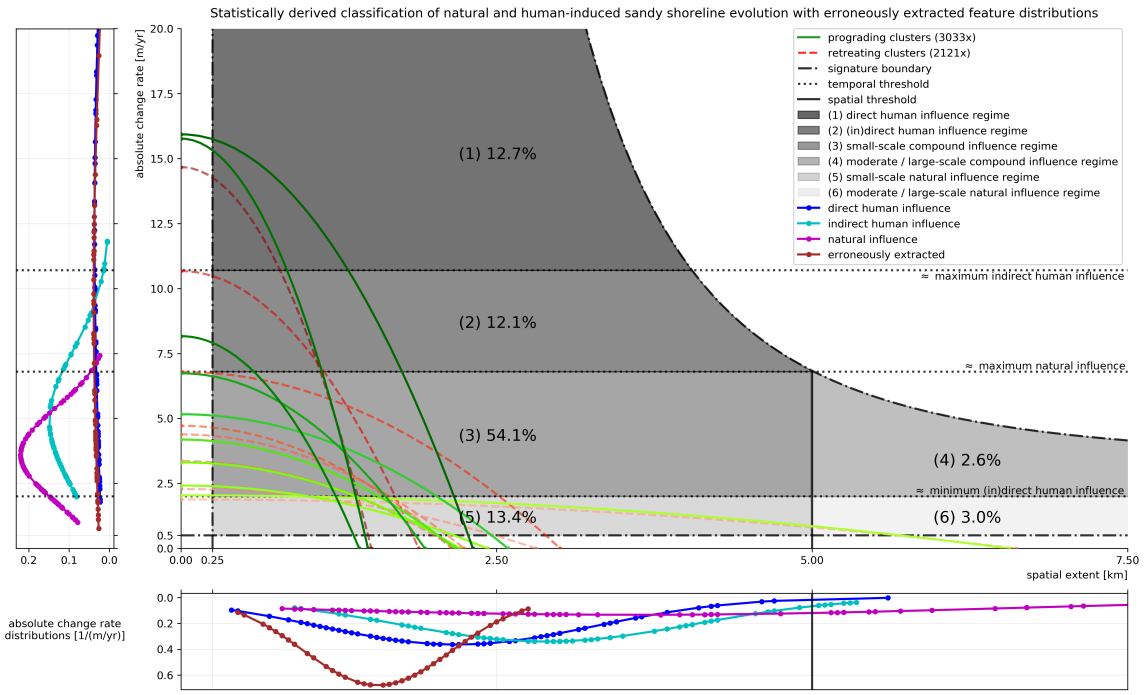


Figure D.2: Visualization of the statistically derived classification of natural and human-induced sandy shoreline evolution for two of the twelve quantifications of the seventeen automatically derived and representative cluster mean spatiotemporal signatures. The greenish and dashed reddish colored lines are related to prograding clusters (nine times) and retreating clusters (eight times) respectively. The number of included features in relation to the total of prograding and retreating clusters are shown between brackets in the legend. The signature boundaries are shown by means of dashed-dotted black lines and temporal thresholds with horizontal dotted black lines with indicative text on the far right of these lines. A vertical black line indicates the spatial threshold. All regimes (see legend) within the signature boundaries are indicated with different grayish colors, dependent on their temporal and spatial threshold splits. Percentages regarding the number of features (from the total of 5154 extracted features) within each of these regimes are also shown. The spatial extent on the x-axis shows the one-sided positioned width and the y-axis the absolute temporal change rate of a cluster. The left and bottom panel show the validated direct human (blue line with dots), indirect human (light blue line with dots) and natural (purple line with dots) influence as well as erroneously extracted feature (brown line with dots) distributions (independent from cluster assignments) for the absolute change rate and the extent respectively.