Characterization of Geomorphological Surface Activities Using Near-Continuous Terrestrial LiDAR Time Series

The classification of 4D objects-by-change

Master Thesis

Daan Hulskemper

TUDelft

Characterization of Geomorphological Surface Activities Using Near-Continuous Terrestrial LiDAR Time Series

The classification of 4D objects-by-change

by

Daan Hulskemper

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Monday February 20th, 2023 at 12:00 PM.

Graduation supervisor:	Dr. R. C. Lindenbergh	Delft University of Technology
Daily supervisor:	Dr. K. Anders	Heidelberg University
Additional supervisor:	Dr. J. A. Á. Antolínez	Delft University of Technology
Additional supervisor:	M. Kuschnerus	Delft University of Technology
Institution:	Delft University of Technology	
Faculty:	Civil Engineering and Geosciences	
Master track:	Geoscience and Remote Sensing	
Project duration:	May 19, 2022 – February 20, 2023	

TUDelft   UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Preface

The following thesis is the result of almost nine months of research conducted as a final part of my studies in Geoscience and Remote Sensing. These months have shown to be the most educational of my life. During this period, I performed my research within the 3DGeo research group at Universität Heidelberg for around seven months. It is safe to say that without all the support from the kind and bright people there, I would not have been able to enjoy it and learn as much as I did. For that, I would like to thank Katharina, Vivien, Bernhard, Hannah, and all the other 3DGeo members. I am very grateful for the opportunity Roderik and Katharina provided to do my research there. It has been a pleasure to work and even perform fieldwork with such an inspiring group of people.

Furthermore, I would like to thank all my supervisors, Roderik, Katharina, José, and Mieke, for all the valuable discussions, ideas, feedback, and enthusiasm throughout the research. Without our bi-weekly meetings, I would likely have struggled to keep myself motivated, confident, and assured. Also, I'd like to thank Roderik, in particular, for the idea of presenting and publishing my work at the O3DM conference. Doing this has been a valuable learning opportunity and extra motivation to keep me going.

Also, aside from studying, I had a fantastic time in Heidelberg, and for that, I would like to thank all the (flat)friends I made during my stay. Their support, kindness, and distractions made my stay unforgettable. The support I received during the last months of writing from my dear friends and family in The Netherlands was also priceless, and I would like to thank them all for that. In particular, I would like to thank the one friend who provided me with madeleines during my most stressed-out period, which truly made a difference in keeping me going.

Daan Hulskemper
Delft, February 2023

Abstract

The Earth's landscapes are shaped by processes eroding, transporting, and depositing material over various timespans and spatial scales. To understand these surface activities and mitigate potential hazards they inflict, knowledge is needed of their occurrences and properties. Near-continuous terrestrial laser scanning (TLS) enables the acquisition of point cloud time series, constituting up to thousands of three-dimensional surface morphology representations. Exploiting the full potential of this large amount of data by extracting and characterizing different types of surface activities inside these point cloud time series is, however, challenging.

This thesis addresses this challenge by developing an automated and unsupervised method for classifying 4D objects-by-change (4D-OBCs). These 4D-OBCs represent the spatiotemporal extent of individual surface activities in a point cloud time series. They are classified using a Self-organizing Map (SOM) and hierarchical clustering, grouping them into different levels of surface activity types. The workflow is tested on its ability to characterize surface activities in two study areas, a sandy beach and a snow-covered Alpine area.

Application of an optimized SOM configuration on the sandy beach results in groups of 4D-OBCs physically interpretable as different types of surface activities. A validation dataset containing 51 manually labeled 4D-OBCs of various surface activity types (e.g., intertidal bar depositions, anthropogenic bulldozer depositions) is distributed over the SOM generally according to the labels provided. The SOM thus enables the identification of 4D-OBCs displaying a particular type of surface activity, as well as subtle differences between events of one surface activity. Hierarchical clustering allows us to find and characterize broader groups of surface activities, even if the same type occurs at different points in space or time. For example, the varying spatial redistribution of sand through the initiation of different types of surface activities after the destruction of an intertidal bar system under different environmental conditions is successfully studied.

A nearly identical workflow configuration applied on the snow cover 4D-OBC set does not result in equal performance. Several groups of surface activity in the SOM contain a combination of 4D-OBCs representing different surface activity types. These results highlight the necessity of a study area specific workflow optimization by selecting specific features and SOM configurations. However, if optimized for the specific environment, the workflow has the potential to be used in long-term automated monitoring of surface activity in systems with complex morphological interactions, increasing the applicability of TLS for studying geomorphological change.

Contents

1

Introduction

The current shape of the Earth's surface results from erosion, transport, and deposition processes. These are naturally (e.g., wind, gravity) or anthropogenically (e.g., bulldozers) forced and occur over various spatial and temporal scales. We, therefore, define surface activities as events where the morphology of a local surface is changing over a certain period. Examples of surface activities in different geographic settings are landslides, avalanches, and sandy beach erosion. They cause severe natural hazards in many settings, e.g., shorelines retreat into densely populated areas due to increased erosion activity (Vousdoukas et al. 2020) and landslides cause fatalities and livelihood damage (Hilker et al. 2009). Knowledge of the impact and occurrence of these activities is therefore essential to predict, mitigate and adapt to the potential hazards they inflict, particularly as projected climate change potentially increases the occurrence and effect of different surface processes (Crozier 2010; Vousdoukas et al. 2020). The large variety in spatial and temporal scale and the often spatiotemporally superimposed and difficult-to-predict nature of surface processes do, however, impose challenges for the observation of surface activities (Anders et al. 2021).

Near-continuous terrestrial laser scanning (TLS) enables monitoring of surface changes over multiple time scales (Eitel et al. 2016). In a near-continuous setup, a TLS device is placed at a fixed location for months to years (e.g., Vos et al. 2017; Campos et al. 2021). 3D point clouds with up to mm-scale accuracy and resolution can then be acquired at (sub)hourly intervals. As such, a substantial amount of data is collected that contains information on the changes in the morphology of the scene's surface. The vast amount of data (e.g., thousands of point clouds) collected through this setup challenges the visual and manual extraction of interpretable and useful information. To exploit the available point cloud time series, methods are needed which identity, segment and characterize occurrences of surface activity from these large four-dimensional (3D + time) datasets.

Anders et al. 2021 developed a fully automated method of extracting spatiotemporal segments of morphological change from these point cloud time series, called 4D objects-by-change (4D-OBCs). This method allows to investigate dynamic environments where the surface morphology changes near-continuously. The technique has already been shown to be useful for extracting spatiotemporal segments of surface activity in two different environments, namely a sandy beach in the Netherlands (Anders et al. 2021), and an Alpine snow-covered region (Anders et al. 2022).

The extracted 4D-OBCs are spatiotemporal segments inside a point cloud time series, which have a certain spatial extent relating to the area over which the surface activity occurred and a temporal extent representing the duration and evolution of height change of the surface activity. Various 4D-OBCs with different spatiotemporal characteristics are present in the dataset, representing different types of surface activity. These different surface activities result from variations in their underlying processes. Thus, to understand the underlying processes and types of surface activities, one has to identify which type of surface activity each 4D-OBC represents. However, this identification is not easily done manually because of the size of the 4D-OBC datasets. E.g., 2,021 4D-OBCs were identified over a period of 5 months in Kijkduin (Anders et al. 2020). Therefore, the automated classification of the

4D-OBC dataset into the various types of surface activity could make the 4D-OBC dataset interpretable and even more helpful for the analysis of surface activity and the processes behind it.

This research aims to make the 4D-OBCs interpretable as different types of surface activities by clustering them into different groups that characterize surface activities. We want to enable the analysis of the impact and correlations of different types of surface activity through near-continuous TLS. Therefore, we develop a method, that automatically identifies different levels of clusters of 4D-OBCs in an unsupervised manner, i.e., low-level clusters that define specific types of erosion and deposition (e.g., sandbar deposition vs. aeolian dune erosion) and high-level clusters (e.g., erosion vs. deposition). The latter is needed as different applications require different surface activity characterization levels. One might for example be interested in the relative impact of storm erosion on the total magnitude of beach erosion (e.g., Callaghan et al. 2009)

Research Questions
Regarding the objective mentioned above, the following question will be answered:
How can different types of surface activity be characterized from a point cloud time series using the spatiotemporal segments derived as 4D objects-by-change?

To answer this question, the following subquestions need to be answered:

- *Which different types of surface activity occur in the two environments studied in this research?*

  We need knowledge of what kind of surface activities occur on the sandy beach, and in the alpine snow-covered regions of our study areas to be able to assess which surface activities the 4D-OBCs might represent. It is also needed to identify which of these has the appropriate spatiotemporal scale, to be represented as a 4D-OBC. This is investigated through a literature study, analysis of the 4D-OBC spatiotemporal segmentation method, and manual investigation of the 4D-OBC dataset.

- *Which unsupervised classification methods are applicable for the grouping of the 4D-OBCs into types of surface activities?*

  A large amount of unsupervised classification methods are present nowadays, and all have advantages for certain applications. It is identified which of these unsupervised classification methods are applicable for the grouping of 4D-OBCs into different surface activities, and subsequently, a (combination of) algorithms is chosen. This is done by identifying the possible types of surface activities, how they are represented in the 4D-OBCs, and comparing the (dis)advantages of various unsupervised classification method.

- *How does the choice of features influence the grouping of the 4D-OBCs into different types of surface activity?*

  We first develop a method to rank a set of features on the basis of their importance, and develop and use methods to test various selections of these features. Furthermore, we inspect the distribution of the features over the various groups of 4D-OBCs we found and identify which features show notable variations over the various groups.

- *How can the unsupervised classification method be optimized for grouping the 4D-OBCs?*

  We develop a method to assess the performance of the unsupervised classification methods using various configurations of the parameters. The assessment is done both data-driven and manually.

- *To what extent can the grouped 4D-OBCs be interpreted as different types of surface activity?*

  The grouped 4D-OBCs are manually inspected, and the feature vectors of the 4D-OBCs in the groups are investigated, to find out what characterizes them, and if these characteristics can be interpreted as representative of a type of surface activity. We manually investigate if the 4D-OBCs in the same groups can be interpreted as part of a single type of surface activity. Furthermore, we label a validation dataset into different types of surface activity and investigate if these are found in the same groups.

- *To what extent is the unsupervised classification workflow transferable to another study area?*

  We optimize and apply the unsupervised classification methods on a dataset obtained at a sandy beach, and after optimization also apply it to the snow cover dataset. We investigate to what extent the groups of 4D-OBCs found for a snow cover dataset can be interpreted as single types of surface activity.

- *How do our methods of automated spatiotemporal extraction and subsequent classification of surface activity compare to existing methods of identifying and classifying surface activity?*

  We compare the final results of our methods to existing methods of extracting surface activity from point cloud time series, and compare (dis)advantages of the methods. Also, a comparison to other (un)supervised data classification methods is performed.

- *To what extent can the grouped 4D-OBCs be used for the analysis of surface activity characteristics and correlations between surface activities and natural drivers?*

  We interpret the obtained groups of 4D-OBCs as different types of surface activity and investigate possible relations between the occurrences of the 4D-OBCs in these groups, which are physically interpretable. Furthermore, these occurrences are compared to several natural drivers. This is done to find correlations that might indicate the validity of the surface activity interpretation, as well as indicating possible underlying processes of surface activities.

In answering these research questions, we will be able to identify how a dataset of 4D-OBCs can be classified, for the use of characterization of surface activities in different environmental settings. This will allow to automatically analyze large 4D point cloud time series, on the occurrence of different types of surface activity, and in this way enable efficient larger scale analysis of surface activity behavior and correlations, in turn, possibly enabling better management and mitigation strategies.

The thesis is structured as follows. Chapter 2 presents the theoretical background of the thesis, here surface activity is further defined, existing work on surface activity monitoring is presented, and several types of classification methods are considered. In Chapter 3 both the point cloud time series data, as well as the 4D-OBCs, used in this research, are discussed in detail. In this chapter, we also describe known types of surface activity present in the two study areas. The methods developed to classify 4D-OBC datasets are provided in Chapter 4. Here, we also present how these methods are optimized. Chapter 5 then shows the resulting characterized surface activities obtained through these methods. This chapter starts with providing the final results, after which the intermediate results are presented. The methods and results are discussed in Chapter 6. We compare our methods and results to existing work on surface activity monitoring and classification, and discuss the potential application of our methods under different scenarios. Chapter 7 then presents the conclusions we can draw from this research by answering the research questions. Here, recommendations for further research are also given.

2

Background

This chapter presents the knowledge and background information for this research, by defining surface activity (Section 2.1) and how surface activity has been monitored up until now (Section 2.2). We also present the advantages and disadvantages of various methods of (un)supervised classification, that could potentially be used for grouping 4D-OBCs to identify different types of surface activity (Section 2.3).

2.1. From geomorphological processes to surface activity

Geomorphological change processes shape the face of the Earth. These processes are different in the various geographical settings of the Earth. In mountainous areas, the geomorphological processes that can be identified are, e.g., landslides and resulting transport through debris flows triggered by hydrodynamic changes or earthquakes (Wieczorek 1996). If this mountainous area is covered in snow, avalanches may be triggered through complex interactions between the terrain, anthropogenic activity, snowpack, and meteorological conditions (Schweizer et al. 2003; Hao et al. 2018). In coastal areas, different morphological processes exist. For example, sandbars can be deposited underwater during storms and afterward migrate shoreward during calmer conditions due to wave working. When exposed to the air during low tide, wind could transport the sand and deposit it in the vegetated dune area (Houser 2009).

These different geomorphological processes found in the various environments result from different (intensities of) underlying forcing and materials. Even within one environment, variations in intensities of forcing also cause variations in geomorphological change processes. The relations between the forcing and the initiation of geomorphological processes are not yet well understood in many of these environments (Houser 2009; Walker et al. 2017; Hao et al. 2018). Understanding these relations and identifying the impact of the geomorphological processes is, however, essential, especially under the changing environmental conditions of the last and coming centuries, as the effects and occurrence of triggering conditions might increase (Gariano and Guzzetti 2016; Vousdoukas et al. 2020).

Identifying and understanding all the geomorphological change processes is challenging, as these processes span multitudes of spatial and temporal scales. Moreover, the different processes can add up to create larger-scale processes. For example, a coastline may retreat as a function of sea level rise and changes in other geophysical forcings on a decadal scale, but on a seasonal scale, the coastline may appear to show growth.

This research focuses on geomorphological processes constrained in time and space. We define these processes as surface activities, which refer to fixed instances of a geomorphological process instead of long-term geomorphological changes. We observe geomorphological processes using a near-continuous terrestrial laser scanning (TLS) setup. In this setup, data is acquired within a limited spatiotemporal span. The surface activities found in our dataset do not extend beyond a fixed spatiotemporal extent. In the particular frameworks studied in this research, geomorphological processes occurring over days to weeks can be observed, but a decadal retreat of the shoreline cannot be moni-

tored. Furthermore, the spatial extent of the observed surface activities is, at maximum, a fraction of a square kilometer.

Surface activity can thus be defined further as events where the morphology of a local surface changes only over a specific period. Different types of surface activity are then defined by differences in their spatial and temporal extent, i.e., the size and location of what is considered local and the length, magnitude, and shape of the surface change. These different types of surface activity relate to differences in their forcing. The intensity of the various forcings at a particular location differs from type to type, or the type of surface activity is defined by different forcings altogether. The known surface activity types found within the fixed spatiotemporal scale of observation at our study areas are discussed in Section 3.3.

2.2. Surface activity monitoring

Past and current work on monitoring surface activity is mainly based on pairwise comparison of topographic measurements. These measurements can be obtained manually through total stations or other leveling techniques (Delgado and Lloyd 2004), which is a time-intensive process. If information is needed on larger areas or over longer timespans, one has to interpolate the height measurements over the whole area or time span of interest. The relatively sparse amount of measurements that are taken through these methods then poses restrictions on the accuracy of the observations of height changes that occur as a result of the dynamic processes. Large-scale assessment, both spatially and temporally, is thus not possible.

2.2.1. Remote sensing methods for surface activity monitoring

Recent advances in laser scanning (LiDAR) and photogrammetry show the potential to overcome these problems. Here, 3D spatial representations (as point clouds) at different epochs are obtained through consecutive acquisitions. These can be acquired through field surveys with TLS devices, fixed equipment (e.g. permanent TLS), airborne acquisitions, or unoccupied aerial vehicles (UAVs). A more detailed description of how LiDAR data is acquired and processed is given in Section 3.1.

Lin et al. 2019 showed that a bitemporal comparison between two height models of a beach obtained through photogrammetric and LiDAR UAV scans could be applied for monitoring larger-scale beach erosion. Other applications of these techniques, as well as multitemporal change identification, in the fields of landslide and glacier studies, have also been studied, demonstrating that these methods can provide efficient as well as rapid monitoring and quantification of surface changes in dynamic natural environments (e.g., Eker et al. 2018; Pellicani et al. 2019; Zahs et al. 2019).

These methods of surface activity monitoring are well applicable if it is known when a geomorphological process has taken place or will take place. When the timing of processes is not known, processes occur superimposed on each other or on smaller time scales than the acquisition interval; these methods do not allow for the analysis of all the individual surface activities. These circumstances often occur on sandy beaches, where various surface processes show complex interactions on various spatial and temporal scales (Walker et al. 2017). Bi- or multitemporal surface measurements can be used for identifying the effects of surface activities that occur on a large temporal scale. However, the effects of smaller-scale surface activities are missed. In these environments, setups that provide permanent or near-continuous acquisition of surface measurements are better suited.

2.2.2. Near-continuous terrestrial laser scanning

Near-continuous TLS is a method that does show the potential of observing surface activities over a broader range of spatiotemporal scales, as it has the advantage of enabling near-continuous topographic observations over long periods of time, without manual interference (Eitel et al. 2016). The amount of data and information on surface dynamics that can be obtained in such a manner is vast. Anders et al. 2019 showed that using near-continuous TLS with shorter time intervals for volume change estimation on a sandy beach, results in significantly different volume estimations than when using longer time intervals between acquisitions. This difference is also observed by Williams et al. 2018: they found that varying observation frequencies hold drastic implications on the number of surface activity events and erosion rates obtained - with higher observation frequencies, the number of identified events increased notably. The dynamics of a system are thus better captured if information obtained over shorter

and larger time scales is incorporated into analyses. At the same time, essential interactions might be missed when only using larger time intervals. Especially for areas with complex morphodynamics, like coastal areas, the use of point cloud time series derived from near-continuous TLS thus offers good potential for deciphering the interaction between the different surface processes.

LiDAR time series methods

Methods that use the entire time series of point clouds obtained through near-continuous TLS to monitor surface activity are still rare. Kuschnerus et al. 2021a used the data of a point cloud time series to characterize different change patterns on a sandy beach. Here, the time series was clustered using k-means, agglomerative, and density-based clustering (DBSCAN). The results showed that these methods enabled the separation of different height change patterns over the entire time interval of the data; for example, areas with deposition followed by erosion could be separated from areas with only erosion. Using bi-temporal change analysis, these groups of change patterns would not have been separated. This indicates that using a full time series gives good possibilities for understanding the sequences of processes at one location. However, the issue with this full time series clustering is that the temporal extent of processes, occurring on a smaller scale than the time series input, are not automatically identified and separated, as the full time series is used as an input. Furthermore, the spatial extent of the various change patterns is not segmented, and one has to identify these manually or through other methods. These factors imply that particular instances of surface activities are not extracted using these clustering methods.

4D objects-by-change

The use of object segmentation in a near-continuous TLS time series brings the possibility to separate individual spatiotemporal processes representing surface activities, independent of the length of the time series analyzed. Anders et al. 2021 proposed a method to segment individual erosion and deposition instances in a near-continuous TLS time series into spatiotemporal segments called 4D objects-by-change (4D-OBC). One 4D-OBC thus shows the spatial and temporal extent of an erosion or deposition surface activity event as a segment of points with a corresponding time series of height change for each point. A detailed description of this method is provided in Section 3.2.

It was found that using this method on a point cloud time series acquired on the sandy beach of Kijkduin, around 95% of the separate erosion and sedimentation events occurring on a sandy beach could be identified automatically, and 84% of the output of the algorithm were indeed surface activity instances. The method was also tested on its ability to quantify the volume of surface activity in an alpine setting, where snow avalanches and anthropogenic snow transport took place (Anders et al. 2022). It was found that the method could separate spatially proximal erosion and deposition surface activities which would be extracted as one using alternative methods (e.g. bi-temporal change detection). Furthermore, surface activities proximal in time, which would, with alternative methods, be accumulated into one process, could be separated. The method gives more accurate estimations of change volumes with data obtained over longer periods than bi-temporal change analysis. The fact that this method works fully automatically without the need for a priori knowledge of the timing of processes means that one can also potentially identify unknown processes, giving it a considerable advantage for use in complex morphodynamic situations.

The 4D-OBC method thus shows good possibilities for large-scale identification and understanding of the complex interactions of processes taking place on the surface of the Earth, but what still needs to be considered is the automated separation and identification of different types of surface activity captured by the 4D-OBCs. Erosion and deposition can occur due to various reasons. For example, anthropogenic displacement of sand might occur on a sandy beach, while other types of erosion might occur due to wave interaction with intertidal sandbars over longer time spans. Identifying and classifying these processes can offer valuable knowledge on the differences in magnitude and occurrence of the various processes. Furthermore, it can, in the end, be used to relate different surface activities spatiotemporally while also identifying the impact of one type of surface activity relative to others. In this way, the complex interactions of surface processes could be unraveled, giving valuable knowledge for the physical modeling of geomorphological change. In turn, enabling more efficient application of mitigation and management strategies (Walker et al. 2017).

2.3. Unsupervised classification methods

This research aims to obtain a method through which a set of 4D-OBCs can be separated into groups of surface activity. This can be done through (un)supervised classification. Supervised classification methods rely on training datasets of labeled data. For each sample in this training dataset, it should be known to which class or value it belongs. A model is then trained such that for an input sample, with its representative feature vector, the model outputs the correct label or value (Sen et al. 2020). In our case of classifying 4D-OBCs, we want to allow for the identification and characterization of possibly unknown surface activities, and it is not known thoroughly which and how the 4D-OBCs represent surface activities. An extensive labeled training dataset can thus not be obtained. Therefore, unsupervised classification methods are needed. Unsupervised classification or clustering methods allow for separating data into different groups independent of any predefined labels. This section introduces several unsupervised classification methods and their (dis)advantages.

One of the most widely applied clustering methods is the k-means algorithm. The k-means clustering algorithm relies on a predefined number of clusters. The centroids of these clusters are initialized in some way, most often by selecting random samples from the dataset. All data samples are added; for each, the closest cluster centroid is computed, and the centroid of each cluster is moved toward the average of the data samples in the cluster. This is done until the cluster centers do not move after adding all the samples. This algorithm is relatively fast for low-dimensional datasets and easy to understand. The disadvantages, however, are that it is susceptible to outliers in the data, one has to know or in some way obtain the number of clusters beforehand, and only linear boundaries between clusters are obtained (Saxena et al. 2017).

A clustering method in which no number of clusters has to be specified beforehand is hierarchical clustering. Two main algorithms exist, which are agglomerative and divisive hierarchical clustering. These rely on similar concepts but work in different directions. Here we only consider agglomerative hierarchical clustering, which works as follows: every data sample is assigned to a separate cluster. The feature distances between these clusters are computed, and the two closest clusters are merged into one cluster. One then computes the new cluster distances and merges the now nearest clusters. This is done until either a desired number of clusters is obtained or a maximum feature distance threshold is reached, i.e., all clusters are further apart than a certain feature distance. The distance between clusters (i.e., linkage criteria) can be computed in various ways. The complete linkage criteria use the furthest or closest samples from each cluster to quantify the feature distance. In contrast, the average linkage criteria use the average distance between all the samples of one cluster to the others. The latter one is less sensitive to outliers (Xu and Wunsch 2005). Overall, the hierarchical clustering algorithm shows some advantages over k-means in that it is unnecessary to specify the number of clusters beforehand. One can visualize and inspect the hierarchical structure of the tree or compute some performance metric at different levels in the tree to decide on the distance threshold to be used. However, a disadvantage is that it is susceptible to outliers, as these will have a considerable feature distance and, thus, only at a large distance threshold be merged. A second disadvantage is that once a sample is assigned to a particular cluster, it will not be removed later in the training run. Furthermore, due to its time complexity, it is not very useful for large datasets (Xu and Wunsch 2005).

A clustering algorithm more useful for larger datasets and higher dimensional data is the DBSCAN algorithm (Ester et al. 1996). This algorithm defines clusters based on density. One initializes a set of core points, which contain more than a certain number of neighbors within a set feature distance. All the neighboring core points are then regarded as one cluster, and non-core points within a certain feature distance are added to the cluster until no more samples are found within the defined feature distance. One thus obtains a number of clusters and a set of samples representing noise, which are not considered a cluster. The resulting clusters are, therefore, less sensitive to noise (Ester et al. 1996), which is an advantage of the DBSCAN algorithm. However, one has to set a number of neighbors and a distance threshold parameter. Furthermore, samples of underrepresented classes in the data might be regarded as noise, as most outlying samples will not be clustered. Also, by setting one threshold of distance, the various groups of samples must show a similar distance between the incorporated points, while the distance between clusters should be larger than this threshold. If a wider feature distribution defines one group, it will be regarded as noise.

Another clustering algorithm less sensitive to outliers is the Self-organizing Map (SOM; Kohonen 1990, Clark et al. 2020). This neural network maps and clusters high-dimensional data onto an n-dimensional grid or lattice. Each grid point contains an initialized weight vector with an equal length to the feature vectors of the data samples. To obtain the clusters, each sample is iteratively matched to the closest weight vector of the n-dimensional grid. After each mapping, this weight vector and its neighbors are adapted toward the feature vector of the mapped sample. After adding all samples over several cycles, one thus obtains weight vectors representing the centroid of a cluster, and all the samples are clustered at the closest of these weight vectors. A more detailed description of the algorithm can be found in Section 4.3. The advantage of this method is that one does not have to predefine a specific number of clusters. Furthermore, neighboring centroids are also adapted when adding samples, enabling the preservation of topological relations of higher dimensional data space, which allows for easier identification of similar clusters. This also means that the algorithm is less influenced by outliers, but in contrast to the DBSCAN algorithm, rare samples are still represented in the SOM clusters. Furthermore, the SOM allows for non-linear separation between clusters (Corne et al. 1999).

Following the exploration of several clustering algorithms, we choose to use a SOM algorithm accompanied by a hierarchical clustering algorithm to cluster the 4D-OBCs extracted from a point cloud time series. This allows for the possible detection of rare surface activities and the detection and characterization of different levels of surface activity clusters present in the two environments studied in this research. In Section 4.3.3 the reasoning behind this choice is further elaborated on.

3

Data and Study Areas

This chapter presents the data and study areas analyzed in this research. Section 3.1 describes how a point cloud time series is acquired, from which the spatiotemporal segments to be classified in this research are extracted. Section 3.2 gives details on the method to extract 4D-OBCs from these point cloud time series. Here, we also describe what the final 4D-OBC data product is. Section 3.3 presents the two study areas over which these point cloud time series and spatiotemporal segments are acquired. Here, details on the specific instrument are given. We identify potential sources of problems in the dataset for each study area and determine which types of surface activity are known to occur in these areas. In Section 3.4, meteorological and hydrological data is presented, which is used to find correlations and drivers between the types of surface activity and geophysical forcings for the sandy beach study area.

3.1. Point cloud acquisition

This research analyzes spatiotemporal segments extracted from two point cloud time series. These point cloud time series are obtained using a near-continuous TLS setup. In such a setup, a TLS device is placed in a fixed position over a period of time. At set intervals, the TLS device scans the area, and a point cloud is obtained. This point cloud serves as a 3D representation of the site. Every point in this point cloud represents an x, y, and z location where the laser pulse, emitted by the TLS device, was reflected after interaction with an object or surface material. This location is determined through the Time-of-Flight principle. The distance from the TLS device to the object is defined as a function of the speed of light and the time between emitting and receiving the laser pulse:

$$Distance = \frac{c \cdot t}{2}$$

(3.1)

where c is the speed of light, and t is the time between emitting and receiving the laser pulse. The 3D location of the reflecting surface can then be obtained by obtaining this distance, combined with knowledge on the angle of emission of the pulse (Beraldin et al. 2010). This angle of emission is measured with respect to a reference system centered in the TLS device. By emitting a series of pulses, the point cloud of the area surrounding the TLS device at one epoch in time is obtained, referenced against this TLS device-centered internal coordinate system (Lichti and Skaloud 2010).

A point cloud time series is obtained by repeating these measurements with a certain temporal interval. The TLS device has to be fixed rigidly to preserve the location of the internal coordinate system as much as possible. This enables the direct comparison between consecutive point clouds. If the location of the TLS device changes, the cloud-to-cloud comparison shows changes not due to changes in the scanned surface but due to the movement of the TLS device. The rigid placement of the TLS device should mitigate this problem. It is, however, found that even though the TLS device is fixed, there is still movement which results in significant changes in the center of the internal coordinate system (Kuschnerus et al. 2021b; Voordendag et al. 2022). To this end, the point clouds must be transformed to refer to a single coordinate reference system. This is done through a rigid body transformation.

In the case of our near-continuous TLS setup, every point cloud is aligned and referenced against the point cloud of the first epoch. Details on the preprocessing methods used for the alignment and referencing are found in Anders et al. 2021. After these preprocessing steps, one obtains a point cloud time series, in which the between point cloud changes due to the nature of the instruments is minimized, and changes above a certain distance are expected to be the result of actual surface change.

3.2. 4D objects-by-change

In this research, spatiotemporal segments extracted from point cloud time series are analyzed and characterized. These are automatically extracted using the 4D-OBC method presented by Anders et al. 2021. The segments represent the temporal and spatial extent of a specific surface activity, e.g., the build-up and consecutive destruction of an intertidal sandbar.

The 4D-OBCs are obtained with the following six steps:

1. Compute space-time array of height change with respect to the first epoch at fixed locations of a regular grid
2. Smooth the height change time series
3. Identify seed candidates representing temporal height change forms
4. Sort seed candidates
5. Grow spatial region from seed candidates
6. Discard spatiotemporal segments not representing surface activity

In each of these steps, slight variations in configurations are used for the two study areas (presented in Section 3.3). The specific designs per study area are shown in Table 3.1.

The first step is obtaining a space-time array of height change by computing the distance of each point cloud to the first point cloud, and representing these on an evenly spaced grid (1). This is done using the Multiscale Model to Model Cloud Comparison (M3C2) method (Lague et al. 2013). For each grid point, a time series of height change relative to the height at the starting epoch is created (Figure 3.1). These time series are smoothed using a temporal window to reduce the effects of errors due to changes in atmospheric conditions (2, Anders et al. 2019). The value is set to the median value within a temporal window.
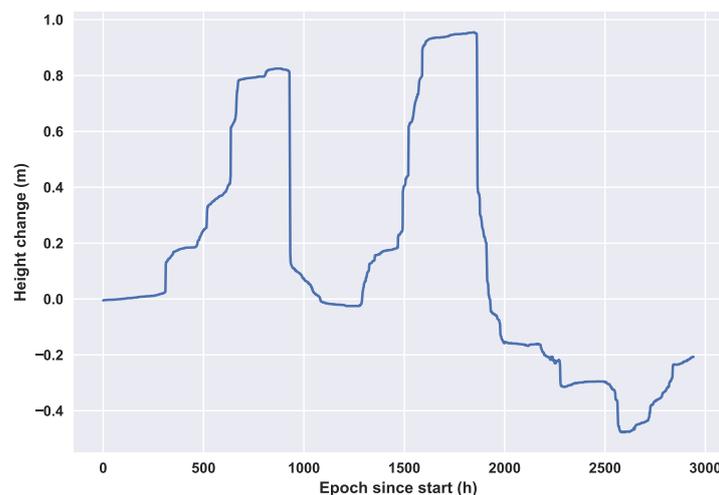

Figure 3.1: Full time series of height change for one grid point of the interpolated point cloud time series.

Hereafter, the temporal segments representing surface activity are identified, which serve as seed candidates for the spatial region growing (3). This is done in two ways, depending on the study area.

For the sandy beach study area, points of change are identified using a sliding temporal window of 24 h (Anders et al. 2021). If there is a discrepancy in the median value between the first 12 h and the second 12 h, the point is detected as a change point. From these change points, temporal segments are extended until the change with respect to the starting point becomes zero again. Here, one temporal segment thus spans the deposition/erosion at a grid point and the recovery to the initial values of the temporal segment after subsequent erosion/deposition (Figure 3.2A). Temporal segments that do not recover to initial values before the end of the observation period are not added.

For the Alpine snow cover area, temporal segments are identified using piecewise linear regression (Anders et al. 2022). A line is fitted to groups of epochs showing similar gradients in height change utilizing least squares fitting. If the difference in height between the start and end epoch of the fitted line is larger than the minimum detectable change of 5 cm, the temporal segment is used for further analysis. These temporal segments thus show only deposition or erosion, without any recovery to the original state, in contrast to the 4D-OBCs of the sandy beach (Figure 3.2B).
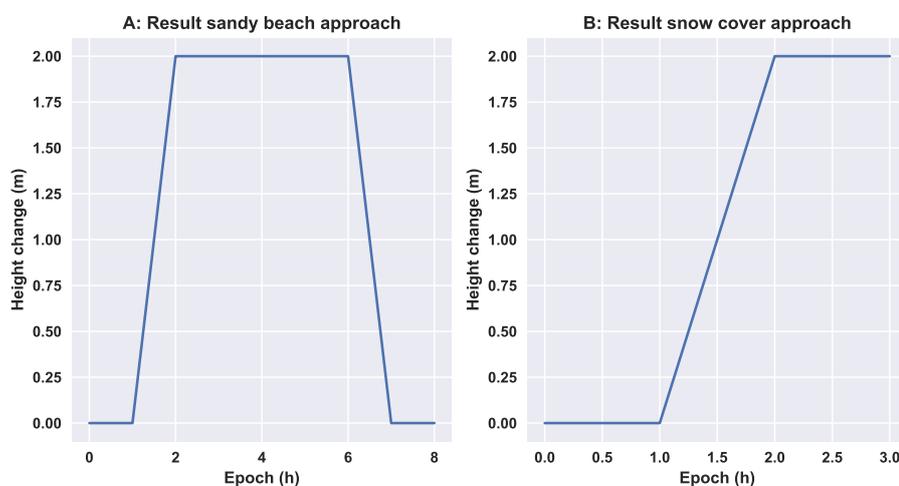

Figure 3.2: The temporal segment resulting from the sandy beach seed detection method (A) and the snow cover seed detection method (B).

For both use cases, these temporal segments serve as seed candidates for a spatial surface activity segment. Before the region growing of this spatial segment, the seed candidates are sorted on their relative importance (4). The seeds are ranked based on their neighborhood similarity determined as the mean normalized Dynamic Time Warping (DTW, Berndt and Clifford 1994) distance between the seed and its eight neighbors. A second ranking is done based on the value of the integral of the absolute height change over the temporal segment of the seed, with higher values ranked first.

From these seed candidates, a region is grown spatially by computing the similarity based on the DTW distance between the seed point's time series and the spatially neighboring points (5). If the similarity between a point within the segment is larger than an adaptive threshold, the neighboring point is added to the segment. This is done until no more points are neighboring with a similarity lower than this threshold. This threshold value is determined for each region growing iteration, testing a range of thresholds and automatically selecting the threshold, before a large increase in segment size occurs. Region growing starts with the first of the ranked seed candidates. Seed candidates are omitted if a seed candidate is already incorporated within the spatiotemporal segment of a previously grown segment.

After the region growing step, segments are discarded that are either too small or too large (6). These thresholds are chosen depending on the study area (Table 3.1). Furthermore, segments that show considerable time series heterogeneity are discarded.

Table 3.1: 4D-OBC configuration and dataset characteristics for the snow cover and sandy beach dataset

Dataset	Sandy beach	Snow cover
Number of epochs	2,942	125
Grid spacing	0.5 m	0.2 m
Smoothing window length	One week	Three hours
Minimum detectable change	0.05 m	0.05 m
Minimum segment size	2.25 m^2	16 m^2
Maximum segment size	-	4,000 m^2
Maximum duration	8 weeks	-

The resulting segments are the 4D-OBCs. Each of these is defined by their three-dimensional geographic space, i.e., the locations and height of all the points incorporated in the segment, and their one-dimensional temporal space, i.e., the change in the height dimension over time. Figure 3.3 shows an example of one of these 4D-OBCs extracted on the sandy beach study area.
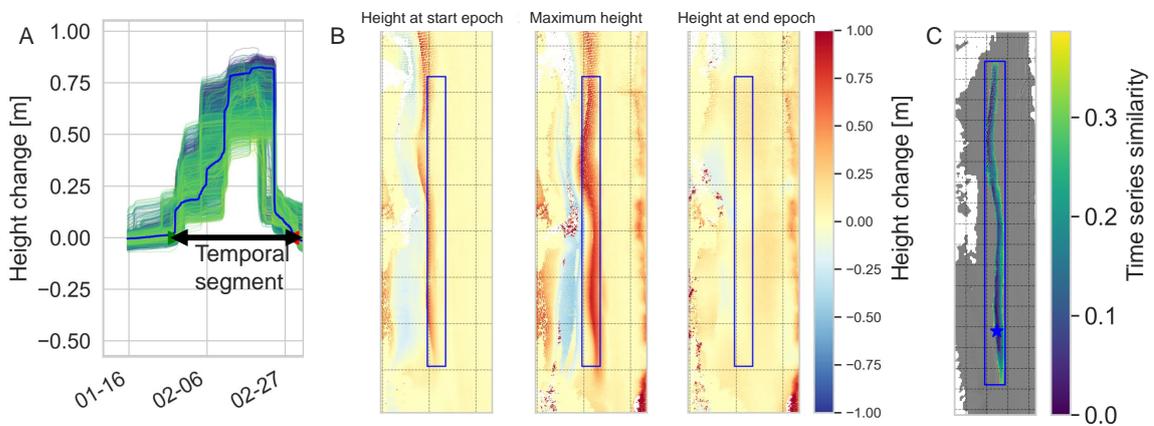

Figure 3.3: Example of a 4D object-by-change (4D-OBC). A) the time series of height change of the points inside the segment. Every line represents the time series in one of the grid points. The colors show the similarity. The blue line is the time series of the seed point. The temporal segment detected as a seed is captured within the arrow. B) Height change at three epochs within the temporal segment of the 4D-OBC. C) The spatial outline of all the points within the 4D-OBC. The blue box is the bounding box of the segment. The star is the location of the seed.

3.3. Study areas

The two study areas of interest in this research have different near-continuous TLS setups and types of surface activity. The following two paragraphs describe the details of both setups and the surface activities that can be found within the spatiotemporal boundaries set by the 4D-OBCs extracted in these areas.

3.3.1. Sandy beach

The main focus of this research is on characterizing the surface activity of a sandy beach on the North Sea located in Kijkduin, The Netherlands (52°04'14" N, 4°13'10" E). This particular beach is monitored using a near-continuous TLS setup with a Riegl VZ-2000 scanner fixed at 30 meters height on a hotel building overlooking the beach (Figure 3.4).

The setup is part of the CoastScan project (Vos et al. 2017). The scanner acquired a point cloud at an hourly interval between November 2016 and May 2017. The 4D-OBCs analyzed in this research are extracted from a subset of this point cloud time series, acquired from 2017-01-15 to 2017-05-26 and containing 2,942 epochs of point cloud acquisitions (Vos et al. 2022). The point clouds are registered against the first epoch, using the ICP method, on manually identified stable surfaces between the dunes and the beaches. The resulting minimum detectable change is 0.05 m (Anders et al. 2019).

Figure 3.4: Photo of the scanner setup mounted on the hotel overlooking the Kijkduin beach. Image courtesy of S. Vos.

The point clouds have point densities between 2 and 20 $points/m^2$, but this point density varies from scan to scan due to surface reflectance properties, tidal level, and meteorological conditions (Vos et al. 2022). The intertidal part of the scan area is submerged by water as a function of the tides. In a submerged situation, the laser pulses emitted by the scanner in this area are not reflected but absorbed by the water (Höfle et al. 2009). This results in gaps in the point cloud time series, which also influence the appearance of the 4D-OBCs.

A spatial subset is used for the extraction of 4D-OBCs, which covers the area shown in Figure 3.5. This area is approximately 300 m by 600 m and spans the dunes, backshore, and intertidal zone. In this area, the distances of the space-time array are computed at a regular grid with a spacing of 0.5 m. The minimum limit on the size of the 4D-OBCs is set to 2.25 m^2, and the maximum duration is 8 weeks. This maximum duration is chosen, as this research focuses on temporary surface changes (Anders et al. 2021). Surface changes of longer durations are easier to decipher by only using bi-temporal methods, whereas this near-continuous TLS setup enables the identification of temporary surface changes. The 4D-OBCs with the shortest duration have a duration of 24 h. We use the derived dataset containing 2,021 4D-OBCs, extracted from the 4D point cloud dataset of Kijkduin, as input for our methods. This dataset is referred to as the beach dataset.

The sandy beach area under consideration is a very dynamic environment where several types of surface activities occur. These surface activities are forced by hydrodynamic (swash, tides), meteorologic (wind, precipitation), and anthropogenic processes (bulldozer work) that interact to displace sand over various spatial and temporal scales (Walker et al. 2017). These surface activities all act on a hierarchical scale, i.e., surface activities on a smaller spatiotemporal scale aggregate into surface activities acting on larger scales (Cowell et al. 2003). The boundary temporal scale the 4D-OBCs represent lies between 24 h and 8 weeks (1344 h). The boundary spatial scale lies between 2.25 m^2 and the full extent of the area of interest, 180,000 m^2. These boundaries indicate that surface activities relating to the micro-scale (Cowell et al. 2003) can be identified. Inside this spatial scale surface activities act altering the state of particular elements of the morphological unit of the area.
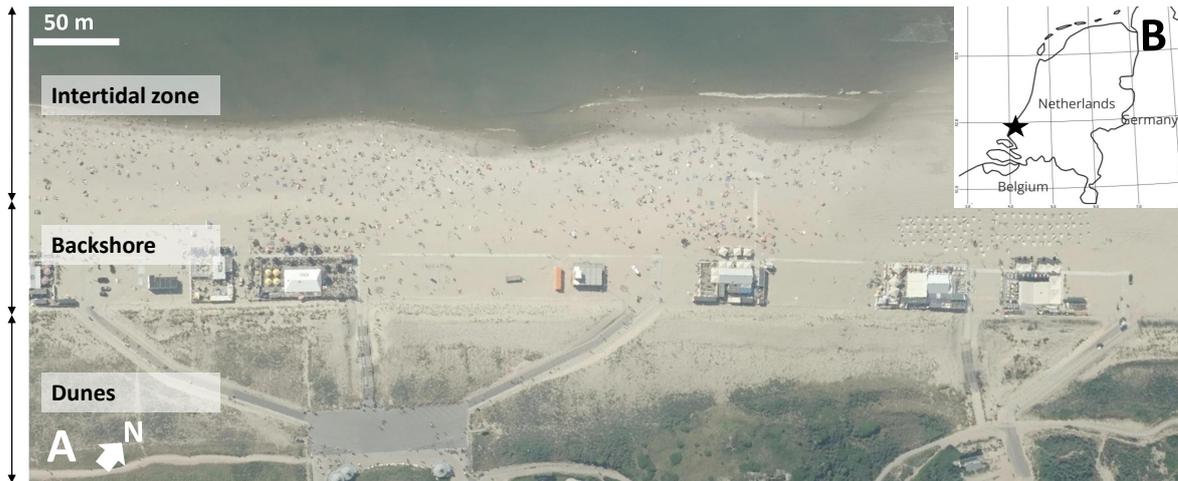
Figure 3.5: Study area (star in B). The point clouds sample the area visualized in A. Data: Aerial imagery ©pdok.nl 2017, borders ©Natural Earth 2022
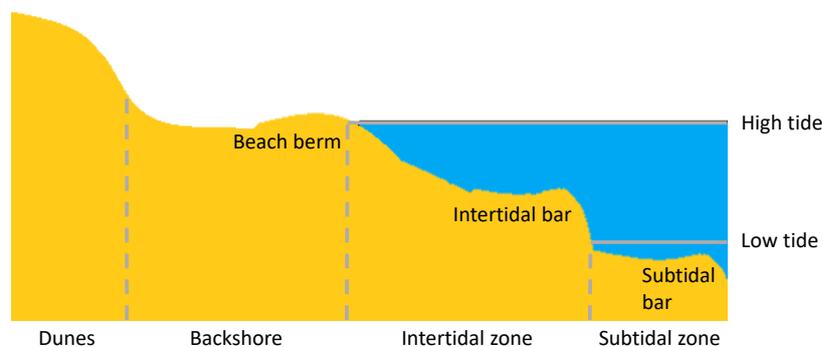

Figure 3.6: Barrier-beach-dune complex, with the different elements indicated, as well as two types of dynamic morphological forms, the beach berm and tidal bars. Adapted from Ruessink and Ranashinge 2014.

The morphological unit comprised by the dataset can be defined as the barrier-beach-dune complex (Cowell et al. 2003). The elements inside this unit are the dune area, backshore, beach berm, and intertidal zone (Figure 3.6). The surface activities to be identified in the 4D-OBCs thus act inside these elements. The intertidal zone is the region between the mean low and mean high tide water levels, and here hydrodynamic and meteorological processes can interact. The beach berm is an accumulation of sand that divides the backshore and this intertidal zone. The backshore is the area of the beach where hydrodynamic processes influence the morphology only during extreme conditions. Meteorological processes, consequently, largely dominate the surface activity here. The boundary between the backshore and dune area is the location where permanent vegetation starts (Ruessink and Ranashinge 2014). This dune area is dominated by meteorological processes. Each of these morphological elements contains surface activities specific and non-specific to those locations. The following sections describe several surface activities known to exist in these morphological elements, which could thus be found when grouping and characterizing the 4D-OBCs. Apart from these surface activities, more types of surface activities could be represented by the 4D-OBCs. Furthermore, several different low-level separations into surface activities might also be found within each of these surface activities, e.g., different types of intertidal bars.

Aeolian erosion, transport, and deposition
Aeolian erosion and deposition can occur in any of the morphological elements of the beach that lie above the mean low tide boundary. The wind speed and the wetness of the sand constrain the magnitude of these surface activities. With lower wind speed, the entrainment of sand by air is smaller,

and rates of erosion and subsequent transport are thus lower (Dong et al. 2003). When the sand is wet, the threshold wind speed for entrainment is larger (Davidson-Arnott et al. 2008). These factors imply several potential characteristics of surface activities. During high wind speed and under dry conditions, erosion surface activities that show sharp decreases in height might exist in all parts of the beach. Under lower wind speeds and wetter conditions, these erosion events might be more gradual or not appear. Deposition and recovery at erosion locations will likely be gradual, except when a site is (temporarily) shielded from the wind. The erosional events will likely be more gradual in the intertidal zone because the intertidal zone can remain wet between the high and low tides. This does, however, depend on the initiation processes of sand transport. If the process is initiated by saltation, i.e., the impact of other sand grains, the wetness shows a lower correlation with transport rates. Consequently, aeolian erosion in the backshore and dune area under dry conditions, with wind directed off-shore, could initiate erosion of the sand in the intertidal zone, even when the sand is still wet at low tides.

Aeolian surface activities are thus likely not easily distinguished from other surface activities by their magnitude, location, or size because local and temporal variations in wind speed and wetness cause variations in these variables. But within the broader aeolian surface activity group, sorting and grouping based on these characteristics could be possible, and correlation with natural drivers could give valuable insights. Furthermore, in the vegetated dune area, the wind force is lower (Mayaud and Webb 2017), and large foredune accumulation events can exist (Hesp et al. 2005). The resulting characteristics of the surface activities in this area might thus also be different.

Beach berm deposition and erosion
The beach berm is a temporary element of the beach. During storms, the berm might be eroded, while during prolonged periods of calm weather, the berm can be deposited (Ruessink and Ranashinge 2014). This beach berm deposition and erosion is a surface activity constrained to a specific location dividing the intertidal zone and the backshore. The spatial extent of a beach berm deposition and erosion surface activity event is not constrained in the along-shore direction. Still, in the cross-shore direction, it will not extend far beyond the boundaries mentioned before. The beach berm is deposited through onshore sediment transport via waves in the swash zone but can also be anthropogenically enhanced as a means of coastal management (Zhu et al. 2022). The magnitude of the beach berm height can range up to several meters. There are no known anthropogenic nourishments to the beach berm during data acquisition, and it is unlikely that anthropogenic nourishment of the beach berm has occurred, as this beach is managed by the mega nourishment of the "sand engine" (Stive et al. 2013).

Intertidal sandbar deposition, transportation, and erosion
Intertidal sandbars are local accumulations of sand deposited during mild wave conditions (Vos et al. 2020), and in particular during post-storm conditions (Ruessink and Ranashinge 2014). The intertidal bars, in most cases, are elongated and positioned parallel to the coast. At the Kijkduin beach, seasonal intertidal bars exist between the winter and summer, reaching a height of about one meter (Vos et al. 2020). These intertidal bars are deposited during the recovery processes after the winter, in which sand is deposited at the beach to create the summer profile. The intertidal bars are deposited during calm conditions and migrate shoreward during moderate to storm conditions (Robin et al. 2009). The destruction of intertidal bars is thought to be related to heavier weather conditions and relatively high water levels (Cohn et al. 2017; Vos et al. 2020). The destruction, thus, is more abrupt than the initial growth of the sandbar. In the same point cloud time series as used in our research, Vos et al. 2020 found that after the formation of an intertidal bar between the 21st and 26th of January 2017, during calm and low water level conditions, the bar migrated shoreward, and grew taller, after which it was destructed rapidly a month later, before the 24th of February, during high water and storm conditions, in between the acquisition of two low tide scans they examined. In this case, the destruction process was thus thought to be due to hydrodynamic processes instead of aeolian processes. Furthermore, they suggest that the dominant transport after destruction, in this case, was off-shore. However, other research implies that after the destruction of intertidal bars, sand deposited in the intertidal zone and beach berm can serve as a source for aeolian transport towards the backshore and dune area (Houser 2009).

This intertidal bar complex, in many cases, contains a bar trough on the landward side of the intertidal bar (Anthony et al. 2009). This trough generally has a lower depth relative to the bar height. It is defined

by erosion during the formation of the intertidal bar and subsequent onshore migration and restoration due to deposition.

Anthropogenic activities
Several known anthropogenic activities exist on the Kijkduin beach. Local sand accumulations are formed, and others are removed through bulldozers to clear parts of the area (Kuschnerus et al. 2021a). This is known to occur in the far backshore and around the entrance paths in the dune area. Local natural accretions are removed in the latter area to clear the path. These human activities show a very steep height change and are small in size. Furthermore, many anthropogenic bulldozer activities might occur around March to clear areas to build pavilions, as these open at that time[1].

3.3.2. Alpine snow cover

The second study area is a snow-covered Alpine area. This area is located at the Schneeferner glacier near the Zugspitze in Germany (47°24'59"N, 10°58'46"E, Figure 3.7). It was monitored using a near-continuous TLS setup with a Riegl VZ-2000i scanner positioned at the environmental research station *Schneefernerhaus*, overlooking a ski resort. The device scanned at hourly intervals for five days between 2018-04-17 and 2018-04-22. The final dataset consists of 125 epochs of point cloud acquisitions (Anders et al. 2022). The point clouds are registered against the scan of epoch 2018-04-18 using the ICP method. After registering, the space-time array of height change is computed at a regular grid with a spacing of 0.2 m.
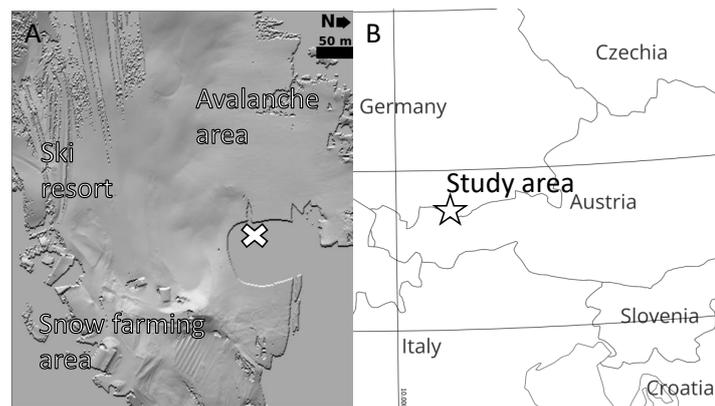
Figure 3.7: A) Hillshade derived from a point cloud obtained at the study area. The cross indicates the location of the scanner. The areas where known types of surface activity occur, adapted from Anders et al. 2022, are indicated. B) Location of the study area, borders ©Natural Earth 2022

The scan position changes between night and day, so the scans from the two positions are coarsely aligned by manually picking points of a stable reference area. The final alignment accuracy is, on average, 2.5 cm. The minimum detectable change for the 4D-OBC method is thus set at 5 cm. The night scans have less spatial coverage than the day scans, resulting in gaps in the point cloud time series, which also influence the appearance of the 4D-OBCs.

From the point cloud time series, 813 4D-OBCs are extracted. We use this dataset as input for our methods. The dataset is further referred to as the snow cover dataset.

Surface activities
In this snow-covered study area, several types of surface activity have visually been observed during the acquisition of the point cloud time series.

First, snow avalanches have been identified. These surface activities occur over a short period, within minutes, and mainly occur on slopes above 30 degrees (Schweizer et al. 2003). The activities involve

---

[1]https://strand-denhaag.nl/overzicht/strandtenten

the transportation of snow from the higher part of the slope towards a lower part and are thus defined by an erosion and deposition zone.

Second, snowfarming has been observed. Here, bulldozers transport snow from one part of the region to another to be used for the preparation of the ski pistes. The transport events have a duration of up to 6 h (Anders et al. 2022).

Third, ablation and compaction of snow due to warm temperatures might have occurred. During the acquisition period, the temperatures were relatively high, combined with extensive solar insolation (Schardt 2018). This causes the snow cover to melt and liquid water to run off or evaporate. The snow cover height then decreases, which results in erosional surface activities that might occur over larger areas but could also be locally bound, through local temperature variations.

A fourth possible surface activity in the Alpine region is snowfall. On 2018-04-17, 6 cm of fresh snowfall was recorded (CDC - Climate Data Center 2018). It is unknown if this fell before the acquisition started at 14:00 on the 17th of April.

A fifth process resulting in surface activity frequently occurring in snow-covered areas is snow transport through wind. During the acquisition, conditions were relatively warm (CDC - Climate Data Center 2018). It is, therefore, not expected that wind-driven transport has occurred.

3.4. Meteorological and hydrodynamic data

Additional datasets of natural drivers influencing the initiation of surface activities on the sandy beach of Kijkduin are investigated to correlate to the occurrences of the obtained groups of surface activities.

The meteorological data is obtained at Hoek van Holland (51°59'31.2"N, 4°07'19.2"E), the closest professional weather station to the Kijkduin beach (Vos et al. 2022). It is provided by the Dutch meteorological institute (KNMI). The data used in this research are the hourly average wind speed and hourly precipitation. For some analysis, we average this data over a daily period.

The hydrodynamic data is obtained at two locations by Rijkswaterstaat. The water level, relating to the tides, is obtained at Scheveningen (52°06'00.3"N, 4°15'32.4"E), measured at a 10-minute interval. The wave height data is measured in front of the coast at IJmuiden (52°32'57.2"N, 4°03'25.1"E), also at a 10-minute interval.

4

Methods

This chapter describes the methods researched and developed to characterize different types of surface activity from the 4D objects-by-change (4D-OBCs) extracted from a point cloud time series. First, the 4D-OBC dataset is split into erosion and deposition subsets (Section 4.1). Second, initial features are extracted, from which a selection of features is used for further development (Section 4.2). Third, the Self-organizing Map (SOM) algorithm is used to provide a low-level sorting and detailed grouping of the 4D-OBCs into surface activities (Section 4.3). The parameters and final feature selection used as input for this algorithm are partially chosen based on literature (Section 4.3.4) and partially optimized by assessing the performance under different configurations (Section 4.5). The groups of 4D-OBCs obtained with the optimized SOM algorithm are evaluated in four ways, presented in subsection 4.3.5. As a final step, the grouped 4D-OBCs of the optimized SOM algorithm are used as an input for a hierarchical clustering algorithm to identify and characterize broader groups of surface activities (Section 4.4). The performance of this clustering algorithm is assessed in terms of the physical interpretability of the clusters as surface activities and the possibility of correlating the clusters with natural drivers.

The final developed methods used to characterize and visualize surface activities from the 4D-OBCs are summarized in four steps (Figure 4.1):

1. Split the 4D-OBC dataset into an erosion and deposition 4D-OBCs dataset

2. Extract spatial and temporal features from the 4D-OBCs to be used in the unsupervised classification

3. For both the erosion and deposition dataset, train a SOM and match all the 4D-OBCs to this SOM to explore the full dataset and organize it into characteristic feature vectors, representing detailed groups of surface activities

4. Use hierarchical clustering to cluster these detailed groups of surface activities and obtain different levels of clustering

4.1. 4D objects-by-change Erosion-Deposition dataset split

The first step in obtaining grouped 4D-OBCs is splitting the dataset of 4D-OBCs into 4D-OBCs representing erosion and deposition surface activities (Figure 4.1). Dividing the dataset in this first stage already enables a grouping of 4D-OBCs into two types of surface activity. Thus, it ensures that the subsequent unsupervised classification methods do not have to cluster these two groups but instead separate the 4D-OBCs into more detailed low-level groups.

The method of extracting 4D-OBCs detects either a negative or a positive change point in the space-time array extracted from the point cloud time series, resulting in two types of 4D-OBCs—one with positive and one with negative height change.

The split into an erosion and deposition dataset is accomplished using the seed time series of each of the 4D-OBCs. First, all seed time series are translated to start at zero by subtracting the height at
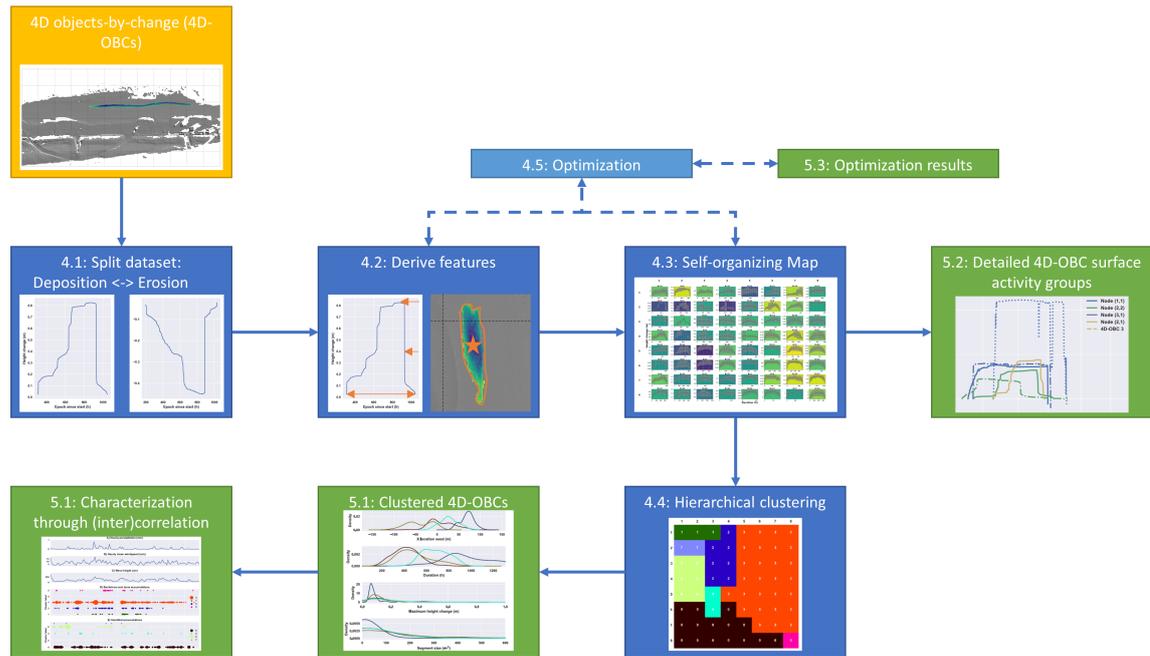
Figure 4.1: Workflow with the four main steps to obtain grouped 4D-OBCs in dark blue and the optimization method in light blue. The output obtained at different levels in the workflow is colored green. The numbers represent the section in which the method or result is presented.

epoch zero from all the heights of a seed time series. Consequently, all the measurements are relative to the first epoch (see Figure 4.2). Then, for each seed time series, the heights at all the epochs are summed, obtaining the area under the curve. If the sum of heights is negative, the 4D-OBC is added to the erosion dataset, and if the sum of heights is positive, the 4D-OBC is added to the deposition dataset. For the beach dataset, we end up with 1205 deposition 4D-OBCs and 816 erosion 4D-OBCs. 278 deposition 4D-OBCs and 535 erosion 4D-OBCs are obtained for the snow cover dataset. In the rest of the thesis, these datasets are referred to as the deposition and erosion (beach and snow cover) datasets.

4.2. Feature processing

Thirty-six features are initially extracted from the 4D-OBCs in the erosion and deposition dataset (Table 4.1), describing spatial and temporal attributes of the 4D-OBCs. These features can be grouped into spatial (3), temporal (32), and spatiotemporal (1) features. In the following paragraphs, details on the features and the extraction methods are given. The final feature selection used for the optimized methods is also given. This selection is obtained through the optimization methods (4.5).

4.2.1. Feature extraction

Temporal features

Thirty-two temporal features are derived, encompassing the largest part of the features. All these features are based on the seed time series of a 4D-OBC. The shape (i.e., the temporal evolution of height change) and magnitude of the seed time series show substantial variability among 4D-OBCs and are therefore expected to be of value for the characterization of surface activities. The seed time series used for feature extraction start at a height change of zero, following the translation presented in Section 4.1. Furthermore, the absolute height change values of the time series are used, as the dataset is already split into erosion and deposition 4D-OBCs, and the sign of the height change consequently does not hold any value for clustering anymore.
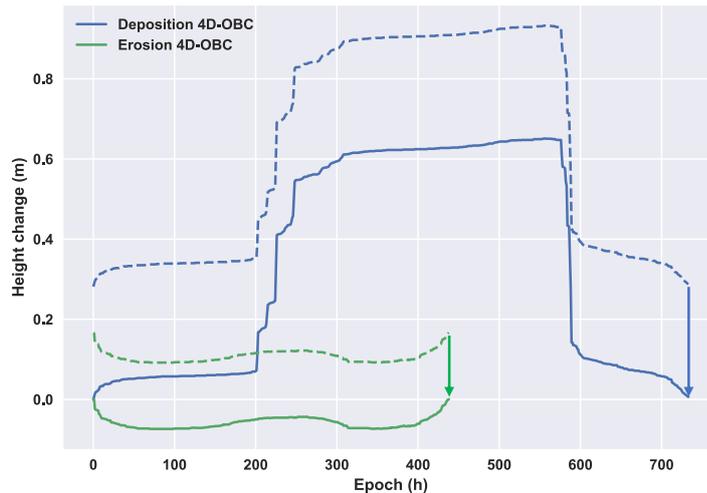
Figure 4.2: The seed time series of an erosion (green) and deposition (blue) 4D-OBC. The dashed lines represent the original seed time series. The solid lines represent the translated seed time series.

A unique feature is the resampled seed time series. The algorithms we use require 4D-OBC feature vectors of equal length. Therefore, using the original time series is impossible, as the duration of the 4D-OBCs varies. We thus resample the seed time series to a fixed number of epochs using linear interpolation. The seed time series show a wide variety in length as a result of variations in the nature of a surface activity, ranging from 24 h to 1,344 h, with a mean of around 500 h for the beach dataset. A resampling size of 500 epochs is therefore chosen. Consequently, the resampled time series feature is represented in the feature vector of the 4D-OBCs by 500 dimensions. The resampling size for the snow cover dataset is 4 h.

Information lost by resampling the time series is added as additional features. The extracted features are divided into two groups, features based on the full seed time series and features based on events in the seed time series. The polynomial features are computed by least square fitting on the seed time series. The first-order polynomial thus adds three features (zeroth order coefficient, first-order coefficient, and residuals), the second-order polynomial adds four, and third order polynomial adds five. We do not explicitly use the zeroth-order polynomial fit as this is equal to the mean height change feature. All other features, except duration and area under the curve, are based on features that have shown to be useful for clustering a point cloud time series dataset of a rockfall-affected site by Winiwarter et al. 2022.

Spatial features
One spatial feature is used for both datasets, namely, the segment size of the 4D-OBC (in m^2), based on the number of points incorporated in the spatial segment. For the beach dataset, we also derive the cross-shore location of the seed (in m) and the cross-shore location of the bounding box center around the spatial segment of the 4D-OBC (in m).

We only incorporate the cross-shore position for the beach dataset as this can distinguish processes occurring in the intertidal zone from processes in the supratidal zone. In contrast, the along-shore position is not considered relevant to any separation in physical processes on the beach. The cross-shore position feature is a case-specific feature, i.e., in other settings, like the alpine setting of the snow cover dataset, specific surface activities might not be constrained to any spatial region based on a coordinate, or the spatial axes over which the surface activities can be grouped is different or not known beforehand. Therefore, we do not use any location features for the snow cover dataset.

Spatiotemporal feature
One feature describing both the spatial and temporal dimensions is extracted: the volume feature. It is computed by summing all the areas under the curve of the temporal segments of the points incorporated in a 4D-OBC.

From the 36 extracted features, we use 8 features as input for the next step of our developed method, highlighted in Table 4.1. This selection is determined based on the optimization methods presented in Section 4.5.

4.2.2. Feature scaling

The features are scaled before further analysis to mitigate any effect of variations in the units of features. The scaling is done by normalizing each feature individually to the range of 0 to 1 using Min-max normalization, i.e., based on the minimum and maximum value present in the dataset, for each feature:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$
(4.1)

Where X is the feature value of the sample under consideration, X_{scaled} is the scaled version of the feature value, and X_{min} and X_{max} are, respectively, the minimum and maximum value found for the feature. When scaling the resampled time series, we set the minimum and maximum feature value as the minimum and maximum occurring height change of all 4D-OBCs in the dataset, with respect to all of the epochs. As such, the resampled time series retains its shape after scaling.

After scaling, all feature values are multiplied by the resample size of the time series, except for the resampled time series feature. Through this, we give equal weight to the different features as to the resampled time series when computing distances between feature vectors, which, in concept, acts the same as scaling the resampled time series between 0 and 1, and all the other features between 0 and the resample size.

Table 4.1: Extracted features. Final selected features (on the basis of Section 4.5) are in bold. Sandy beach specific features are in italics.

Feature name:	Description:	Type
Resampled time series	Height change values at 500 (resampled) epochs of the seed time series (m)	Temporal, full seed time series
Duration	Duration of the seed time series (h)	Temporal, full seed time series
Cross-shore location of seed	The cross-shore location of the seed (m)	Spatial
Timing of min. acc.	Epoch of minimum acceleration event (hs after start of seed time series)	Temporal: event based
Max. change	Magnitude of the maximum height change value (m)	Temporal, event based
Segment size	Number of grid cells in a 4D object-by-change multiplied by grid cell size (m^2)	Spatial
Area under curve	Sum of height change values in the seed time series ($m \cdot h$)	Temporal, full seed time series
Volume	Sum of height change time series of all grid cells in a 4D object-by-change ($m^2 \cdot h$)	Spatiotemporal
Mean change	Mean height change (m)	Temporal, full seed time series
Mean slope	Mean slope of the height change (m/h)	Temporal, full seed time series
Mean acceleration	Mean acceleration of the height change (m/h^2)	Temporal, full seed time series
Mean absolute slope	Mean of the absolute slope of the height change (m/h)	Temporal, full seed time series
Total curvature	Sum of the absolute acceleration of the height change (m/h^2)	Temporal, full seed time series
first-order polynomial fit	Coefficients of the least squares first-order polynomial fit to the time series (three features)	Temporal, full seed time series
second-order polynomial fit	Coefficients of the least squares second-order polynomial fit to the time series (four features)	Temporal, full seed time series
third-order polynomial fit	Coefficients of the least squares third-order polynomial fit to the time series (five features)	Temporal, full seed time series
Timing of max change	Epoch of maximum height change event (hs after start of seed time series)	Temporal, event based
Max. slope	Magnitude of the maximum slope of the height change (m/h)	Temporal, event based
Timing of max. slope	Epoch of maximum slope event (hs after start of seed time series)	Temporal, event based
Min. slope	Magnitude of the minimum slope of the height change (m/h)	Temporal, event based
Timing of min. slope	Epoch of maximum slope event (hs after start of seed time series)	Temporal, event based
Max. acc.	Magnitude of the maximum acceleration of the height change (m/h^2)	Temporal, event based
Timing of max. acc.	Epoch of maximum acceleration event (hs after start of seed time series)	Temporal, event based
Min. acc.	Magnitude of the minimum acceleration of the height change (m/h^2)	Temporal, event based
Acc. at min. change	Acceleration at the minimum height change (at epoch zero, m/h^2)	Temporal, event based
Acc. at max. change	Acceleration at the maximum height change (m/h^2)	Temporal, event based
Cross-shore location of bounding box	The cross-shore location of the center of the bounding box of a 4D object-by-change (m)	Spatial

4.3. Self-organizing Map (SOM)

4.3.1. SOM overview

The first level of unsupervised classification is done using the SOM. The SOM (or Kohonen map) is an artificial neural network designed by Teuvo Kohonen (Kohonen 1990). Since then, SOMs have shown to be suitable for data exploration and clustering in various fields of research (e.g., Skupin et al. 2013, Kohonen 2013, Hagenauer and Helbich 2013, Clark et al. 2020). The central concept behind the SOM algorithm is mapping high-dimensional data onto an n-dimensional grid of neurons, which is done using a competition network. In such a competition neural network, neighboring neurons compete through lateral interactions with input data, and consequently, the different neurons develop to become recipient of specific data patterns (Kohonen 1990). After training a network, each grid point (i.e., neuron or node) is represented by two vectors, the axes on the grid (defined by x,y,...N-dimensions) and the axes in the data space, represented by the weight vector (v with length n equal to the number of features). Each node represents a group of similar samples present in the data, and nodes or groups of nodes on the SOM can be interpreted as clusters if the distance in feature space (dissimilarity) of the mapped samples is relatively small and the distance to the samples in other groups of nodes is large. A trained version of a SOM is visualized in Figure 4.3. Here, samples of a three-dimensional toy dataset are used for training a SOM, and afterward, the samples are matched to their closest weight vector. One can identify six distinct regions of the SOM representing the clusters. In between these regions, the distance in the feature space of the trained weight vectors is large, while within these regions, the distance is smaller, and only subtle variations are apparent.
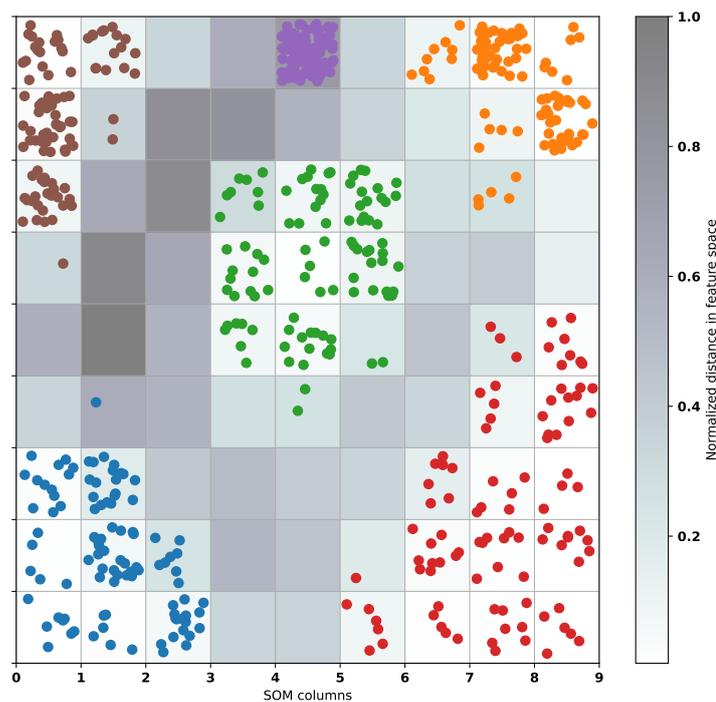
Figure 4.3: Toy example of a Self-organizing Map (SOM, 9x9 nodes) with six distinct clusters, represented as either a region of the SOM or a SOM node. The colored dots represent the different clusters. The location within a grid point has no meaning but is randomized for visualization purposes.

4.3.2. SOM algorithm

The SOM is a neural network that maps and clusters high-dimensional data onto an n-dimensional grid or lattice (Kohonen 1990). All grid points contain a weight vector v_j (with $j = 1,...M$, $M = No.\ grid\ points$) with a length equal to the number of features of the input samples. During each of the training cycles $t = 1,...T$, all samples x_i (with $i = 1,...n, n = No.\ samples$) in the training dataset

are iteratively and in fixed order mapped to the closest node, and the weight of the node is updated. As a result, the final variance between the weight vector and the mapped samples is minimized.

The SOM algorithm consists of the following steps:

1. Initialize weight vectors, v_j with $j = 1, ...M$

2. Select for sample x_i the closest weight vector v_j as best matching unit (BMU)

3. Update the weight vector and the surrounding weight vectors:

$$v_j = v_j + \alpha_t h_{i,j}(t)(x_i - v_j) \tag{4.2}$$

here $h_{i,j}$ is a Gaussian kernel function defining the magnitude of influence of the sample x_i on the weight vectors in the grid:

$$h_{i,j}(t) = e^{\frac{-d_{i,j}^2}{2\sigma_t^2}} \tag{4.3}$$

where $d_{i,j}$ is the grid distance between v_j and BMU, in grid units; σ_t is the standard deviation of the Gaussian kernel at cycle t, indicating the radius of influence of the sample; and α_t is the learning rate at cycle t.

4. Repeat step 2 and 3 for every sample in the dataset

5. Repeat step 4 for a given amount of cycles T

The initial values of the learning rate and radius are predefined and decrease with the number of cycles to achieve convergence and global and local data ordering. The values at cycle t are computed using an asymptotic decay function:

$$(\alpha_t, \sigma_t) = (\alpha_{t-1}, \sigma_{t-1})\frac{1}{1 + \frac{2t}{T}} \tag{4.4}$$

After the training cycles, all training and non-training samples in the dataset are again matched to their BMU to obtain the final grouping of the samples.

4.3.3. SOM architecture advantages

The SOM has several advantages which are not altogether present in other unsupervised classification methods. First, the exact number of clusters to be found in the data does not have to be determined beforehand. In contrast to the k-means clustering algorithm, where one defines the number of clusters, in a SOM, one may define a large number of nodes and let the algorithm itself converge to a certain number of distinct SOM regions (i.e., clusters) between which the distance in feature space is considerable. One can then manually identify how many clusters are found in the data.

Second, a SOM can preserve the topological order of higher dimensional feature space in its lower dimensional grid representation. Consequently, a SOM shows the characteristic feature vectors and distinct groups of these vectors in the data, but also which of these characteristic vectors are neighboring in the feature space. This contrasts with unsupervised classification methods like k-means, hierarchical clustering, and DBSCAN and enables a more straightforward identification of similarities between samples not grouped in the same node while also allowing the identification of subtle differences between samples in neighboring nodes.

Third, by preserving topological order, the SOM shows gradual patterns of variation in the data. SOM nodes might represent transitional groups of 4D-OBCs between two or more distinct groups of nodes (i.e., SOM regions). The SOM can thus, in a sense, be seen as a fuzzy clustering algorithm (Han et al. 2019). Hence, the SOM enables the identification of samples representing combinations of groups.

Fourth, the SOM algorithm causes areas in the feature space with a larger data density to represent a larger area of the SOM (Clark et al. 2020). A SOM is thus overfitted on samples that occur more often. This can be seen as both a benefit and a drawback. If one wants to represent the data without the influence of rare samples, the SOM can do so and is thus less affected by outliers than other unsupervised classification methods. However, if there are rare samples in the data that are important, or some potential clusters have a lower data density or considerable variance, the SOM does not always represent these well. This is a common problem in classification using machine learning, related to

class imbalance (Weiss 2012). It can be mitigated by training an algorithm on a specific subset that better represents the rare samples of the dataset. A maximum dissimilarity sampling algorithm (MDA, Kennard and Stone 1969) or Synthetic Minority Over-sampling Technique (SMOTE, Chawla et al. 2002) could, for example, be used.

The SOM thus inhibits good potential for the characterization and visualization of the 4D-OBCs dataset, as we cannot know which and how many types of surface activities are captured by the 4D-OBCs. Furthermore, the surface activities captured by the 4D-OBCs contain gradual variations, and boundaries between their spatiotemporal properties are thus not necessarily distinct. On top of this, by inspecting the 4D-OBCs in regions of a SOM, one can identify 4D-OBCs the slight variations in properties that characterize the different surface activities.

4.3.4. SOM parameter initialization

The SOM requires 14 parameters to be set (see Table 4.2). The configuration of these parameters is not conventional. Different applications require different settings; determining the optimal configuration beforehand or based on heuristics is, therefore, in many cases, not possible (Clark et al. 2020; Kohonen 2013). Moreover, the SOM does not evaluate to a single objective function during training. As such, the choice of parameters cannot be tweaked based on one objective, like in other neural networks (Yin 2008). This also challenges the choice for the optimal set of features using a wrapping method (see Section 4.5.1). We implement the SOM algorithm using the Python *MiniSOM v2.3.0* implementation (Vettigli 2018).

Methods for parameter selection

In literature, the parameters of the SOM are often chosen by evaluating different configurations, using specific quality measures and visual inspection of the SOM. Quality measures most commonly used are the quantization error and topographic error (Clark et al. 2020; Camus et al. 2011). The quantization error measures the average distance between each input sample and the BMU. It is comparable to an intracluster distance measure; lower values indicate better performance. The topographic error measures the preservation of the topology of the higher dimensional feature space in the lower dimensional SOM grid. It is computed by selecting for each sample the BMU and second-BMU. If these two are adjacent in the SOM, it is counted as no error, and if they are not adjacent, it counts as an error. The total topographic error is given as the error count divided by the total number of samples. Section 4.5 provides a more detailed description of the performance metrics.

We use these two measures, and two others introduced in 4.5, to assess the quality of different configurations of the SOM and choose an optimal configuration. Three parameters are tweaked and tested: the number of nodes, the initial kernel size, and the training subset. On top of this, we also test different subsets of the features introduced in Section 4.2. We do not perform a complete SOM algorithm optimization on all possible configurations as this is too computationally intensive and beyond the scope of this thesis. Section 4.5 describes the optimization methods. The optimization is done based on the beach dataset, and all the other parameters are set based on literature or small empirical tests.

Parameter values

We use a **two-dimensional** SOM with **a square shape**. The SOM can be computed with any number of dimensions, but for intuitive visualization, a two-dimensional SOM is used. The **grid cells** used in our SOM have a rectangular shape. The grid cell shape is an important parameter of the SOM as this determines the topological structure of the SOM. With rectangular-shaped cells, one cell has four nearest neighbors, while a hexagonal-shaped cell has six. As one increases the number of sides of the shapes, the topological preservation of the higher dimensional feature space will be larger. We choose a rectangular-shaped grid cell as this provides easier visualization.

The **number of nodes** is an important parameter as this defines the amount of information extracted by the SOM, the visual interpretability, and the possible distance between the weight vectors in the nodes. With a smaller number of nodes, the distance between individual nodes is larger, and the visual interpretability is likely better. However, a lower amount of information is extracted. With larger SOMs, the opposite happens, i.e., there is a trade-off between the accurateness of the data representation and how generalized the results are. This can be seen as a trade-off between over-representation and

Table 4.2: Parameters of the Self-organizing Map, and their respective values used in this research.

Dimensions:	2
Shape:	Square
No. nodes:	On the basis of SOM algorithm optimization (Section 4.5.2)
Grid connection shape:	Rectangular
Decay function:	Asymptotic
Learning rate at cycle t_0 (α_0):	1.0
Kernel shape:	Gaussian
Std. dev. of kernel at cycle t_0 (σ_0):	On the basis of SOM algorithm optimization (Section 4.5.2)
No. training cycles (T):	20,000
Activation distance metric:	Manhattan
Weight initialization method:	PCA
Order of input:	On the basis of SOM algorithm optimization (Section 4.5.3)
Training subset:	On the basis of SOM algorithm optimization (Section 4.5.3)
Feature set:	On the basis of SOM algorithm optimization (Section 4.5.2)

under-representation of the dataset (Clark et al. 2020). To inspect which number of nodes is useful for our datasets, we train and investigate SOMs with 4, 16, 64, 256, and 1024 nodes.

The **decay function** to determine the values of the learning rate and kernel size at each training cycle is chosen to decrease asymptotically (Equation 4.4). Using an asymptotic function to decrease these values ensures a global and local ordering of the SOM. In the first part of the training, both the learning rate and radius of influence of a sample are big, and the samples are globally ordered. At later cycles, the radius of influence and learning rate become smaller, causing only local fine-tuning and ordering. In the later stages of the training, the algorithm approaches a k-means algorithm, as a matched sample primarily influences the weight vector to which it is matched. The initial learning rate can have a value between 0 and 1. It is chosen to start at 1 to obtain an efficient initial global ordering.

We use a Gaussian **kernel shape**. The shape of this function is asymptotic and causes all the weight vectors in the SOM to be influenced in each training iteration, which results in the smoothest SOMs, thus revealing subtle and gradual differences between 4D-OBCs (Clark et al. 2020). Different values for the **standard deviation** of this kernel σ will be tested, ranging from half the number of rows and columns to a quarter of the number of rows and columns. A larger kernel size means that the shaping of the map (in the initial phase) is more rigid, and the topological structure is thus better preserved. This also means that the relative influence of each 4D-OBC on the SOM is smaller, and the quantization error will be larger as the local ordering of the data and shaping of the SOM is less intense. The influence of rare 4D-OBCs is then also smaller. It is nevertheless proposed by (Kohonen 2001) to use an initial kernel width or standard deviation of half of the largest dimension of the SOM to make sure that the opposite does not happen, namely, the SOM ending up in a local minimum of quantization and topographic error, i.e., the SOM is overfitted on one specific part of the data. As a result, we test both a smaller and larger initial neighborhood and identify if one enables both the identification of rare 4D-OBCs and global ordering.

The **number of training cycles** is set to 20,000. All 4D-OBCs in the dataset are matched to the SOM and update its weight vectors 20,000 times. This number of cycles is chosen as it should be set as high as computationally possible (Clark et al. 2020).

We use the Manhattan distance as a **metric** for the **activation distance** and all other subsequent performance metrics. This metric is chosen as it reduces the influence of dimensions in the data that are not directly comparable to each other due to differences in scaling or weight. This is the case for the resampled time series feature, compared to the other spatial and temporal features. The resampled time series feature consists of 500 separate features, scaled from 0 to 1. Therefore, the other features are scaled from 0 to 500 to obtain equal weight as the resampled time series. For that reason, the other features would get a larger weight if the distances were computed using the more common Euclidean

distance metric. This is explained by the following example:

$$EC = \sqrt{d_{t1}^2 + d_{t1}^2 + ... + d_{t500}^2 + d_{f1}^2} \qquad (4.5)$$

$$M = |d_{t1}| + |d_{t1}| + ... + |d_{t500}| + |d_{f1}| \qquad (4.6)$$

Here EC is the Euclidean distance and M the Manhattan distance, d_{t1} to d_{t500} are the differences at every epoch of the resampled time series between two 4D-OBCs and d_{f1} is the difference between one of the other features. d_{f1} can range between 0 and 500, while d_{tn} can range between 0 and 1. If the two 4D-OBCs were at a maximum difference, the resulting EC would be 500.5, and M would be 1,000. If the resampled time series were equal and the other feature would be at a maximum difference, the resulting EC and M would be 500. However, if the resampled time series were at a maximum distance and the other feature were equal, the EC would be $\sqrt{500}$ while the M would be 500. As such, using the EC, both features do not get equal weight, and the resulting SOM and performance would be overfitted on the non-time series features.

The SOM is initialized using a principal component **weight initialization method**. All the weight vectors in the SOM are initially set such that their values are uniformly distributed over the first two principal components of the dataset. The influence of the weight initialization is not regarded as large and could also be randomly set if the amount of training cycles is large enough. However, the PCA initialization method ensures a faster algorithm convergence while lowering the chances of converging into a local optimum (Kohonen 2001). Therefore, this initialization method can use a lower initial learning rate and kernel size.

As noted before, the **order of input** in the SOM algorithm can considerably influence the final results, as the SOM can be seen as a greedy algorithm. The 4D-OBC dataset contains several rare types of 4D-OBCs, which are significantly different in terms of feature vectors compared to a large part of the dataset (e.g., sandbar depositions, anthropogenic activity), and these might thus be underrepresented in the SOM if the order of input is random (Clark et al. 2020). We, therefore, test the effect of inputting the data based on a ranking determined by a maximum dissimilarity sampling algorithm (MDA; (Kennard and Stone 1969)) in an optimization method, further explained in Section 4.5.3. The MDA algorithm works as follows:

1. Compute the distance matrix between all samples in the dataset based on the Manhattan distance
2. Select the two most distant samples and add them to the ranked list
3. Select the next sample as the sample that maximizes the smallest distance to any of the samples already in the ranked list
4. Repeat step 3 until all data samples have been considered

A ranked list of 4D-OBCs based on dissimilarity is obtained, rated from most dissimilar to least dissimilar. Thus if this ranked list is used as input order for the SOM, the first 4D-OBCs added are the most dissimilar, i.e., the samples at the edges of the data distribution, while at the end of the training cycle, the 4D-OBCs from a denser part of the feature space are added. This ensures that the data is not dominated by what is more often occurring in the dataset, enabling better representation of rare surface activities (Bakker et al. 2022). One could take this even further to allow better identification of rare and outlying 4D-OBCs by only training on a subset of the data (e.g., only the first 300 of the ranked MDA subset). This is also tested using the optimization methods presented in Section 4.5.3.

4.3.5. SOM evaluation

The performance of the trained SOM, with a parameter and feature configuration, determined through the SOM algorithm optimization methods (Section 4.5), is further analyzed and evaluated in its ability to distinguish, identify, and sort groups of 4D-OBCs representing different types of surface activity in four ways.

A global evaluation is done by inspecting the mean feature vectors and weight vectors of the nodes in the SOM and identifying patterns and regions in the SOM showing significant feature distance to the other regions. After identification, the different types of surface activity in the SOM regions are interpreted. This is done for both the beach and snow cover datasets.

A validation dataset is used as a second method of evaluation. It is inspected if the various groups of surface activity are found in the same regions of the SOM. We also identify what characterizes the 4D-OBCs of similar groups in terms of their features if they are (not) found in the same region, i.e., we recognize the reasoning behind the sorting of the SOM algorithm. This is done only for the beach dataset. A subset of this dataset is labeled into groups that should be clustered separately. We do not know what specific classes of surface activity are found in the 4D-OBCs dataset, i.e., there is no ground truth available. Therefore, we only label broad groups of 4D-OBCs from which we find that the 4D-OBCs in the different groups should not be grouped. In this sense, the algorithm's performance with a particular configuration can still be good, even though not all 4D-OBCs with the same label are found in the same cluster. The labeling is based on visual inspection of an animated evolution of height change over the time span of which a 4D-OBC is extracted. Figure 4.4 shows three frames of such an animation, representing the growth of an intertidal bar. We label the 4D-OBCs used as a validation dataset in Anders et al. 2021. From this dataset, 51 4D-OBCs are labeled into the groups presented in Table 4.3. Only the 4D-OBCs are labeled, of which we are confident that they should be clustered into different groups.
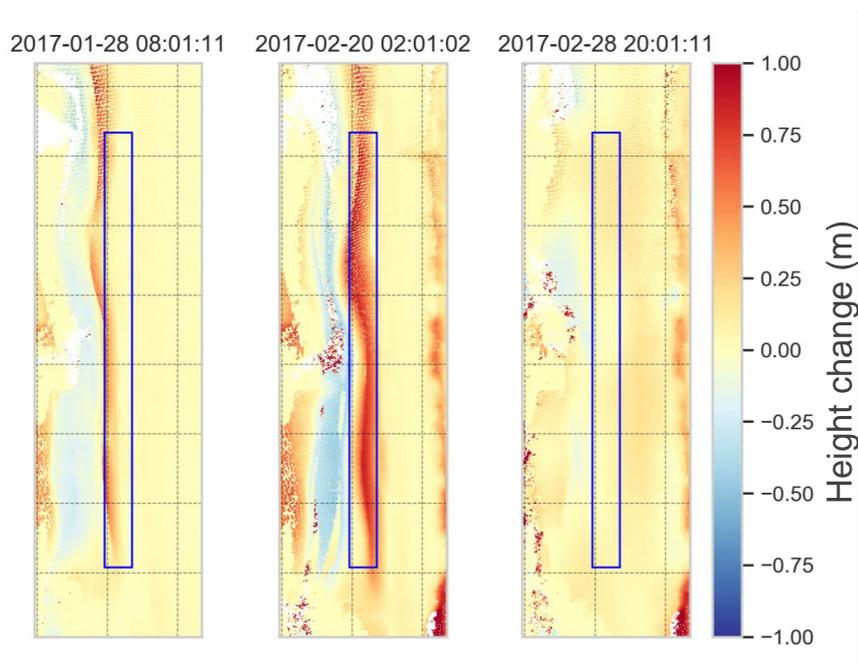

Figure 4.4: Visualization of one of the animations used for labeling the validation dataset. The subfigures represent a frame of the animation at the first, maximum height, and final epoch. The blue box is the bounding box of the extracted 4D object-by-change.

Table 4.3: Number of 4D objects-by-change per labeled group of the validation dataset

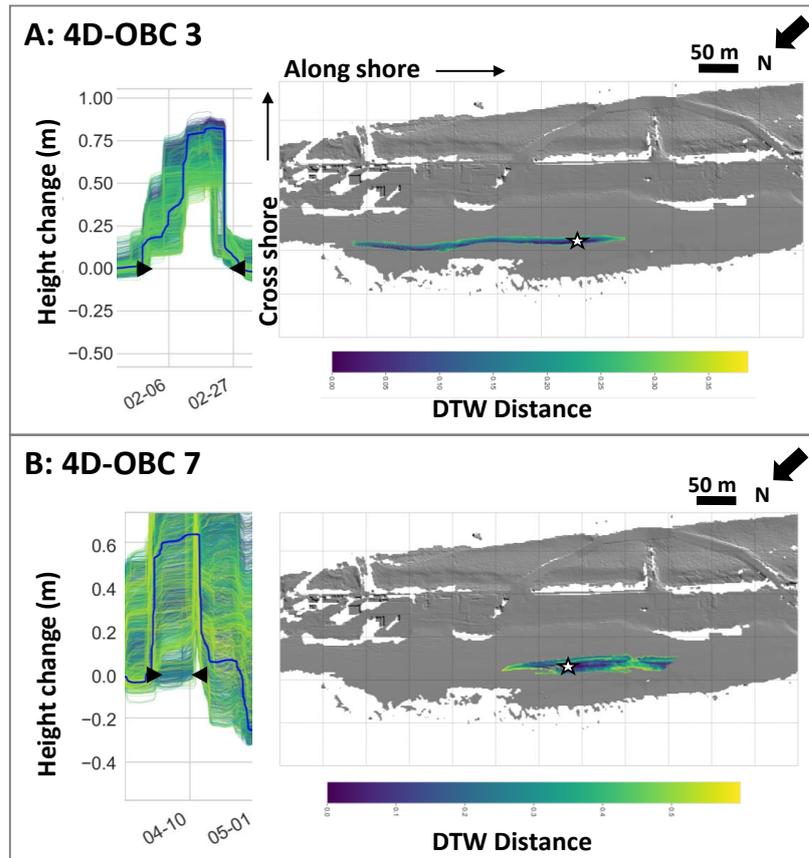| Group	Count
Intertidal erosion	9
Anthropogenic erosion and deposition	9
Foredune erosion	2
Beach berm erosion	2
Beach berm deposition	10
Intertidal sandbar deposition	18
Deposition on road	1

Figure 4.5: Two 4D objects-by-change (4D-OBCs), manually identified as parts of intertidal bar depositions. For each subfigure, the left figure shows the temporal segment of the 4D-OBC, with the seed time series of height change in blue. The other plots represent the time series of additional points incorporated in the spatial segment, colored by their dynamic time warping (DTW) distance. The extent of the temporal segment is captured between the two black triangles. The right figure of each subfigure represents the spatial segment of the 4D-OBC. Each point is colored according to its DTW distance. The star is the location of the seed.

The third evaluation method is performed by inspecting the SOM region around two 4D-OBCs from the validation dataset. These 4D-OBCs are manually identified as parts of intertidal sandbar deposition surface activities. Figure 4.5 shows the spatial and temporal outline of these 4D-OBCs. Here, 4D-OBC 3 represents a full intertidal bar, and 4D-OBC 7 shows a large part of one. We investigate the 4D-OBCs matched to the nodes surrounding these 4D-OBCs and identify what characterizes them in terms of their feature vectors. This evaluation is used to assess if the 4D-OBCs assigned to neighboring nodes in the SOM are related to comparable surface activities and what distinguishes them. Furthermore, this evaluation shows if the combination of 4D-OBCs and SOM is useful for exploring the characteristics of one type of surface activity of interest.

The fourth evaluation method tests the transferability of our methods to different study areas. We train a SOM on the snow cover dataset, using the same parameters and features as provided in Table 4.2, as obtained with the SOM algorithm optimization methods (Section 4.5). Only the SOM size and feature selection are slightly altered. We test a few different sizes and visually inspect the results. For each size, it is identified if groups of 4D-OBCs in each node and distinct region are interpretable as different types of surface activity. We use the same features except for the cross-shore location feature. It is not likely that in the mountainous region over which the snow cover dataset is obtained, the location with respect to one axis holds any separation into different types of surface activity.

4.4. Hierarchical clustering

The SOM nodes serve as the input for an agglomerative hierarchical clustering algorithm, through which we can identify different levels of clusters of surface activity present in the datasets (Scott et al. 2020). This algorithm also automatically determines which SOM regions can be seen as a cluster representing a type of surface activity. We compute the full hierarchical tree using the mean feature vectors of the 4D-OBCs assigned to each SOM node, configured based on the optimization methods. We, therefore, start with all mean feature vectors in separate clusters and iteratively merge these based on the intra-cluster distances. In this way, we obtain a specific clustering level of the dataset per distance threshold. We use an average linkage criterion based on the Manhattan distance to determine if two clusters are merged. The Python *sklearn v1.0.2 agglomerative clustering* implementation is used for clustering. The hierarchical clustering is only done for the beach dataset.

The distance thresholds used to obtain the final numbers of clusters are determined based on the mean silhouette score s_{sil} (Rousseeuw 1987), computed as follows:

$$s_{sil} = \frac{1}{n} \sum_{i=1}^{n} \frac{a(x_i) - b(x_i)}{max(b(x_i), a(x_i))} \tag{4.7}$$

where x_i is the feature vector of a sample in a cluster, $a(x_i)$ is the mean feature distance between a sample and all other samples in the cluster it is assigned to, and $b(x_i)$ is the mean distance between the sample and all the samples belonging to the closest cluster it is not assigned to. In this case, a sample represents the mean feature vector of the 4D-OBCs in a SOM node, as the hierarchical clustering algorithm groups these nodes. The mean silhouette score is close to 1 if the separation between clusters is large, while the intra-cluster variability is low. The score is close to 0 if many clusters overlap. If the score is smaller than 0, many samples are assigned to the wrong cluster.

Using the mean silhouette scores, we can estimate at which distance threshold clusters appear, representing surface activities. If at a certain distance threshold, a local optimum in silhouette score exists (i.e., with increasing distance threshold, the silhouette score drops again), this indicates that the clusters at this threshold show a more prominent separation and smaller intra-cluster distance than after merging. These clusters might therefore hold a physical value and show clusters of high-level surface activities. If from one distance threshold to the other, the silhouette score jumps and stabilizes with increasing distance threshold, the clustering at that threshold is also of interest, as there is a considerable distance between all the clusters at this threshold. We thus identify relevant distance thresholds using these two methods and further evaluate the clusters of SOM nodes found with these thresholds. The evaluation is done in two ways.

First, a global evaluation is done by inspecting the distribution of features from the 4D-OBCs in the clusters. Using these distributions, we characterize the different clusters in terms of their characteristic feature values. With these characterizations, we then interpret what types of surface activity the 4D-OBCs in the clusters represent.

After this interpretation, the second evaluation method is through correlation with natural drivers (presented in Section 3.4). The various clusters are plotted against the drivers, and we try to find correlations that indicate the correctness of the interpretation. For example, if one of the clusters interpreted as aeolian dune erosion contains 4D-OBCs only initiating during high wind velocity, this suggests that these 4D-OBCs indeed represent aeolian dune erosion.

Apart from the evaluation, we look for correlations between different clusters and with natural drivers. We focus on the destruction of two intertidal bars occurring in the same beach area at different epochs. To identify where and how the sand of the bar is transported after the collapse, we look into which clusters contain 4D-OBCs initiating at the moment of this destruction. The different environmental conditions (waves, wind, water level, and precipitation) are also considered.

4.5. Feature selection and Self-organizing Map (SOM) optimization methods

In Section 4.2 and 4.3, several configuration settings were identified for which the final choice of values used in this research is based on the results of optimization methods. These are the following settings:

- The final selection of features
- The size of the SOM
- The standard deviation of the SOM kernel
- The order of input and training subset of the SOM

The optimized final selection of the features is obtained as follows. An initial selection of the 36 features, presented in Table 4.1, is chosen by selecting features with a large variance and low correlation (Section 4.5.1). From this initial selection, we test several subsets of features as input for the SOM (Section 4.5.2). Based on the performance attained with the different subsets, we choose a final optimized selection of features. The optimized size of the SOM and standard deviation of the SOM kernel are obtained in the same way (Section 4.5.2). Several sizes and kernel standard deviations are tested, and the optimal configurations are chosen based on the performance under the different configurations. The order of input and training subset of the SOM is optimized by testing four different configurations in combination with the optimized version of the other settings (Section 4.5.3). An optimal choice is determined by comparing the distribution in the feature space of the obtained weight vectors to the distribution of the entire dataset.

The optimization methods are only performed for the beach dataset, and the results are transferred to the snow cover dataset. With these methods, we identify the best configuration to group the 4D-OBCs in the beach dataset. Furthermore, we investigate how sensitive the outcome of the SOM is to variations in the settings mentioned above. The results of these analyses are also used to recommend what possible configurations are useful when applying the methods to different 4D-OBC sets.

4.5.1. Initial feature selection

In classification problems, the determination of importance and subsequent selection of features is of large importance, as the use of irrelevant and redundant features can lower the quality of the classification, even if relevant features are also incorporated (Dy and Brodley 2004; Hancer et al. 2020). Furthermore, the use of fewer features enhances the interpretability of the classification results. To this end, only a portion of the 36 features is selected for final use in the developed methods. These features are chosen based on the attributes of the 4D-OBCs from the beach dataset. These selected features are then also used for the snow cover dataset. In principle, one can use any set of features as input for step 3 in our main workflow (Figure 4.1), depending on the 4D-OBC set under consideration. However, as we want to optimize our selection, we develop and apply a method to select an optimized set of features from the 36 features we extract.

In supervised classification, the selection and relevance of features are more easily obtained than in clustering, as it is known what the target label of each sample is, and one can thus determine which features are most relevant for the different labels. In clustering, however, in most cases, it is not known what the target labels are, and as such, the determination of feature relevance is more challenging.

Methods for measuring feature importance and selecting relevant features for unsupervised learning are a filter approach, a wrapping approach, or combinations thereof. In filter approaches, the importance of a feature is based on statistical measures in the data itself. In wrapper approaches, one evaluates the output of the clustering algorithm with various subsets of features. This process is thus more computationally expensive than filter methods. Especially in high-dimensional datasets, the methods become impractical, as they use a search space of 2^d with d for the number of dimensions (Li et al. 2017). As such, we use a filter approach to determine an importance ranking and selection of the features, which are then iteratively added and evaluated as input for our SOM.

The feature ranking and selection are made using the Relevance Redundancy Feature Selection (RRFS) approach, as proposed by (Ferreira and Figueiredo 2012). The RRFS approach consists of a relevance and redundancy (similarity) measure to rank the features based on their relevance and subsequently

discard redundant features. We use a simple variance-based ranking measure (Liu et al. 2005). This assumes that features that show a higher variance also inhibit more information for clustering (Solorio-Fernández et al. 2020). Thus, the features are ranked from the highest to the lowest variance. The similarity measure we use is Pearson's correlation coefficient (Ferreira and Figueiredo 2012), computed as follows:

$$\rho(x_i, x_j) = \frac{cov(x_i, x_j)}{\sqrt{var(x_i)var(x_j)}} \tag{4.8}$$

where, x_i and x_j are two features, cov and var are the covariance and variance, respectively.

The algorithm for ranking and selection of the features consists of the following steps:

1. Compute variance for each feature

2. Rank features from high to low variance in a ranked list

3. Add the highest ranked feature to the selected features list

4. Compute Pearson's correlation between the last added feature in the selected features list and highest ranked feature in the ranked list

5. If correlation is larger than threshold ρ discard feature from ranked list

6. Else, add the feature to the selected features list and discard the feature from the ranked list

7. Repeat step 4 to 6 until the ranked list is empty

The resampled time series feature is not considered in the feature selection. This feature consists of as many dimensions as its resample size, which makes its variance and similarity incomparable to the other features. From the resulting selection of features, a smaller subset of features is again chosen based on the optimization method presented in the next section.

4.5.2. Methods for optimization of SOM size, kernel size, and feature selection

We test various configurations (SOM size, kernel standard deviation, and feature subset) to determine the optimal configuration of the SOM and further investigate the relative importance of the different features for the grouping of the 4D-OBCs. One parameter value or feature configuration is changed in every iteration of a test run, resulting in 324 SOMs. Figure 4.6 shows the workflow. We use one routine in which one variable is changed at a time. We hypothesize that with more features, a larger SOM could be more useful and achieve higher performance, as there's more room for sorting, while the contrary is the case when using fewer features.

The test is performed as follows. We start with the ranked list and initial selection of features determined through the methods in Section 4.5.1. From this list, the first, most important, feature is added to the feature subset, together with the resampled time series. We then train several SOMs with this feature subset, with 10 parameter configurations (5 different sizes with each 2 kernel sizes, see Figure 4.6). This is done for the beach dataset's erosion and deposition datasets. The training is done with all the 4D-OBCs, without any MDA subsampling, but with an input order according to the MDA ranking based on the feature subset. This results in 20 different SOMs for the feature subset. Then, the next feature in the ranked list is added, and 20 more SOMs are trained using the same parameter configurations. This is done until all the features of the ranked list have been added to the feature subset.

The performance of the resulting 324 SOMs is quantified using four metrics. The mean silhouette score, mean quantization error, normalized mean quantization error, and topographic error. The mean silhouette score s_{sil} (Rousseeuw 1987) is already introduced in Section 4.4, and is computed as follows:

$$s_{sil} = \frac{1}{n} \sum_{i=1}^{n} \frac{a(x_i) - b(x_i)}{max(b(x_i), a(x_i))} \tag{4.9}$$

where x_i is the feature vector of a 4D-OBC, $a(x_i)$ is the mean distance between the 4D-OBC and all other 4D-OBCs in the cluster it is assigned to, and $b(x_i)$ is the mean distance between the 4D-OBC and all the 4D-OBCs belonging to the closest cluster it is not assigned to.

In the ideal case, the best configuration of the SOM algorithm should result in the largest silhouette score, as we want to obtain clusters of 4D-OBCs that are significantly similar to each other but different
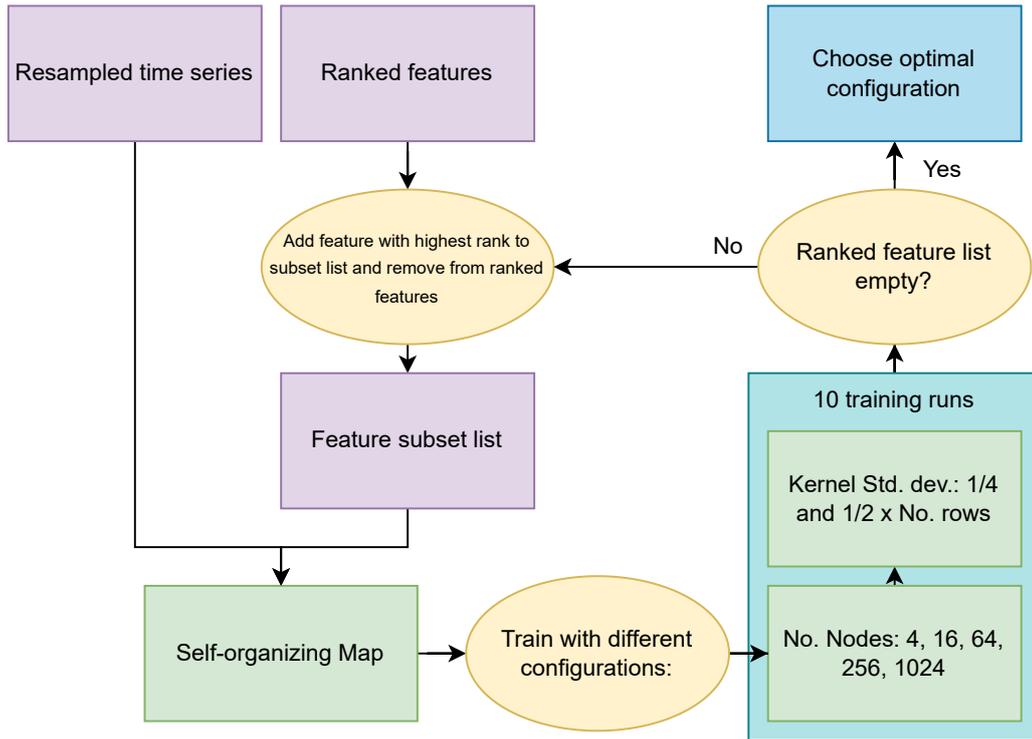
Figure 4.6: Steps to optimize the configuration of the Self-organizing Map algorithm, regarding the SOM size (as the number of nodes), kernel standard deviation, and the subset of features.

from the 4D-OBCs not in the cluster. However, a lower silhouette score is not necessarily bad. As mentioned in Section 4.3.1, a SOM region, instead of only a SOM node, can also represent a cluster (and be manually interpreted as a type of surface activity) if the distance in feature space of the 4D-OBCs in this SOM region to the other SOM regions is large. Within this cluster, one can then identify the different variations of this specific surface activity. This implies that if the SOM were to show, e.g., six regions of relatively high similarity with large intra-region dissimilarity (as in Figure 4.3), the resulting silhouette score would not be high, whereas, in reality, the SOM performs very well. Therefore, to assess the performance, three other metrics are used for evaluation.

The mean quantization error QE measures the within-node variance around the weight vector and shows how well the SOM can represent the data. It is computed as follows:

$$QE = \frac{1}{n} \sum_{i=1}^{n} \|x_i - BMU\|_{L_1} \tag{4.10}$$

where n is the number of samples, x_i is the feature vector of a sample and BMU is the closest weight vector in feature space x_j. It is thus computed by taking the average distance between each feature vector and its BMU. As such, lower values of the quantization error are preferable, as this means the within-node variance is low, and the map nodes are thus representative of the dataset. It is notable that this error measure does not perform well when comparing SOMs with different numbers of features, as a larger feature set by itself will increase the L_1-norm. We, therefore, also use a normalized version of the quantization error nQE, computed through normalizing within the maximum distance possible between two feature vectors:

$$nQE = \frac{QE}{max(\|x_i - BMU\|_{L_1})} \tag{4.11}$$

For example, with two features (resampled time series and duration), the maximum possible distance would be 1,000. 500 for the resampled time series, scaled from 0 to 1, and 500 for duration, scaled from 0 to 500.

The topographic error TE measures how well the SOM preserves the topology of the feature space. It is computed as follows:

$$TE = \frac{1}{n} \sum_{i=1}^{n} e_{x_i} \tag{4.12}$$

where n is the number of samples, e_{x_i} is the topographic error of a sample. This error has a value of 1 if the BMU and second-BMU are not neighboring in the SOM and a value of 0 when they are neighboring. A topographic error of zero thus means that the topology is fully preserved. An ideal SOM thus results in a topographic error of zero, making it possible to interpret neighboring nodes as most similar and thus allow for identifying gradual patterns in the data.

Apart from using these metrics, we manually investigate a selection of the 324 SOMs and explore if the grouping and sorting show some underlying physical reasoning. This, in combination with the performance in terms of metrics, is used to choose the optimal configuration, of which the analysis and interpretation are presented in Chapter 5.

4.5.3. Methods for optimization of SOM training subset size and order of input
The configuration of the SOM size, kernel standard deviation, and feature selection obtained with the optimization methods presented in the previous section is used to identify the effect of the order of input and training subset on the SOM. We run the SOM on the deposition dataset with four training input configurations. A training dataset comprising the entire dataset; with random input order, an input order based on the MDA ranking, and an input order based on the flipped version of the MDA ranking– least dissimilar samples first. Lastly, we use a subset of the entire dataset based on the MDA ranking, as a training dataset, in combination with MDA ranked input order. The size of this training dataset is based on the distribution of the data along the first two principal components of the entire dataset. We want to represent every part of the feature space of the dataset evenly to ensure that the SOM fits the entire data distribution evenly instead of overfitting on denser data space. As such, we choose a subset size such that the distribution along the first two principal components of the full dataset is approximately uniform. These principal components are computed with the Python sklearn v1.0.2 PCA decomposition function, using the randomized truncated singular value decomposition (Halko et al. 2009).

The four configurations are analyzed by investigating the distribution of the weight vectors of the trained SOMs compared to the distribution of the 4D-OBCs in the entire dataset. In this sense, we identify how well the different configurations represent the rare and outlying 4D-OBCs found at the edges of the data space. The preferred method used for our further analysis of SOM performance and hierarchical clustering is chosen such that the weight vectors are not overfitted on the denser part of the data space but are also not overfitted on the outliers and rare 4D-OBCs.

4.6. Conclusions
In this chapter, the methods to characterize 4D-OBC as types of surface activities are described.

- 4D-OBCs extracted from point cloud time series are characterized in a developed workflow of four steps.

- The 4D-OBC dataset is split into erosion and deposition subsets (1).

- 36 initial features are extracted, and a selection of 8 features based on optimization methods, is used for further development (2).

- The SOM algorithm is used to group the 4D-OBCs into detailed surface activities (3).

- The parameters and final feature selection used as input for SOM are partially chosen based on literature and partially optimized by assessing performance under different configurations.

- The grouped 4D-OBCs obtained with the optimized SOM algorithm are evaluated and used as input for a hierarchical clustering algorithm to identify and characterize broader clusters of surface activities.

- The performance of the hierarchical clustering algorithm is assessed in terms of physical interpretability and correlation with natural drivers (4).

5

Results

This chapter describes the results of the previously explained methods. It is structured as follows. First, the results of the final step in the workflow (Figure 5.1) are presented, which is the hierarchical clustering of the SOM nodes (Section 5.1). This final step was only performed for the deposition and erosion beach dataset. Here, we thus present two sets of clusters representing broad clusters of surface activities on the sandy beach. Four clusters are investigated and characterized at each of two identified distance threshold levels for the erosion dataset. All clusters at two distance threshold levels are investigated, interpreted, and characterized for the deposition dataset. These deposition clusters are further investigated regarding their inter-correlations and correlations with natural drivers.

Second, the results of the preliminary step in the workflow (Figure 5.1) are presented, namely, the detailed 4D-OBC surface activity groups obtained using the SOM algorithm (Section 5.2). This step is applied to both the beach and snow cover datasets. We thus present the detailed surface activity groups identified in two SOMs for the beach dataset and two SOMs for the snow-cover dataset. As for both datasets, we obtain erosion and deposition SOMs. We evaluate the performance of the SOM trained on the beach dataset (1) globally through interpretation of the attributes of the groups, (2) concerning intertidal sandbar characterization, and (3) with the labeled validation dataset. The snow cover SOMs are evaluated only globally. The optimized configuration used to achieve these results, in terms of size, kernel standard deviation, feature selection, training subset, and input order, is obtained through the optimization methods (Section 4.5).

The third section describes the results of these optimization methods (Section 5.3). Here, we investigate the performance of the SOMs under different configurations of these parameters. This section thus presents the reasoning behind the final choice for the optimal configuration. The optimization methods are only applied to the beach dataset.
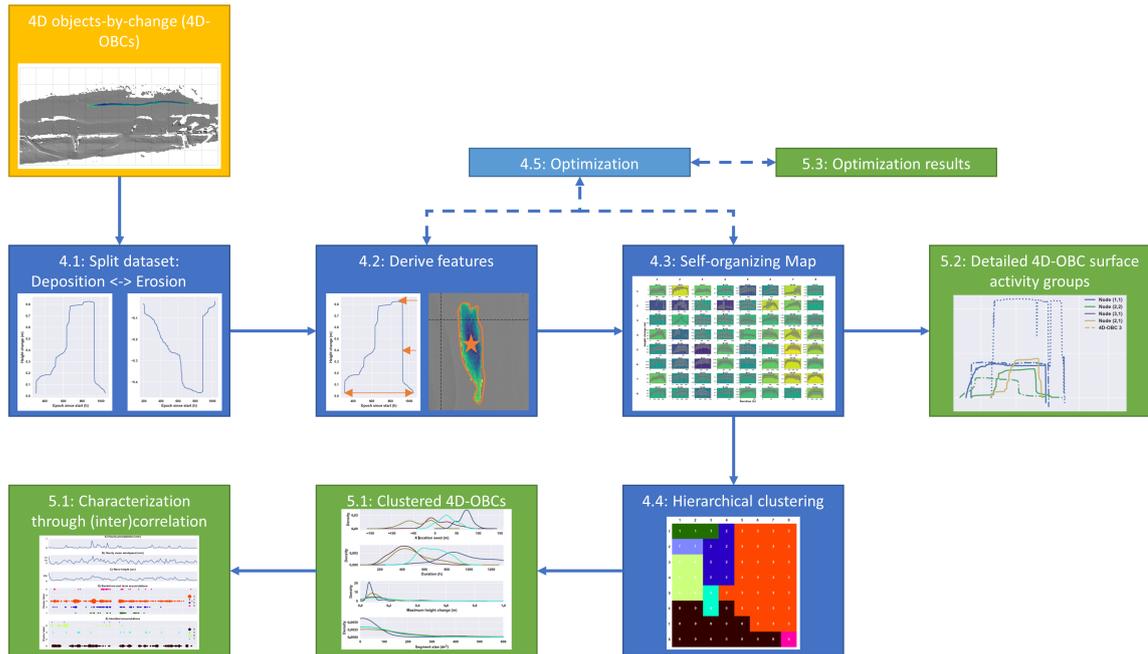
Figure 5.1: Workflow with the four main steps to obtain grouped 4D-OBCs in dark blue and the optimization method in light blue. The output obtained at different levels in the workflow is colored green. The numbers represent the section in which the method or result is presented.

Table 5.1: Optimized parameter values for the Self-organizing Map (SOM). The feature only used for the beach dataset is in italics

Dimensions:	2
Shape:	Square
Number of nodes:	64 (beach dataset), 25 (snow cover dataset)
Grid connection shape:	Rectangular
Decay function:	Asymptotic
Learning rate at cycle t_0 (α_0):	1.0
Kernel shape:	Gaussian
Std. dev. of kernel at cycle t_0 (σ_0):	Quarter of SOM width
Number of training cycles (T):	20,000
Activation distance metric:	Manhattan
Weight initialization method:	PCA
Order of input:	Maximum dissimilarity sampling algorithm ranking
Training subset:	Full dataset
Feature set:	Resampled time series, duration, *cross-shore location*, maximum height change, timing of minimum acceleration, segment size, area under curve, volume

5.1. 4D objects-by-change clusters obtained using hierarchical clustering on Self-organizing Maps

This section presents the results concerning the fourth step in the optimized workflow (Figure 5.1). This step results in two sets of clusters of 4D-OBCs, each representing a type of surface activity. These are obtained by applying a hierarchical clustering algorithm on the output of the third step in the workflow (Figure 5.1). This step is applied only to the beach dataset. Thus, after first splitting the dataset of 2,021 4D-OBCs into the 816 erosion and 1,205 deposition 4D- OBCs, extracting features, and training an optimized SOM for each subset, the mean feature vectors of all the SOM nodes are hierarchically clustered. The final output then contains two sets of clusters of SOM nodes, one for the deposition beach dataset and one for the erosion beach dataset. Each of these sets consists of clusters obtained at two distance thresholds. In this way, we enable the identification and characterization of different levels of surface activities present in the datasets. In the following three subsections, we first identify the distance threshold levels, after which the characteristics of the 4D-OBCs in the clusters are discussed. Finally, the inter-correlations between the clusters and correlations with natural drivers are investigated. The latter is done only for the deposition subset.

5.1.1. Identified hierarchy levels

The identified distance threshold levels for the erosion and deposition beach SOM are determined based on the evolution of the silhouette score with increasing distance thresholds (Figure 5.2). For both the SOM trained on the deposition dataset (D-SOM) and erosion dataset (E-SOM), two levels of local optimum silhouette score are chosen, of which the obtained clusters are further analyzed. At a distance threshold of 270 and 410, 22 and 12 clusters are obtained for the E-SOM, respectively. At a distance threshold of 280 and 440, 20 and 8 clusters are obtained for the D-SOM, respectively.
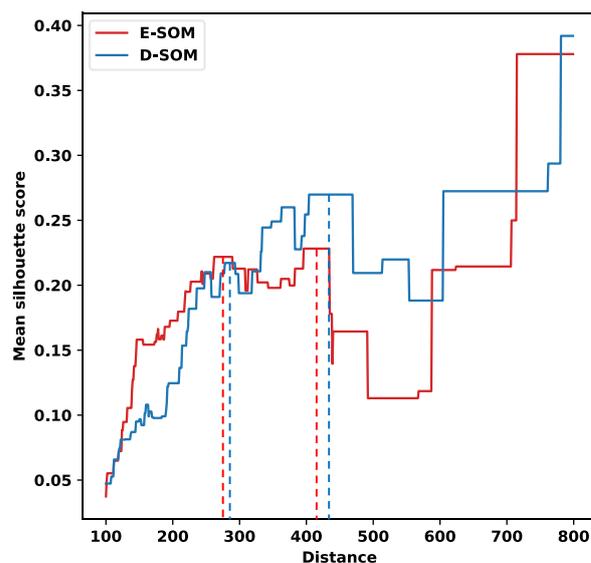

Figure 5.2: Mean silhouette score at different distance thresholds of the hierarchical clustering algorithm for the erosion and deposition Self-organizing Map of the beach dataset. The dashed lines indicate the chosen distance thresholds.

5.1.2. Characterization of the clusters at different cluster levels

Erosion clusters at small distance threshold

Figure 5.3B shows which nodes are grouped at the distance threshold of 270. This threshold yields 20 clusters, of which some encompass only one node (e.g., clusters 13, 17, and 15). Several clusters contain nodes not neighboring in the SOM (e.g., clusters 1, 3, and 6). The mean feature vectors of the 4D-OBCs are thus closest in feature space but not in SOM space. Cluster 3 encompasses all the nodes, where 4D-OBCs identified as dune erosion are matched (see section 5.2). We further inspect the characteristics of this cluster and clusters 2, 10, and 4 to characterize what slight differences in surface activity characteristics occur, as these all contain nodes with similar mean feature vectors.
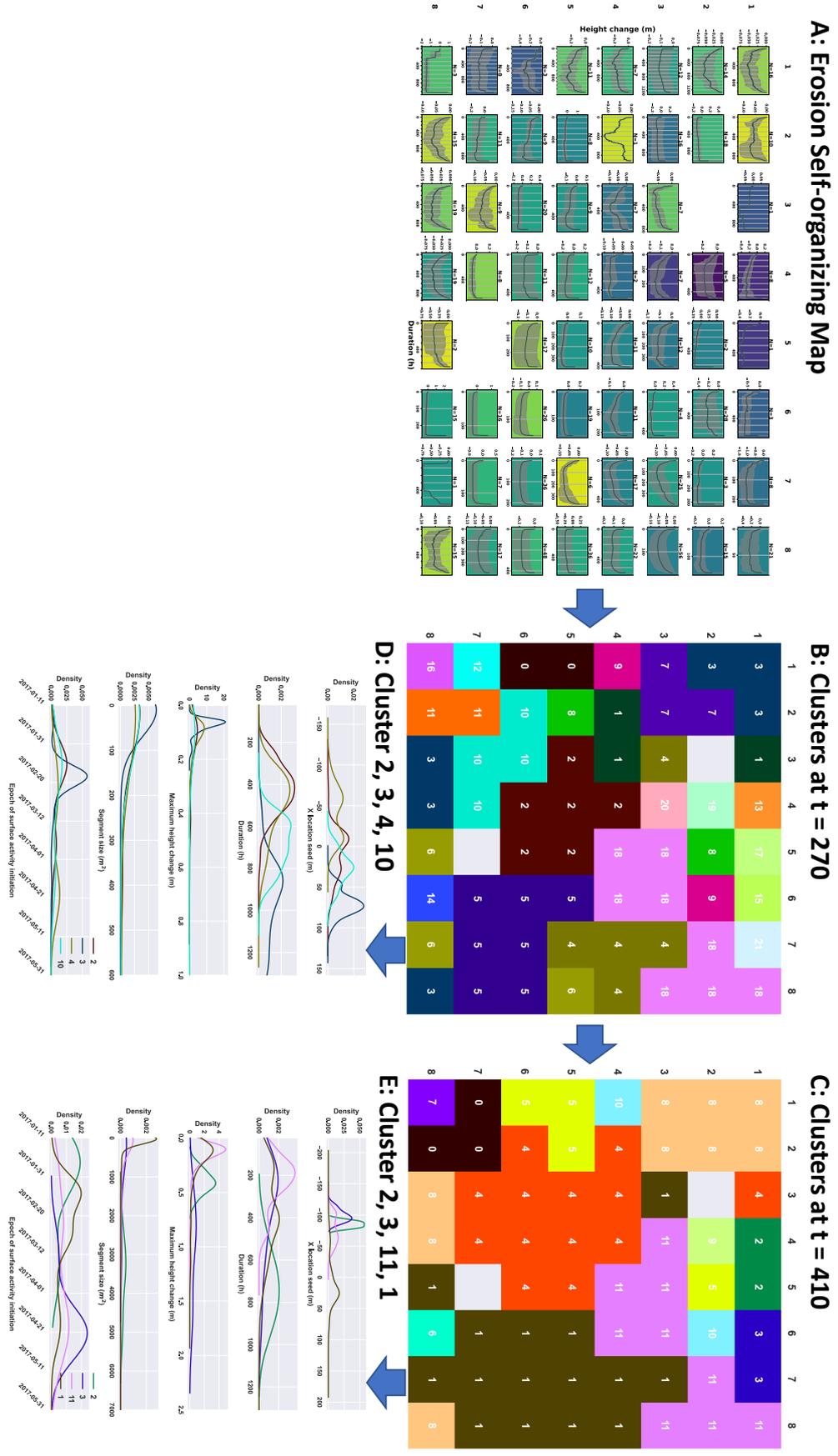
Figure 5.3: Clusters obtained at two hierarchy levels for the erosion Self-organizing Map (E-SOM). A) The E-SOM. Details of this visualization are explained in Figure 5.8. B) The clusters obtained at a distance threshold *t* of 270. Each grid point refers to a node in A, and each color and number represents one cluster of nodes. C) The clusters obtained at *t* is 410. D) Density plots of four features and the start epoch of the 4D objects-by-change grouped in clusters 2, 3, 4, and 10 at *t* is 270. The colors represent the same clusters as in B. E) Density plots of four features and the start epoch of the 4D objects-by-change grouped in clusters 2, 3, 11, and 1 at *t* is 410. The colors represent the same clusters as in C.

A: Erosion Self-organizing Map

B: Clusters at t = 270

C: Clusters at t = 410

D: Cluster 2, 3, 4, 10

E: Cluster 2, 3, 11, 1

Figure 5.3D shows the density plots of four features and the distribution of the start epoch of the 4D-OBCs in these clusters. Cluster 3 contains 4D-OBCs primarily located around the dune area. These 4D-OBCs have a long duration ($\sim 800\,h$) and low maximum height change (<0.1 m). The peak segment size is smaller than the other clusters visualized here ($\sim 20\,m$). These variables indicate that the 4D-OBCs in this cluster are low-magnitude dune erosion surface activities. Most of these dune erosion 4D-OBCs initiate around 2017-02-17. Around this date, mean hourly wind speed values increase from 5 to 18 m/s, the highest mean hourly wind speed throughout data acquisition. Furthermore, around 2017-02-22 precipitation starts under similar wind speeds, which is accompanied by a drop in the density of initiation of the 4D-OBCs in this cluster. This indicates that these particular dune erosion surface activities are mainly initiated during periods of strong winds with low precipitation.

Clusters 2, 4, and 10 contain 4D-OBCs located in the backshore or intertidal zone. The 4D-OBCs in clusters 10 and 2 appear in the backshore area, but the 4D-OBCs of cluster 10 have a longer duration, whereas the segment size, maximum height change, and epoch of initiation are comparable. Furthermore, cluster 2 contains 4D-OBCs that predominantly occur in the intertidal zone around the beach berm. These two clusters are thus separable based on their location and duration, but there is a slight spatial overlap between them. This suggests that these define backshore and beach berm erosion types where the underlying process displays different periods of forcing.

Cluster 4 also shows this overlap in location and duration. Still, in this cluster, a large portion of the 4D-OBCs appear further into the intertidal zone, and the maximum height change distribution has a narrower range, which means this cluster is better characterized by its magnitude than clusters 2 and 10. This cluster thus defines intertidal and berm erosion of particular low magnitude.

Erosion clusters at large distance threshold
A distance threshold of 410 results in a larger amount of nodes per cluster (Figure 5.3C). Some of the previously single-node clusters are merged at this distance threshold. Several clusters containing these merged nodes are interpreted as intertidal erosion surface activities. We further inspect these (2, 3, 11, and 1) to characterize the different types of intertidal erosion surface activities in the 4D-OBC dataset.

Figure 5.3E shows the density plots of four features and the distribution of the start epoch of the 4D-OBCs in these clusters. Clusters 2, 3, and 11 only contain 4D-OBCs located in the intertidal zone, whereas cluster 1 also contains 4D-OBCs situated in the backshore and only encompasses the inland part of the intertidal zone. Clusters 2 and 3 appear far into the intertidal zone but are distinguished by their duration, maximum height change, and segment size. These aspects thus indicate that these clusters represent different intertidal and near-subtidal erosion types. Moreover, the 4D-OBCs of the larger, lower magnitude and longer duration cluster 2 mostly initiate around the start of the measurement period, whereas the 4D-OBCs of the smaller, short duration and higher magnitude cluster 3 initiate in the second half of the measurement period. This suggests that the forcing of intertidal erosion changes throughout the measurement period.

Cluster 11 also contains 4D-OBCs located in the intertidal zone, but these appear slightly more inland and have lower maximum height change and segment size than the 4D-OBCs in clusters 2 and 3. Furthermore, they initiate approximately evenly over the measurement period, which implies that more stable underlying processes force these intertidal erosion surface activities.

Deposition clusters
Figure 5.4B and C show which nodes of the D-SOM are grouped at the distance threshold of 280 and 440, respectively. At a threshold of 280, several clusters contain only one node, just as in the E-SOM. However, a larger cluster share contains more nodes, and fewer clusters contain nodes not neighboring in the SOM. At a distance threshold of 440, several larger clusters are obtained.

In Table 5.2, the characteristics and interpretations of the clusters obtained at both levels for the D-SOM are presented. These characteristics are defined based on inspection of the density distributions of all features, using visualizations similar to the ones in Figure 5.3D and E, as well as the mean seed time series and spatial shapes of the 4D-OBCs in the several clusters.
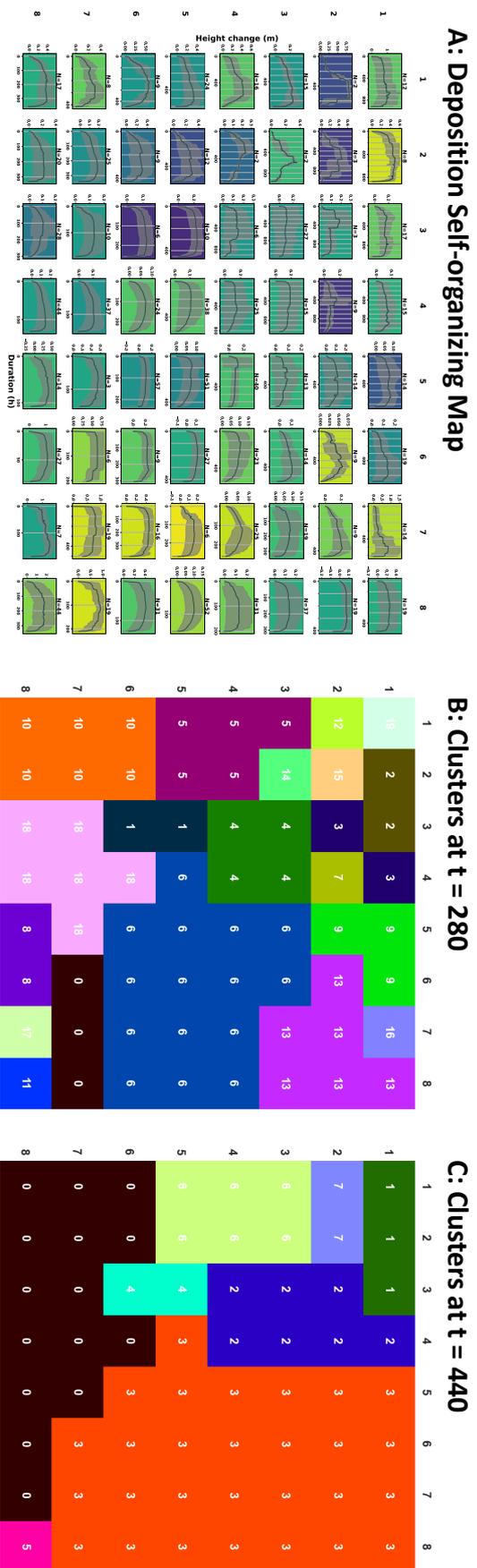
Figure 5.4: Clusters obtained at two hierarchy levels for the deposition Self-organizing Map (D-SOM). A) The D-SOM. Details of this visualization are explained in Figure 5.8. B) The clusters obtained at a distance threshold t of 280. Each grid point refers to a node in A, and each color and number represents one cluster of nodes. C) The clusters obtained at t is 440.

Table 5.2: Characteristics and interpretations of the clusters of 4D objects-by-change obtained with the hierarchical clustering distance threshold set at 280 and 440. Clustering the mean feature vectors of the deposition Self-organizing Map of the beach dataset.

Cluster at threshold 280	Characteristics	Interpretation	Count	Cluster at threshold 440	Interpretation	Count
11	High magnitude, instant deposition and erosion, small size, backshore location slightly inland of beach berm, medium duration	Anthropogenic deposition	44	5	Anthropogenic deposition	44
10	Various magnitudes, gradual deposition, medium size, various shapes, intertidal zone, medium duration	Intertidal deposition	88	0	Intertidal and subtidal deposition, not, or only a part, of intertidal bars	282
18	Low magnitude, gradual deposition, various sizes, various shapes, elongated segments, intertidal zone, short duration	Low magnitude intertidal deposition	146			
8	High magnitude, instant deposition and erosion, small size, rounded shape, at the far border of the observation area, very short duration	High magnitude subtidal deposition	41			
17	High magnitude, instant deposition and erosion, small size, rounded shape, at the far border of the intertidal zone, short duration	High magnitude subtidal deposition	7			
1	Low magnitude, instant deposition and erosion, large size, elongated wide rounded shape, stretching from the bar far into the intertidal zone, medium to long duration	Low magnitude large scale intertidal deposition	16	4	Low magnitude large scale intertidal deposition	16
5	Low to medium magnitude, gradual deposition with instant erosion, various sizes, elongated perpendicular shapes, mostly in the center of the intertidal zone, long duration	Low-medium magnitude (parts of) intertidal bar depositions	69	6	Low-medium magnitude (parts of) intertidal bar depositions	71
14	Low magnitude, gradual deposition and erosion, small size, at the far border of the intertidal zone, long duration	Low magnitude (parts of) inter or subtidal bar depositions	2			
12	High magnitude, gradual deposition with instant erosion, large size, elongated perpendicular shapes, center of the intertidal zone, medium duration	High magnitude intertidal bar depositions	2	7	Medium to high magnitude intertidal bar depositions	4
15	Medium to high magnitude, gradual deposition with instant erosion, large size, elongated perpendicular shapes, center of the intertidal zone, medium duration	Medium to high magnitude intertidal bar depositions	3			
3	Low magnitude, gradual deposition and erosion with various short decreases/increases of height, medium size, rounded shapes, backshore, very long duration	Low magnitude aeolian deposition	18	2	Low magnitude large scale aeolian backshore and berm deposition	100
4	Low magnitude, gradual deposition and erosion, medium to large size, triangled shapes, mostly in the backshore and at beach berm, very long duration	Low magnitude, medium to large scale aeolian deposition	73			
7	Low magnitude, gradual deposition and erosion, some very large in size, segments stretching from intertidal zone to dune area, very long duration	Low magnitude, large scale aeolian deposition	9			
2	Low to medium magnitude, fairly instant deposition and erosion, very small size, rounded shapes, backshore area and on roads, long duration	Low to medium magnitude aeolian deposition	25	1	Long duration, anthropogenic and aeolian backshore deposition	37
19	Medium to very high magnitude, gradual and instant deposition and erosion, small size, rounded shapes, backshore area, long duration	Medium magnitude aeolian deposition and long duration anthropogenic deposition	12			
0	High magnitude, instant deposition with some more gradual erosion, small size, rounded to elongated shapes, backshore area and beach berm, short to medium duration	High magnitude berm and backshore depositions. Possibly anthropogenically or hydrodynamically forced	44	3	Various, mostly natural, deposition processes in the berm, backshore and dune area	643
6	Low magnitude, instant deposition and erosion, various sizes, various shapes, all over area, mostly backshore, short to medium duration	Low magnitude instant backshore depositions	433			
9	Low magnitude, gradual deposition and erosion, large to very large size, various shapes, backshore and start of intertidal zone, medium to long duration	Low magnitude large scale aeolian and hydrodnamic depositions	47			
16	High magnitude, gradual initiation with instant deposition and instant erosion, small size, rounded shapes, dune-backshore border, medium to long duration	Superimposed aeolian and anthropogenic depositions	7			
13	Low magnitude, gradual deposition and erosion, various sizes, various shapes, dune and backshore area, short duration	Low magnitude gradual aeolian depositions in the backshore and dune area	112			

Inter-correlation of deposition clusters and correlation with natural drivers

Figure 5.5 shows the timing of initiation of the 4D-OBCs in the different clusters at a threshold of 440. The 4D-OBCs in cluster 5, identified as anthropogenic deposition, initiate mostly in between a specific span at the start of March when it is known that preparations of the beach for the summer season begin. Cluster 1, containing partially anthropogenic and aeolian deposition 4D-OBCs, has an equivalent distribution over time, whereas cluster 2 appears approximately evenly over the measurement period. All these clusters contain few 4D-OBCs with initiations at the final part of the measurement period as these 4D-OBCs represent longer duration surface activities and are thus likely not finished before the end of acquisition. Cluster 3, representing various berm, backshore, and dune depositions, shows significant variation over time. Around 2017-02-23, a considerable amount of these 4D-OBCs initiate, accompanied by large hourly precipitation, wind speed, and wave height. Furthermore, on this day, the first of two intertidal bar 4D-OBCs matched in node D(2,1) of the D-SOM is destroyed. The other intertidal bar destruction is not accompanied by a considerable wind speed nor the initiation of as many 4D-OBCs from cluster 3.
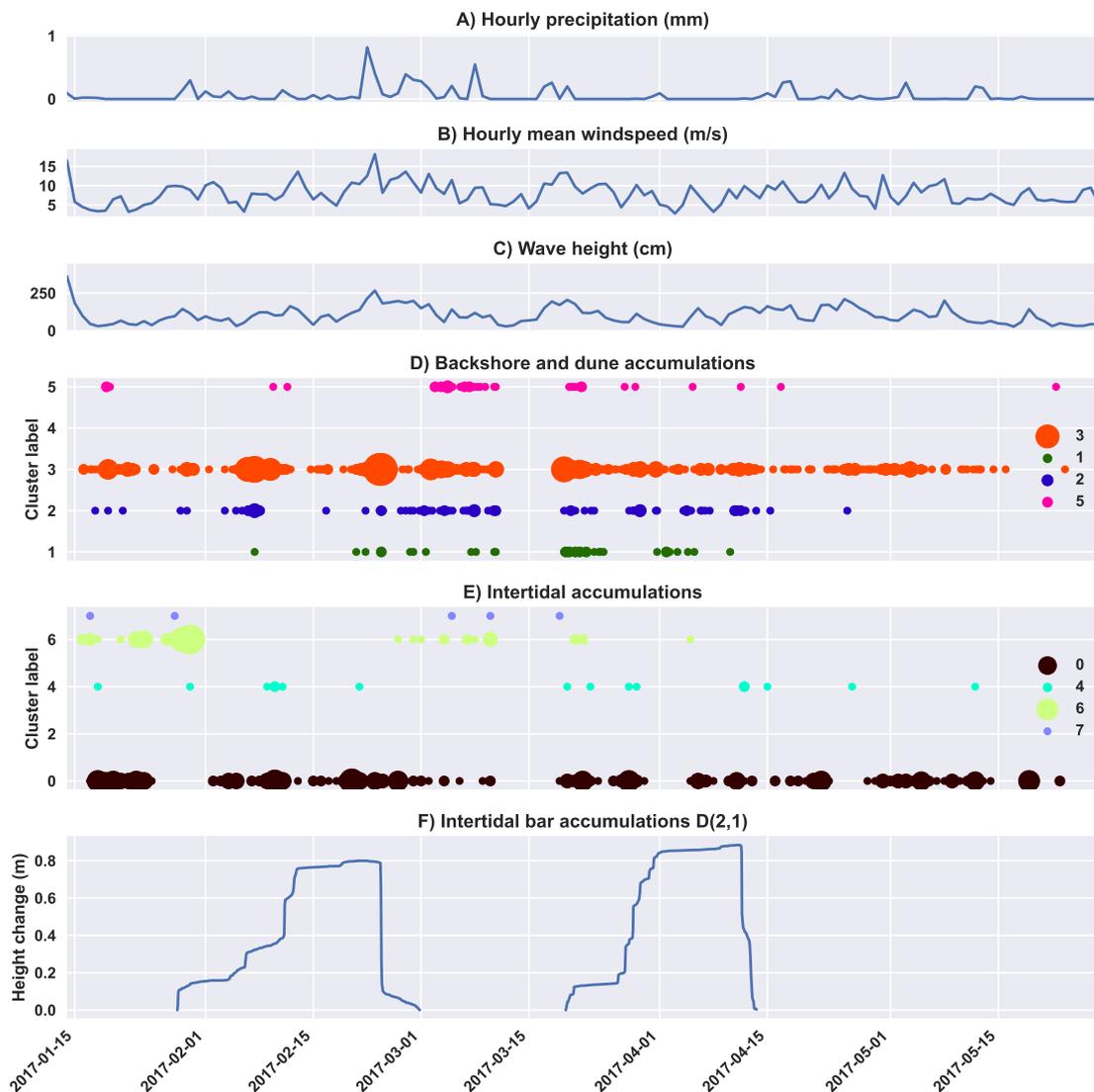

Figure 5.5: A-C) Values of various natural drivers over the measurement period. D-E) Initiation count per day (as the size of the dot) of the 4D objects-by-change (4D-OBCs) in clusters at a threshold of 440 identified as backshore and dune depositions (D) and intertidal depositions (E). F) seed time series of height change of the two 4D-OBCs matched to node (2,1) of the deposition Self-organizing Map, identified as intertidal bar depositions.

To further investigate the destruction of these intertidal bars, we focus on the timespan surrounding this destruction, particularly the initiation of the 4D-OBCs of the clusters obtained at a threshold of 280 within this timespan.

Figure 5.6 shows the timing of initiation and cluster labels of the 4D-OBCs surrounding this destruction, as well as the precipitation, wind speed, wave height, and water level. At the most significant drop in the height of the intertidal sandbar deposition, the water level is relatively low, the wave height is at its highest, the mean wind speed lies around 10 m/s, and there is little to no precipitation. Furthermore, a large number of 4D-OBCs from cluster 6 initiate. This cluster is characterized by its low magnitude instant deposition, mainly in the backshore. Several other aeolian backshore and dune deposition clusters also initiate. At the exact time of destruction, the only intertidal zone surface activity cluster initiating is cluster 18, indicating that most of the sand is transported instantly to the backshore and dune area, especially since a large amount of instant deposition events are initiated. This is in contrast with the destruction of the second intertidal bar.
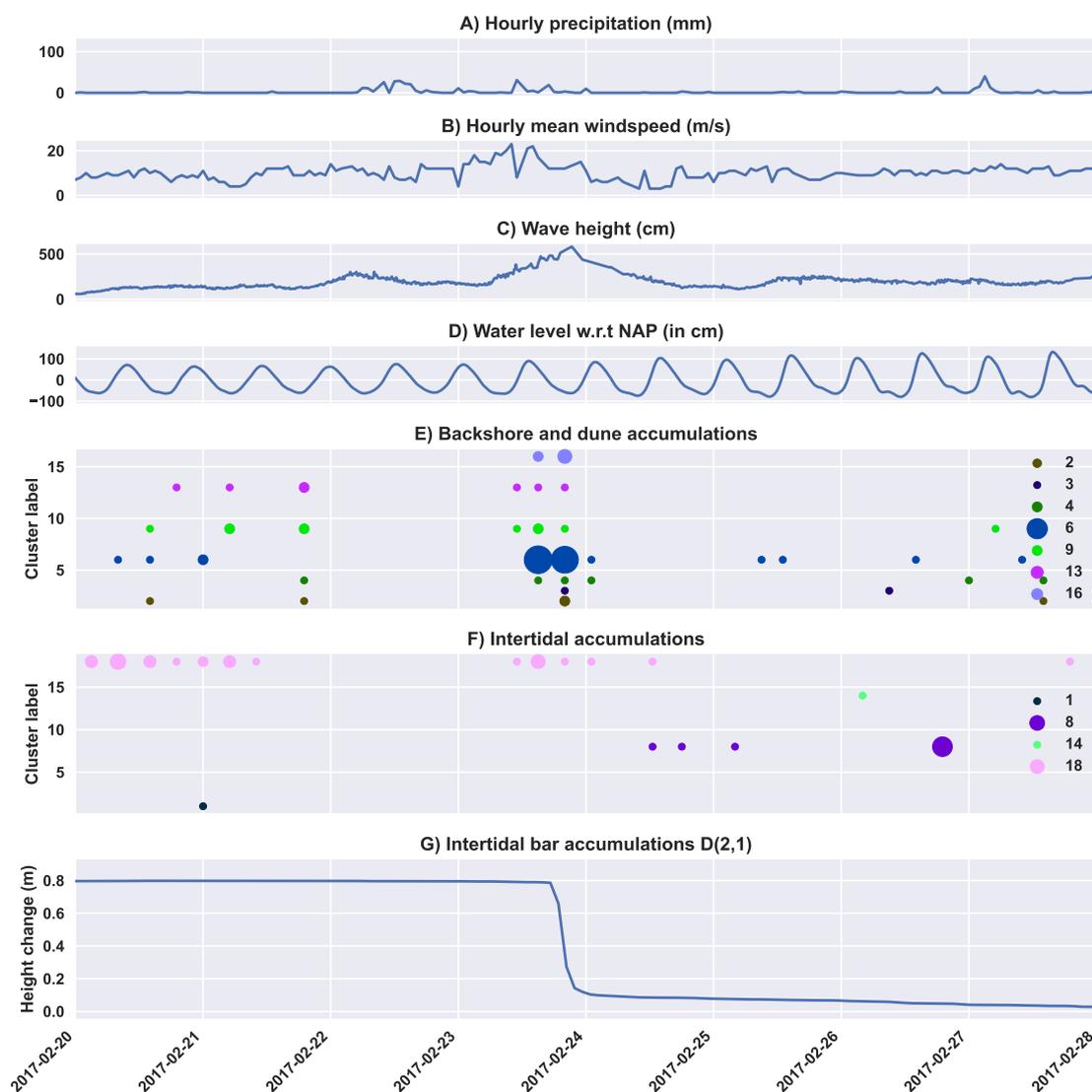

Figure 5.6: A-D) Values of various natural drivers over the measurement period. E-F) Initiation count per day (as the size of the dot) for the 4D objects-by-change (4D-OBCs) in clusters at a threshold of 280 identified as backshore and dune depositions (D) and intertidal depositions (E). F) seed time series of height change of the chronologically first 4D-OBC matched to node (2,1) of the deposition Self-organizing Map, identified as intertidal bar depositions.

Figure 5.7 shows the timing of initiation and cluster labels of the 4D-OBCs surrounding the destruction of this second intertidal bar and the precipitation, wind speed, wave height, and water level. Here, it is first noticed that this destruction occurs in two steps. The first step is more instant and occurs at a high water level. The precipitation, mean wind speed, and wave height are all relatively low at this epoch. Initiation of several backshore and dune deposition 4D-OBCs occurs. This number is, however, smaller than for the first intertidal bar, and the timings of initiation are more spread out over the temporal window.

Furthermore, at the epoch of partial intertidal bar destruction, one 4D-OBC of cluster 8 is initiated. This cluster represents instant high-magnitude deposition events at the far border of the observation area. The initiation of these 4D-OBCs suggests that during this initial destruction of the intertidal bar under a high water level, the sand is transported partially off-shore, as indicated by the relatively low amount of backshore and dune depositions. The second part of the destruction is more gradual and partly occurs during low tide. Even so, only a few aeolian backshore and dune depositions initiate.

In contrast, at the epoch the final destruction sets in, two 4D-OBCs of cluster 1 initiate. These represent low-magnitude large-scale intertidal depositions, stretching from the berm into the intertidal zone. This indicates that a larger portion of the sand during the destruction of this intertidal bar is transported and distributed over the intertidal zone, compared to the destruction of the first intertidal bar. Less sand is transported to the backshore, which could be due to lower wind speeds, wave height, and a higher water level during the first part of the destruction. The low amount of precipitation and low soil moisture does not increase aeolian deposition surface activities.
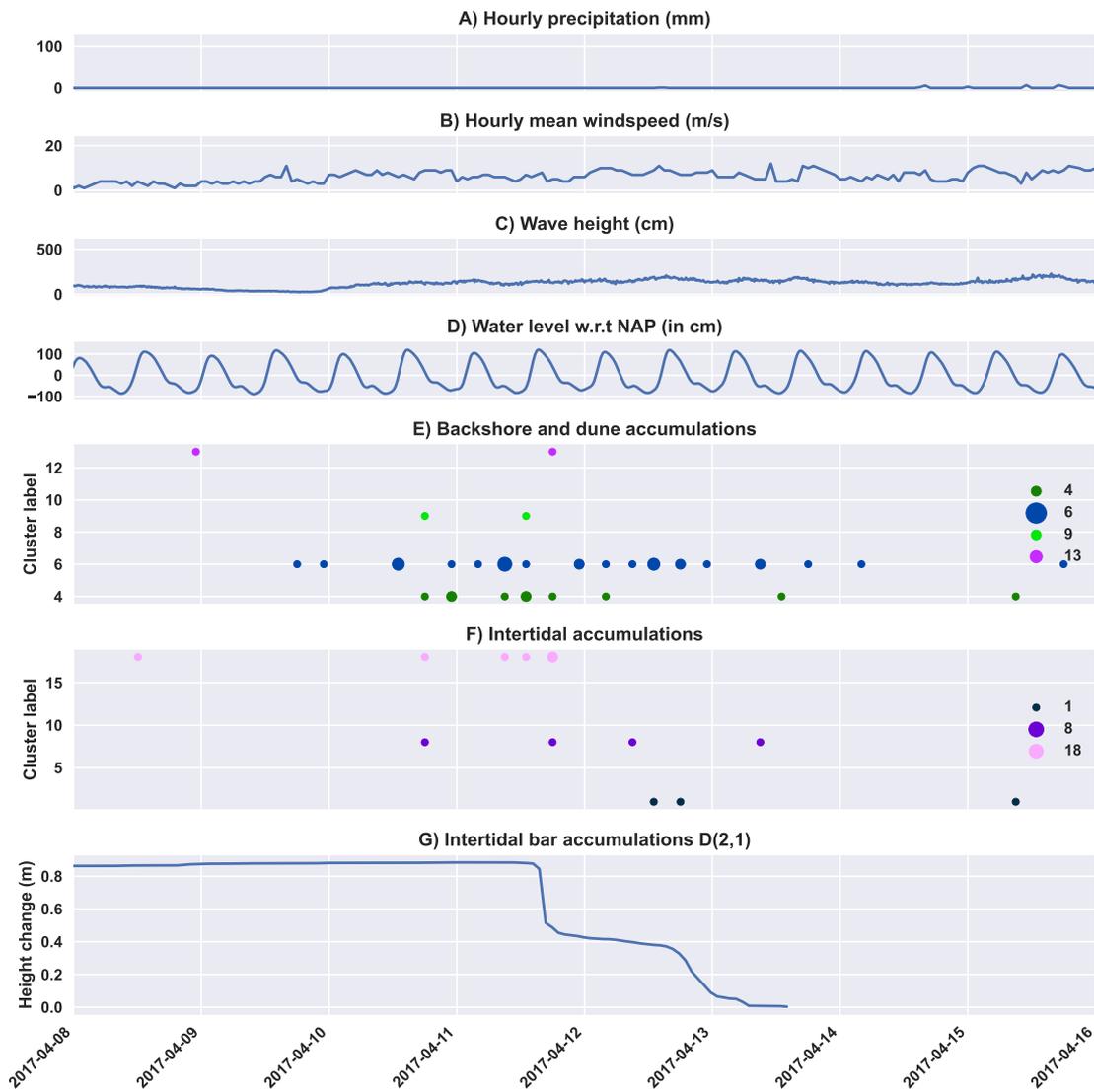
Figure 5.7: A-D) Values of various natural drivers over the measurement period. E-F) Initiation count per day (as the size of the dot) for the 4D objects-by-change (4D-OBCs) in clusters at a threshold of 280 identified as backshore and dune depositions (D) and intertidal depositions (E). F) seed time series of height change at the destruction of the chronologically second 4D-OBC of node (2,1) of the deposition Self-organizing Map, identified as intertidal bar depositions.

5.2. Detailed 4D objects-by-change surface activity groups obtained with Self-organizing Maps (SOMs)

This section presents the results concerning the third step in the workflow (Figure 5.1). These results are detailed groups of surface activity obtained with the SOM algorithm as an intermediate step before the broader clusters of surface activity are obtained. The detailed groups of surface activity are computed for both the beach and snow cover datasets. Thus, for the beach dataset, after (1) splitting the dataset of 2,021 4D-OBCs into the subsets of 816 erosion and 1,205 deposition 4D-OBCs, (2) extracting the optimized selection of features, (3) we train two SOMs. One for each subset, using the optimized configuration presented in Table 3.1. This results in two times 64 SOM nodes, each representing a detailed group of surface activity. For the snow cover dataset, the 813 4D-OBCs are split into subsets of 278 depositions and 535 erosion 4D-OBCs, after which the same features except the beach-specific cross-shore location feature are extracted. For each subset, a SOM is computed, resulting in two times 25 nodes, all representing a detailed group of surface activity. We thus discuss the detailed groups of surface activity obtained with two SOMs for the beach dataset (Section 5.2.1) and two SOMs for the snow cover dataset (Section 5.2.2). For the beach dataset, this discussion is divided into a global evaluation, an evaluation concerning the validation dataset, and an evaluation concerning intertidal sandbar characterization. For the snow cover dataset, only a global evaluation is performed.

5.2.1. SOMs for the beach dataset

Figure 5.8 and Figure 5.10 show visualizations of the mean 4D-OBC feature vectors in the beach SOMs for the deposition (D-SOM) and erosion (E-SOM) dataset, respectively. The values for the performance metrics are presented in Table 5.3.

Table 5.3: Performance scores of the deposition Self-organizing Map (D-SOM) and erosion Self-organizing Map (E-SOM), trained on the beach dataset

	Mean silhouette score	Quantization error	Topographic errors
D-SOM	0.11	152.04	0.14
E-SOM	0.14	138.84	0.14

Deposition SOM

The D-SOM shows several patterns of sorting and grouping (Figure 5.8). First, it is noticeable that from columns 1 to 8, the mean time series of height change in the 4D-OBCs changes from more instant increases in height to more gradual height increases with a smaller magnitude, and, for most of the 4D-OBCs, the drop in height at the end of the temporal segment is more instant than the increase in height at the start. Second, a clear sorting of duration is notable. From rows and columns 1 to 8, the duration of the mean 4D-OBCs generally decreases from as long as $800\,h$ to $100\,h$. There is a noticeable sorting on segment size. On the right side of the SOM, columns 6 to 8, the mean 4D-OBCs are smaller, and from there to the left, columns 6 to 1, the mean segment size increases on average, with a few exceptions (e.g., D(1,2), D(1,1), D(1,4)).

Figure 5.9 shows the distribution of each of the mean feature values of the 4D-OBCs (A-G), as well as the number of matched 4D-OBCs (H) and the mean feature distance of each node to the nearest neighboring nodes (I). The 4D-OBCs are sorted on duration, with very long duration 4D-OBCs being matched to the nodes D(1-3,2-4) (Figure 5.9A).

The sorting on segment size is also noticeable, but the global sorting is less rigid, as we find more exceptions (Figure 5.9E). Generally, the first three features (duration, cross-shore location of seed, and timing of minimum acceleration) exhibit a distinct global distribution pattern. In comparison, the other four features show distributions with more local variations in mean feature values of the 4D-OBCs.
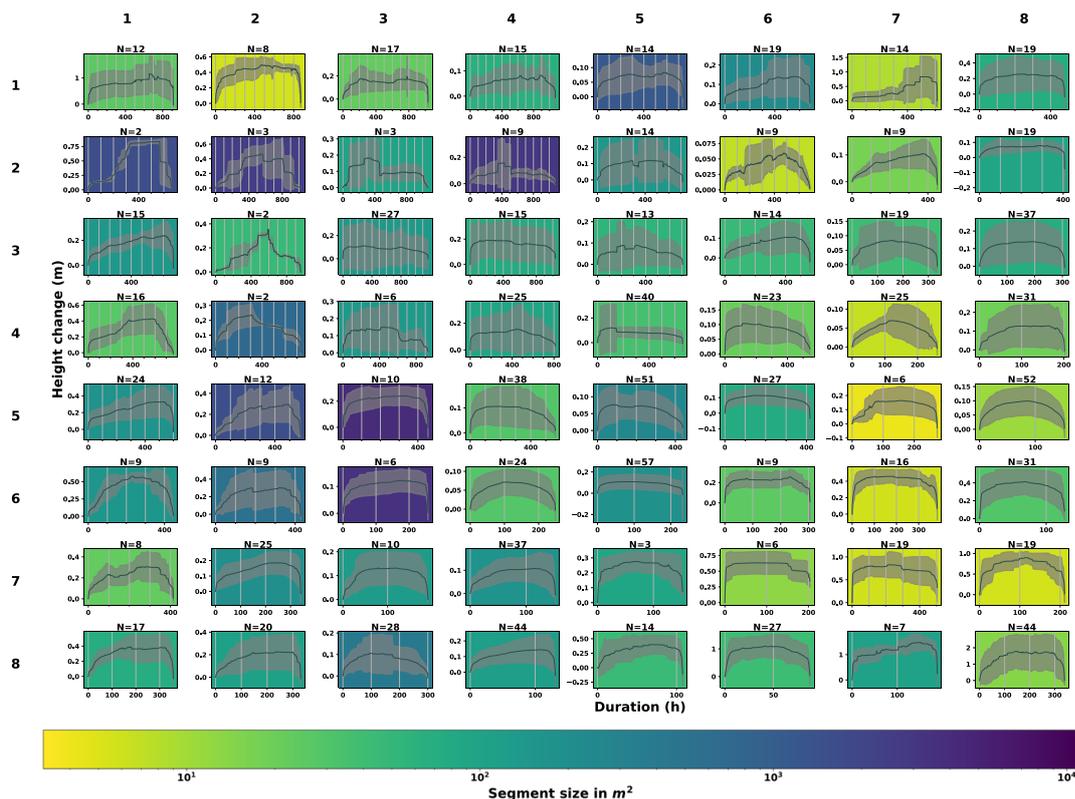
Figure 5.8: Visualization of the deposition Self-organizing Map (D-SOM) trained with the 1,205 deposition 4D objects-by-change (4D-OBCs) of the beach dataset. The resampled time series features + the first seven ranked features (Table 3.1) are used. Each plot represents a SOM node. The x-axis of each plot represents the mean duration of the 4D-OBCs in the node. The y-axis of each plot represents the mean height change of the 4D-OBCs in the node. The limits of both vary between each graph. The black curves show the mean time series of the 4D-OBCs in each node, and the grey area is the standard deviation. The background colors represent the mean segment size per node (on a logarithmic scale). Nodes without a graph represent nodes where no sample is matched after training.

Regarding this global ordering, one can identify a clear sorting on cross-shore location (Figure 5.9B). From the top right to the lower left, the mean cross-shore location of the seeds appears further towards the sea. One SOM region containing mean 4D-OBCs far inland is visible (D(2-4,6-7)); these 4D-OBCs have a low maximum height change and area under the curve (Figure 5.9D, F).

The timing of minimum acceleration (largest negative acceleration) also shows a clear sorting (Figure 5.9C). With this feature, surface activities are characterized that do not have gradual build-ups but are likely more correlated to immediate initiation processes. The first two rows and columns (1+2) generally contain mean 4D-OBCs with a late epoch of minimum acceleration, $\sim 400\,h$ or later. As expected, this is, in most cases, associated with longer durations. The nodes outside these two rows and columns display relatively earlier moments of minimum acceleration. An interesting group of 4D-OBCs is the SOM region D(3-5,3-4). Here, an early moment of acceleration is associated with longer durations ($\sim 400\,h$ - $\sim 1000\,h$). The mean time series of height change shows a relatively rapid increase in height, followed by an instant stagnation around the 100th epoch. The final drop in height, however, occurs more gradually. This pattern of change is comparable to the mean time series and timing of minimum acceleration in the lower right nodes of the D-SOM (D(6-8,6-8)).

The mean maximum height change distribution can be used to separate these two comparable SOM regions into surface activities with a large and small magnitude. The mean maximum height change is relatively high ($> \sim 0.6\,m$) in the 4D-OBCs in the lower right (D(6-8,6-8)). The previously identified region (D(3-5,3-4)) shows significantly lower maximum height change ($< \sim 0.4\,m$). We thus identify two instantaneous surface activity types with different characteristic intensities. One fairly outlying node with a large maximum height change is node D(1,7). Here, the mean maximum height change of the 4D-OBCs is around $0.8\,m$, larger than its surroundings nodes, and comparable to the region (D(6-

8,6-8)) and (D(1-2,1-2)). The mean time series of height change of the 4D-OBCs in this node display little height increase during the first 300 epochs, after which a significant increase in height occurs. The shape of the latter part of the time series is comparable to the 4D-OBCs in regions D(6-8,6-8) and D(1-2,1-2) (Figure 5.8). This indicates that 4D-OBCs in node D(1,7) describe surface activities resulting from a combination of processes. At first, the surface activity initiates like surface activities represented by the 4D-OBCs in its surrounding nodes. However, at some point, another process is superimposed on this surface activity, comparable to the shorter duration, higher magnitude surface activities represented by the 4D-OBCs of SOM region D(6-8,6-8).

The area under the curve feature displays a distribution similar to the maximum height change distribution, though with subtle differences. The mean 4D-OBCs in nodes with a relatively large mean maximum height change and a rather long duration also display a large area under the curve (e.g., D(7,7), D(1,1)). The distribution of the volume feature is similar to the segment size features, with some subtle differences. These differences occur in nodes where the 4D-OBCs have large values for the area under the curve and maximum height change feature (e.g., D(7,7)).

In Figure 5.9H, 4D-OBCs representing rare and common surface activities can be distinguished. The least amount of 4D-OBCs appear in the top left corner of the SOM in D(2,1), D(2,2), D(2,3), D(3,2), D(4,2), where per node only 2 or 3 4D-OBCs are matched. These 4D-OBCs show a relatively large segment size (>100 $m2$), long duration (>700 h) and relatively gradual pattern of height change (see Figure 5.8). Three SOM nodes with the largest number of 4D-OBCs (>51 4D-OBCs) are D(5-6,5) and D(5,8). These 4D-OBCs thus represent the most common types of surface activity. The common surface activities in D(5-6,5) are characterized by their relatively short duration ($\sim$200 h to $\sim$400 h), moderate size ($\sim$100 m^2), position in the backshore, and relatively low maximum height change, area under the curve and volume. The 4D-OBCs in both nodes show a relatively abrupt increase in height (Figure 5.8). The other node representing common surface activities D(5,8) is characterized by its gradual pattern of height change, both during deposition and erosion (Figure 5.8). It represents an almost convex shape, which is not observed as explicitly in any of the other D-SOM nodes. The 4D-OBCs matched here show a short duration ($\sim$150 h), small maximum height change ($\sim$0.1 m), small size ($\sim$10 m^2) and backshore location.

Several distinct surface activities are identified in the SOM distance plot (Figure 5.9I). For example, D(1,1) shows a considerable feature distance to its surrounding nodes and thus contains 4D-OBCs representing distinct surface activities. The same goes for the 4D-OBCs in D(2,1). D(5-8,1-2) and D(1,4-8) are SOM regions with low feature distance between the neighboring nodes; all nodes in each of these regions thus contain similar 4D-OBCs and are representative of similar surface activities.
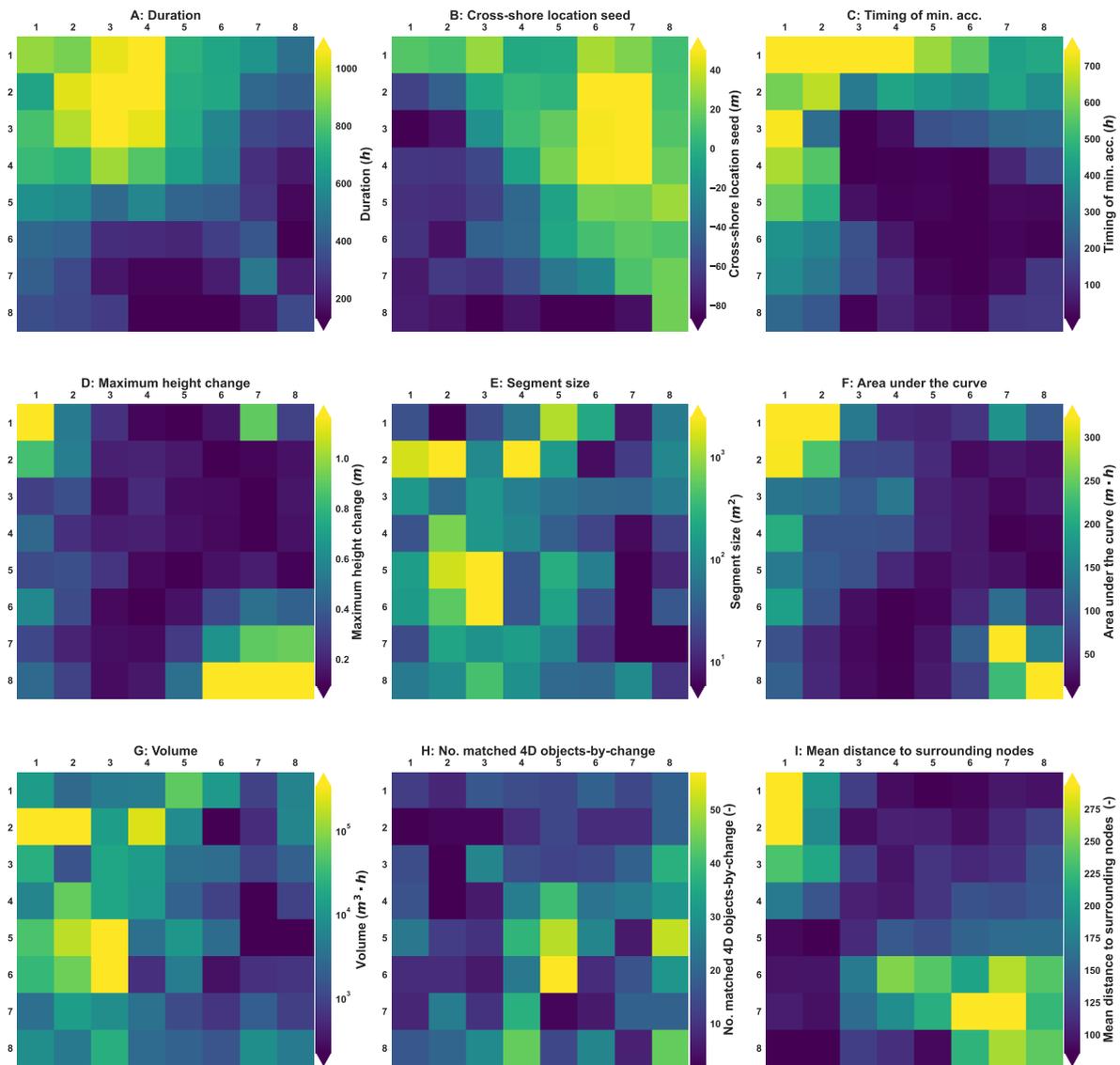
Figure 5.9: (A-G) Distributions of the mean feature values over the nodes of the Deposition SOM. The cross-shore location is zero at the approximate location of the beach berm. The dune area starts around +50 m. (H) The number of matched 4D objects-by-change per node. (I) The mean feature distance between each node and its 2-4 neighboring nodes.

Erosion SOM

The pattern of sorting and grouping based on the resampled time series in the E-SOM are less apparent in Figure 5.10 than for the D-SOM. Most nodes show gradual seed time series, like E(1,1) and E(4,5). The graduality exists in the erosion and deposition parts of the seed time series of height change of the 4D-OBCs. Some nodes do, however, contain 4D-OBCs with more abrupt height changes. This abrupt height change occurs at the end of the temporal segments for most nodes (e.g., E(7,6), E(2,7)), although some nodes contain 4D-OBCs with both abrupt height change at the start and end (E(8,1), E(6,4). In contrast to the D-SOM, the SOM does not seem to be sorted on the shape of the time series, as gradual and more instant mean 4D-OBCs occur over the whole SOM.

The patterns of global and local sorting of the various features, apart from the seed time series, are analogous to what was identified for the D-SOM (Figure 5.11). The E-SOM thus enables the identification and characterization of comparable surface activities as identified with the D-SOM. These are, therefore, not further discussed.

Several interesting surface activity groups, different from what was identified for the D-SOM, can nevertheless be distinguished. Several types of high-magnitude erosion surface activities are identified. Node E(8,1), E(1,7) and E(7-8,6) all contain 4D-OBCs with large values of 0.9 m maximum height change. These nodes thus represent high-magnitude erosion surface activities, at different elements of the beach morphology, with various durations and timings of minimum acceleration. The mean seed time series of height change in E(8,1) (Figure 5.10) exhibits instant and step-like erosion, while E(1,7) contains gradual erosion. This implies that in the intertidal zone, the high-magnitude erosion events occur more gradually, as the 4D-OBCs of E(1,7) are located there, while in the backshore, the erosion is more instantaneous and can be related to anthropogenic activity. Nodes E(7,6) and E(8,6) show instantaneous recovery after the erosion, where the height increases to more than the initial height. This indicates that these nodes contain erosion events, after which high-magnitude deposition surface activities often occur.
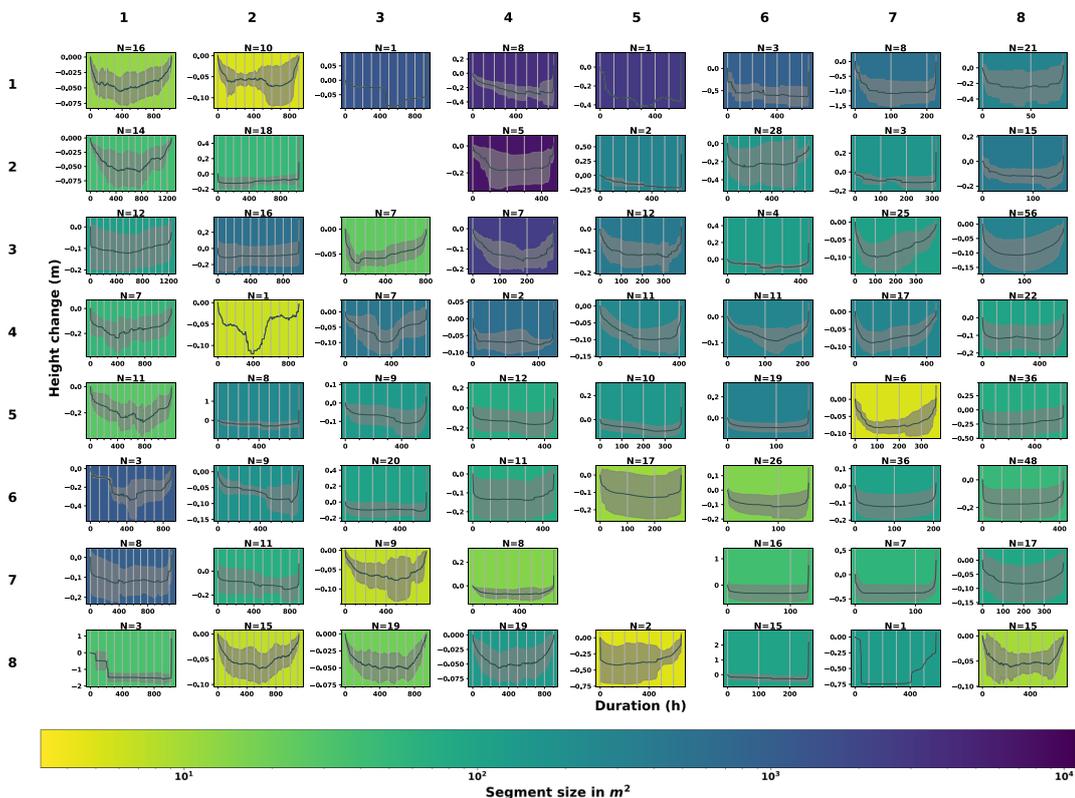

Figure 5.10: Visualization of the erosion Self-organizing Map (E-SOM) trained with the 816 erosion 4D objects-by-change (4D-OBCs) of the beach dataset. The resampled time series features + the first seven ranked features are used (Table 3.1). For a detailed explanation, refer to the caption of Figure 5.8.

A few notable nodes, e.g., E(1,2), E(1,4), and E(8,5), display surface activities characterized by their small sizes of less than 10 m^2. These 4D-OBCs are located in the backshore or dune area, whereas the largest 4D-OBCs appear in nodes E(1-3,4) and E(1,5) and are located in the intertidal zone. This thus implies that intertidal erosion events are more often characterized by their large size, and backshore and dune erosion events are characterized by smaller sizes.

The most often occurring erosion surface activities are situated in E(3,8) and E(6,8) (Figure 5.11H). These surface activities are characterized by their relatively small magnitude ($\sim$0.15 m), steep erosion at the start of the activities, and more gradual recovery through deposition (Figure 5.10). This is in contrast to the nodes representing less common surface activities (<10 4D-OBCs, e.g., E(1,3-6)). Here the erosion occurs more gradually, with a fast recovery afterward.

More nodes and groups representing distinct surface activities are identified using the mean distance plot (Figure 5.11I). One node with a distinct weight vector compared to its surrounding is node E(5,8). This node represents surface activities characterized by their relatively large amount of sand displacement over a long period. These characteristics are different compared to what characterizes other backshore erosion surface activities (e.g., E(5,7), E(4,8), E(6,6); Figure 5.11B).
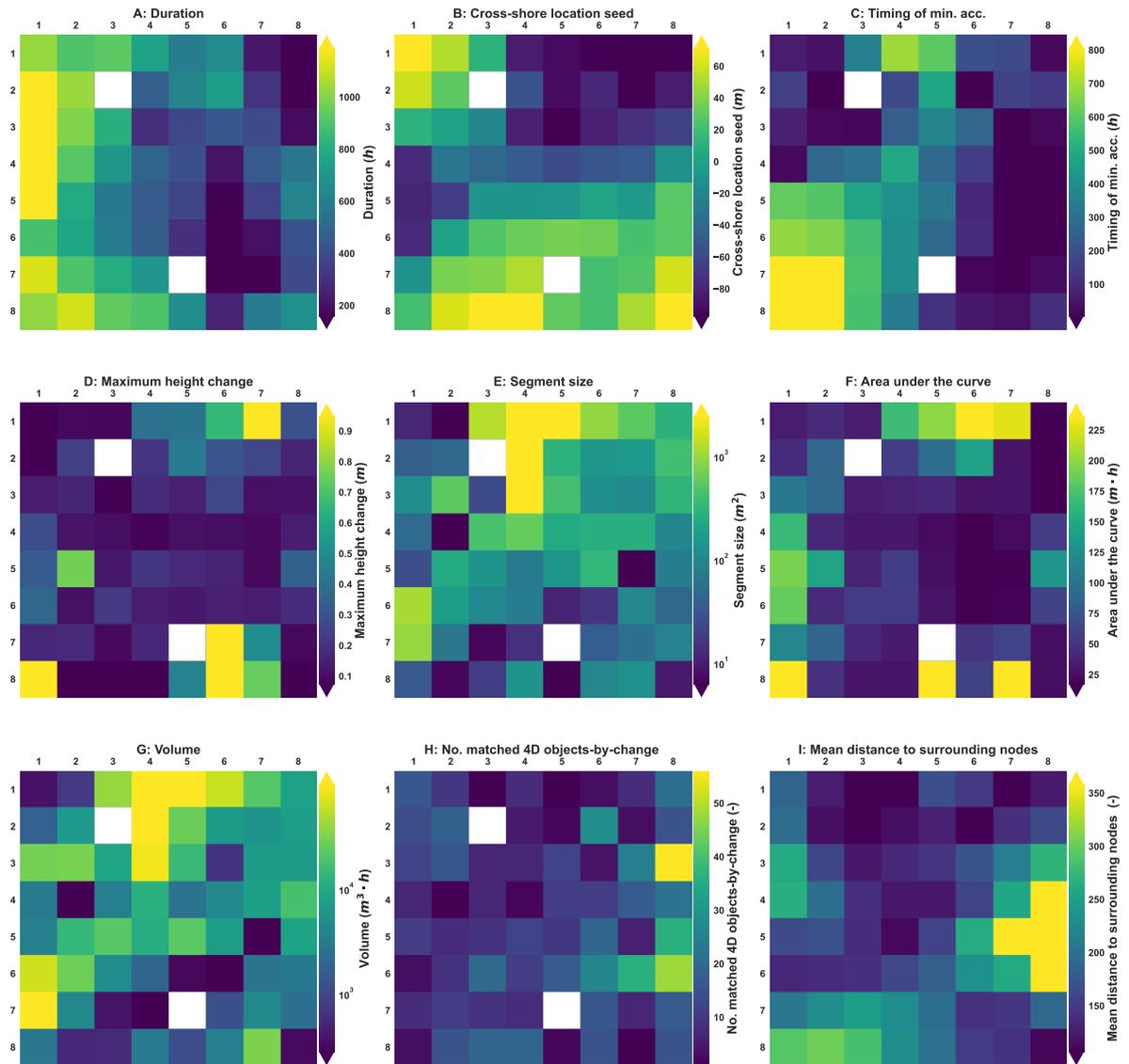
Figure 5.11: (A-G) Distributions of the mean feature values over the nodes of the Erosion SOM. The cross-shore location is zero at the approximate location of the beach berm. The dune area starts around +50 m. (H) The number of matched 4D objects-by-change per node. (I) The mean feature distance between each node and its 2-4 neighboring nodes.

Evaluation of the beach SOM using the validation dataset

To further evaluate the performance of the SOMs in terms of their ability to group 4D-OBCs, representing similar surface activities, the distribution of the validation dataset (Section 4.3.5) over the SOM nodes is investigated. We thus identify if similarly labeled 4D-OBCs also appear in nearby SOM nodes. This is done for the D-SOM and E-SOM of the beach dataset. Figure 5.12A and Figure 5.12B show to which nodes the 4D-OBCs of the validation beach datasets are matched, for the E-SOM and D-SOM, respectively.
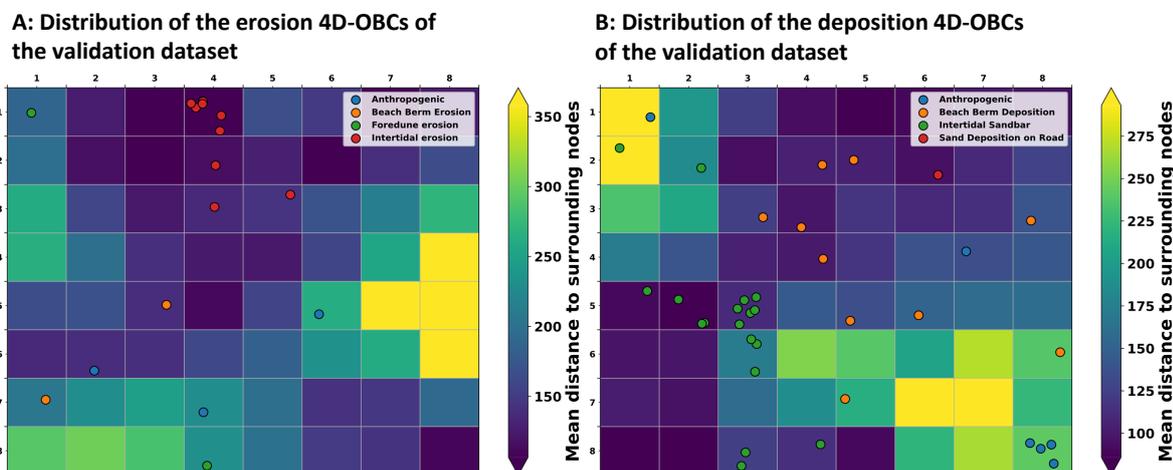

Figure 5.12: Distribution of the labeled validation 4D-OBCs over (A) the erosion Self-organizing Map (SOM) and (B) the deposition Self-organizing Map. Every colored marker represents one 4D-OBC, and each color represents a surface activity type. The mean feature distance map is visualized as background.

The intertidal erosion 4D-OBCs are located in one SOM region (E(1,4), E(2,4), E(3,4), and E(3,5)) with a small feature distance between the nodes. The two beach berm erosion 4D-OBCs are also reasonably close in nodes E(7,1) and E(5,3). Between these two nodes, one of the anthropogenic erosion 4D-OBCs is matched (E(6,2)). This 4D-OBC is not situated close to the other anthropogenic erosion 4D-OBCs (E(7,4) and E(5,6)). The two foredune erosion 4D-OBCs do not appear close together either (E(1,1) and E(8,4)). However, these nodes display similar mean time series, durations, and magnitudes (Figure 5.10). The mean time series show an even shape, with an equal slope for the erosion and recovery part. The other features also have similar values in both nodes. Only the segment size and timing of minimum acceleration differ (Figure 5.11). Thus, two nodes with comparable characteristics are not located close in the SOM.

The nodes with foredune erosion and beach berm erosion 4D-OBCs show distinct mean time series. Node E(7,1), E(5,3), and the node in between E(6,2), likely representing beach berm erosion all display relatively gradual erosion, followed by fast recovery through deposition. In E(6,2) and E(5,3), the time series of the 4D-OBCs shows a two-step pattern. The nodes with foredune erosion 4D-OBCs exhibit gradual height change during the erosion and recovery phase.

The 4D-OBCs manually identified as intertidal bar deposition events are located in the same part of the SOM, on the left side (Figure 5.12B). A large spread is, however, noticeable. Two of the intertidal bar 4D-OBC appear in a distinct SOM region (D(2,1) and D(2,2)), and the other intertidal bar 4D-OBCs are situated in another SOM region (D(5-8,1-4)). Thus, several representations of intertidal bar deposition events exist in the 4D-OBC dataset. The intertidal bar 4D-OBCs of nodes D(8,3) and D(8,4) are matched to nodes with a relatively short duration (<300 h, Figure 5.9A) and lower volume (Figure 5.9G), compared to the other nodes with intertidal bar 4D-OBCs. Two nodes with intertidal bar 4D-OBCs (D(2,1) and D(2,2)) have an earlier timing of minimum acceleration (Figure 5.9C), but comparable durations (Figure 5.9A). Furthermore, the mean time series in these nodes shows a step-like deposition, with a stagnation around 100 h (Figure 5.8). This is not the case in the other nodes with intertidal bar 4D-OBCs. Those nodes exhibit a single slope of height change increase. The differences in height change patterns thus distinguish these types of intertidal bar depositions.

The 4D-OBCs identified as anthropogenic events are also situated in distant SOM regions, namely nodes D(1,1) and D(8,8). These two nodes have different mean durations of the 4D-OBCs (Figure 5.9A), whereas the mean time series appears similar in both nodes (Figure 5.8), implying that both nodes characterize anthropogenic surface activity, and are distinguished by how long the deposited sand remains at the same location.

The beach berm deposition 4D-OBCs all appear in the same SOM region, namely the top right above the diagonal, with a few exceptions (D(7,5), D(3,8), and D(6,8). All these 4D-OBCs occur in nodes with cross-shore locations around the beginning of the backshore area (Figure 5.9B). These nodes, thus, characterize beach berm depositions, with various attributes defined by the other features. The 4D-OBC identified as sand deposition on the road lies close to the 4D-OBCs identified as beach berm deposition (D(2,6)) but inside a node characterized by a more inland position (Figure 5.9B).

Performance of the beach SOM for the characterization of intertidal depositions
We perform a third evaluation by examining the effectiveness of using the SOMs to explore the characteristics of a specific surface activity. We study the features of 4D-OBCs in nodes and nearby nodes where two 4D-OBCs, identified as intertidal bar deposits (Figure 4.5), are matched. We aim to determine if these 4D-OBCs represent intertidal deposits and what characterizes and distinguishes them.

Intertidal bar 4D-OBC 3 is assigned to node D(2,1). In Figure 5.13, the spatial and temporal outline of the 4D-OBCs assigned to this node and its three nearest neighboring nodes is visualized. For each node, a maximum of three 4D-OBCs are shown.
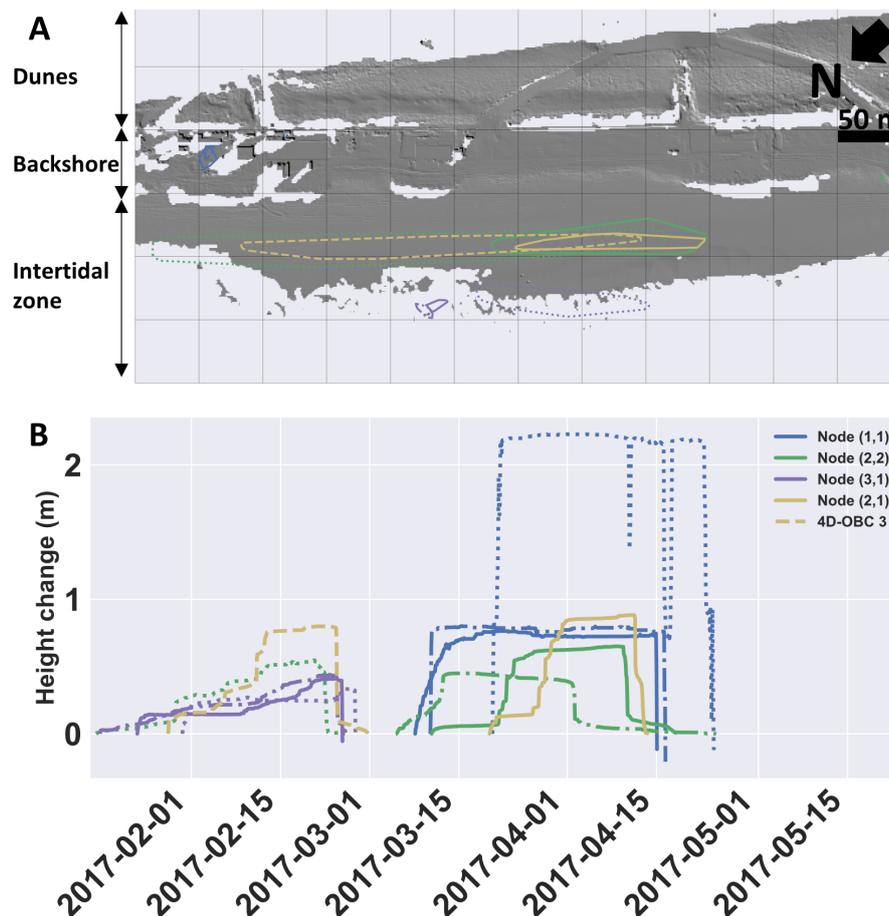

Figure 5.13: A) The convex hulls of the 4D objects-by-change (4D-OBCs) in the node where the intertidal sand bar deposition 4D-OBC 3 is matched and its nearest neighboring nodes. For every node, a maximum of three 4D-OBCs are visualized. B) The seed time series and timing of the same 4D-OBCs as in A. The line style relates the spatial outline to the seed time series of the 4D-OBC.

The 4D-OBCs in nodes D(2,1), D(2,2), and D(3,1) are all located in the intertidal zone, whereas the 4D-OBCs in node D(1,1) are situated in the backshore area (Figure 5.13A). These 4D-OBCs are small and have angled shapes. Node D(2,2) and node D(2,1) show similar elongated segments parallel to the shore. Each contains one smaller and one larger segment. D(2,2) includes one 4D-OBC located at the beach berm. Node D(3,1) contains 4D-OBCs occupying the furthest observed part of the intertidal zone. These segments have irregular shapes.

The 4D-OBCs in node D(2,1) initiate at different moments (2017-01-28 and 2017-03-19, Figure 5.13B). The events do, however, have similar time series shapes. First, there is a small height increase over a short period, after which the height is stable for around 5 days, then gradually increasing to reach the same maximum height of around $0.8\ m$. The 4D-OBCs end with a sharp drop in height over a short period. The 4D-OBCs in node D(2,2) are spatially similar, but the time series differs. The two 4D-OBCs located in the intertidal zone also initiate at two points in time slightly earlier than the 4D-OBC in node D(2,1) (2017-01-16 and 2017-03-09). The height change for the first part only increases slightly, after which the increase accelerates when the 4D-OBCs in node D(2,1) initiate. The height then gradually increases but stays lower than the 4D-OBCs in node D(2,1). The decrease in height occurs slightly before the decrease of the 4D-OBCs in node D(2,1). This implies that the 4D-OBCs in these nodes likely are part of the same intertidal sandbar deposition events. However, the 4D-OBCs in node D(2,1) have a larger maximum height and are slightly smaller than the intertidal bar 4D-OBCs from node D(2,2). This indicates that the 4D-OBCs from D(2,1) represent the tallest parts of the same intertidal bar, which is destructed the latest. The one 4D-OBC from node D(2,2), located at the beach berm, displays a comparable time series initiated between the two other 4D-OBCs, but its segment size and location are different. This 4D-OBC thus does not represent an intertidal bar, but because the other features are similar to the weight vector of this node, it is grouped here.

The seed time series of the 4D-OBCs in node D(3,1) show a gradual build-up with a relatively abrupt decrease in height at the final epoch. All 4D-OBCs initiate around the same time in between the first 4D-OBCs of node D(2,2) and D(2,1) and have time series comparable to the 4D-OBCs of D(2,2). The maximum height lies around $0.4\ m$. These 4D-OBCs thus have similar temporal characteristics, while their spatial outline differs.

The time series of the 4D-OBCs in node D(1,1) differ from the other 4D-OBCs discussed here. The initiations all occur between 2017-03-07 and 2017-03-18. After initiation, the height increases instantly and stays at one level over the whole span of the 4D-OBC before decreasing abruptly. This node thus characterizes a different type of surface activity than the other three nodes discussed here, which is also apparent by the large mean feature distance between the weight vector of D(1,1) and its neighbors (Figure 5.9I).

4D-OBC 7, identified as intertidal bar deposition, is assigned to node D(5,3). Figure 5.14 shows the spatial and temporal outline of the 4D-OBCs assigned to this node and its four nearest neighboring nodes. For each node, a maximum of three 4D-OBCs are visualized.

Figure 5.14A shows that the 4D-OBCs in all these nodes are located in the intertidal zone. The 4D-OBCs in nodes D(5,2) and D(5,3) are all elongated and occur at the same cross-shore location. The 4D-OBCs in node D(5,3) are smaller in the along-shore direction. The 4D-OBCs in node D(6,3) are also elongated but appear wider in the cross-shore direction. Two of the three 4D-OBCs are located more toward the backshore area. The 4D-OBCs in nodes D(4,3) and D(5,4) have smaller segments with more rounded shapes than the other 4D-OBCs discussed here.

Figure 5.14B shows the time series and timing of the same 4D-OBCs. The 4D-OBCs in node D(5,3) contain instantaneous deposition to a height of $0.2\ m$. After which the height slightly increases gradually and drops abruptly. The instant height increase of two of the three 4D-OBCs initiates around the same time when the 4D-OBCs in node D(2,1) have their last jump toward their maximum height (Figure 5.13). The other 4D-OBC in this node display a different timing of initiation, twice the duration, but with a similar end epoch as one of the other 4D-OBCs in this node. These two 4D-OBCs thus describe parts of the same intertidal bar deposition surface activity.

The first two 4D-OBCs in node D(5,2) display comparable duration, start epoch, end epoch, and maximum height change as the longest duration 4D-OBC in node D(5,3), but the increase in height over

time is more step-like. The other 4D-OBC has a different outline, with a more gradual increase towards a larger height followed by a slightly gradual drop in height.
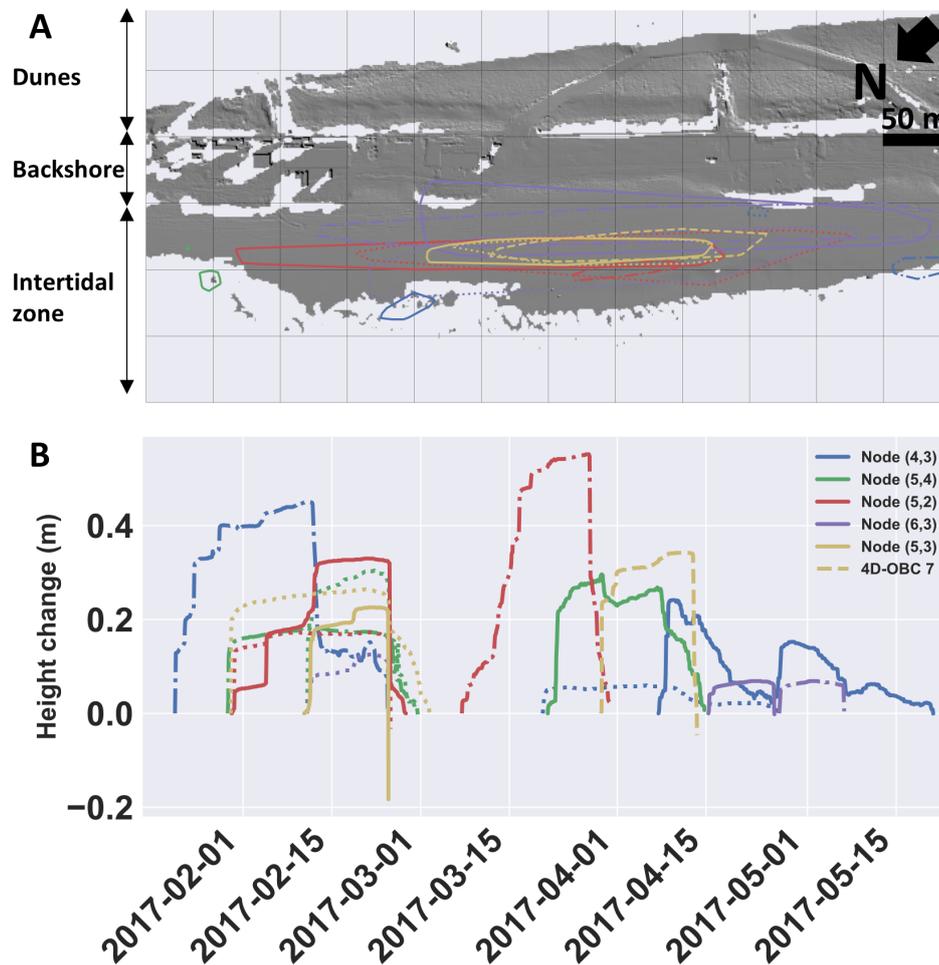

Figure 5.14: A) The convex hulls of the 4D objects-by-change (4D-OBCs) in the node where the intertidal sand bar deposition 4D-OBC 7 is matched and its nearest neighboring nodes. For every node, a maximum of three 4D-OBCs are visualized. B) The seed time series and timing of the same 4D-OBCs as in A. The line style relates the spatial outline to the seed time series of the 4D-OBC.

The 4D-OBCs in node D(6,3) displays a similar time series shape as 4D-OBC 7, but the maximum height changes are smaller. One of the 4D-OBCs in this node initiates directly after the local deposition captured by 4D-OBC 7 is destructed, and another directly after that one. These 4D-OBCs are also located in front of 4D-OBC 7 (with respect to the coast, Figure 5.14A). The third 4D-OBC in node D(6,3) spans the same temporal segment as one of the 4D-OBCs in node D(5,3), while it also partially overlaps its spatial segment. This 4D-OBC also spatially overlaps with two of the 4D-OBCs from node D(5,2), indicating that they are all part of the same larger-scale surface activity.

The first two initiated 4D-OBCs of node D(5,4) have similar time series as the first two elongated 4D-OBCs in node D(5,2), but with a slightly more gradual erosion at the end. These are the small segments on the left side of the area shown in Figure 5.14A. This indicates that these are part of the same intertidal bar deposition surface activity, but due to limitations of the scanner with high water, these are not captured in their entire spatial extent. The last initiated 4D-OBC is located on the far right side of the area. The time series of this 4D-OBC is comparable to the other 4D-OBCs in this node, but the timing of initiation differs from any of the 4D-OBCs here, except one from node D(4,3). The end epoch is similar to the one from 4D-OBC 7.

The time series of the 4D-OBCs from node D(4,3) are all different in shape and not comparable to any of the other 4D-OBCs under consideration. They are located in the intertidal area, from far seaward to the beach berm. These 4D-OBCs can thus be identified as intertidal depositions, but their characteristics are not easily interpretable as one type of surface activity. Instead, they represent various intertidal and berm depositions, irrespective of the intertidal sandbars.

5.2.2. SOMs for the snow cover dataset

Two SOMs are computed for the 813 4D-OBCs of the snow cover dataset—one for the deposition (278 4D-OBCs) and one for the erosion subset (535 4D-OBCs). The size of the SOM is set at 25 nodes after an investigation of SOMs with a size of 16, 25, 36, 49, and 64 nodes (Appendix B). The SOMs are trained using the resampled time series, duration, timing of minimum acceleration, maximum height change, area under the curve, and volume feature, selected through our optimization method applied on the beach dataset (Section 5.3). All other SOM parameters are set identically to the parameters for the beach dataset (Table 5.1).

Figure 5.15 shows the SOM for the deposition 4D-OBCs (SD-SOM). The lower right nodes contain 4D-OBCs with a gradual increase in height over a longer period (SD(4,5), SD(5,4), and SD(5,5)). The remaining nodes contain 4D-OBCs with more instantaneous height change and a mean duration of 3 h. A sorting on segment size is visible—the nodes in the top right show 4D-OBCs with the largest segment size. Around SD(5,3), another SOM region with a large mean segment size is observable. The maximum height change varies strongly over the SOM.
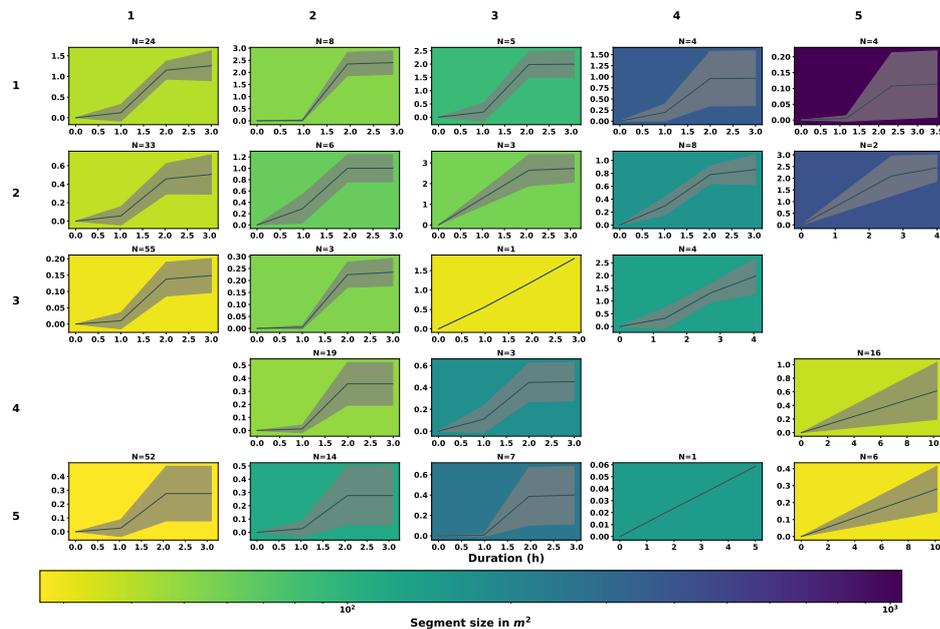

Figure 5.15: Visualization of the deposition Self-organizing Map of the snow cover dataset (SD-SOM). Trained on the resampled time series, duration, timing of minimum acceleration, maximum height change, area under the curve, and volume feature. Each plot represents a SOM node. The x-axis of each plot represents the mean duration of the 4D-OBCs in the node. The y-axis of each plot represents the mean height change of the 4D-OBCs in the node. The limits of both vary between each graph. The black curves show the mean time series of the 4D-OBCs in each node, and the grey area is the standard deviation. The background colors represent the mean segment size per node (on a logarithmic scale). Nodes without a graph represent nodes where no sample is matched after training.

The gradual processes on the bottom right (SD(4,5), SD(5,4), and SD(5,5)) display both low and high maximum height change. The timing, time series, and spatial extent of a part of the 4D-OBCs in these nodes are visualized in Figure 5.16. These 4D-OBCs appear in the same region around the ski resort. One of the 4D-OBCs in SD(4,5) initiates and ends simultaneously with all the 4D-OBCs in SD(5,5). All the 4D-OBCs in these nodes start in a periodic instance, around 21:30 on three consecutive days, lasting until 07:40 the next morning. The periodicity and locality of these 4D-OBCs and the knowledge that bulldozer work has occurred in this area imply that these 4D-OBCs represent anthropogenic preparation. It should be noted that the area in which these 4D-OBCs are located is part of the area not acquired at night. The 4D-OBC in the other node displaying gradual deposition over a long period, node SD(5,4), initiates at 11:00 and lasts until 16:00. The magnitude is lower than the magnitude of 4D-OBCs in surrounding nodes. This 4D-OBC might represent small daytime work on the ski resort preparations.
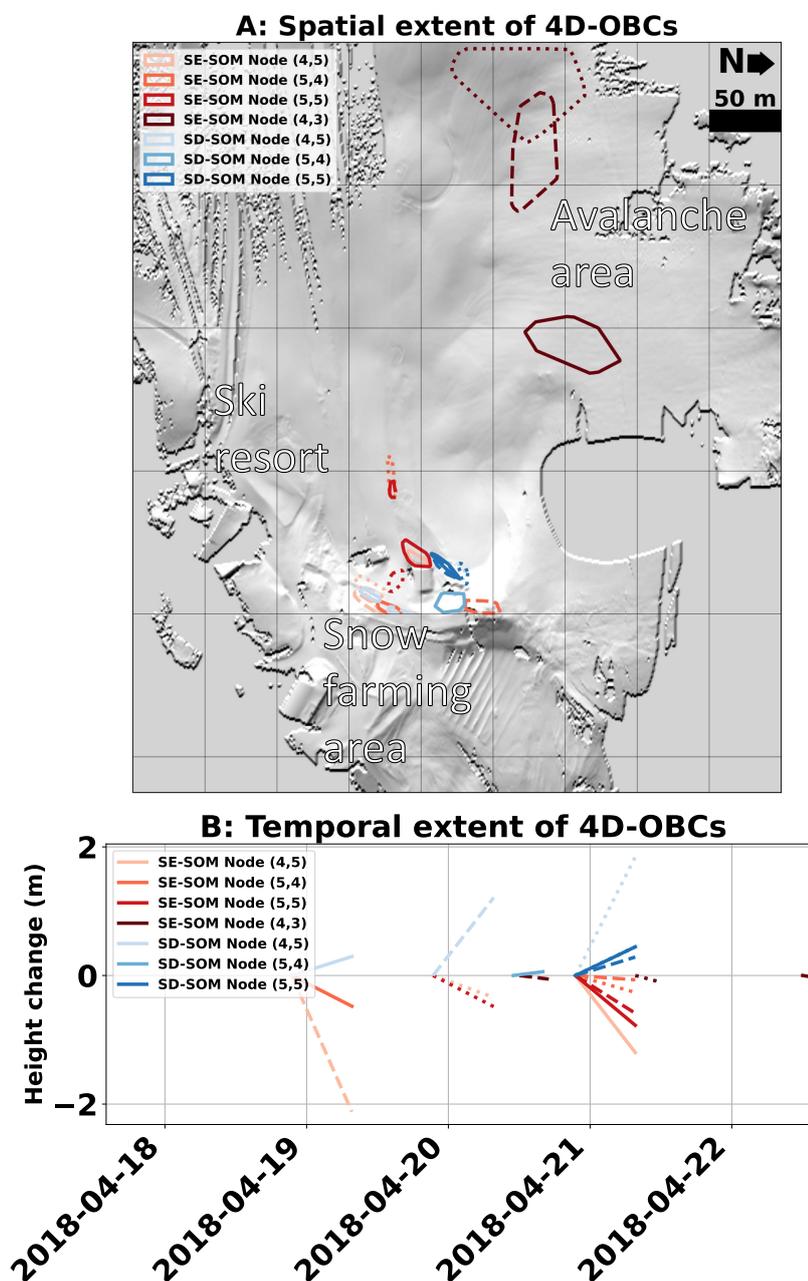
A: Spatial extent of 4D-OBCs

B: Temporal extent of 4D-OBCs

Figure 5.16: A) The convex hulls of the 4D objects-by-change (4D-OBCs) matched in several nodes of the erosion Self-organizing Map (SOM) of the snow cover dataset, representing gradual, long-duration 4D-OBCs. For every node, a maximum of three 4D-OBCs are visualized. B) The seed time series and timing of the same 4D-OBCs of A. The line style relates the spatial outline to the seed time series of the 4D-OBC.

Figure 5.17 shows the SOM for the erosion subset (SE-SOM). Here, the lower right part of the SOM (SE(4,5), SE(5,4), SE(5,5)) also contains the longest and most gradual 4D-OBCs. The graduality and duration decrease from the lower right up to the left. The duration decreases towards a mean duration of 3 h. The nodes representing gradual erosion events (SE(4,5), SE(5,4), SE(5,5)) occur at the same time and around the same location as the gradual deposition events, visible in Figure 5.16. The difference between the 4D-OBCs in these nodes is mainly the result of differences in the magnitude of change and subtle differences in the timing of minimum acceleration. As such, some time series from 4D-OBCs in different nodes appear similar but are not in the same node. Node SE(4,2) also shows 4D-OBCs with similar graduality as nodes SE(4,5), SE(5,4), and SE(5,5), but the magnitude is lower, and the size is larger. These 4D-OBCs initiate during the day and are located on a south-facing slope further away from the ski resort, near the avalanche area. These 4D-OBCs, therefore, likely represent the compaction or ablation of snow.
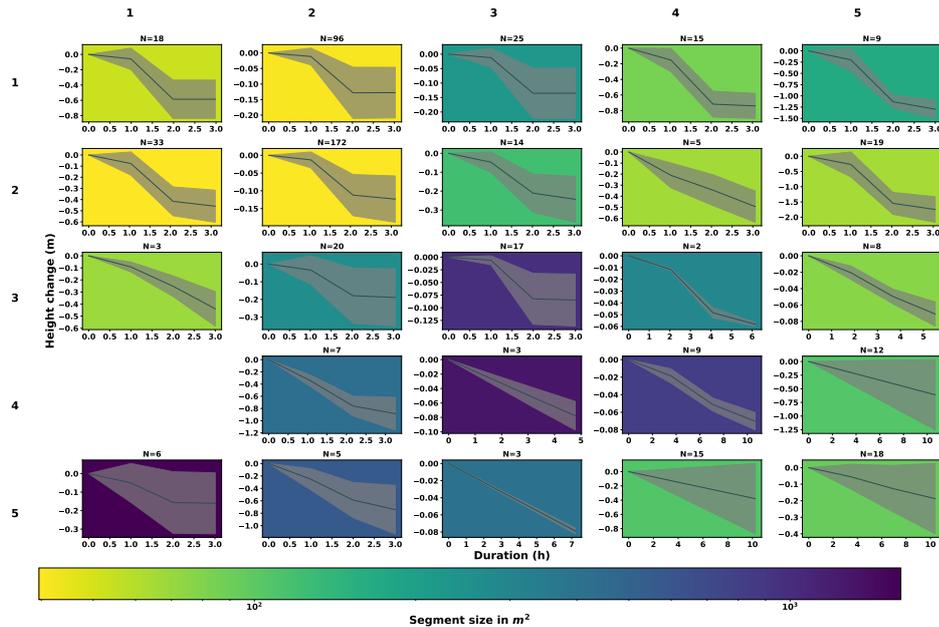

Figure 5.17: Visualization of the erosion Self-organizing Map of the snow cover dataset (SE-SOM). Trained on the resampled time series, duration, timing of minimum acceleration, maximum height change, area under the curve, and volume feature. For a detailed explanation, refer to the caption of Figure 5.15.

The distinction between different types of instant short-duration 4D-OBCs is less clear than the distinction between gradual, longer-duration, and instant, short-duration 4D-OBCs. Figure 5.18 shows the spatial and temporal extent of three 4D-OBCs matched to a selection of nodes of the SD and SE-SOM. Several noteworthy groups of 4D-OBCs are visible. First, erosion and deposition 4D-OBCs with a high magnitude (>1 m) and short duration are grouped in nodes SE(1,5) and SD(1,1). These all appear in regions of the ski resort, where snow farming and ski resort works have been observed. These 4D-OBCs thus likely represent bulldozer work. SE(5,1) and SD(1,5) contain relatively large, low-magnitude instant erosion and deposition 4D-OBCs. Figure 5.18A shows that the 4D-OBCs in both nodes are spread out over the area, appearing in the avalanche area. The spatial outlines represent the shape of avalanche transport, represented by one 4D-OBC from SE(5,1) and one 4D-OBC from SD(1,5). These 4D-OBCs occur simultaneously (Figure 5.18B). One 4D-OBC from node SE(5,1) is also located in the same avalanche area, with a shape comparable to the gradual erosion 4D-OBCs in node SE(4,3) (Figure 5.16). Some 4D-OBCs from these nodes are situated in the ski resort.

SE(2,2) and SD(3,1) contain most of the 4D-OBCs. The 4D-OBCs in these nodes are characterized by their instant deposition or erosion of relatively small magnitude ($\sim$0.15 m) and size. The locations of the 4D-OBCs in these nodes are spread out over the area (Figure 5.18A), whereas the time series of the 4D-OBCs in these nodes are comparable (Figure 5.18B). The 4D-OBCs thus represent small-scale deposition and erosion events. As they are not fixed to one location, the underlying processes are likely variable, from compaction/ablation to small-scale avalanches, to anthropogenic activity.
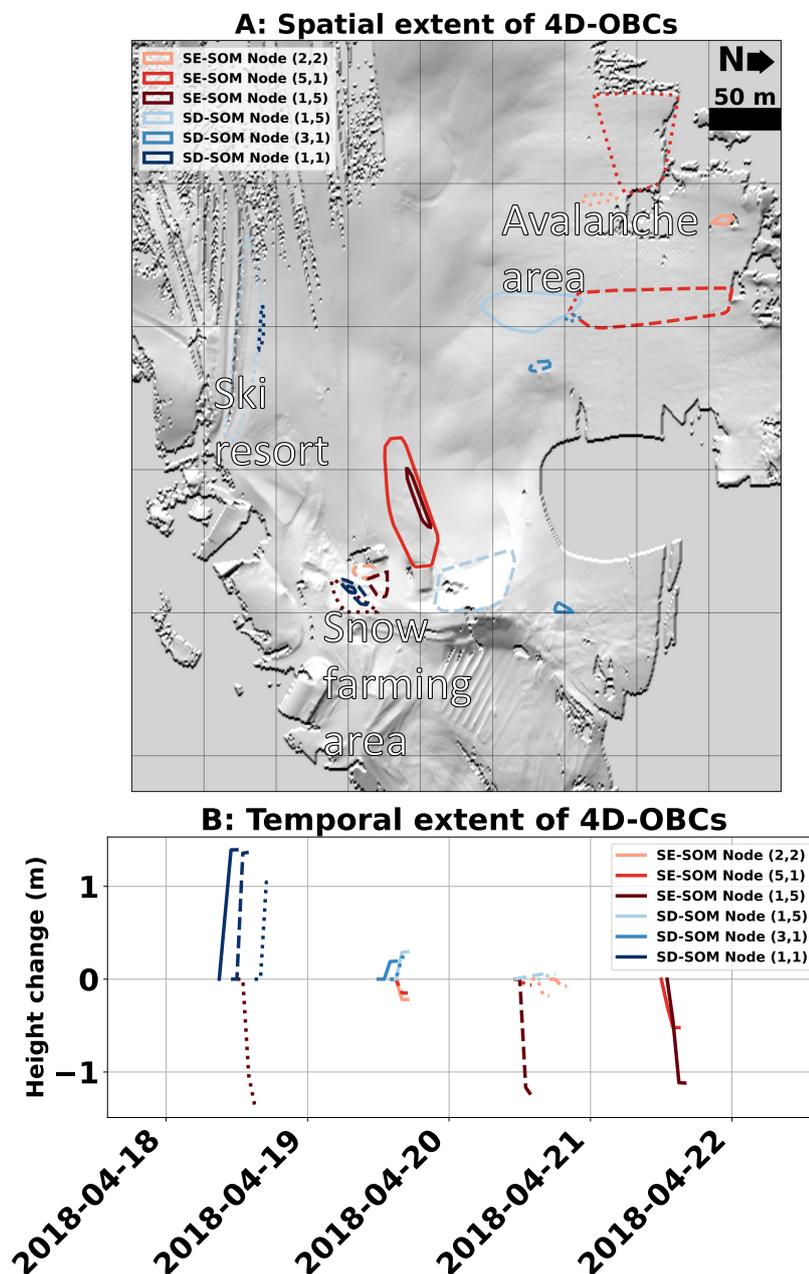
A: Spatial extent of 4D-OBCs

B: Temporal extent of 4D-OBCs

Figure 5.18: A) The convex hulls of the 4D objects-by-change (4D-OBCs) matched in several nodes of the erosion Self-organizing Map (SOM) of the snow cover dataset, representing instant, short-duration 4D-OBCs. For every node, a maximum of three 4D-OBCs are visualized. B) The seed time series and timing of the same 4D-OBCs as in A. The line style relates the spatial outline to the seed time series of the 4D-OBC.

5.3. Optimization of feature and Self-organizing Map (SOM) configuration

This section presents the results concerning the optimization step of the workflow (Figure 5.1). Here, the reasoning behind the final SOM configuration used to obtain the results presented in the previous sections is explained. These optimization results are all based on the beach dataset. The initial selection of 18 features from the 36 features from Section 4.2 is presented in Section 5.3.1, after which the performance under different subsets of this selection of 18 features is shown, and the final choice for features is given (Section 5.3.2). Hereafter, the performance concerning the SOM size and kernel standard deviation, obtained by iterative training of the SOM in different configurations, are presented in Sections 5.3.3 and 5.3.4, respectively. Section 5.3.5 then demonstrates the performance of the SOM algorithm concerning the choice of the subset size of the 4D-OBC dataset and input order during the training of the SOM.

5.3.1. Initial feature selection

We select an initial subset of the features presented in Section 4.2, based on variance and correlation between the features, using the methods presented in Section 4.5.1. Figure 5.19 shows the normalized variance and correlation for each feature.

There is considerable variation in the variance of the features. The duration feature has the largest variance. This feature, together with the cross-shore location of the seed, cross-shore location of the bounding box center, the timing of minimum acceleration, the timing of minimum velocity, the timing of maximum acceleration, and the timing of maximum change show a large variance, after which the largest drop in variance is seen.

The threshold of correlation ρ, above which features are discarded, is set at 0.4. We chose this value, as Ferreira and Figueiredo 2012 found that a value between 0.4 and 0.9 produces adequate feature subsets. A specific value of 0.4 was selected as we want to minimize the number of features to lower the chance of redundant features and increase the interpretability of the results. This means only 18 of the 36 features are being selected. The features based on the timing of events and the polynomial features, in particular, have a large correlation. A large part of these features is thus discarded. The resulting selection of features (Table 5.4) is used to optimize the SOM configuration further.

Table 5.4: Ranked list of selected features based on variance

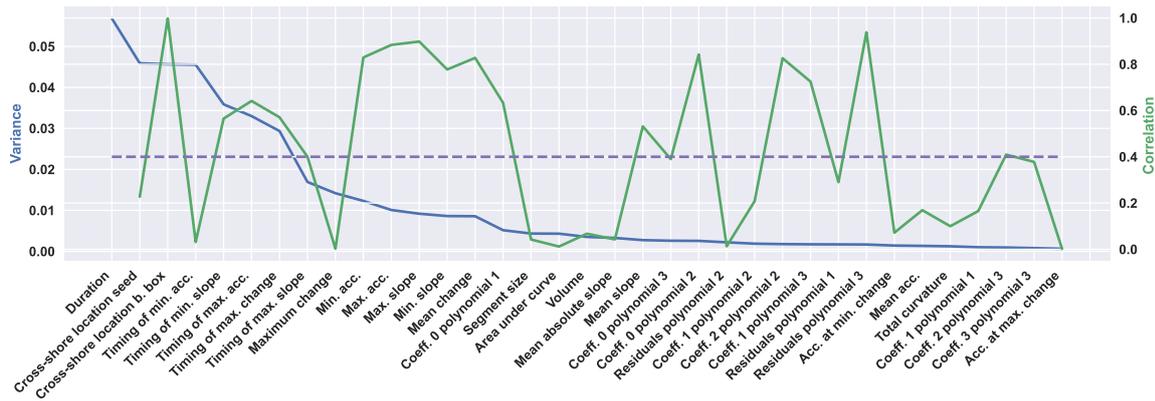| Rank	Features	Normalized variance
1.	Duration	0.057
2.	Cross-shore location seed	0.046
3.	Timing of minimum acceleration	0.046
4.	Maximum height change	0.014
5.	Segment size	0.004
6.	Area under curve	0.004
7.	Volume	0.004
8.	Mean absolute slope	0.003
9.	Residuals polynomial 2	0.002
10.	Coeff. 1 polynomial 2	0.002
11.	Residuals polynomial 1	0.002
12.	Acceleration at minimum change	0.001
13.	Mean acceleration	0.001
14.	Total curvature	0.001
15	Coeff. 1 polynomial 1	0.001
16.	Coeff. 1 polynomial 2	0.001
17.	Coeff. 3 polynomial 3	0.001
18.	Acceleration at maximum change	0.001

Figure 5.19: Normalized variance of the features (in blue) and Pearson's correlation of each feature with the first higher-ranked feature not discarded through the correlation threshold (in green). The purple line indicates the threshold value of feature correlation, above which features are discarded.

5.3.2. Optimization of feature selection

The iterative addition of the ranked features Table 5.4 in combination with the resampled time series feature, results in the mean silhouette scores (S_{sil}), topographic errors (TE), and (normalized) quantization errors ((n)QE) shown in Figure 5.20. These only represent the scores based on the deposition beach dataset. The erosion beach dataset scores appear very similar and are found in Appendix C. Substantial differences are noted in the text.

Relation between silhouette score and feature selection

For all SOM sizes and kernel standard deviations, the S_{sil} is largest when only using the resampled time series and duration feature or combined with the cross-shore location feature. With an increasing number of features, the S_{sil} generally decreases. There is no increase in S_{sil} when adding more features in combination with a larger SOM size. This is in contrast to what is hypothesized, as a larger SOM offers more room for distributing the increasing amount of feature combinations possible. Furthermore, there is no clear peak in S_{sil} after adding the first two features. Meaning that based on the S_{sil}, only using resampled time series and duration or both and cross-shore location, the SOM performs optimally.

Relation between topographic error and feature selection

The TE indicates a different optimal configuration of features. The TE shows large variations when adding extra features to the configuration (Figure 5.20a). No clear general trend in TE is notable. However, the TE is larger when only using the resampled time series and duration combined with a large SOM, indicating that this set of features is not optimal. With an increasing number of features, the TE deviates between approximately 0.1 and 0.2 for most SOM configurations, except for SOMs with 4 and 16 nodes, which under most feature combinations, combined with smaller kernel sizes, result in larger TEs. The smallest SOM shows fluctuations of the TE ranging from 0 to more than 0.6. The small values with some feature configurations are explained by the fact that, with such a small SOM, two of the three possible second BMUs of a sample are neighboring. With this large ratio of neighboring nodes, there is a higher chance that the second BMU is neighboring the sample. The large TE scores we find in some configurations are explained by the fact that if the SOM in the training phase attains a local optimum configuration of weight vector configurations, where two slightly different types of 4D-OBCs are matched to diagonally opposing SOM nodes, there is little room for adaptation in later cycles of training, to bring the two kinds of 4D-OBCs together, as the two other nodes are likely already fitted to different types of 4D-OBCs. These high TE scores thus relate to the SOMs disadvantage of fitting to a local optimum. The TE does not clearly indicate the best set of features but implies that only using resampled time series and duration is not a useful feature set for correctly representing the beach dataset's topological relations.

(a) Silhouette score and topographic error scores for the deposition beach dataset
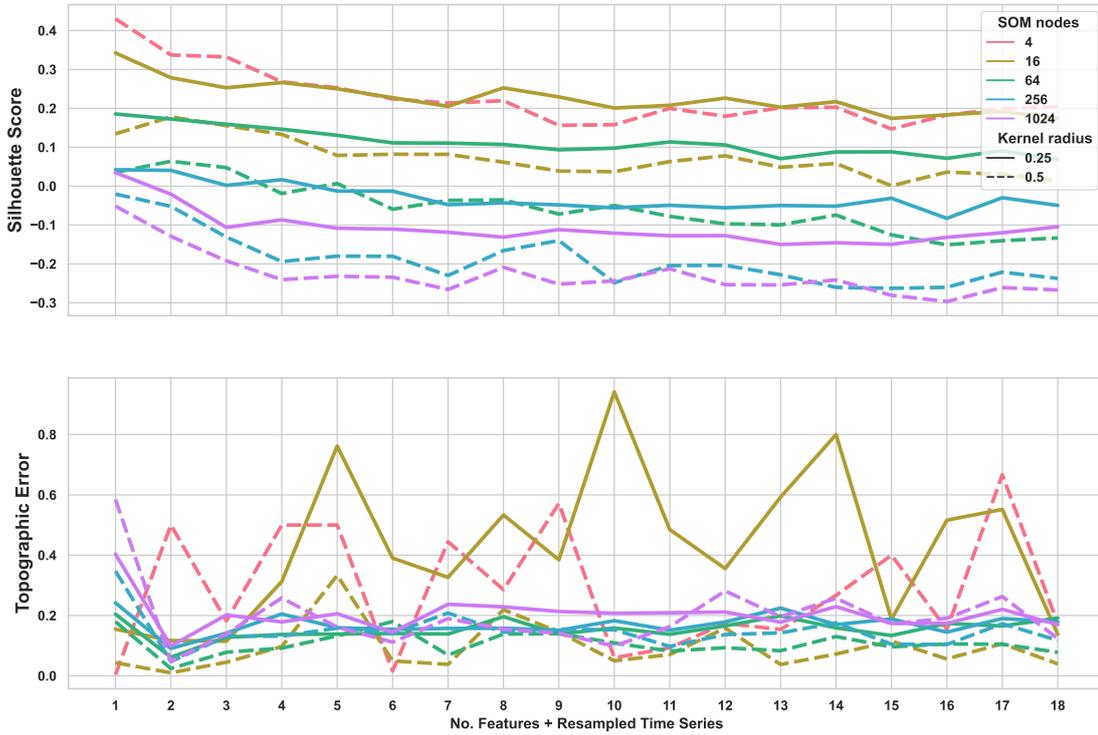
(b) Quantization errors and normalized quantization errors for the deposition beach dataset
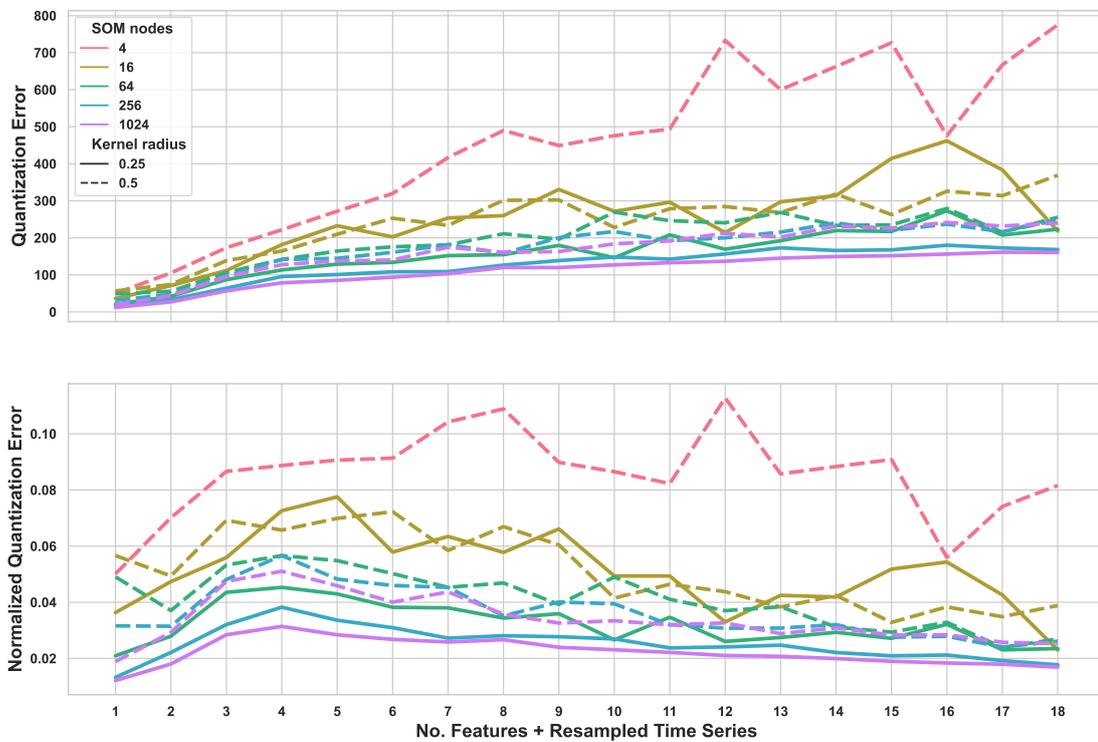

Figure 5.20: Performance metrics in relation to the number of features, SOM size (colors), and kernel standard deviation (line type).

Relation between (normalized) quantization error and feature selection
The QE and nQE show an inverse trend compared to the S_{sil} (Figure 5.20b). The QE increases with an increasing number of features, which was hypothesized, as an increasing number of features adds dimensions to the distance computation, thus increasing the distance between samples per definition. The nQE, however, rises until 3-6 features are added on top of the resampled time series and then drops again with an increasing number of features. For example, for a SOM with 16 or 64 nodes and a quarter of the map width as kernel standard deviation, the nQE is smallest when using all features, which can, therefore, be regarded as an optimal feature selection.

Proposed feature selection based on performance metrics
The S_{sil}, TE, QE, and nQE indicate that either a feature selection of resampled time series, duration, and cross-shore seed location or a selection of all features are best used for our dataset.

The SOMs constructed with a feature selection of resampled time series, duration, and cross-shore location (SOM A, Figure 5.21), and all features (SOM B, Figure 5.23) are evaluated for their effectiveness in characterizing surface activities on a sandy beach. SOM A shows notable characteristics such as sorting on duration, increasing diagonally from the bottom left to the top right corner, and a correlation between resampled time series, duration, and segment size. However, a large variance in the time series of height change within the 4D-OBCs grouped in nodes A(6-7,1-8) occurs, which indicates that these nodes might not represent the same surface activities (assuming equal surface activities have similar time series).

In contrast, SOM B has a lower variance in the time series of most nodes and shows better performance in grouping and sorting on segment size and time series but worse on duration. Therefore, using all features is deemed the better choice for representing surface activities in the beach dataset. However, this configuration is not necessarily optimal for the use case, as the temporal dimension dominates the organization of the SOM, resulting in similar intertidal sandbars being separated in the SOM (Figure 5.24). Thus, this configuration may not be suitable for investigating variations between different types of intertidal sandbar activities. It can, nevertheless, be beneficial for examining surface activities less defined by their spatial characteristics.

Final optimized feature selection
Altogether, these observations indicate that our data-driven selection of features based on S_{sil}, TE, and (n)QE does not result in a proper feature selection for our use case. We, therefore, manually inspect the SOM following an intuitive selection of features, namely resampled time series + the first 7 features: duration, cross-shore location, timing of minimum acceleration, maximum height change, segment size, area under the curve, and volume. With this selection, the ratio between spatial, temporal, and spatiotemporal features is as equal as possible when selecting based on feature ranking. Thus, both dimensions get an approximately similar weight.

In Figure 5.22, the SOM trained on the resampled time series + 7 features is visualized (SOM C). One can identify a similar pattern of sorting on duration as in SOM A. However, the duration of the mean 4D-OBCs increases from row 7 upwards. Furthermore, there is no distinct region of nodes in which the mean resampled time series shows a large variance and high magnitude. There is a region with a large maximum height change in the average 4D-OBCs (see C(7-8,6-8)), but here the variance of the resampled time series is significantly smaller.

Additionally, clear grouping on segment size occurs, with some large mean segment sizes (e.g., C(5-6,3)). This is comparable to what was noted in SOM C. If we extend the comparison to SOM B by identifying in which SOM region the three intertidal bars end up, we find that they are close together (Figure 5.24a). 4D-OBCs 7 and 111 are near in the SOM and have the most similar seed time series of height change, shape, and cross-shore location. 4D-OBC 3 appears slightly further away, but this is appropriate, as its seed time series of height change also has a somewhat different magnitude and shape. Therefore, this SOM enables the investigation of characteristics of different intertidal bar activities, as the intertidal bar 4D-OBCs appear close in the SOM. This implies that other types of surface activity are likely also closer together in the SOM. The intuitive selection of the first 7 features is therefore shown to be the most useful for our use case and is consequently used for further analyses.
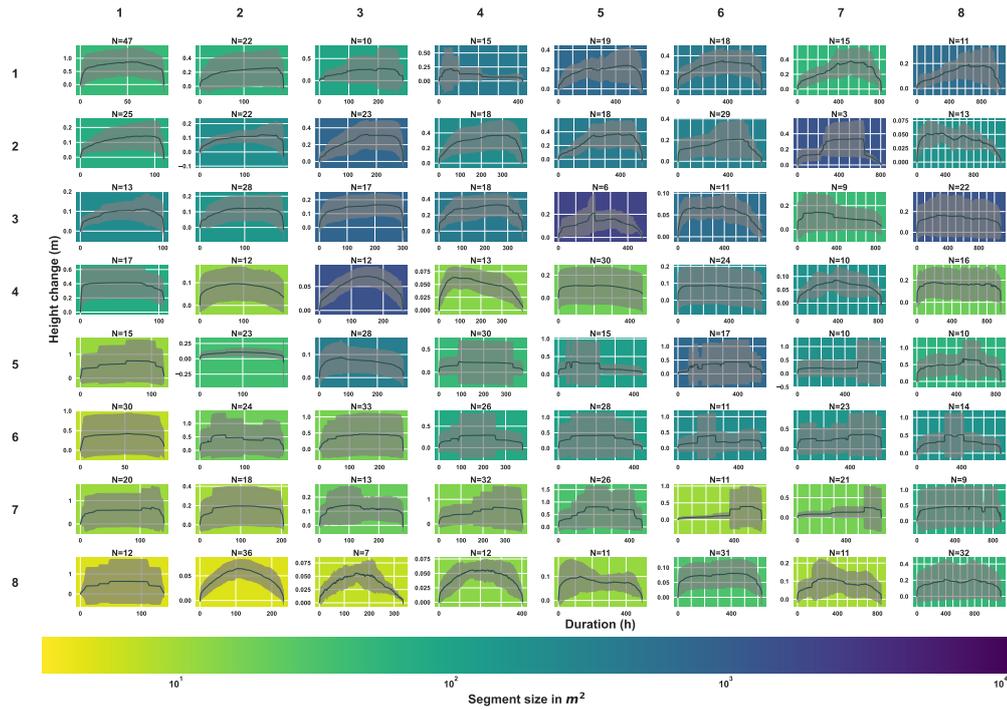
Figure 5.21: Visualization of the deposition Self-organizing Map (SOM), trained using the resampled time series, duration, and cross-shore location of the seed as features. Each plot represents a SOM node. The x-axis of each plot represents the mean duration of the 4D objects-by-change (4D-OBCs) in the node. The y-axis of each plot represents the mean height change of the 4D-OBCs in the node. The limits of both vary between each graph. The black curves show the mean time series of the 4D-OBCs in each node, and the grey area is the standard deviation. The background colors represent the mean segment size per node (on a logarithmic scale). Nodes without graphs represent nodes where no sample is matched after training.
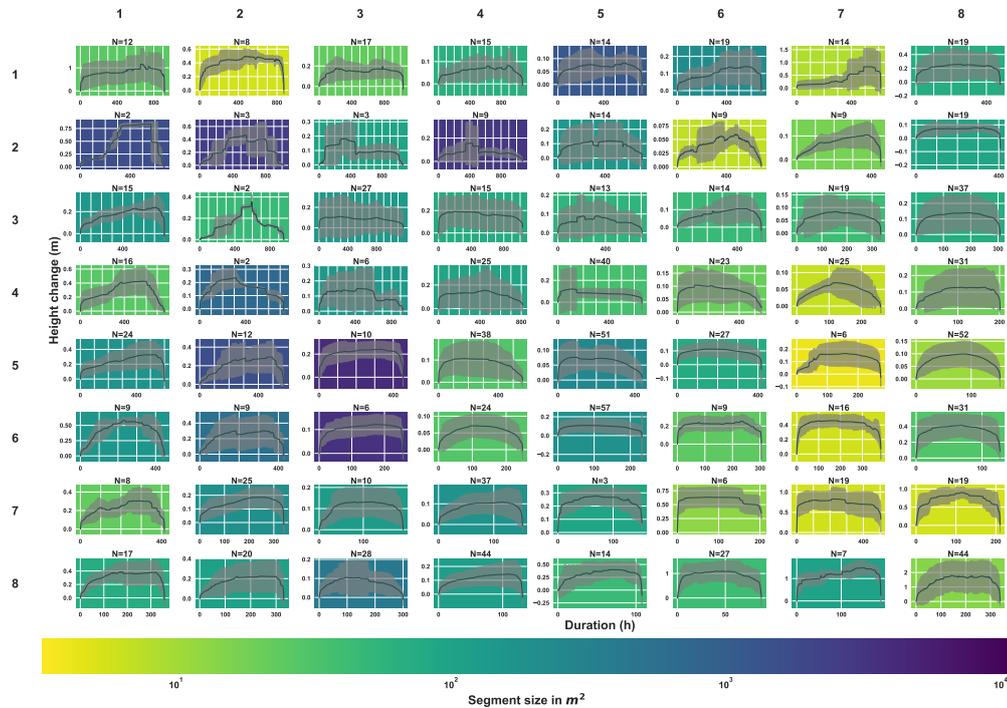

Figure 5.22: Visualization of the deposition Self-organizing Map (SOM), trained using the resampled time series and first 7 features in the variance ranking (see Table 5.4). Refer to Figure 5.21 for a detailed description.
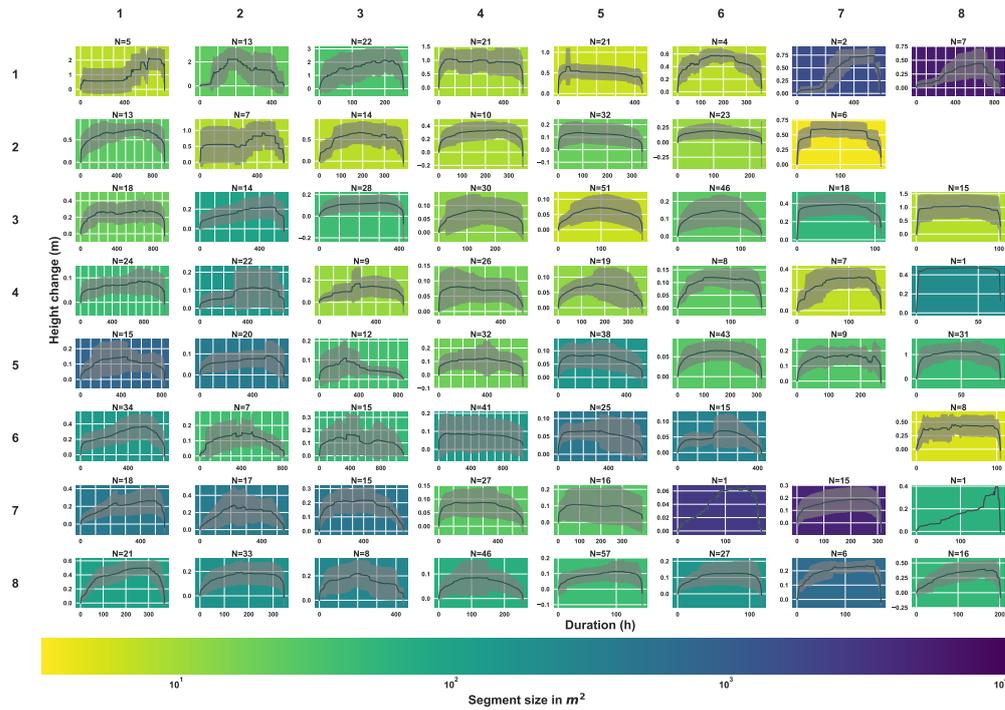
Figure 5.23: Visualization of the deposition Self-organizing Map (SOM), trained using the resampled time series and all 18 features obtained after the first features selection (see Table 5.4). Refer to Figure 5.21 for a detailed description.
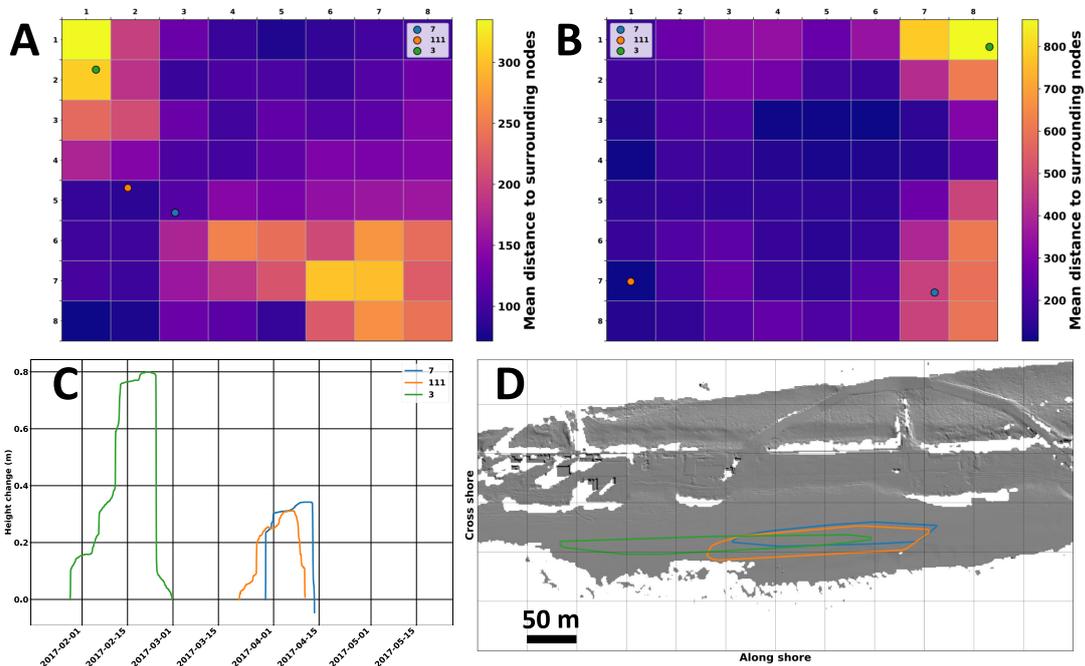

Figure 5.24: A: the locations of the best matching unit (BMU) of three 4D objects-by-change (4D-OBCs) manually identified as representing intertidal bars, in the SOM trained with the resampled time series + the first 7 features presented in Table 5.4. The grid color indicates the mean feature distance of the node to its neighboring nodes. B: the locations of the same three 4D-OBCs, on a SOM trained on the resampled time series + all features presented in Table 5.4. D: the time series of height change of the seed of the three 4D-OBCs, 7 and 111, appear similar, while 3 shows a similar shape to 111 but with a higher magnitude. D: the cross-shore locations and convex hulls of the spatial segments of the same 4D-OBCs.

5.3.3. Optimization of SOM size

The five SOM sizes; with 4, 16, 64, 256, and 1024 nodes; attain the S_{sil}s, TEs, and (n)QEs as visualized in Figure 5.20. Again, only the scores for the deposition beach dataset are visualized, but the scores for the erosion dataset are comparable, and differences are noted in the text.

Relation between silhouette score and SOM size

The S_{sil} generally decreases with an increasing SOM size. This is logical as a larger SOM inhibits a larger area over which the data can be distributed, and thus 4D-OBCs that appear in similar nodes in smaller SOMs can get assigned to close but different nodes. This decreases the internode distance, accompanied by only a small decrease in intranode distance, which is not necessarily a poor performance or a reason for which one should choose a smaller SOM. It shows that one has to find a balance between either achieving great internode separation, accompanied by a lower amount of detail, or vice versa. What is, however, notable is that with a SOM larger than 64 nodes, the S_{sil} becomes negative for most configurations, indicating that most of the 4D-OBCs are not assigned to the node of which the mean feature vector of 4D-OBCs is closest in features space. Thus, the trained weight vector of the node to which most of the 4D-OBCs are assigned to lies further in feature space than the mean feature vector of another node, while the weight vector of this other node is further away. Therefore, the weight vectors of the SOM do not represent the dataset well, as they do not represent centroids of clusters in the dataset. It is thus clear that SOMs of these sizes are not useful for our beach dataset, as most of the weight vectors of these SOMs show only slight variations, which are identified as not characterizing surface activities. The fact that the S_{sil} does not increase or even stabilize with increased SOM sizes implies that the beach dataset is well distributed, and more nodes only add seemingly random nodes, not representing 4D-OBCs and thus not representing different types of surface activity. Sixty-four nodes is deemed the maximum valuable size for the beach dataset.

Relation between topographic error and SOM size

The TE is unstable at smaller sizes (4 and 16 nodes). This relates to the larger chance of attaining a local optimum with these smaller SOMs (see Section 5.3.2). The SOM of size 16, with an initial kernel standard deviation of 0.5, does not show this unstable behavior, likely due to the large initial kernel standard deviation, creating a lower chance for local optimum configuration. From SOMs with a size of 64 and larger, the TE increases in most feature and kernel configurations. This is, however, not the case for every feature configuration (e.g., TE for the deposition dataset, with resampled time series + 6 features and initial kernel standard deviation 1/4), and for any configuration, the performance also varies between the erosion and deposition dataset. The TE thus indicates that the dataset is well represented either by a SOM with a large initial kernel standard deviation and 16 nodes or larger sizes with any of the two kernel radii.

Relation between (normalized) quantization error and SOM size

The QE and nQE are both decreasing with increasing SOM size, which was hypothesized, as larger SOMs contain more nodes and thus more area over which the 4D-OBCs can be distributed. Fewer 4D-OBCs are matched on each node, thus lowering the chance of a large internode variance, with the resulting large QE and nQE. For the initial kernel standard deviation of 1/4, with resampled time series+7 features, the largest decreases in QE and nQE occur between sizes 4 and 16, and 16 and 64. The drop in QE and nQE from 64 to 256 and 256 to 1024 are approximately equal and significantly smaller. Thus, the larger SOMs show decreasing intranode distance, but this increase is not proportional to the increase in size.

Final optimized SOM size

The SOM size of 64 offers the best representation of the dataset, as the S_{sil} remains above 0, the TE is comparable to other sizes, and the QE does not decrease significantly with larger sizes. On top of this, a SOM with 64 nodes is still easily visualized and investigated, in contrast to a SOM with 256 or more nodes.

5.3.4. Optimization of SOM kernel standard deviation

A larger kernel size, in combination with equal SOM size and features, in general, shows a lower S_{sil} and TE and a higher QE and nQE (Figure 5.20). This is because, with a larger kernel size, every 4D-OBC has a larger radius of influence on the weight vectors of the SOM. Thus, all the weight vectors deviate more towards feature vectors that occur more often in the data, i.e., the denser dataspace. Therefore, topological preservation is better, hence the lower TE. However, the weight vectors of each node also appear more similar, hence the lower S_{sil} and higher QE and nQE. Figure 5.25 shows a SOM with 64 nodes trained on resampled time series+7 features, with an initial kernel width of 1/2 times the SOM width. A lot of the nodes here (e.g., (7,6), (5,4), and (1,3)) have a large variance in the resampled time series. With the same configurations but using an initial kernel standard deviation of 1/4 times SOM width (Figure 5.22), this variance is lower in most nodes. Furthermore, a larger amount of nodes with extreme segment sizes and durations appear, thus indicating that outliers and rare 4D-OBCs are also represented better in a SOM with a smaller kernel standard deviation. We, therefore, choose to use an initial kernel standard deviation of 1/4 times the SOM width, as we want to enable the detection and characterization of rare surface activities.
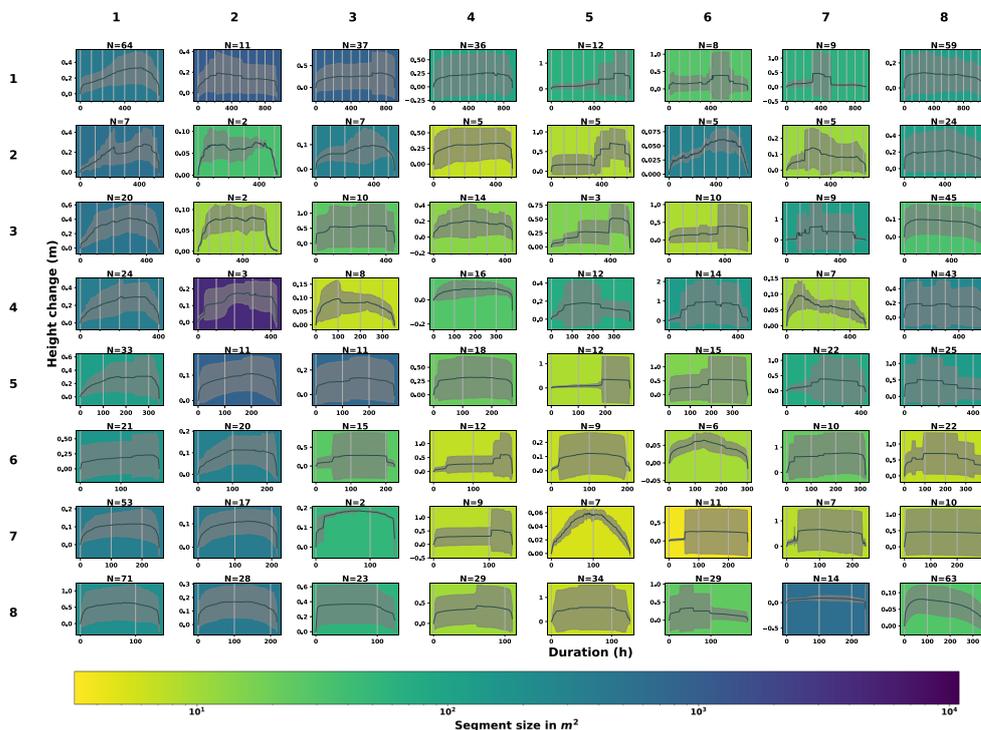

Figure 5.25: Visualization of the Self-organizing Map (SOM) trained on the resampled time series + first 7 features from Table 5.4 using a kernel size of 1/2 of the map width. For a detailed explanation of the visualization and color scheme, see Figure 5.21.

5.3.5. Optimization of SOM training subset size and order of input

The SOM configuration, with 64 nodes, initial kernel standard deviation of 1/4 times the SOM width, and resampled time series + 7 features, is used to assess the influence of training using an MDA subset and MDA order on the sorting and grouping of the SOM, with the model optimization methods as presented in Section 4.5.3.

The MDA subset size tested contains 300 of the 1205 deposition 4D-OBCs. In Figure 5.26, it is visible that this subset size shows an approximately even distribution over the first two principal components of the full deposition beach dataset. These first two principal components explain only 66% of the total variance of the data; how well the subset represents the total variance of the dataset can thus not be determined based on only these two. However, it was found that the subset still displayed an approximately even distribution for the third principal component. The subset should thus represent rare and outlying 4D-OBCs and more common 4D-OBCs evenly, ensuring no overfit on denser dataspace.
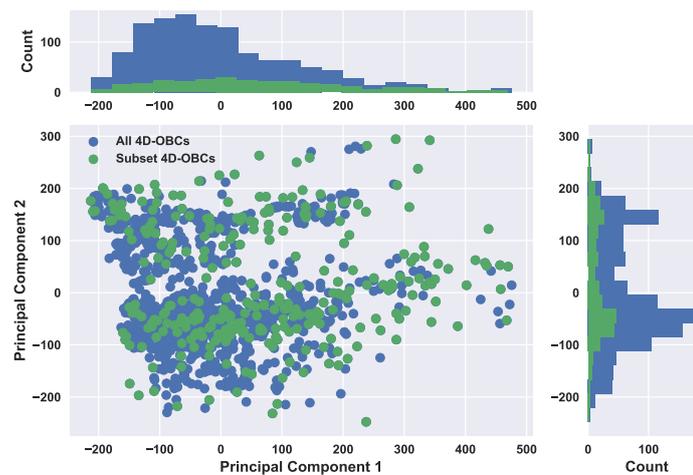

Figure 5.26: Distribution of the 4D objects-by-change in the beach deposition dataset, along the first two principal components, in blue. Distribution of the subset of this dataset, sampled through maximum dissimilarity sampling, in green.

In Figure 5.27, the weight vectors of the SOMs obtained using the full deposition beach dataset for training, with a random input order, MDA ranked input order, and flipped MDA ranked input order; and using the MDA subset and order for training, are projected onto the first two principal components of the full dataset. These weight vectors should represent characteristic feature vectors of the 4D-OBCs in the deposition dataset. It is noticeable that for none of the configurations, the weight vectors span the whole distribution of the dataspace, i.e., the most outlying weight vectors are not representing the most outlying 4D-OBC feature vectors. The use of an MDA subset, however, does result in the widest distribution of weight vectors, which was expected. The flipped version of the MDA order results in the narrowest weight vector distribution.

In Figure 5.28, the density plots of the same weight vectors over the first three principal components explaining 81% of the variance, are shown together with the density plot of the full 4D-OBC beach deposition dataset to investigate further how well the weight vectors represent the dataset. The distribution over the first principal component of the weight vectors trained with the MDA subset best represents a normal distribution with a center around the center value of the principal components of the full dataset. Meaning the weight vectors are the most evenly distributed. When using MDA ranked input order and the full dataset during training, the distribution of the weight vectors along the first principal component also has a wider and less skewed distribution. However, the highest data density still approaches the peak data density of the full dataset. When a random input order is used, this peak is even stronger, but it is still noticeable that the weight vectors do not represent the exact data distribution. The same goes for the SOM trained on the full dataset with flipped MDA ranked input order, but this distribution represents the full dataset the closest. For the second and third principal components, the relations between the distributions are comparable to the first. However, it is noticeable that the MDA subset
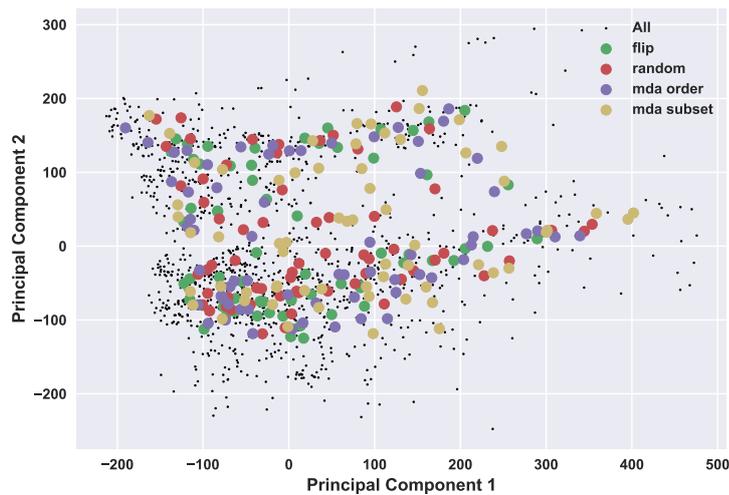
Figure 5.27: Distribution of the 4D objects-by-change (4D-OBCs) in the beach deposition dataset, over the first two principal components, in black. Distribution of the weight vectors of the Self-organizing Maps; trained using the maximum dissimilarity subset (MDA, in yellow), the full 4D-OBC dataset with MDA ranked input order (in purple), the full 4D-OBC dataset with random input order (in red), and the full 4D-OBC dataset with the flipped version of the MDA ranked input order (in green).

causes the weight vector distribution over the third principal component to show a large density at the higher values of the principal component, where the data itself has a low density, indicating that these weight vectors better represent the outlying 4D-OBCs.

Using an MDA subset effectively mitigates the problem of overfitting on a denser dataspace, however the SOM then is largely dominated by anthropogenic surface activities, and show the wide range of characteristics of these high magnitude outlying events (Appendix D). As we still want to enable the identification of rare and outlying 4D-OBCs, but are more interested in variations in surface activities different than anthropogenic we choose the MDA order, as this offers a balance between overfitting on the denser data space and overfitting on outliers.
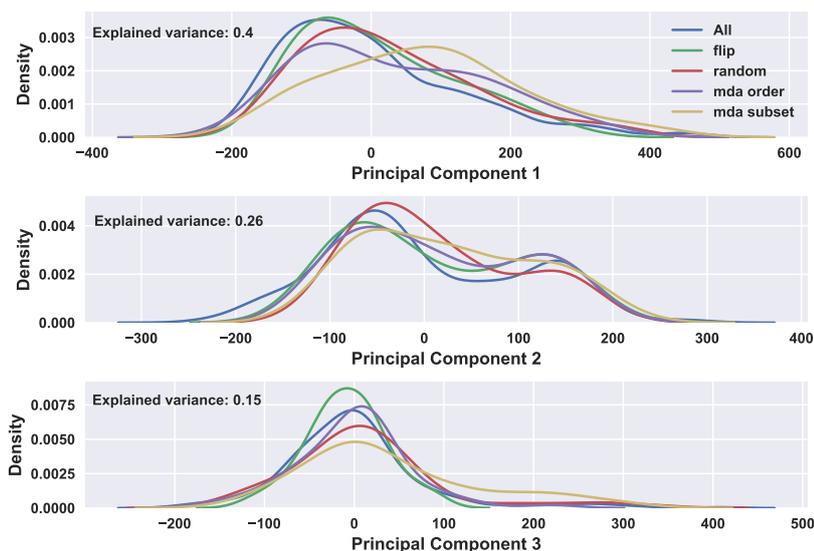
Figure 5.28: Density plots of the distribution of the 4D objects-by-change (4D-OBCs) in the beach deposition dataset, over the first three principal components, in blue. Density plots of the distribution of the weight vectors of the Self-organizing Maps; trained using the maximum dissimilarity algorithm subset (MDA, in yellow), the full 4D-OBC dataset with MDA ranked input order (in purple), with random input order (in red), and with the flipped version of the MDA ranked input order (in green).

5.4. Conclusions

In this chapter, the results at two levels of the workflow established in the previous chapter are presented. We also show the results of the SOM optimization methods.

- The hierarchical clustering of the SOM nodes with two distance threshold levels result in the identification of 22 and 12 clusters of surface activity for the erosion SOM, and 20 and 8 clusters of surface activity for the deposition SOM of the beach dataset.

- The clusters show interpretable relations and characterizations with respect to natural drivers. The clusters are successfully used to study the destruction of intertidal bar systems at two points in time.

- An optimized version of the SOM algorithm is used to obtain and characterize detailed groups of surface activity from the 4D-OBCs of the beach and snow cover dataset.

- For the beach dataset, these detailed groups of surface activity are physically interpretable and offer a good grouping of the validation dataset. The SOM is useful for exploring variations in intertidal sandbar deposition surface activities. The preservation of topological relations of higher dimensional feature space is slightly worse in the erosion SOM compared to the deposition SOM.

- The optimized methods applied on the snow cover dataset result in poorer performance, as 4D-OBCs displaying different types of surface activity are grouped in one node.

- The optimization methods result in the selection of an optimal configuration of the SOM based on the beach dataset. For the beach dataset this optimal configuration uses a feature selection of 8 features, 64 nodes, a kernel standard deviation of 1/4 of the SOM width, and an MDA input order during training. The number of nodes is set at 25 for the snow cover dataset, after investigation of a number of sizes.

- The optimized feature selection is not determined based on the performance metrics, as these do not indicate a useful feature selection, but rather on the basis of manual investigation of the SOM under various configurations.

Discussion

In this chapter, the methods and results of our workflow are discussed. In Section 6.1, the detailed groups of surface activities obtained with the SOM are discussed, as well as a sensitivity analysis of the parameters and the transferability of the SOM method. Section 6.2 presents the discussion concerning the hierarchical clustering of the SOM nodes. Here we also discuss the intercorrelations between the clusters and the correlations with natural drivers. In Section 6.3, we compare our results to previous studies on monitoring surface activity and present a hypothetical comparison to other classification methods that could be used.

6.1. Self-organizing Map (SOM) for the identification and characterization of detailed surface activity groups

SOM for exploring the feature distribution in 4D-OBC sets

A SOM was used to sort, group, and characterize different 4D-OBCs to identify detailed groups of surface activities and explore what characterizes the data space of the 4D-OBCs. The results of the beach dataset show that using a SOM, we can identify the distribution of features among the 4D-OBC dataset. The SOM enables exploration of the relative importance of the features for the characterization of 4D-OBCs, as the more important features exhibit the largest global ordering and grouping over the SOM. Both for the deposition SOM (D-SOM, Figure 5.9) and erosion SOM (E-SOM, Figure 5.11), the amount of global ordering decreases with an increasingly lower ranking of a feature (Table 5.4). The lower-ranked features define local ordering. Thus, through inspection of the feature distributions, one can identify how different surface activities are characterized. This is especially distinct for the D-SOM, and for the E-SOM, the sorting on some features is weaker.

The sorting based on the resampled time series is weaker, as similar shapes of time series appear throughout the SOM (as shown in Figure 5.10). This could imply that the resampled time series is not crucial in characterizing erosion surface activities. However, this idea is disproven through hierarchical clustering of the E-SOM nodes. This results in the formation of clusters that bring together nodes not located near each other in the E-SOM. The nodes in this cluster are well characterized by their gradual time series shape (Figure 5.10). This indicates that the E-SOM is not sorted as well in terms of its topological order in feature space as the D-SOM, even though their topographic errors are very similar (Table 5.3). This might be because the choice for SOM configuration is equal for the deposition and erosion beach datasets, which show different characteristics. The number of 4D-OBCs in both datasets differ (816 vs. 1205). The kernel standard deviation chosen as 1/4 might, therefore, be too small for the erosion dataset, causing it to converge to a local optimum. This problem is encountered more often in SOMs (Kohonen 2001).

By exploring these feature distributions (Figure 5.9, Figure 5.11), we can also identify which of the selected features are likely redundant for the grouping of 4D-OBCs into different types of surface activity. These are the area under the curve and volume feature, which considerably correlate with the segment size and maximum height change feature. Thus, these features should have been discarded by our

feature selection method. However, as this method only measures the correlation between two features neighboring in the feature ranking, some high correlations might be missed. Discarding these correlated features might increase the performance, as it lowers the apparent fit to the segment size and maximum height change feature. Exploring the SOM's feature distribution, one can also identify features that show some correlation but are not as global as the previously mentioned ones. For example, both large and small 4D-OBCs occur at any cross-shore location, but larger segment sizes are found more often further from the shore (Figure 5.9 and Figure 5.11).

SOM for characterization of surface activities

By examining the mean feature vectors in a SOM node, we can identify detailed characteristics of different surface activities. Combining this information with the number of matched 4D-OBCs, we can determine both common and rare surface activities. For example, in the D-SOM, we can identify two frequent 4D-OBCs in the backshore area with moderate size and short duration (nodes D(6,5) and D(5,5) in Figure 5.8), but with different time series shapes. These likely represent aeolian deposition and erosion activities but with different recovery intensities. This could be due to one being sheltered from the wind for some time until strong wind occurs (D5,5), while lighter wind impacts the other throughout deposition and erosion. This demonstrates the ability of the SOM to study the different characteristics of seemingly similar surface activities. If more detailed characterization is desired, a new SOM can be trained on only the 4D-OBCs situated within one of these highly populated nodes, creating a hierarchy of SOMs for more in-depth exploration of the surface activities, like a workflow proposed by Palomo et al. 2012.

SOM for identification of similar surface activities

Types of surface activities similar to one type of surface activity of interest can be characterized by inspecting the 4D-OBCs in the SOM region surrounding the node where the 4D-OBC, describing the surface activity of interest, is situated. When the feature distance between the weight vectors of the SOM nodes is considered, the SOM becomes even more useful. This is indicated by our evaluation using the validation dataset and the evaluation regarding the characterization of intertidal bars. The performance concerning this notion is different for the D-SOM and the E-SOM due to variations in the preservation of topological order.

Intertidal bar depositions are well separated from backshore and dune depositions, as indicated by the distribution of the validation dataset over the D-SOM (Figure 5.12B). These intertidal bar 4D-OBCs appear in different parts of the SOM and are well divided by regions of relatively large feature distance between the SOM nodes. One main apparent issue with this SOM is that two nodes contain anthropogenic deposition, but these are not neighboring. This is not necessarily a large problem as these nodes are still well separated from non-anthropogenic 4D-OBCs, by the large feature distance between the neighboring SOM nodes. However, if one wants to inspect the variations in anthropogenic depositions through inspection of a SOM node surrounding one 4D-OBC of interest, not all anthropogenic 4D-OBCs can be found. This is caused by the different durations of the 4D-OBCs in these two nodes. One contains anthropogenic depositions with a long duration, whereas the other contains anthropogenic events with a short duration. This variation in duration indicates that one bulldozer pushed sand to one position, and in some cases, the sand is transported again after a short time and, in other cases, after a long time. Lowering the weight of the duration feature in the computation of the SOM or adding a rectangular shape feature could bring these nodes closer together, as the rest of the features are very similar (Figure 5.9). Nevertheless, it is clear that the SOM allows for the distinction and investigation of different characteristics of surface activities in the broad group of anthropogenic surface activities.

All the 4D-OBCs manually identified as intertidal bars, as well as the beach berm deposition events, appear in close proximity in the SOM, in contrast to the broader group of anthropogenic depositions. One can thus obtain characterizations of different types of intertidal bar depositions through inspection of a SOM region surrounding a known intertidal bar deposition 4D-OBC (Figure 5.13 and Figure 5.14). All 4D-OBCs shown here are related to similar types of surface activity, namely intertidal depositions, except for one node D(1,1). However, this node shows a considerable feature distance compared to the other nodes (Figure 5.12B). In future work, this feature distance could be more appropriately visualized by computing and visualizing the actual feature distance between the weight vectors of two nodes instead of the mean distance to all nodes. In this way, the identification of similar nodes would be easier.

The neighboring nodes show similar surface activities but with variations in size, location, duration, magnitude, and time series shape. The SOM thus allows investigation of the variety of characteristics that a surface activity and the 4D-OBCs representing a particular surface activity can exhibit while also allowing to find similar activities occurring at different points in time by investigating the nodes surrounding a sample of interest.

Differences between erosion and deposition SOM

The performance concerning the validation dataset of the E-SOM is lower than for the D-SOM (Figure 5.12A). Like for the D-SOM, intertidal 4D-OBCs appear close together and are thus correctly grouped. However, anthropogenic, beach berm, and foredune erosion are all not situated in neighboring SOM nodes. The different nodes in which each of the 4D-OBC types are located show large similarities in their mean feature vectors even though they are not neighboring in the SOM. This is confirmed by clustering the nodes using hierarchical clustering. Several nodes far apart in the SOM are clustered, even at a low distance threshold level. As mentioned in the first paragraph, this likely relates to the fact that the E-SOM does not converge to a global optimum. This relates to the selected SOM configuration being equal for the deposition and the erosion beach dataset.

Issues related to 4D objects-by-change extraction

The spatial extents of the 4D-OBCs in some neighboring nodes differ considerably even though the other features are comparable. This can occur due to incorrect growing of the 4D-OBCs, likely due to missing data. In the case of, e.g., intertidal bars, some 4D-OBCs show similar shapes in their temporal dimension as full intertidal bars 4D-OBCs, whereas their size is not visually remnant of an intertidal bar (i.e., no elongated shape parallel to the coast). Some of these 4D-OBCs are very small, occurring on the edges of the study area or in locations submerged at high tide. These, therefore, represent similar surface activities as larger 4D-OBCs but are not grown to their full extent. This is not necessarily a problem, as one could discard these by not incorporating specific nodes of the SOM that contain many of these undergrown segments.

The spatial and temporal extents of several 4D-OBCs in different nodes overlap in both space and time (e.g., the 4D-OBCs from node D(2,2) and D(2,1), Figure 5.13). This might be due to the method of 4D-OBC extraction. The highest-ranked seed is first used to grow a spatial region. Afterward, a lower-ranked seed not incorporated in the previous segment might be selected. From this seed, a region that incorporates the full extent of the previous segment is grown. One thus obtains two 4D-OBCs representing part of the same surface activity, as noted by Anders et al. 2021. The processing of the 4D-OBCs could be enhanced by merging such segments.

6.1.1. Sensitivity analysis

The SOM optimization results show how sensitive the SOM algorithm is to features and parameter variations. It was found that feature selection is a defining factor for the performance of the SOM. The separation into detailed surface activity groups tends to be dominated by a specific dimension of the 4D-OBCs when using too many features derived from one dimension. Therefore, depending on the environment under consideration, one has to obtain a balance in the selection of features. If it is known that the spatial dimension, e,g, cross-shore location, can separate a large number of surface activities, one could, in contrast to what was done here, add more weight to this feature while still incorporating a large number of features from the temporal dimension. Knowledge of the types of surface activity in the environment under consideration is thus essential for accurately grouping a set of 4D-OBCs.

The optimization results also indicate that the performance of the SOM largely depends on the size and kernel standard deviation. However, in contrast to what is proposed by Kohonen 2013, a larger SOM size, in this case, does not necessarily result in better performance. The optimal size largely depends on the dataset size, and with this, in particular, the number of possible surface activities represented by the dataset. Furthermore, the level of detail needed for a particular use case also influences the needed size. Though, when applying a hierarchical clustering algorithm after the SOM, this factor is of minor importance, as one can then identify less detailed levels of separation with the hierarchical clustering algorithm.

The kernel standard deviation is of large importance for obtaining topological order of higher dimensional data space into the two-dimensional grid, as well as reducing or increasing the influence of

outliers on the grouping of the SOM. As is identified by Clark et al. 2020 and Kohonen 2013, if one wants to obtain global ordering, i.e., neighboring nodes contain 4D-OBCs neighboring in feature space, the kernel width is best set as large as possible. However, it observed that this reduces the influence of outliers and how distinct each node's characteristics are. In other words, with a larger kernel standard deviation, the distribution of characteristic weight vectors becomes smoother, and the surface activity groups are thus less well defined. In datasets with fewer rare and outlying samples, a larger initial kernel standard deviation could enhance the ability to identify slight variations in the surface activity characteristics through better topological preservation and prevent the SOM from reaching a local optimum fit to the dataset.

The order of input and training subset selection also largely influences the effect of 4D-OBCs representing rare surface activities. With a random input order, the SOM is overfitted on denser data space and thus more often occurring surface activities. Whereas, with an input order according to the maximum dissimilarity sampling algorithm (MDA) ranking, rare surface activities are better represented by the SOM. When only selecting a subset of the full dataset of the most dissimilar samples, this effect is even more substantial. This implies that when selecting the configuration, one has to consider how important the particular rare surface activities are for the study. If these rare surface activities are indeed outlying events not of importance, the input order is of no importance. However, in our case, rare surface activities, like intertidal bar depositions, are of considerable importance, and thus an input order is essential. Nevertheless, a subset was not used, as this meant overfitting on the most dissimilar samples, which were anthropogenic events, with large magnitude.

It should be noted that the selection of subset size for the training is made based on the first two principal components, whereas this does not explain all the variance. A smaller subset could more evenly capture the full data spread in its weight vectors, and a larger dataset will likely only push the weight vector distribution closer to the denser part of the data space.

6.1.2. Transferability of SOM configuration
The results of the SOM applied on the snow cover dataset show that the optimized methods and features are only partially transferable to a scenario where the 4D-OBCs display different surface activities and their temporal seed segments are obtained differently. We can explore the gradual variations concerning the magnitude of the events with the SE- and SD-SOM and obtain characteristics of rare, high-magnitude surface activities. However, the separation of known surface activities is only achieved in some cases. The avalanches and anthropogenic works are hard to separate with our optimized workflow, as they show comparable sizes and magnitudes. Some geographical constraints or features might be integrated to separate these events, like slope (Schweizer et al. 2003), aspect, and possibly meteorological variables like the amount of solar irradiation at the time of initiation.

Longer-duration and instant short-duration processes are separated by the SE- and SD-SOM. However, a large part of the gradual processes, identified as anthropogenic work during the night, might not be gradual but appear gradual due to the lack of data in this area during the night (Section 3.3.2). This shows that this duration feature does not necessarily distinguish well between different types of surface activity, which was also the case for the anthropogenic deposition events on the beach.

The poorer performance on the snow cover dataset could also be that surface activities in this alpine snow cover area are generally less easily distinguishable by their attributes. In this case, it can be helpful to set a spatial constraint to group the different surface activities, primarily since it is known where avalanches and anthropogenic activity occurs in this particular study area during acquisition. In this sense, solutions like a GeoSOM (Bação et al. 2005) could be more helpful in characterizing the dataset into different surface activities. The GeoSOM first matches samples based on their geographical location and then sorts based on other features.

Application of the SOM on new 4D-OBC sets
Thus, knowledge of which features are likely to be characteristic of certain surface activities, like the cross-shore location for the beach, is essential for accurately grouping the 4D-OBCs into different surface activities. Furthermore, it is beneficial to tune not only the features but also parts of the SOM configuration for every new dataset, as it is shown that the performance differs between the erosion and deposition SOM, even within the beach setting. This tuning should be done based on training using

a few configurations and inspecting the results. It was found that this tuning is preferably done using a small labeled 4D-OBC set, of which it is known which 4D-OBCs should be grouped close together. For smaller datasets, configurations with a larger initial kernel standard deviation are proposed in order to reduce the chance of topographic errors. With smaller datasets, fewer moments of weight vector tuning occur, as the kernel standard deviation is lowered every training cycle (i.e., going through the whole dataset once). Thus a larger chance of local optimum configuration occurs. It should, however, be tested if this larger kernel standard deviation does not cause the different nodes to show weak distinctive characteristics, which is a possibility identified in the sensitivity analysis. If this is the case, a larger number of training cycles might also slightly increase the performance in terms of topological preservation of higher dimensional data space. In that case, more moments of weight vector tuning occur with larger kernel standard deviations.

In conclusion, the results imply that it is best to optimize the SOM size, and kernel standard deviation, for both the erosion and deposition SOM separately, considering differences in the subset sizes and an indication of the expected amount of types of surface activities. When transferred to another sandy beach dataset, it is likely unnecessary to tune all the parameters except kernel standard deviation relating to the dataset size, as the surface activities are expected to be similar. Nevertheless, at other locations, different surface activities may initiate with different frequencies. This is, however, accounted for when using a case-specific MDA order during training. When transferred to a completely different environment it is advised to select a new set of features based on knowledge of the characteristics of the study area-specific surface activities.

6.2. Hierarchical clustering for the identification of broader surface activity clusters

The use of hierarchical clustering for the automated identification of higher-level groups of surface activity is shown to be effective and valuable. Especially for the E-SOM of the beach dataset, it helps bring together groups of nodes close in feature space but wrongly separated in the SOM. With the two levels of thresholds set for the hierarchical clustering algorithms, we can characterize different levels of separation in surface activity.

For the E-SOM, as an example, we inspected four clusters obtained with this method at each distance threshold. At the lower level, the characteristic features of different clusters still showed slight overlaps in feature space, but different variations in backshore and dune erosion are identifiable. For example, one distinct group of dune erosion 4D-OBCs initiates only during high wind speed. At a higher threshold, previously separate groups showing very distinct feature vectors, though still of comparable surface activity, are merged. This allows for the identification of characteristics of broader groups of surface activity.

For the D-SOM, the characteristics and interpretations of all clusters are presented in Table 5.2. At the higher distance threshold of 440, some clusters are well-defined (e.g., clusters 1,4, 5, and 7) by their tight range of characteristic feature values. In contrast, other clusters show a broader spectrum of features (e.g., cluster 3). The largest amount of 4D-OBCs are in clusters with gradual deposition and erosion in the berm, backshore, and intertidal zone. Several clusters represent specific intertidal bar depositions that can be recognized as a large part of an intertidal bar. In contrast, others show 4D-OBCs that contain only parts of intertidal bars (clusters 6 and 7 at t = 440). At the lower threshold level, the clusters show more detailed and specific characteristics, and many of these detailed clusters are still presentable as a specific type of surface activity, characterized mainly by their location, magnitude, shape, and duration. Furthermore, it indicates that we can identify and distinguish the different groups of surface activity that are known to exist on sandy beaches (Section 3.3.1).

In future studies, the two levels of clustering can be used together for subsequent surface activity analysis. For the E-SOM, two clusters (2 and 10), one containing beach berm erosion 4D-OBCs, and the other anthropogenic 4D-OBCs, are merged at the higher level. In contrast, the nodes containing intertidal 4D-OBCs are all in separate nodes at the lower level and merge into broader intertidal erosion clusters at a higher level. Thus, to obtain the same detail for all different types of surface activities, one could use the intertidal clusters at a higher threshold level and the backshore and dune clusters at a lower level. This difference in the size of the hierarchical clusters occurs because some surface

activities are defined by a larger range of higher feature values (intertidal erosion activities). In contrast, others are characterized within a tight range of feature values (beach berm and anthropogenic erosion). Thus, if the same amount of detail in surface activity events is needed over the whole area, one has to choose different levels of feature thresholds.

6.2.1. Intercorrelation of clusters of surface activity and correlation with natural drivers

The results of the correlation between the hierarchical clusters and natural drivers (Section 5.1) depict that for the sandy beach, physically interpretable relations between the 4D-OBCs of the different clusters can be obtained, as well as relations between the 4D-OBCs of different clusters and natural drivers. These relations are, in most cases, comparable to what was expected based on existing knowledge of surface activity on a sandy beach but also suggests some relations deviating from what is known.

The 4D-OBCs classified as intertidal sandbar deposits were initiated and grew in height during lower wave conditions, according to what is expected based on literature (Robin et al. 2009). Two periods with the initiation of intertidal bar 4D-OBCs are observed. However, contrary to expectations, the destruction of the intertidal bars did not only occur during heavy weather conditions. The first of the destructions did, which is in accordance with the observations by Vos et al. 2020. Some slight differences are, however, observed. Our results suggest that the destruction of the intertidal bar occurred during low water instead of high water. Furthermore, our results imply that the sediment transport after destruction mainly results in deposition activity in the backshore and beach berm due to hydrodynamic or aeolian transport. In contrast to what is suggested by Vos et al. 2020, the dominant sand transport thus appears to be onshore. It should, however, be noted that deposition events that might have occurred further offshore are lacking in the 4D-OBC data as they cannot be observed with the TLS because these parts are mostly submerged.

The second intertidal bar 4D-OBC initiation and growth period also occurs during low wave conditions. Here, however, the destruction is not accompanied by heavy weather conditions. The destruction occurs in two steps, and in contrast to the previous bar, the initial destruction occurs during high water. The results suggest that the destruction of an intertidal bar complex during low wave and wind conditions is associated with larger sand redistribution in the intertidal zone, identified through a larger amount of intertidal deposition 4D-OBCs being initiated, and less aeolian and hydrodynamic beach berm and backshore 4D-OBCs. This can be related to lower wave energy and wind energy available for sand transport. The actual reason for the intertidal bar destruction should be further researched in future studies.

6.3. Comparison to existing studies

Comparison to existing studies on monitoring surface activity

We can cluster groups of spatiotemporal segments that are not initiated around the same time but represent the same type of surface activity, in contrast to existing point cloud time series analysis methods. Existing studies in which full time series clustering was used to group similar change patterns (Kuschnerus et al. 2021a; Winiwarter et al. 2022) allowed for the identification and characterization of different change patterns occurring over the whole time-span of TLS acquisition. However, if comparable surface activities occurred at different points in time, these were not identified as part of the same type of change pattern.

Furthermore, in other existing studies using the near-continuous TLS to observe the accumulation, transportation, and destruction of an intertidal bar, the intertidal bar deposition event had to be manually identified (Vos et al. 2020). Using the combination of 4D-OBC extraction and clustering with the SOM and hierarchical clustering, we can automatically identify the emergence of an intertidal bar system at two different points in time and using the other clusters of surface activities, the post-destruction sand redistribution can be efficiently assessed.

Comparison to other classification methods

The SOM requires parameter tuning, which might indicate a disadvantage compared to other clustering algorithms when applying it on 4D-OBC sets in new environments. However, with little knowledge of the exact number, characteristics, and types of surface activities, we are able to tune these parame-

ters in order to identify physically interpretable and useful groups of surface activity. The advantage of the SOM in this sense is that it allows for easy visual navigation through the groups of surface activity obtained with a certain configuration, i.e., a configuration performs well if nodes with similar and physically interpretable feature vectors are situated close in the SOM. This enables simple tuning of these parameters, deemed less complicated than tuning the cluster number for k-means clustering or the distance thresholds for DBSCAN. However, further studies could explore their potential for characterizing 4D-OBCs.

Furthermore, the SOM works as expected in terms of its ability to represent outlying 4D-OBCs, while not being overfitted on these groups of 4D-OBCs. Using a SOM combined with hierarchical clustering reduces the influence of outliers while still representing them. The direct application of hierarchical clustering on the 4D-OBC set would likely result in the representation of outliers in separate clusters up until a large distance threshold, as it is known that the hierarchical clustering algorithm is sensitive to outliers (Xu and Wunsch 2005).

The 4D-OBCs could, in theory, also be grouped based on setting thresholds in a decision tree. This decision tree could be established manually or automatically. When grouping 4D-OBCs manually, knowledge is needed of specific values of features that distinguish the surface activities of interest. Optimizing the SOM and hierarchical clustering workflow established in this thesis also requires knowledge of which features possibly distinguish the different surface activities. However, knowledge of the actual values of these features per surface activity is not necessary. As such, the workflow is advantageous when extensive knowledge of surface activities in an environment is absent. For automated configuration of decision trees and random forests (Breiman 2001) or other supervised classification methods, a training dataset is necessary, capturing all different possible surface activities. This dataset should exhibit most of the distribution in feature values per surface activity group present in the dataset. This requires more particular knowledge of which surface activities are presented by the 4D-OBCs. In contrast, the optimization of our workflow requires only an indication of the number of surface activities present in the 4D-OBCs. Furthermore, creating a training dataset requires more work before training. It should be noted, however, that our method requires more work after training in order to interpret the different groups than a supervised classification method.

7

Conclusions and Recommendations

This chapter presents the conclusions drawn from this study as well as recommendations for improvements in the workflow and further research.

7.1. Conclusions

This research aims to make spatiotemporal segments of morphological surface activity derived from point cloud time series, called 4D objects-by-change (4D-OBCs), interpretable as different types of surface activities. This is done by developing an unsupervised classification workflow to cluster them into different groups, characterizing types of surface activities. We want to enable the analysis of the impact and correlations of different types of surface activity through near-continuous terrestrial laser scanning (TLS). Regarding this aim, the main research question to answer is:

How can different types of surface activity be characterized from a point cloud time series using the spatiotemporal segments derived as 4D-OBCs?

This research shows that using a Self-organizing Map (SOM), together with hierarchical clustering, various levels of groups of 4D-OBCs can be identified and characterized from a point cloud time series. In this thesis, 4D-OBCs obtained on a sandy beach are analyzed. These groups of 4D-OBCs are physically interpretable as different types of surface activity and show logical correlations to natural drivers. Furthermore, they can be used to find intercorrelations between the occurrences of different types of surface activity. When transferring the method to the different environment of an Alpine area covered by snow, the results are not as easily interpretable. In different environments, the parameters and features used as input for the SOM need to be fine-tuned based on knowledge of the surface activity processes in the area.

The subquestions are:
Which different types of surface activity occur in the two environments studied in this research?

In the sandy beach and Alpine snow cover study areas, various surface activities occur within the spatial and temporal span of the point cloud and subsequent 4D-OBC acquisition. At the beach, surface activities are in literature, often defined by their cross-shore location. We identify intertidal bar depositions; aeolian dune and backshore erosion, transport and deposition processes; beach berm deposition and erosion; and anthropogenic surface activities. In the Alpine snow cover area, the surface activities are less well defined by one geographic axis but are constricted by geographic features like slope and aspect. Here, snow avalanches, comprising erosion, transport, and deposition; anthropogenic snow farming and ski resort works; and ablation or compaction of the snow cover are identified. The characteristics that define and differentiate various surface activities vary between different environments. This implies that, when analyzing 4D-OBCs for different study areas, unique challenges arise in classifying them into different surface activities, which should be considered.

Which unsupervised classification methods are applicable for the grouping of the 4D-OBCs into types of surface activities

In this research, we use a Self-organizing Map (SOM) algorithm to first sort, group, and characterize the 4D-OBCs into detailed groups of surface activity, after which hierarchical clustering is used to identify and characterize broader clusters of surface activity. This method is deemed useful and possibly advantageous compared to other algorithms for the grouping and characterization of 4D-OBCs for several reasons. First, the SOM can be used without specifying an exact amount of clusters present in the data. Second, The SOM allows for identifying patterns in the 4D feature space of the 4D-OBCs. Third, by combining the SOM with a maximum dissimilarity sampling algorithm (MDA), we can obtain a balance between overfitting groups on denser data space while also ensuring that 4D-OBCs representing rare surface activities are grouped separately. Fourth, the hierarchical clustering algorithm allows for the identification of broader clusters of surface activity, which can enable the automated identification of large-scale correlations between clusters of surface activity and natural drivers.

How does the choice of features influence the grouping of the 4D-OBCs into different types of surface activity?

The identification and choice of useful features are essential for the performance of the SOM and subsequent hierarchical clustering. A wide range of features can be extracted from the 4D-OBCs as they contain both a spatial and temporal dimension. By ranking, selecting, and testing different sets of features, it is found that when too many features extracted from one of these dimensions are used, the SOM tends to be dominated by one specific dimension. One must find a balance between the two dimensions suitable for the surface activities characteristic of the study area. For the sandy beach, the best performance occurs when a set of features is used that are known to define different types of surface activity. A data-driven approach to feature selection does not perform optimally in our case.

How can the unsupervised classification method be optimized for grouping the 4D-OBCs?

The optimization of the SOM is best performed using expert knowledge of surface activities and manual investigation of different configurations, of features and parameters. The data-driven optimization methods tested here only produce a distinctive choice for features or parameters when combined with a manual investigation of the obtained groups. The optimization is thus best performed using a small labeled 4D-OBC set, of which it is known which 4D-OBCs should be grouped close together. The SOM size, kernel standard deviation, and input order are essential parameters to be considered when grouping the 4D-OBCs. The choice of parameter values depends on the dataset size, the number of potential surface activities present, the level of detail needed, and the need to identify rare surface activities.

To what extent is the optimized unsupervised classification workflow transferable to another study area?

According to our results, the optimized methods and features are not fully transferable from one study area to the other. Within one study area, different configurations of the SOM parameters for the deposition and erosion part of the dataset might also result in better separation between different surface activities and more global ordering of the SOM. In different environments, apart from choosing case-specific features, three parameters should be fine-tuned. The SOM size should be set larger if more detailed surface activity characterization is needed and for datasets inhibiting more types of surface activity. A larger kernel size should be used if the influence of outliers needs to be reduced or if the dataset size is smaller. However, this also causes the SOM to show more gradual variations; thus, separate groups of surface activity are less well-defined. Another way of representing or not representing rare surface activities is by training on only a subset of maximum dissimilar samples or using the (flipped) MDA ranking as input order. An indication of the specific values of these parameters as a function of the characteristics of a dataset cannot be given based on this research.

To what extent can the grouped 4D-OBCs be interpreted as different types of surface activity?

Using an optimized version of the SOM, different groups of 4D-OBCs can be identified, of which the respective mean feature vectors can be interpreted as known surface activity types. This holds for the sandy beach, where most of the 4D-OBCs manually interpreted as one type of surface activity are grouped in adjacent parts of the SOM. Furthermore, one can find and characterize comparable surface activities by inspecting the SOM nodes surrounding a SOM node where a 4D-OBC representative of a surface activity of interest is matched. It is essential to consider the feature distance between the weight vectors of the SOM in defining where comparable surface activities are to be situated. The clusters of 4D-OBCs obtained with hierarchical clustering of the SOM nodes are interpretable as broader types of surface activity. The initiation of the 4D-OBCs of different clusters can be related to variations in natural drivers or known occurrences of anthropogenic work, showing that the grouping at this level is indeed physically interpretable.

How do our methods of automated spatiotemporal extraction and subsequent grouping of surface activity compare to other methods of identifying and classifying surface activity?

In contrast to existing studies on surface activity or geomorphological change monitoring, we can automatically identify different types of surface activity present in a study area, independent of their relative timing. This allows larger scale and more efficient studies on characteristics of surface activities, as well as their underlying processes of initiation and finalization at different moments in time. Furthermore, the optimized workflow allows to study surface activities independent of specific knowledge of feature values that define the different surface activities. On top of this, it is not necessary to identify all the groups of surface activities present in the dataset before applying the workflow, as the SOM finds these groups automatically.

To what extent can the grouped 4D-OBCs be used for the analysis of surface activity characteristics and correlations between surface activities and natural drivers?

The grouped 4D-OBCs can, to a high extent, be used for the analysis of surface activity characteristics and correlations between surface activities and natural drivers on a sandy beach. Our results of the inter-correlations of the hierarchical clusters and correlations with natural drivers indicate that the results of our workflow can be used to identify physically interpretable relations between different clusters and natural drivers, both in accordance with and deviating from existing knowledge. We can identify clusters of 4D-OBCs representing intertidal bar depositions, which all show the expected initiation and growth of intertidal sandbar deposits during lower wave conditions, but with deviations in their expected destruction patterns. The results suggest that applying our unsupervised classification workflow on the 4D-OBCs can provide valuable insight into the complex relations between surface activities and natural drivers and merit further research.

7.2. Recommendations

The developed methods are shown to be of use for the characterization of different surface activity groups found on a sandy beach. Nevertheless, several limitations and recommendations are identified for which further research can be advantageous.

Use of case-specific features

The developed method for the sandy beach 4D-OBC set does not perform optimally on the snow cover 4D-OBC set. In further research on surface activity characterization in the alpine snow cover area, more investigation is needed into case-specific features, defining the surface activities, like the slope. This could vastly improve the performance in this case. Also, when applying the workflow to even different environments, the specific features defining the surface activities must be examined beforehand.

Test other data-driven approaches of feature selection

The data-driven selection method for the features does not result in a clear choice for features, and a manual selection through investigation of a selection of SOMs is still needed. Further studies on surface activity characterization in different locations might benefit from relying on more extensive expert

knowledge about the surface activities in the study area for feature selection or employing alternative feature selection techniques like more comprehensive wrapper methods.

Increase topological preservation

The use of a hexagonal SOM could increase the topological preservation of higher dimensional feature space without decreasing the distinctness of the characteristic feature vectors in each SOM node. It is identified that the performance of the erosional SOM of the sandy beach for surface activity characterization is inferior to the depositional SOM due to worse topological preservation. However, the common practice of increasing the kernel standard deviation of the SOM to increase topological preservation would mean that the nodes themselves show less distinct characteristics. Thus, instead of increasing this standard deviation, one can use a connected hexagonal lattice, as this would increase the interconnection between nodes during the training phase with an equal standard deviation.

Merging of 4D-OBCs segments, before or after classification

The merging of 4D-OBCs representing the same surface activity event will increase the reliability of the found relations between surface activity initiations. The identified presence of 4D-OBCs representing the same surface activity can cause the number of initiations of one surface activity to seemingly go up, even though not more surface activities of that type were, in reality, initiated. The merging of these segments by, e.g., setting limitations on the number of 4D-OBCs that can exist within one spatiotemporal interval could thus increase the applicability and validity of the complete workflow.

More detailed investigation of surface activity clusters

Further analysis of the sandy beach clusters of surface activity can be beneficial to make a stronger case for the identified redistribution of sand after intertidal sandbar destruction. Here, the initiation of erosion and deposition clusters could be compared to each other instead of only comparing the initiation of one of the two. Furthermore, future research could further relate the threshold natural driver forcing necessary for specific surface activity initiation by comparing the driver intensity to the number of initiations per cluster.

Application of the methods in an online setting

The workflow developed here could be tested and used in an online setting. This requires the development of a Python pipeline integrating point cloud acquisition, co-registration, 4D-OBC extraction, and characterization. For the case of a sandy beach, one could use the groups and clusters identified in this study and match new 4D-OBCs to the closest of the detailed surface activity groups found in the SOM. In this way, it enables the instantaneous identification of surface activity types of interest. Furthermore, new 4D-OBCs obtained in this online setting could be used to train the SOM further and fine-tune the characteristics of the detailed groups of surface activity.

References

Anders, Katharina, Roderik Lindenbergh, Sander Vos, Hubert Mara, Sierd De Vries, and Bernhard Höfle (May 2019). "High-Frequency 3D Geomorphic Observation Using Hourly Terrestrial Laser Scanning Data Of A Sandy Beach". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 4. 2/W5. Copernicus GmbH, pp. 317–324. DOI: `10.5194/isprs-annals-IV-2-W5-317-2019`.

Anders, Katharina, Lukas Winiwarter, and Bernhard Hofle (2022). "Improving Change Analysis from Near-Continuous 3D Time Series by Considering Full Temporal Information". In: *IEEE Geoscience and Remote Sensing Letters* 19. ISSN: 15580571. DOI: `10.1109/LGRS.2022.3148920`.

Anders, Katharina, Lukas Winiwarter, Roderik Lindenbergh, Jack G. Williams, Sander E. Vos, and Bernhard Höfle (Jan. 2020). "4D objects-by-change: Spatiotemporal segmentation of geomorphic surface change from LiDAR time series". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 159, pp. 352–363. ISSN: 09242716. DOI: `10.1016/j.isprsjprs.2019.11.025`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0924271619302850`.

Anders, Katharina, Lukas Winiwarter, Hubert Mara, Roderik Lindenbergh, Sander E. Vos, and Bernhard Höfle (Mar. 2021). "Fully automatic spatiotemporal segmentation of 3D LiDAR time series for the extraction of natural surface changes". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 173, pp. 297–308. ISSN: 09242716. DOI: `10.1016/j.isprsjprs.2021.01.015`.

Anthony, Edward J., Marie Hélène Ruz, and Stéphane Vanhée (Apr. 2009). "Aeolian sand transport over complex intertidal bar-trough beach topography". In: *Geomorphology* 105.1-2, pp. 95–105. ISSN: 0169555X. DOI: `10.1016/j.geomorph.2007.12.013`.

Bação, Fernando, Victor Lobo, and Marco Painho (Mar. 2005). "The self-organizing map, the Geo-SOM, and relevant variants for geosciences". In: *Computers and Geosciences* 31.2, pp. 155–163. ISSN: 00983004. DOI: `10.1016/j.cageo.2004.06.013`.

Bakker, Tije M., José A.A. Antolínez, Tim W.B. Leijnse, Stuart G. Pearson, and Alessio Giardino (Sept. 2022). "Estimating tropical cyclone-induced wind, waves, and surge: A general methodology based on representative tracks". In: *Coastal Engineering* 176. ISSN: 03783839. DOI: `10.1016/j.coastaleng.2022.104154`.

Beraldin, J.-Angelo, Francois Blais, and Uwe Lohr (Dec. 2010). *Airborne and Terrestrial Laser Scanning: Laser Scanning Technology*. Ed. by George Vosselman and Hans-Gerd Maas. Dunbeath: Whittles Publishing, pp. 1–5.

Berndt, Donald J and James Clifford (1994). "Using Dynamic Time Warping to Find Patterns in Time Series". In: *AAAI-94 Workshop Knowledge Discovery Databases*.

Breiman, Leo (2001). *Random Forests*. Tech. rep., pp. 5–32.

Callaghan, David P., Ranasinghe Roshanka, and Short Andrew (Jan. 2009). "Quantifying the storm erosion hazard for coastal planning". In: *Coastal Engineering* 56.1, pp. 90–93. ISSN: 03783839. DOI: `10.1016/j.coastaleng.2008.10.003`.

Campos, Mariana Batista, Paula Litkey, Yunsheng Wang, Yuwei Chen, Heikki Hyyti, Juha Hyyppä, and Eetu Puttonen (Jan. 2021). "A Long-Term Terrestrial Laser Scanning Measurement Station to Continuously Monitor Structural and Phenological Dynamics of Boreal Forest Canopy". In: *Frontiers in Plant Science* 11. ISSN: 1664462X. DOI: `10.3389/fpls.2020.606752`.

Camus, Paula, Fernando J. Mendez, Raul Medina, and Antonio S. Cofiño (June 2011). "Analysis of clustering and selection algorithms for the study of multivariate wave climate". In: *Coastal Engineering* 58.6, pp. 453–462. ISSN: 03783839. DOI: `10.1016/j.coastaleng.2011.02.003`.

CDC - Climate Data Center, Deutscher Wetterdienst (Apr. 2018). *Daily station observations of sum of fresh snow in cm at Zugspitze*. URL: `https://cdc.dwd.de/portal/202209231028/view1`.

Chawla, Nitesh V, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. Tech. rep., pp. 321–357.

Clark, Stephanie, Scott A. Sisson, and Ashish Sharma (Sept. 2020). "Tools for enhancing the application of self-organizing maps in water resources research and engineering". In: *Advances in Water Resources* 143. ISSN: 03091708. DOI: `10.1016/j.advwatres.2020.103676`.

Cohn, N, P Ruggiero, S De Vries, and G García-Medina (2017). *Beach growth driven by intertidal sandbar welding*. Tech. rep., pp. 12–16.

Corne, Simon, Tavi Murray, Stan Openshaw, Linda See, and Ian Turton (1999). *Using computational intelligence techniques to model subglacial water systems*. Tech. rep., pp. 37–60.

Cowell, Peter J, Marcel J F Stive, Alan W Niedoroda, Huib J De Vriend, Donald J P Swift, George M Kaminsky, and Michele Capobianco (2003). *The Coastal-Tract (Part 1): A Conceptual Approach to Aggregated Modeling of Low-Order Coastal Change*. Tech. rep. 4, pp. 812–827.

Crozier, M. J. (Dec. 2010). "Deciphering the effect of climate change on landslide activity: A review". In: *Geomorphology* 124.3-4, pp. 260–267. ISSN: 0169555X. DOI: `10.1016/j.geomorph.2010.04.009`.

Davidson-Arnott, Robin G D, Yanqi Yang, Jeff Ollerhead, Patrick A Hesp, and Ian J Walker (2008). "Earth Surface Processes and Landforms Earth Surf". In: *Earth Surf. Process. Landforms* 33, pp. 55–74. DOI: `10.1002/esp`. URL: `https://onlinelibrary.wiley.com/doi/10.1002/esp.1527`.

Delgado, Irene and Graham Lloyd (2004). "A simple low cost method for one person beach profiling". In: *Journal of Coastal Research* 20.4, pp. 1246–1252. ISSN: 07490208. DOI: `10.2112/03-0067r.1`.

Dong, Zhibao, Xiaoping Liu, Hongtao Wang, and Xunming Wang (2003). "Aeolian sand transport: a wind tunnel model". In: *Sedimentary Geology*. DOI: `10.1016/S0037-0738(03)00396-2`. URL: `www.elsevier.com/locate/sedgeo`.

Dy, Jennifer G and Carla E Brodley (2004). *Feature Selection for Unsupervised Learning*. Tech. rep., pp. 845–889.

Eitel, Jan U.H. et al. (Dec. 2016). "Beyond 3-D: The new spectrum of lidar applications for earth and ecological sciences". In: *Remote Sensing of Environment* 186, pp. 372–392. ISSN: 00344257. DOI: `10.1016/j.rse.2016.08.018`.

Eker, Remzi, Abdurrahim Aydın, and Johannes Hübl (Jan. 2018). "Unmanned aerial vehicle (UAV)-based monitoring of a landslide: Gallenzerkogel landslide (Ybbs-Lower Austria) case study". In: *Environmental Monitoring and Assessment* 190.1. ISSN: 15732959. DOI: `10.1007/s10661-017-6402-8`.

Ester, Martin, Hans-Peter Kriegel, Jiirg Sander, and Xiaowei Xu (1996). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. Tech. rep. URL: `www.aaai.org`.

Ferreira, Artur J. and Mário A.T. Figueiredo (Sept. 2012). "An unsupervised approach to feature discretization and selection". In: *Pattern Recognition*. Vol. 45. 9, pp. 3048–3060. DOI: `10.1016/j.patcog.2011.12.008`.

Gariano, Stefano Luigi and Fausto Guzzetti (Nov. 2016). *Landslides in a changing climate*. DOI: `10.1016/j.earscirev.2016.08.011`.

Hagenauer, Julian and Marco Helbich (Oct. 2013). "Hierarchical self-organizing maps for clustering spatiotemporal data". In: *International Journal of Geographical Information Science* 27.10, pp. 2026–2042. ISSN: 13658816. DOI: `10.1080/13658816.2013.788249`.

Halko, Nathan, Per-Gunnar Martinsson, and Joel A. Tropp (Sept. 2009). "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions". In: URL: `http://arxiv.org/abs/0909.4061`.

Han, Liang, Guijun Yang, Huayang Dai, Hao Yang, Bo Xu, Haikuan Feng, Zhenhai Li, and Xiaodong Yang (July 2019). "Fuzzy clustering of maize plant-height patterns using time series of UAV remote-sensing images and variety traits". In: *Frontiers in Plant Science* 10. ISSN: 1664462X. DOI: `10.3389/fpls.2019.00926`.

Hancer, Emrah, Bing Xue, and Mengjie Zhang (Aug. 2020). "A survey on feature selection approaches for clustering". In: *Artificial Intelligence Review* 53.6, pp. 4519–4545. ISSN: 15737462. DOI: `10.1007/s10462-019-09800-w`.

Hao, Jian sheng, Fa rong Huang, Yang Liu, Chukwudi Amanambu Amobichukwu, and Lan hai Li (July 2018). "Avalanche activity and characteristics of its triggering factors in the western Tianshan Mountains, China". In: *Journal of Mountain Science* 15.7, pp. 1397–1411. ISSN: 16726316. DOI: `10.1007/s11629-018-4941-2`.

Hesp, Patrick A., Robin Davidson-Arnott, Ian J. Walker, and Jeff Ollerhead (Feb. 2005). "Flow dynamics over a foredune at Prince Edward Island, Canada". In: *Geomorphology* 65.1-2, pp. 71–84. ISSN: 0169555X. DOI: `10.1016/j.geomorph.2004.08.001`.

Hilker, N, A Badoux, and C Hegg (2009). "The Swiss flood and landslide damage database 1972-2007". In: *Natural Hazards and Earth System Sciences* 9, pp. 913–925.

Höfle, Bernhard, Michael Vetter, Norbert Pfeifer, Gottfried Mandlburger, and Johann Stötter (Sept. 2009). "Water surface mapping from airborne laser scanning using signal intensity and elevation data". In: *Earth Surface Processes and Landforms* 34.12, pp. 1635–1649. ISSN: 01979337. DOI: `10.1002/esp.1853`.

Houser, Chris (2009). "Synchronization of transport and supply in beach-dune interaction". In: *Progress in Physical Geography* 33.6, pp. 733–746. ISSN: 03091333. DOI: `10.1177/0309133309350120`.

Kennard, R. W. and L. A. Stone (1969). "Computer Aided Design of Experiments". In: *Technometrics* 11.1, pp. 137–148. ISSN: 15372723. DOI: `10.1080/00401706.1969.10490666`.

Kohonen, Teuvo (1990). "The Self-Organizing Map". In: *Proceedings of the IEEE* 78.9, pp. 1464–1480. ISSN: 15582256. DOI: `10.1109/5.58325`.

— (2001). *Self-organizing Maps*. Tech. rep. Springer. DOI: `https://doi.org/10.1007/978-3-642-97610-0{\_}6`. URL: `http://www.springer.de/phys/`.

— (Jan. 2013). "Essentials of the self-organizing map". In: *Neural Networks* 37, pp. 52–65. ISSN: 08936080. DOI: `10.1016/j.neunet.2012.09.018`.

Kuschnerus, Mieke, Roderik Lindenbergh, and Sander Vos (2021a). "Coastal Change Patterns from Time Series Clustering of Permanent Laser Scan Data". In: *Earth Surface Dynamics* 9, pp. 89–103. DOI: `10.5194/esurf-2020-34`.

Kuschnerus, Mieke, D. Schröder, and Roderik Lindenbergh (June 2021b). "Environmental influences on the stability of a permanently installed laser scanner". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. Vol. 43. B2-2021. International Society for Photogrammetry and Remote Sensing, pp. 745–752. DOI: `10.5194/isprs-archives-XLIII-B2-2021-745-2021`.

Lague, Dimitri, Nicolas Brodu, and Jérôme Leroux (2013). "Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z)". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 82, pp. 10–26. ISSN: 09242716. DOI: `10.1016/j.isprsjprs.2013.04.009`.

Li, Jundong, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang, and Huan Liu (Dec. 2017). *Feature selection: A data perspective*. DOI: `10.1145/3136625`.

Lichti, Derek and Jan Skaloud (Dec. 2010). *Airbrone and Terrestrial Laser Scanning: Registration and Calibration*. Ed. by George Vosselman and Hans-Gerd Maas. Dunbeath: Wittles Publishing, pp. 83–125.

Lin, Yi Chun, Yi Ting Cheng, Tian Zhou, Radhika Ravi, Seyyed Meghdad Hasheminasab, John Evan Flatt, Cary Troy, and Ayman Habib (Dec. 2019). "Evaluation of UAV LiDAR for mapping coastal environments". In: *Remote Sensing* 11.24. ISSN: 20724292. DOI: `10.3390/rs11242893`.

Liu, Luying, Jianchu Kang, Jing Yu, and Zhongliang Wang (2005). "A comparative study on unsupervised feature selection methods for text clustering". In: *Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering, IEEE NLP-KE'05*. Vol. 2005, pp. 597–601. ISBN: 0780393619. DOI: `10.1109/NLPKE.2005.1598807`.

Mayaud, Jerome R. and Nicholas P. Webb (Sept. 2017). "Vegetation in drylands: Effects on wind flow and aeolian sediment transport". In: *Land* 6.3. ISSN: 2073445X. DOI: `10.3390/land6030064`.

Palomo, E. J., J. North, D. Elizondo, R. M. Luque, and T. Watson (Aug. 2012). "Application of growing hierarchical SOM for visualisation of network forensics traffic data". In: *Neural Networks* 32, pp. 275–284. ISSN: 08936080. DOI: `10.1016/j.neunet.2012.02.021`.

Pellicani, Roberta, Ilenia Argentiero, Paola Manzari, Giuseppe Spilotro, Cosimo Marzo, Ruggero Ermini, and Ciro Apollonio (June 2019). "UAV and airborne LiDAR data for interpreting kinematic evolution of landslide movements: The case study of the montescaglioso landslide (Southern Italy)". In: *Geosciences (Switzerland)* 9.6. ISSN: 20763263. DOI: `10.3390/geosciences9060248`.

Robin, Nicolas, Franck Levoy, and Olivier Monfort (May 2009). "Short term morphodynamics of an intertidal bar on megatidal ebb delta". In: *Marine Geology* 260.1-4, pp. 102–120. ISSN: 00253227. DOI: `10.1016/j.margeo.2009.02.006`.

Rousseeuw, Peter J (1987). "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65.

Ruessink, Gerben and Roshanka Ranashinge (June 2014). *Coastal Environments and Global Change*. Ed. by Gerd Masselink and Roland Gehrels. 1st ed. American Geophysical Union, pp. 149–176.

Saxena, Amit, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding, and Chin Teng Lin (Dec. 2017). "A review of clustering techniques and developments". In: *Neurocomputing* 267, pp. 664–681. ISSN: 18728286. DOI: `10.1016/j.neucom.2017.06.053`.

Schardt, Robert (2018). *Monatsrückblick der Wetterwarte Zugspitze für April 2018*. Tech. rep. Garmisch-Partenkirchen: Deutscher Wetterdienst. URL: `https://schneefernerhaus.de/fileadmin/user_files/Downloads/Daten/Archiv/MORZG_2018-04.pdf`.

Schweizer, Jürg, J. Bruce Jamieson, and Martin Schneebeli (2003). "Snow avalanche formation". In: *Reviews of Geophysics* 41.4. ISSN: 87551209. DOI: `10.1029/2002RG000123`.

Scott, Fred, Jose A.A. Antolinez, Robert McCall, Curt Storlazzi, Ad Reniers, and Stuart Pearson (May 2020). "Hydro-Morphological Characterization of Coral Reefs for Wave Runup Prediction". In: *Frontiers in Marine Science* 7. ISSN: 22967745. DOI: `10.3389/fmars.2020.00361`.

Sen, P. C., M. Hajra, and M. Ghosh (2020). "Supervised Classification Algorithms in Machine Learning: A Survey and Review". In: *Emerging technology in modelling and graphics*. Singapore: Springer, pp. 99–111.

Skupin, André, Joseph R. Biberstine, and Katy Börner (Mar. 2013). "Visualizing the Topical Structure of the Medical Sciences: A Self-Organizing Map Approach". In: *PLoS ONE* 8.3. ISSN: 19326203. DOI: `10.1371/journal.pone.0058779`.

Solorio-Fernández, Saúl, J. Ariel Carrasco-Ochoa, and José Fco Martínez-Trinidad (Dec. 2020). "A systematic evaluation of filter Unsupervised Feature Selection methods". In: *Expert Systems with Applications* 162. ISSN: 09574174. DOI: `10.1016/j.eswa.2020.113745`.

Stive, Marcel J.F. et al. (Sept. 2013). "A new alternative to saving our beaches from sea-level rise: The sand engine". In: *Journal of Coastal Research* 29.5, pp. 1001–1008. ISSN: 07490208. DOI: `10.2112/JCOASTRES-D-13-00070.1`.

Vettigli, Giuseppe (2018). "MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map". In: URL: `https://github.com/JustGlowing/minisom/%20v2.3.0`.

Voordendag, Annelies B et al. (2022). "The Stability of a Permanent Terrestrial Laser Scanning System- a Case Study with Hourly Scans". In: DOI: `10.3929/ethz-b-000582544`. URL: `https://doi.org/10.3929/ethz-b-000582544`.

Vos, Sander, Katharina Anders, Mieke Kuschnerus, Roderik Lindenbergh, Bernhard Höfle, Stefan Aarninkhof, and Sierd de Vries (Dec. 2022). "A high-resolution 4D terrestrial laser scan dataset of the Kijkduin beach-dune system, The Netherlands". In: *Scientific Data* 9.1. ISSN: 20524463. DOI: `10.1038/s41597-022-01291-9`. URL: `10.1038/s41597-022-01291-9`.

Vos, Sander, Roderik Lindenbergh, Sierd De Vries, T Aagaard, R Deigaard, and D Fuhrman (2017). "Coastscan: Continuous monitoring of coastal change using terrestrial laser scanning." In: *Proceedings of Coastal Dynamics* 2017.233, p. 115.

Vos, Sander, Lennard Spaans, Ad Reniers, Rob Holman, Robert McCall, and Sierd de Vries (Nov. 2020). "Cross-shore intertidal bar behavior along the dutch coast: Laser measurements and conceptual model". In: *Journal of Marine Science and Engineering* 8.11, pp. 1–21. ISSN: 20771312. DOI: `10.3390/jmse8110864`.

Vousdoukas, Michalis I., Roshanka Ranasinghe, Lorenzo Mentaschi, Theocharis A. Plomaritis, Panagiotis Athanasiou, Arjen Luijendijk, and Luc Feyen (Mar. 2020). "Sandy coastlines under threat of erosion". In: *Nature Climate Change* 10.3, pp. 260–263. ISSN: 17586798. DOI: `10.1038/s41558-020-0697-0`.

Walker, Ian J., Robin G.D. Davidson-Arnott, Bernard O. Bauer, Patrick A. Hesp, Irene Delgado-Fernandez, Jeff Ollerhead, and Thomas A.G. Smyth (Aug. 2017). *Scale-dependent perspectives on the geomorphology and evolution of beach-dune systems*. Tech. rep., pp. 220–253. DOI: `10.1016/j.earscirev.2017.04.011`.

Weiss, Gary M (2012). *FOUNDATIONS OF IMBALANCED LEARNING*. Tech. rep.

Wieczorek, Gerald F (June 1996). *Landslides: investigation and mitigation. Ch. 4: Landslide triggering mechanisms.* Vol. 247. Washington, D.C.: National Academy Press, pp. 76–90.

Williams, Jack G., Nick J. Rosser, Richard J. Hardy, Matthew J. Brain, and Ashraf A. Afana (Feb. 2018). "Optimising 4-D surface change detection: An approach for capturing rockfall magnitude-frequency". In: *Earth Surface Dynamics* 6.1, pp. 101–119. ISSN: 2196632X. DOI: `10.5194/esurf-6-101-2018`.

Winiwarter, Lukas, Katharina Anders, Daniel Schröder, and Bernhard Höfle (2022). "Full 4D Change Analysis of Topographic Point Cloud Time Series using Kalman Filtering". In: DOI: `10.5194/esurf-2021-103`. URL: `https://doi.org/10.5194/esurf-2021-103`.

Xu, Rui and Donald Wunsch (May 2005). *Survey of clustering algorithms*. Tech. rep. 3, pp. 645–678. DOI: `10.1109/TNN.2005.845141`.

Yin, Hujun (2008). "The Self-Organizing Maps: Background, Theories, Extensions and Applications". In: pp. 715–762. DOI: `10.1007/978-3-540-78293-3{\_}17`. URL: `http://link.springer.com/10.1007/978-3-540-78293-3_17`.

Zahs, Vivien, Martin Hämmerle, Katharina Anders, Stefan Hecht, Rudolf Sailer, Martin Rutzinger, Jack G. Williams, and Bernhard Höfle (2019). "Multi-temporal 3D point cloud-based quantification and analysis of geomorphological activity at an alpine rock glacier using airborne and terrestrial LiDAR". In: *Permafrost and Periglacial Processes* 30.3, pp. 222–238. ISSN: 10991530. DOI: `10.1002/ppp.2004`.

Zhu, Jun, Fengyan Shi, Feng Cai, Qing Wang, Hongshuai Qi, Chao Zhan, Jianhui Liu, Gen Liu, and Gang Lei (Apr. 2022). "Influences of beach berm height on beach response to storms: A numerical study". In: *Applied Ocean Research* 121. ISSN: 01411187. DOI: `10.1016/j.apor.2022.103090`.

A
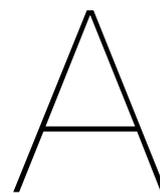
Published O3DM Conference Paper

CHARACTERIZATION OF MORPHOLOGICAL SURFACE ACTIVITIES DERIVED FROM NEAR-CONTINUOUS TERRESTRIAL LIDAR TIME SERIES

Daan Hulskemper [a,b]*, Katharina Anders[b], José A. Á. Antolínez[c], Mieke Kuschnerus[a], Bernhard Höfle[b], Roderik Lindenbergh[a]

[a] Dept. of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geoscience, Delft University of Technology, The Netherlands - d.c.hulskemper@student.tudelft.nl, (m.kuschnerus, r.c.lindenbergh)@tudelft.nl
[b] 3DGeo Research Group, Institute of Geography, Heidelberg University, Germany - (katharina.anders, hoefle)@uni-heidelberg.de
[c] Dept. of Hydraulic Engineering, Faculty of Civil Engineering and Geoscience, Delft University of Technology, The Netherlands - j.a.a.antolinez@tudelft.nl

Technical Commission II

KEY WORDS: 4D objects-by-change, Self-organizing Map, Terrestrial laser scanning, Coastal monitoring, Surface Activity

ABSTRACT:

The Earth's landscapes are shaped by processes eroding, transporting and depositing material over various timespans and spatial scales. To understand these surface activities and mitigate potential hazards they inflict (e.g., the landward movement of a shoreline), knowledge is needed on the occurrences and impact of these activities. Near-continuous terrestrial laser scanning enables the acquisition of large datasets of surface morphology, represented as three-dimensional point cloud time series. Exploiting the full potential of this large amount of data, by extracting and characterizing different types of surface activities, is challenging. In this research we use a time series of 2,942 point clouds obtained over a sandy beach in The Netherlands. We investigate automated methods to extract individual surface activities present in this dataset and cluster them into groups to characterize different types of surface activities. We show that, first extracting 2,021 spatiotemporal segments of surface activity using an object detection algorithm, and second, clustering these segments with a Self-organizing Map (SOM) in combination with hierarchical clustering, allows for the unsupervised identification and characterization of different types of surface activities present on a sandy beach. The SOM enables us to find events displaying certain type of surface activity, while it also enables the identification of subtle differences between different events belonging to one specific surface activity. Hierarchical clustering then allows us to find and characterize broader groups of surface activity, even if the same type of activity occurs at different points in space or time.

1. INTRODUCTION

The current shape of the Earth's surface is the result of combinations of erosion, transport, and deposition processes. These are either naturally (e.g., wind, gravity) or anthropogenically forced, and occur over various spatial and temporal scales. Examples of surface activities related to these processes in different geographic settings are rockfalls, avalanches, and sandy beach erosion. Surface activities can therefore be defined as events where the morphology of a local surface is changing over a certain timespan. These surface activities cause severe natural hazards in many settings, such as shoreline retreat (Vousdoukas et al., 2020). Knowledge of the impact and occurrence of these activities is therefore essential to predict, mitigate and adapt to the potential hazards they inflict. The large variety in spatial and temporal scale, and the often spatiotemporally superimposed and difficult to predict nature of surface processes do impose challenges for the observation of surface activities (Anders et al., 2021).

Near-continuous terrestrial laser scanning (TLS) enables monitoring of surface changes over multiple time scales (Eitel et al., 2016). In a near-continuous setup, a TLS device is placed at a fixed location for months to years. 3D point clouds with up to mm-scale accuracy and resolution can then be acquired at (sub)hourly intervals. As such, a substantial amount of data is collected that contains information on the changes in morphology of the scene's surface. The vast amount of data (e.g., thousands of point clouds) collected through this setup brings challenges for the visual and manual extraction of interpretable and useful information. To exploit the available point cloud time series, methods are needed which identify, segment and characterize occurrences of surface activity from these large four-dimensional (3D + time) datasets.

Previous research on morphological change extraction from such four-dimensional (4D) datasets, was focused on full time series clustering, and as such, identifying areas of homogeneous surface change over the full time series (Kuschnerus et al., 2021; Winiwarter et al., 2022). A drawback of these methods is that individual occurrences of surface activity are not separated. Furthermore, it is impossible to cluster two similar surface activities together if they do not occur around the same time.

The purpose of this research is to develop a method that enables the identification of characteristic clusters of spatiotemporal segments in the point cloud time series dataset of a sandy beach, that represent specific surface activities (e.g., sandbar formation, aeolian storm erosion on a beach) in an unsupervised manner. This large dataset is first reduced into individual surface activities using a spatiotemporal segmentation method presented by Anders et al. (2021). Next, the surface activities are grouped into characteristic types using unsupervised classification methods. We identify different levels of clusters present in the point cloud time series dataset, i.e., low-level clusters that define specific types of erosion and deposition (e.g., sandbar deposition vs. aeolian dune formation) and high-level clusters

---
*Corresponding author

(e.g., erosion vs. deposition). This is needed as different applications require different levels of characterization of surface activity. One might for example be interested in the relative influence of storm erosion on the total magnitude of beach erosion (e.g., Callaghan et al., 2009).

2. DATA

2.1 Study area and data acquisition

This research is focused on characterizing surface activity using a time series of TLS point clouds of a sandy beach on the North Sea located in Kijkduin, The Netherlands (52°04'14" N, 4°13'10" E). On such sandy beaches, a combination of hydrological processes (swash, tides), aeolian processes and anthropogenic processes interact to erode, transport and deposit sand. The tide causes part of the beach to be exposed periodically in the intertidal zone (Figure 1). On the supratidal part of the beach (backshore and dunes, see Figure 1), aeolian and anthropogenic processes dominate, but with severe weather, hydrological processes also influence the morphology of the beach. The resulting surface activities on a sandy beach show temporal scales ranging from seconds to years and even longer, while the spatial extents range from millimeter to more than dozens of kilometers. Our research focuses on processes that occur in the range of days to weeks, over spatial scales of several meters to hundreds of meters.

This particular beach is monitored using a near-continuous TLS setup with a Riegl VZ-2000 scanner fixed on a hotel building overlooking the beach. The setup is part of the CoastScan project (Vos et al., 2017). We use a subset of the dataset acquired from January to May 2017 (Vos et al., 2022). The scans cover the dunes, backshore and intertidal zone. The point clouds at the beach itself (ranging from 100 m to 600 m from the sensor) have point densities between 2 and 20 $points/m^2$.

Figure 1. Study area (star in B). The point clouds sample the area visualized in A. The blue box represents the subset of the study area as shown in Figure 2. Data: Aerial imagery ©pdok.nl 2017, borders ©Natural Earth 2022

2.2 Surface activities extracted as 4D objects-by-change

From the point cloud time series, spatiotemporal segments representing individual surface activities need to be extracted automatically, as a first step to reduce the large amount of data. These segments represent the temporal and spatial extent of a distinct surface activity, e.g. the build-up and consecutive destruction of an intertidal sandbar. In this paper we use spatiotemporal segments, named as 4D objects-by-change (4D-OBCs), extracted with the method presented by Anders et al. (2021). The 4D-OBCs are obtained through first interpolating each point cloud to a regular grid with a spacing of 0.5 m

and computing for each grid point the distance to the first point cloud, using the M3C2 method (Lague et al., 2013). Hereafter, points of significant change are identified using a sliding temporal window. From these starting points, temporal segments are extended until the change with respect to the starting point becomes zero again. These temporal segments serve as seed candidates for a spatial surface activity segment. A region is grown spatially by computing the similarity between the time series of the seed point and the time series of the spatially neighboring points, on the basis of the dynamic time warping (DTW) distance (Berndt and Clifford, 1994). If the similarity is larger than an adaptive threshold, the neighboring point is added to the segment. The full details of the method can be found in Anders et al. (2021). For this use case, the resulting 4D-OBCs have been shown to represent 95 % of manually identified surface activities (Anders et al., 2021). We use the derived dataset containing 2,021 4D-OBCs, extracted from the 4D point cloud dataset of Kijkduin, as input for our method. An example of a 4D-OBC is shown in Figure 2. This 4D-OBC represents an erosion form due to tidal activity in the intertidal area.
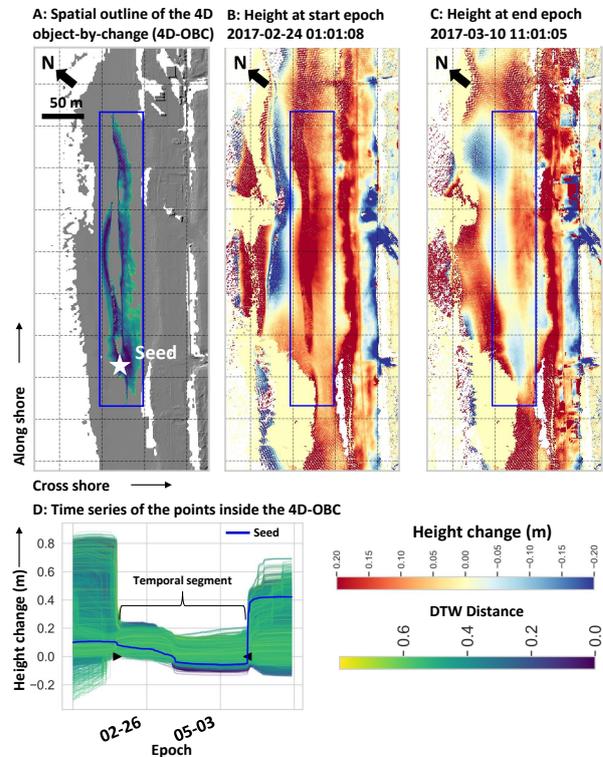

Figure 2. Temporal evolution and extracted spatial segment of an intertidal sandbar erosion surface activity.

3. METHODS

The methods used to characterize and visualize surface activities from the 4D-OBCs are summarized in five steps (Figure 3): (1) split the 4D-OBC dataset into erosion and deposition surface activities, (2) extract spatial and temporal features from the 4D-OBCs to be used in the unsupervised classification methods, (3) for both the erosion and deposition dataset create a training subset of the 4D-OBCs that show a maximum dissimilarity with regards to the derived features, (4) with each subset train a Self-organizing Map (SOM) and match all the 4D-OBCs to this SOM, to explore the full dataset and organize it into characteristic feature vectors. Lastly (5), use hierarchical clustering

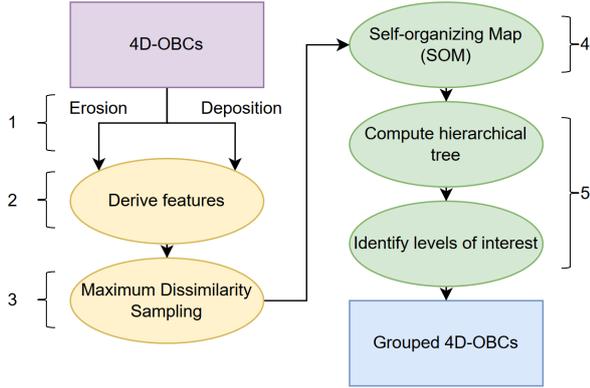to group these characteristic feature vectors and obtain different levels of grouping.

Figure 3. Workflow with the five main steps to obtain grouped 4D-OBCs based on their features.

3.1 Feature extraction

We characterize the 4D-OBCs at first by deriving five features. The nature of the 4D-OBCs allows for the incorporation of both spatial (2) and temporal (3) features, as each 4D-OBC contains time series of height change (one for each point incorporated in the segment) and locations of all the points belonging to the 4D-OBC.

3.1.1 Spatial features We derive the size of the segment (as the number of grid points in the final spatial outline of the 4D-OBC) and the position of the seed of the segment relative to the cross-shore axis. We only incorporate the cross-shore position as this can distinguish processes occurring in the intertidal zone from processes in the supratidal zone, while the along-shore position is not considered important to any separation in physical processes.

3.1.2 Temporal features The three temporal features are based on the temporal evolution of height of the seed of a 4D-OBC. The shape and magnitude of the seed time series shows strong variability among 4D-OBCs and is therefore expected to be of value for the characterization of surface activities. We resample the seed time series to a fixed number of epochs using linear interpolation, as the algorithms we use require a fixed amount of features as input. The seed time series show a wide variety in length as a result of variations in the nature of a surface activity, ranging from 24 h to 1,344 h with a mean of around 500 h. We therefore choose a resampling size of 500 epochs.

To incorporate information that is removed when resampling the seed time series, we also extract features based on the original time series of the seeds. The features obtained from the original time series are the area under the curve of the seed time series and the duration of the seed time series.

Next, we extract the sign of each seed time series and use this to split the dataset into erosion and deposition subsets. Here, 4D-OBCs with a negative sign are regarded as erosion and 4D-OBCs with a positive sign are regarded as deposition. This results in a dataset of 1,205 deposition 4D-OBCs and 816 erosion 4D-OBCs. The sign feature itself is not used as input for the algorithms.

3.1.3 Feature scaling The features are scaled before analysis as to mitigate any effect of variations in the units of features. The scaling is done by normalizing each feature individually to the range of 0 to 1 using min-max normalization, i.e., on the basis of the minimum and maximum value present in the dataset, for each feature. When scaling the resampled time series, we set the minimum and maximum feature value as the minimum and maximum occurring height change of all 4D-OBCs in the dataset, with respect to all epochs. As such, the resampled time series retains its shape after scaling.

After scaling, all feature values, apart from the resampled time series, are multiplied by 500 (the resample size of the time series). Through this, we give equal weight to the separate features as to the time series when computing distances between feature vectors.

3.2 Self-organizing Map (SOM)

SOMs have proven to be suitable for data exploration and characterization in various fields of research (see e.g.,Skupin et al., 2013; Clark et al., 2020). The advantage of using a SOM over other unsupervised classification methods, like k-means or principal component analysis, is that a SOM can, to a certain degree, preserve topological order of higher dimensional space into lower dimensions. As such, a SOM shows not only the characteristic feature vectors of the data, but also which of these characteristic vectors are neighboring in the feature space. It can consequently show gradual patterns of variation in the data. The SOM thus inhibits good potential for the characterization and visualization of the 4D-OBCs dataset, as the surface activities we characterize contain gradual variations, and boundaries between their spatiotemporal properties are not necessarily distinct.

The SOM is a neural network which is used to map and cluster high-dimensional data onto an n-dimensional grid or lattice (Kohonen, 1990). All grid points contain a node (i.e., weight vector, v_j with $j = 1, ...M$, $M = No.\ grid\ points$) with a length equal to the number of features of the input samples. During each of the training cycles $t = 1, ...T$, all samples x_i (with $i = 1, ...n, n = No.\ samples$) in the dataset are iteratively and in fixed order mapped to the closest node (our order is based on a maximum dissimilarity ranking, see section 3.3) and the weight of the node is updated. As a result, the final variance between the weight vector and the mapped samples is minimized.

The SOM algorithm consist of the following steps:

1. Initialize weight vectors, v_j with $j = 1, ...M$
2. Select for sample x_i the closest weight vector v_j as best matching unit (BMU)
3. Update the weight vector and the surrounding weight vectors:

$$v_j = v_j + \alpha_t h_{i,j}(t)(x_i - v_j) \qquad (1)$$

here $h_{i,j}$ is a Gaussian kernel function defining the magnitude of influence of the sample x_i on the weight vectors in the grid:

$$h_{i,j}(t) = e^{\frac{-d_{i,j}^2}{2\sigma_t^2}} \qquad (2)$$

where $d_{i,j}$ is the grid distance between v_j and BMU, in grid units; σ_t is the standard deviation of the Gaussian kernel at cycle t, indicating the radius of influence of the sample; and α_t is the learning rate at cycle t.

4. Repeat step 2 and 3 for every sample in the dataset
5. Repeat step 4 for a given amount of cycles T

The initial values of the learning rate and radius are predefined, and decrease with the number of cycles as a means to achieve convergence, and both global and local ordering of the data. The values at cycle t are computed using an asymptotic decay function:

$$(\alpha_t, \sigma_t) = (\alpha_{t-1}, \sigma_{t-1}) \frac{1}{1 + \frac{2t}{T}} \qquad (3)$$

SOMs can be generated with an arbitrary amount of dimensions and shapes. As to obtain a visually interpretable representation of the data, we use a two-dimensional rectangular grid of 8 by 8 nodes. This is deemed to be large enough to describe the distribution in the feature space of the 4D-OBCs, while still allowing visual interpretability. The other parameters required for the SOM generation were either empirically determined or based on literature (Kohonen, 1990; Clark et al., 2020). All the parameter settings can be found in Table 1. We implement the SOM algorithm using the Python *MiniSOM v2.3.0* implementation (Vettigli, 2018). We train the algorithm using the subset described in section 3.3 and afterwards assign all the 4D-OBCs to a node by matching them to their closest weight vector.

No. nodes:	64
Shape:	8x8
Learning rate at epoch t_0 (α_0):	1.0
Std. dev. of kernel at t_0 (σ_0):	2.0
Distance metric:	Manhattan
Weight initialization method:	PCA
Order of input:	Based on MDA ranking
No. training cycles (T):	20,000

Table 1. Parameters for the Self-organizing Map

3.3 Subset selection for training

We select a subset of the 4D-OBCs dataset to be used in the training phase of the SOM algorithm, using a maximum dissimilarity sampling algorithm (MDA; Kennard and Stone, 1969). Areas in the feature space with a larger data density represent a larger area of the SOM (Clark et al., 2020). Selecting a maximum dissimilar subset prevents that the SOM is dominated by surface activities that occur more often and enables a better representation and identification of rare surface activities (Bakker et al., 2022).

The selection of the most dissimilar samples is executed as follows: (1) compute the distance matrix between all samples in the dataset based on the Manhattan distance, (2) select the two most distant samples as initial subset, (3) select the next sample as the sample that maximizes the smallest distance to any of the samples already in the subset, and (4) repeat step three until the desired size of the subset is reached. A ranked subset based on dissimilarity is then obtained. We choose the final subset sizes such that the samples in the subsets show an approximately uniform distribution along the first two principal components of the full datasets.

3.4 Hierarchical clustering

The SOM nodes serve as the input for a hierarchical clustering algorithm, through which we can identify different levels of separation present in the datasets (Scott et al., 2020). We compute the full hierarchical tree using the mean feature vectors of the 4D-OBCs assigned to each SOM node. We therefore start with all mean feature vectors in separate clusters and iteratively merge these together on the basis of the intracluster distances. In this way, we obtain a specific clustering level of the dataset per distance threshold. We use an average linkage criterion based on the Manhattan distance to determine if two clusters are merged. The Python *sklearn v1.0.2 agglomerative clustering* implementation is used for clustering.

3.5 Evaluation

We evaluate the performance of the SOMs in terms of their ability to characterize and visualize the dataset, through visual inspection of the mean feature vectors of the 4D-OBCs in each of the SOM nodes and their variance. The example 4D-OBC shown in Figure 2 is used to assess if the 4D-OBCs assigned to neighboring nodes in the SOM are indeed related to comparable surface activities, and what distinguishes them. This is done through inspection of the feature vectors of the 4D-OBCs in the node where the example is assigned to, and its surrounding nodes.

The performance of the hierarchical clustering algorithm at each distance threshold is evaluated using the mean silhouette score s_{sil} over all samples (Rousseeuw, 1987). Each sample here represents the mean feature vector of a SOM node. For each of these, s_{sil} is computed as follows:

$$s_{sil}(x) = \frac{b(x) - b(a)}{max(b(x), a(x))} \qquad (4)$$

where x is a sample, $a(x)$ is the mean distance between the sample and all other samples in the cluster it is assigned to, and $b(x)$ is the mean distance between the sample and all the samples belonging to the closest cluster it is not assigned to. We then take the mean of all the silhouette scores to obtain one representative value. The silhouette score has a value close to 1 if the separation between clusters is large, while the intra-cluster variability is low. The score has a value close to 0 if many clusters overlap. If the score is smaller than 0, many samples are assigned to the wrong cluster.

Using the mean silhouette scores we can therefore estimate at which distance thresholds clusters appear that represent groups of surface activities. If at a certain distance a local optimum in silhouette score exists (i.e., with increasing distance threshold the silhouette score drops again), this indicates that the clusters at this threshold show a larger separation and smaller intra-cluster distance than after merging. These clusters might therefore hold a physical value and show clusters of high level surface activities. If from one distance threshold to the other the silhouette score jumps and stabilizes with increasing distance threshold, the clustering at that threshold is also of interest, as there is a large distance between all the clusters at this threshold.

4. RESULTS

4.1 Self-organizing Maps

The training subsets of the two datasets are found to be appropriate at a size of around a quarter of the full dataset. For the erosion and deposition dataset this results in a subset of 200 and 300 4D-OBCs, respectively. With each subset, we train a SOM algorithm, using the settings provided in Table 1. After training, we match all the 4D-OBCs in the datasets to the closest weight
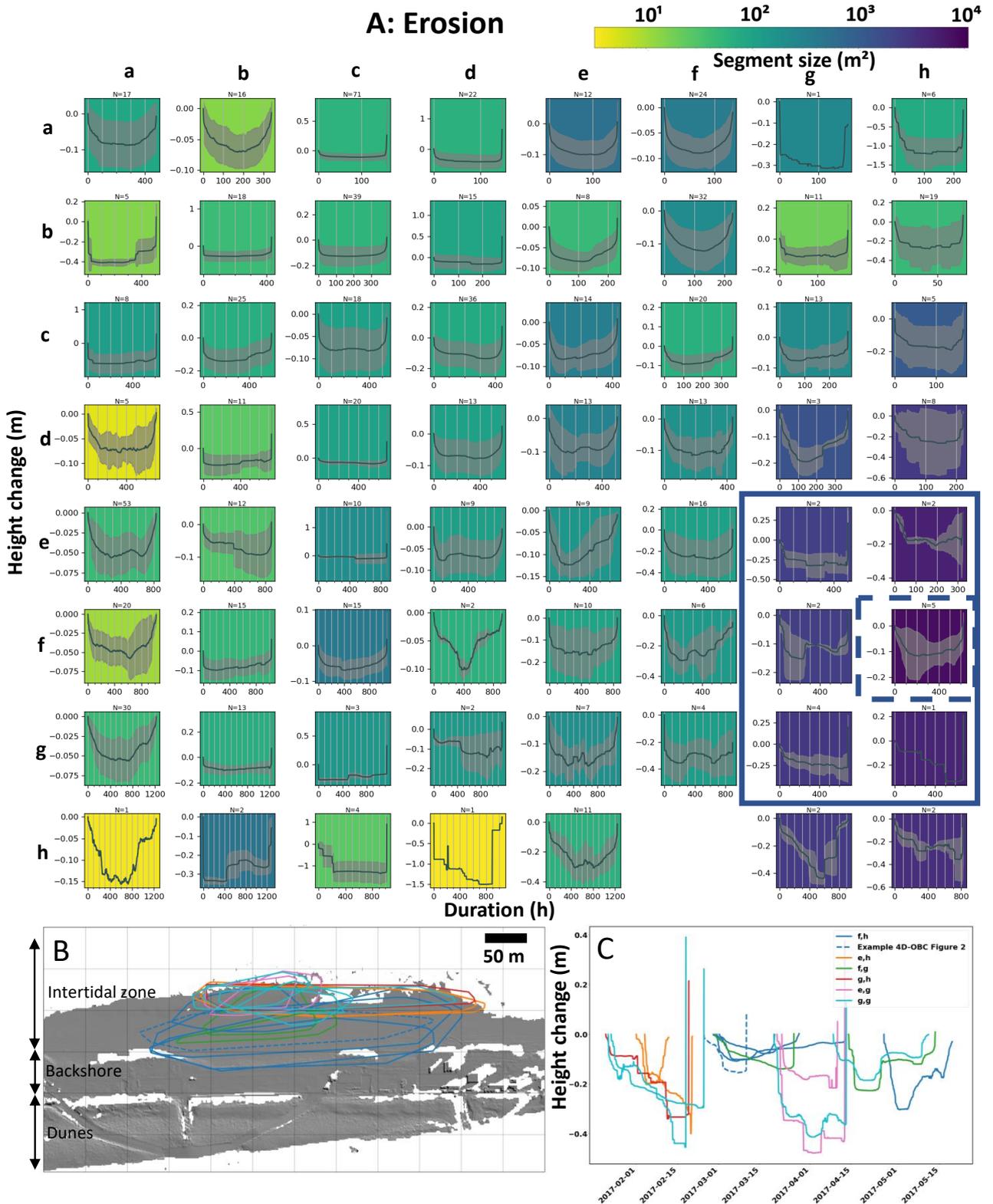
Figure 4. A) Visualization of the erosion Self-organizing Map (SOM E). Each plot represents a SOM node. The x-axis of each plot represents the mean duration of the 4D objects-by-change (4D-OBCs) in the node. The y-axis of each plot represents the mean height change of the 4D-OBCs in the node. The limits of both vary between each graph. The black curves show the mean time series of the 4D-OBCs in each node, and the grey area the standard deviation. The background colors represent the mean segment size per node (on a logarithmic scale). Nodes where no graph is visible represent nodes where no sample is matched after training. B) shows the convex hulls of the 4D-OBCs found in the nodes inside the blue polygon in A. The example shown in Figure 2 is visualized with a blue dotted line. C) shows the seed time series and timing of the 4D-OBCs as in B.

vector. In Figure 4A and Figure 5A we visualize the SOMs for the erosion and deposition dataset, respectively.

In both SOMs there is a clear sorting on the duration of the surface activities. For the deposition SOM (Figure 5A, referred to as SOM D for Deposition) from columns a to h and rows h to a the mean duration in the nodes increases, with a few exceptions (e.g., node D(h,g)). There is also noticeable sorting on segment size, though the pattern is less explicit than for duration. There is a clear pattern where on the bottom right the nodes show sharp increases in height (instant deposition) followed by fairly steep erosion, with a low number of 4D-OBCs being matched to these nodes. From here, following the columns and rows towards a, the mean time series of height change in the nodes becomes more gradual. Several groups of similar nodes can also be identified. One clear group of nodes, where few 4D-OBCs are matched, are the previously mentioned instantaneous change nodes with high magnitude and small size. This group roughly encompasses the bottom right quarter of SOM D (although D(h,a) and D(h,b) could also be included, despite their larger size). Another group of nodes, where few 4D-OBCs are matched, is characterized by 4D-OBCs with a long duration ($\sim$400 h - 1200 h), large size ($\sim$100 m^2 - 10000 m^2) and small magnitude (up to $\sim$0.5 m). This group is visible in the top row D(a,a-h). From column a-d downwards, this group gradually transforms into nodes with shorter durations. The latter SOM region also possesses the nodes where the largest amount of 4D-OBCs are matched (e.g., D(c,c), D(c,d), D(h,d) and D(h,e)). These nodes show 4D-OBCs with a duration between $\sim$100 h and 400 h, with a size of around 100 m^2 and magnitude of maximum $\sim$0.5 m. The variance of the seed time series of the 4D-OBCs is fairly high in these nodes and a distinct shape is thus less pronounced. Several outlying nodes (i.e., a low amount of matched 4D-OBCs) are D(a,h), D(a,g), D(d,c), D(e,b) and D(f,d). The 4D-OBCs in most of these outlying nodes are large ($\sim$1000 m^2 - 10000 m^2) and of relatively low magnitude, but show a variety of durations.

For the erosion SOM (Figure 4A, referred to as SOM E for Erosion) the mean segment size in the nodes approximately increases from the top left rows a-h and columns a-h, with some exceptions (e.g., E(c,f)). Most of the nodes are more gradual, and nodes that show a sharper decrease in height tend to have a lower magnitude than the instant deposition surface activities in SOM D (e.g., E(d,a) and E(g,a)). Groups of nodes are less easily distinguished than in SOM D, as there are more gradual variations from node to node. Some groups of nodes are nevertheless noticeable. One group is comprised by the highly populated node E(e,a) and its neighbors (E(a-b,e-g); group I). These nodes show 4D-OBCs with a low magnitude (minimum of $\sim$0.1 m), relatively gradual erosion, and relatively long duration ($\sim$400 h - 1200 h). To the right, this duration decreases, and the amount of 4D-OBCs matched there also decreases. A second group is a highly populated area of node E(a,c) and its neighbors (group II). Here, the low magnitude 4D-OBCs are also present, but the height changes over time are more instant. Furthermore, the duration is shorter, and at the end of the time series, the height increase is substantial and instantaneous. This indicates that a deposition activity commonly occurs directly after such an erosion activity. A third interesting group is visible on the right side of SOM E (E(d-h,g-h), group III), here 4D-OBCs are matched that show a large size (up to $\sim$10000 m^2), with mean magnitudes slightly larger than most nodes in group I and II, and a more gradual erosion and recovery phases. In this group, the mean durations vary over

the full range of durations present in the dataset. There are a few outlying nodes (E(h,c), E(h,d), and E(a,h)) that represent fairly instantaneous or step-like 4D-OBCs with a large magnitude (down to -1.5 m). The 4D-OBCs represented by E(a,h) are distinguishable from the others by their shorter duration.

4.1.1 Example of the identification of intertidal erosion

The intertidal erosion activity shown in Figure 2 is assigned to node E(f,h). Figure 4B and C show the spatial and temporal outline of the 4D-OBCs assigned to this node and the surrounding nodes. Figure 4B shows that all the 4D-OBCs grouped in these nodes occur in the intertidal zone. The 4D-OBCs assigned to E(f,h) are generally slightly larger and stretch out towards the backshore. The nodes E(e,h) and E(g,h) contain fairly elongated segments, whereas the segments of the 4D-OBCs in E(e,g) and E(f,g) are shorter. E(g,g) shows both elongated and short segments.

Figure 4C shows the time series and timing of the same 4D-OBCs. There is variation in duration of the 4D-OBCs contained in node E(f,h). Some of the durations are similar to the duration of the 4D-OBCs in the other nodes. What distinguishes the 4D-OBCs in node E(f,h) is the magnitude of the change. The minimum height change (for three of the four 4D-OBCs visualized) is closer to zero, than for the 4D-OBCs in the other nodes. Moreover, the 4D-OBCs captured in node E(f,h) shows more gradual recovery than most of the 4D-OBCs captured by the other nodes, with the exception of the example 4D-OBC shown in Figure 2.

4.2 Hierarchical clustering

We hierarchically cluster the mean feature vectors of the 4D-OBCs in the nodes of SOM E and SOM D separately. It is found that the silhouette score generally increases with an increasing distance threshold, there are nonetheless several plateaus where the silhouette stabilizes. What is more, the silhouette score related to the clustering of SOM D exhibits a local optimum around a distance of 200. We investigate what characterizes the clusters found at this threshold and identify if this indeed results in a physically relevant grouping of surface activities. At this distance threshold the SOM nodes of the deposition dataset are clustered into 16 clusters. The nodes incorporated in each of the clusters are visualized in Figure 5B.

We further inspect cluster 0, cluster 1 and cluster 12, as these contain nodes with visually comparable mean feature vectors. The bottom right cluster is also investigated which contains notably different mean feature vectors (cluster 5). Figure 5C-F shows density plots of the cross-shore location with respect to the edge of the data array (Figure 5C), duration (Figure 5D), and starting epoch of the 4D-OBCs allocated to the respective cluster (Figure 5F). We also show the density plots of the maximum height change derived from the seed time series (Figure 5E).

The three apparently similar clusters (0, 1 and 12) show notably different characteristics. Cluster 0 and 12 both contain 4D-OBCs mostly found in the intertidal zone (between -150 m and - 300 m). The durations of the 4D-OBCs do however differ. Cluster 12 is defined by shorter durations, with a density peak around 170 h, whereas cluster 0 shows a peak around 610 h, while also containing activities of longer durations. It is suggested that these define two types of intertidal deposition where the underlying process displays different periods of forcing. This is also indicated by the fact that
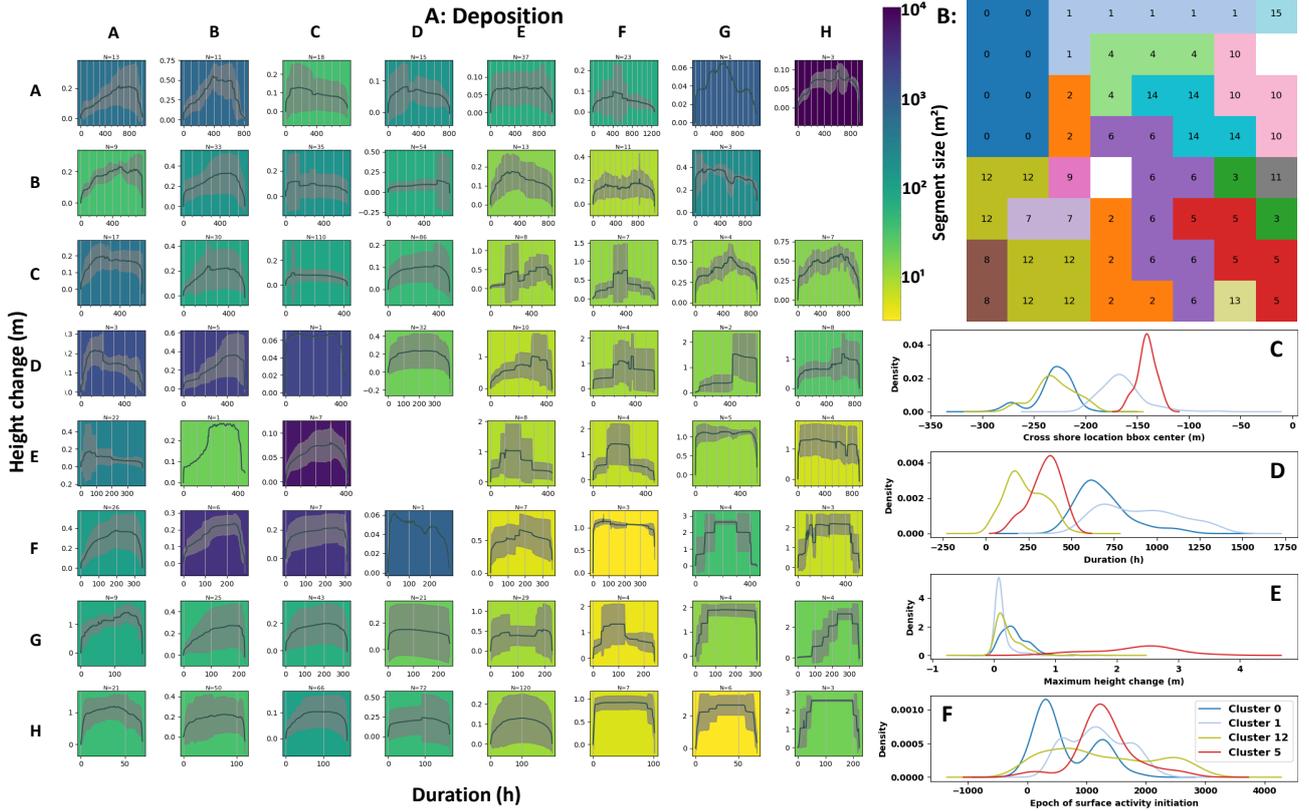
Figure 5. A) Visualization of the deposition Self-organizing Map (i.e., SOM D). For detailed explanation refer to the caption of Figure 4. B) The clusters found with a distance threshold of 230. Each grid point refers to a node in A, each color and number represents one cluster of nodes. C-F) Density plots of four features of the 4D objects-by-change grouped in cluster 0, 1, 5 and 12.

cluster 0 shows two distinct peaks of initiation in the density plot (see Figure 5F), whereas the peaks of cluster 12 are less distinct, indicating that the initiation occurs more frequently. Moreover, the difference in characteristics between cluster 1 and the previously mentioned clusters is considerable, as the 4D-OBCs in cluster 1 predominantly occur closer to the backshore area (between -100 m and -150 m), suggesting that these surface activities are more influenced by aeolian drivers. The 4D-OBCs in this cluster on average also show a longer duration, and a larger peak at low maximum change than the aforementioned clusters. These last two characteristics also distinguish the cluster from cluster 5. 4D-OBCs in cluster 5 predominantly occur in the backshore area, but slightly land inward. They are characterized by a large maximum height change and shorter duration. These aspects indicate that cluster 5 shows anthropogenic deposition by bulldozers, as it has already been identified that such activities occur in this part of the beach (Kuschnerus et al., 2021).

5. DISCUSSION

From the analysis of our dataset, we can see that types of surface activities similar to one type of surface activity of interest can be found by inspecting the 4D-OBCs in the SOM region surrounding the node where the 4D-OBC, describing the surface activity of interest, is found. All 4D-OBCs shown in Figure 4B and C are related to the same type of surface activity, namely intertidal sandbar erosion, but with variations in location, extent, duration, magnitude and time series shape. The 4D-OBC shown in Figure 2 is a slightly outlying surface activity, as it shows a similar magnitude, location and size as the sur-

face activities captured in the same node, but with a shorter duration and faster recovery to initial state. The SOM thus allows to investigate the variety of characteristics that a certain surface activity can exhibit, while also allowing to find similar activities occurring at different points in time, by investigating the nodes surrounding a sample of interest. It should be noted that this was investigated only for one example of intertidal sandbar erosion in the scope of this paper.

We find that some 4D-OBCs that are not grouped together actually show a lot of similarities. The two 4D-OBCs with a center epoch around 2017-04-01 (cyan and pink in Figure 4C) are matched in different nodes, even though their visualized spatial and temporal features are similar. Furthermore, there is one 4D-OBC in node E(g,g), that shows the same elongated outline as the 4D-OBCs in node E(e,h) and E(g,h), with a similar time series. These 4D-OBCs thus show very similar surface activities while not being grouped together. It has to be explored in further work what is the exact reason of this behavior.

A more extensive parameter tuning of the SOM algorithm might enhance the performance of the SOM. So far, most of the settings were determined on the basis of visual inspection or literature. The use of the full seed time series, or even all the time series of grid points in a 4D-OBC, instead of only the resampled seed time series, in combination with a DTW distance metric instead of Manhattan distance, might also improve the performance. In future work, these settings will be tested and performance will be assessed using e.g., a labelled test dataset and quantification of the intra-node distances.

The use of a hierarchical clustering algorithm for the automated identification of higher level groups of surface activity is shown to be very effective and valuable for our dataset. As an example we inspected four of the clusters found with this method, which showed characteristic types of surface activity with considerable separation in the feature space. In contrast to previous methods of point cloud time series analysis, in which full time series clustering was used to group similar change patterns, we are able to cluster groups of spatiotemporal segments that are not initiated around the same time, but do represent the same type of surface activity. It may be explored if using a hierarchical clustering algorithm directly on the 4D-OBCs instead of first organizing it in a SOM achieves better or comparable results.

6. CONCLUSION

The aim of this research is to develop a method that enables the identification and characterization of types of surface activity present in large near-continuous point cloud time series.

A Self-organizing Map (SOM) in combination with a maximum dissimilarity sampling algorithm (MDA) can sort 4D-objects-by-change (4D-OBCs) that represent single surface activities. The SOM enables to identify gradual patterns existing in the 4D feature space, as well as groups of surface activities represented in the dataset. The combination with the MDA allows for the identification of both rarely and frequently occurring surface activities. Furthermore, one can identify and characterize comparable occurrences of a certain type of surface activity of interest, by means of investigating the respective SOM node and its surroundings.

Hierarchically clustering these SOM nodes is shown to be a promising method of identifying distinct groups of surface activity. It further allows for the identification of broader but distinct groups of surface activity present in the dataset.

The presented methods allow exploration of the different properties of the 4D-OBCs and group them together. For example, in the case of bulldozer works and erosion in the intertidal zone, we could group similarly behaving time series and show what distinguishes these groups from one another. Even for 4D-OBCs that appear in the same location, e.g., in the intertidal zone, or around the same time, our methods allow to group them separately and identify why these 4D-OBCs represent different surface activities. A next step is investigating what the driving mechanisms behind the found groups of 4D-OBCs are.

References

Anders, K., Winiwarter, L., Mara, H., Lindenbergh, R., Vos, S. E., Höfle, B., 2021. Fully automatic spatiotemporal segmentation of 3D LiDAR time series for the extraction of natural surface changes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 297–308.

Bakker, T. M., Antolínez, J. A., Leijnse, T. W., Pearson, S. G., Giardino, A., 2022. Estimating tropical cyclone-induced wind, waves, and surge: A general methodology based on representative tracks. *Coastal Engineering*, 176.

Berndt, D. J., Clifford, J., 1994. Using Dynamic Time Warping to Find Patterns in Time Series. *AAAI-94 Workshop Knowledge Discovery Databases*.

Callaghan, D. P., Roshanka, R., Andrew, S., 2009. Quantifying the storm erosion hazard for coastal planning. *Coastal Engineering*, 56(1), 90–93.

Clark, S., Sisson, S. A., Sharma, A., 2020. Tools for enhancing the application of self-organizing maps in water resources research and engineering. *Advances in Water Resources*, 143.

Eitel, J. U., Höfle, B., Vierling, L. A., Abellán, A., Asner, G. P., Deems, J. S., Glennie, C. L., Joerg, P. C., LeWinter, A. L., Magney, T. S., Mandlburger, G., Morton, D. C., Müller, J., Vierling, K. T., 2016. Beyond 3-D: The new spectrum of lidar applications for earth and ecological sciences. *Remote Sensing of Environment*, 186, 372–392.

Kennard, R. W., Stone, L. A., 1969. Computer Aided Design of Experiments. *Technometrics*, 11(1), 137–148.

Kohonen, T., 1990. The Self-Organizing Map. *Proceedings of the IEEE*, 78(9), 1464–1480.

Kuschnerus, M., Lindenbergh, R., Vos, S., 2021. Coastal Change Patterns from Time Series Clustering of Permanent Laser Scan Data. *Earth Surface Dynamics*, 9, 89–103.

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 10–26.

Rousseeuw, P. J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

Scott, F., Antolinez, J. A., McCall, R., Storlazzi, C., Reniers, A., Pearson, S., 2020. Hydro-Morphological Characterization of Coral Reefs for Wave Runup Prediction. *Frontiers in Marine Science*, 7.

Skupin, A., Biberstine, J. R., Börner, K., 2013. Visualizing the Topical Structure of the Medical Sciences: A Self-Organizing Map Approach. *PLoS ONE*, 8(3).

Vettigli, G., 2018. MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map. https://github.com/JustGlowing/minisom/ v2.3.0.

Vos, S., Anders, K., Kuschnerus, M., Lindenbergh, R., Höfle, B., Aarninkhof, S., de Vries, S., 2022. A high-resolution 4D terrestrial laser scan dataset of the Kijkduin beach-dune system, The Netherlands. *Scientific Data*, 9(1). 10.1038/s41597-022-01291-9.

Vos, S., Lindenbergh, R., De Vries, S., Aagaard, T., Deigaard, R., Fuhrman, D., 2017. Coastscan: Continuous monitoring of coastal change using terrestrial laser scanning. *Proceedings of Coastal Dynamics*, 2017(233), 115.

Vousdoukas, M. I., Ranasinghe, R., Mentaschi, L., Plomaritis, T. A., Athanasiou, P., Luijendijk, A., Feyen, L., 2020. Sandy coastlines under threat of erosion. *Nature Climate Change*, 10(3), 260–263.

Winiwarter, L., Anders, K., Schröder, D., Höfle, B., 2022. Full 4D Change Analysis of Topographic Point Cloud Time Series using Kalman Filtering. *Earth Surface Dynamics Discussions*, 2022, 1–25.

B

Deposition SOMs of snow cover dataset with various sizes

The following deposition SOMs are inspected to determine the appropriate size of the SOM, to be used for the transferability test.
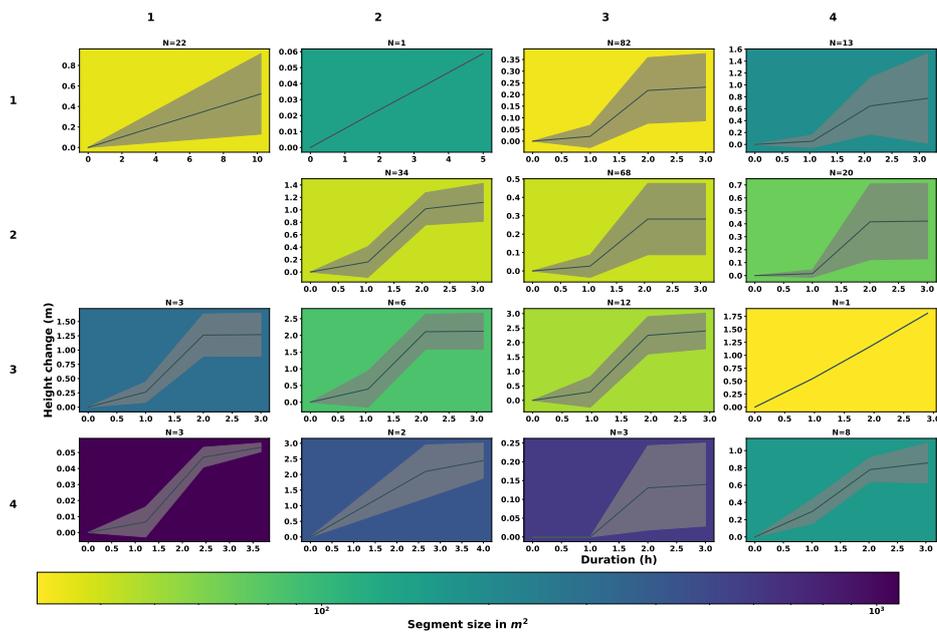

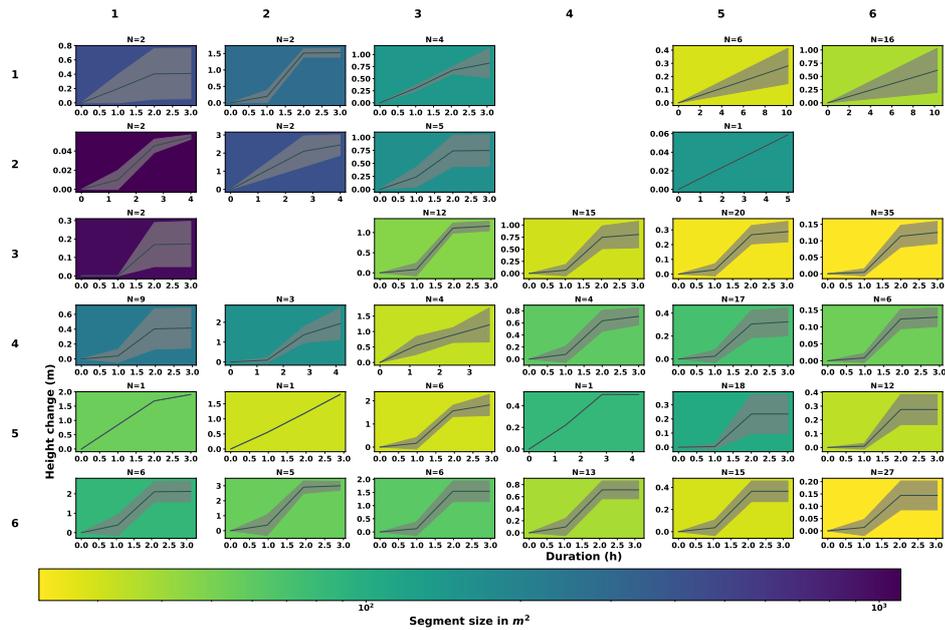**Figure B.1:** Visualization of the deposition Self-organizing Map of the snow cover dataset with 16 nodes. Trained on the resampled time series, duration, timing of minimum acceleration, maximum height change, area under the curve, and volume feature. Each plot represents a SOM node. The x-axis of each plot represents the mean duration of the 4D-OBCs in the node. The y-axis of each plot represents the mean height change of the 4D-OBCs in the node. The limits of both vary between each graph. The black curves show the mean time series of the 4D-OBCs in each node, and the grey area is the standard deviation. The background colors represent the mean segment size per node (on a logarithmic scale). Nodes without a graph represent nodes where no sample is matched after training.

Figure B.2: Visualization of the deposition Self-organizing Map of the snow cover dataset with 36 nodes. For a detailed explanation of the visualization and color scheme, see Figure B.1.
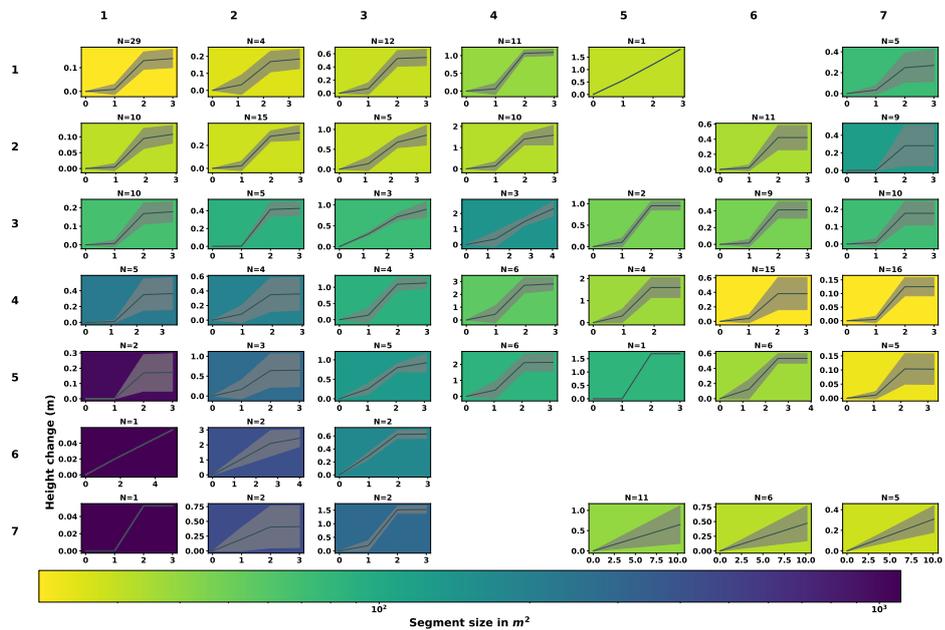
Figure B.3: Visualization of the deposition Self-organizing Map of the snow cover dataset with 49 nodes. For a detailed explanation of the visualization and color scheme, see Figure B.1.
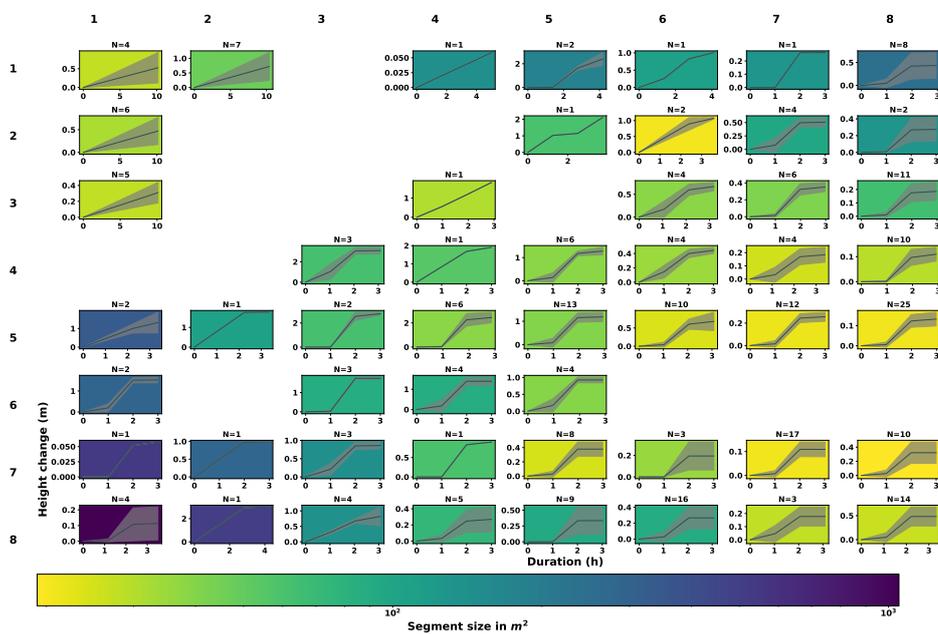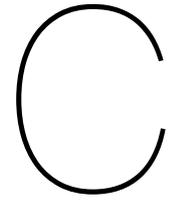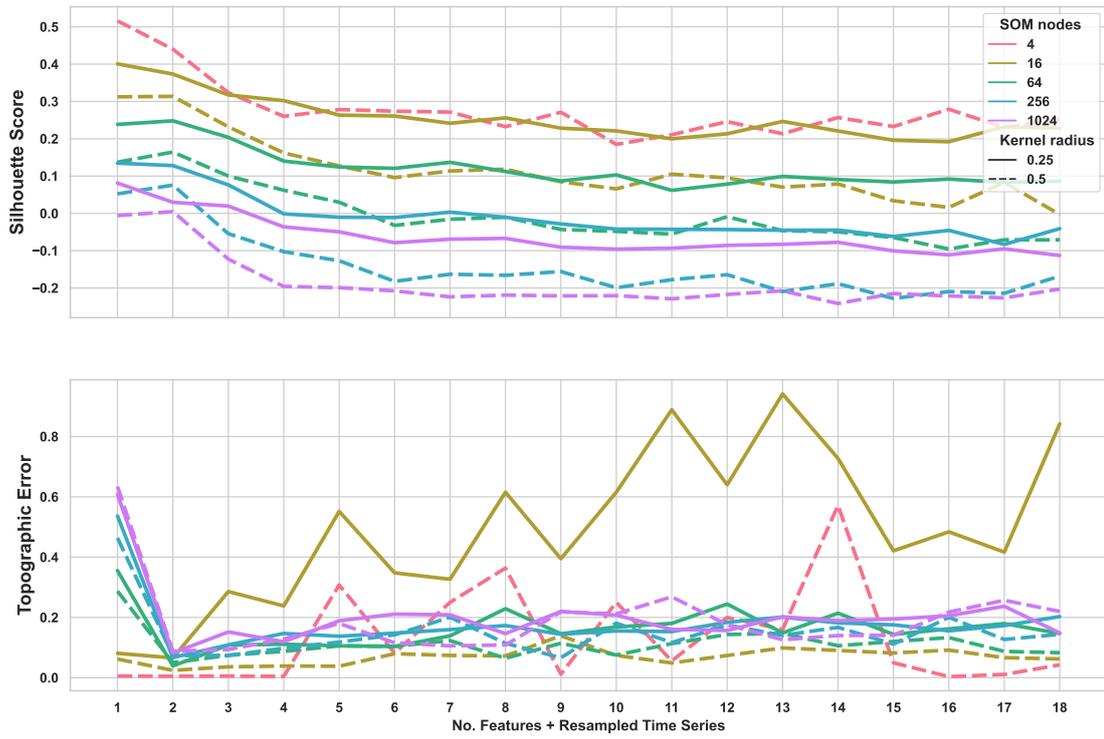
Figure B.4: Visualization of the deposition Self-organizing Map of the snow cover dataset with 64 nodes. For a detailed explanation of the visualization and color scheme, see Figure B.1.

C

Performance scores for the optimization of the erosion SOM

(a) Silhouette score and topographic error scores for the erosion beach dataset

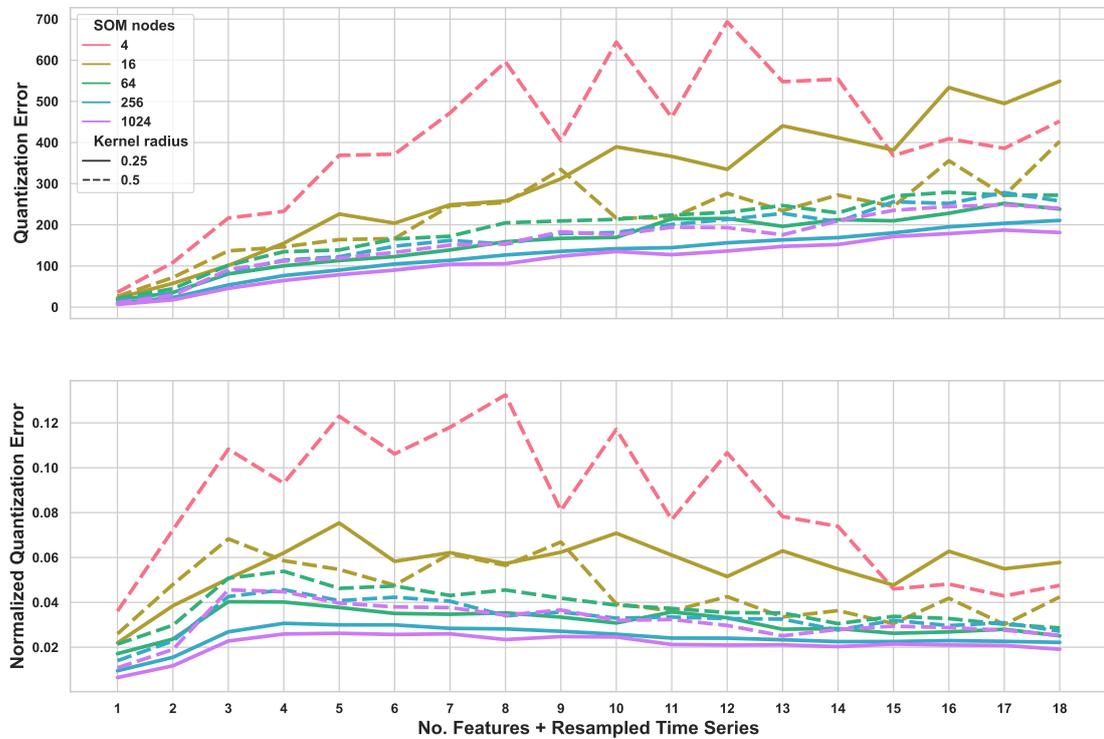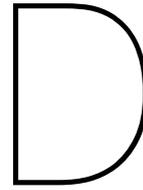**(b)** Quantization errors and normalized quantization errors for the erosion beach dataset

Figure C.1: Performance metrics in relation to the number of features, SOM size (colors), and kernel standard deviation (line type).

D

Appearance of the beach deposition SOM with various training order inputs

In the optimization methods various input orders during training of the SOM are tested, as well as the use of a subset determined on the basis of a maximum dissimilarity sampling algorithm. The resulting optimal choice for input order and subset is partially determined through inspection of the following SOM visualizations.
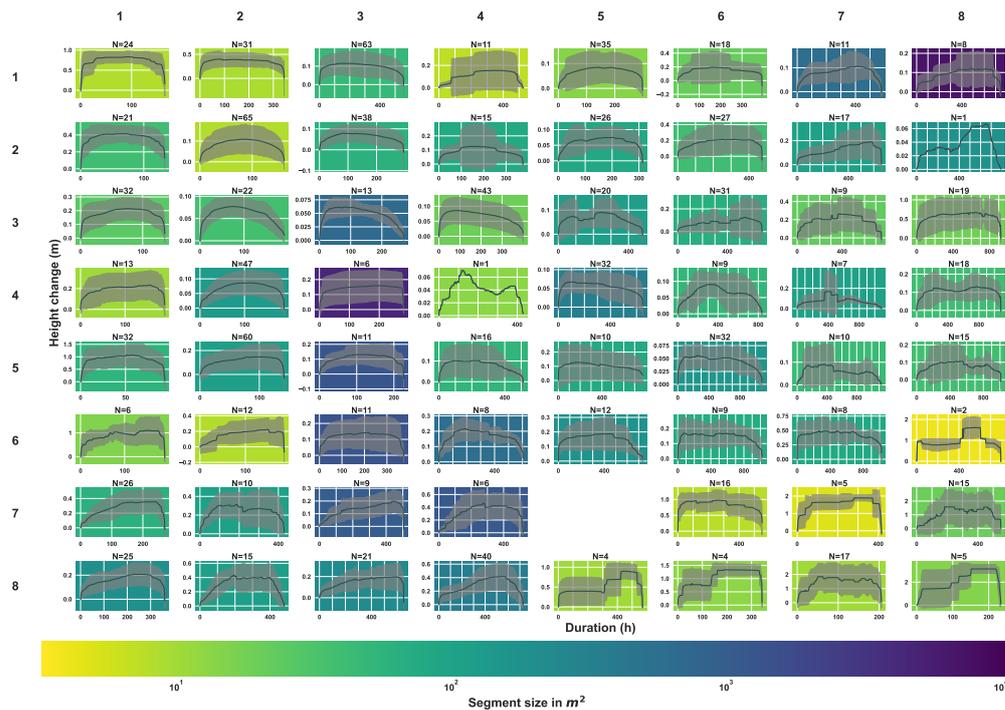

Figure D.1: Visualization of the deposition Self-organizing Map (SOM), trained using the resampled time series + first 7 features from Table 5.4 and a random input order of the full dataset. Each plot represents a SOM node. The x-axis of each plot represents the mean duration of the 4D objects-by-change (4D-OBCs) in the node. The y-axis of each plot represents the mean height change of the 4D-OBCs in the node. The limits of both vary between each graph. The black curves show the mean time series of the 4D-OBCs in each node, and the grey area is the standard deviation. The background colors represent the mean segment size per node (on a logarithmic scale). Nodes without graphs represent nodes where no sample is matched after training.
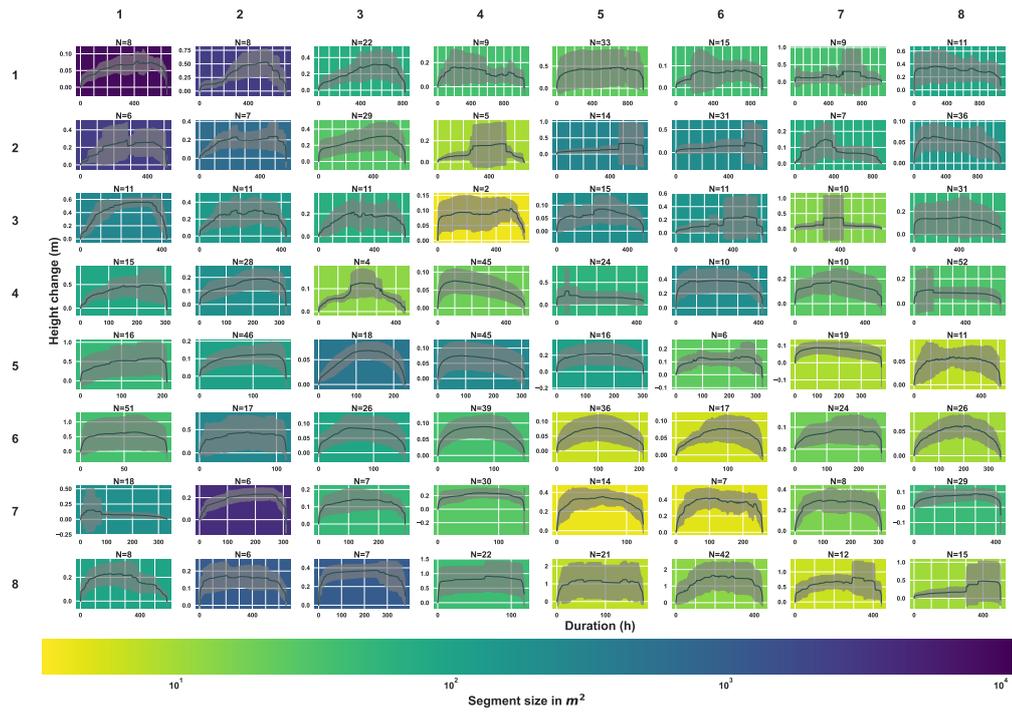
Figure D.2: Visualization of the deposition Self-organizing Map (SOM), trained using the resampled time series + first 7 features from Table 5.4 and a flipped maximum dissimilarity sampling input order of the full dataset. For a detailed explanation of the visualization and color scheme, see Figure D.1.
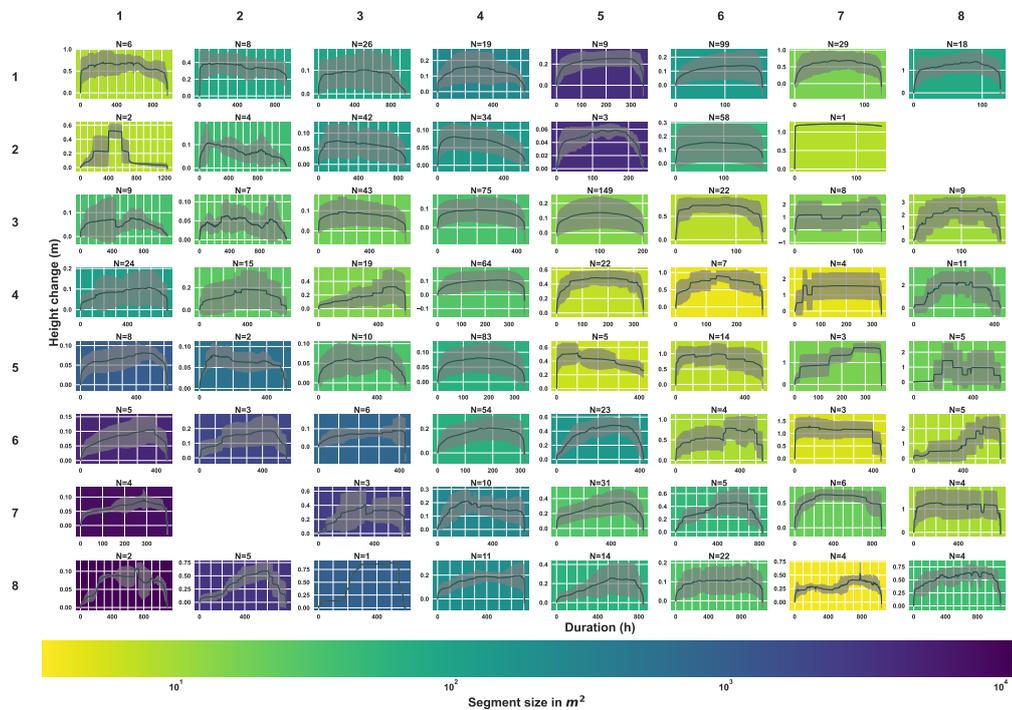

Figure D.3: Visualization of the deposition Self-organizing Map (SOM), trained using the resampled time series + first 7 features from Table 5.4 and a maximum dissimilarity sampling input order and subset of 300 samples. For a detailed explanation of the visualization and color scheme, see Figure D.1.