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.

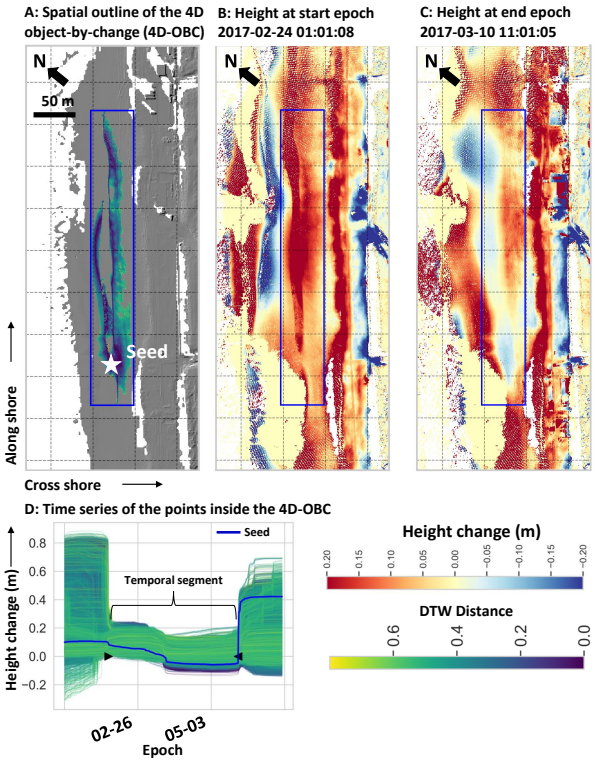

Figure 2. Temporal evolution and extracted spatial segment of an intertidal sandbar erosion surface activity.

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.

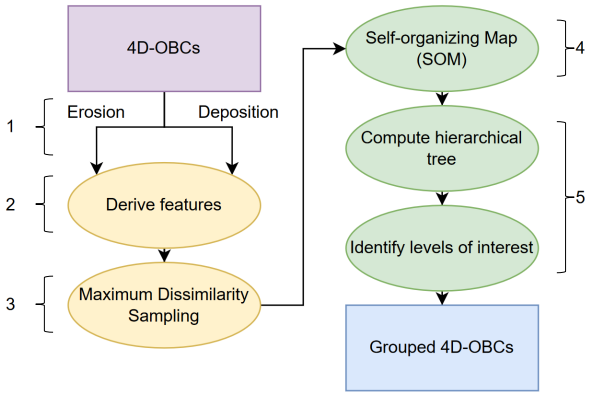

Figure 3. Workflow with the five main steps to obtain grouped 4D-OBCs based on their features.

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:

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

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:

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

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:

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

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:

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

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

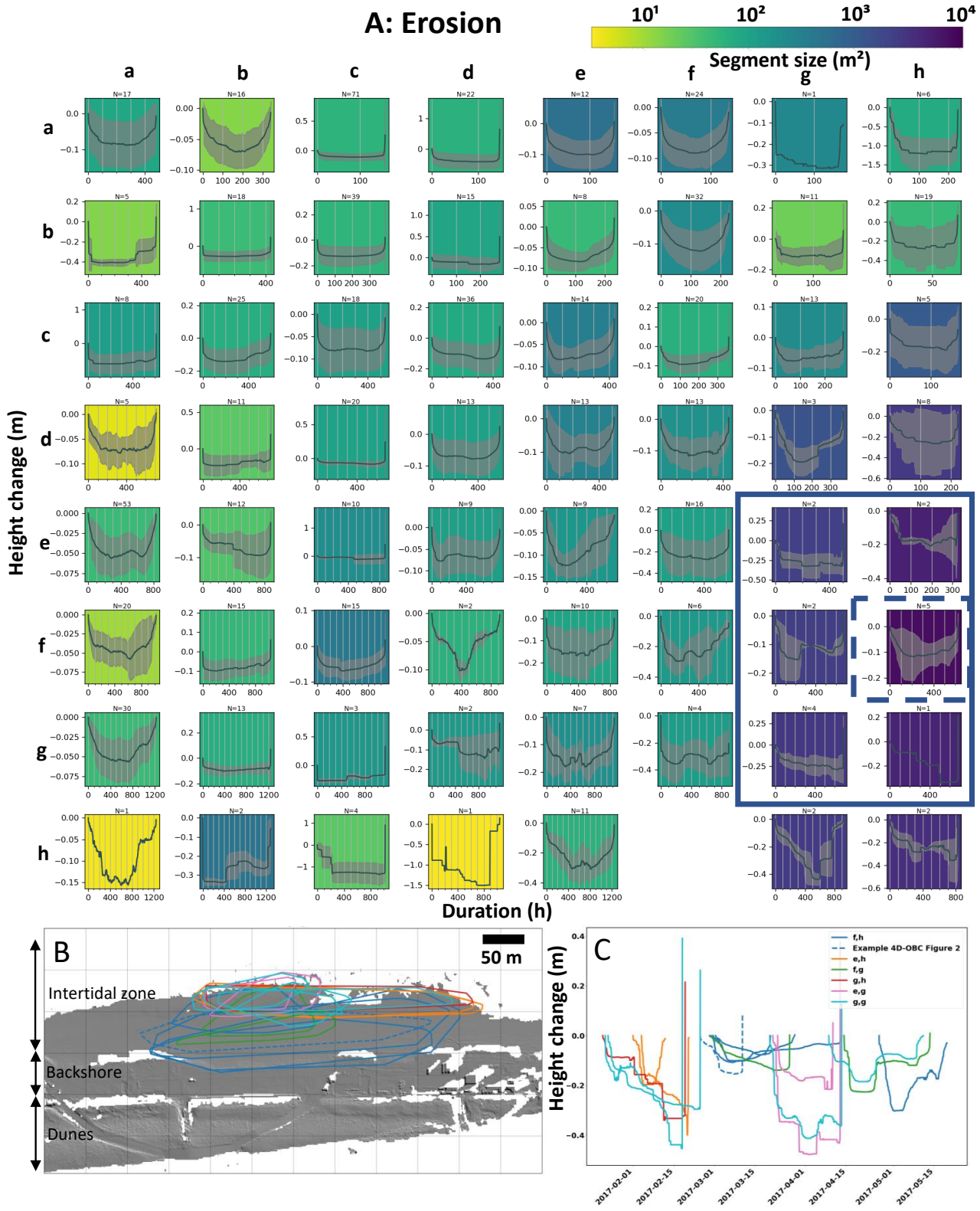

Figure 4. A) Visualization of the erosion Self-organizing Map (SOM E). Each plot represents a SOM node. The x-axis of each plot represents the mean duration of the 4D objects-by-change (4D-OBCs) in the node. The y-axis of each plot represents the mean height change of the 4D-OBCs in the node. The limits 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

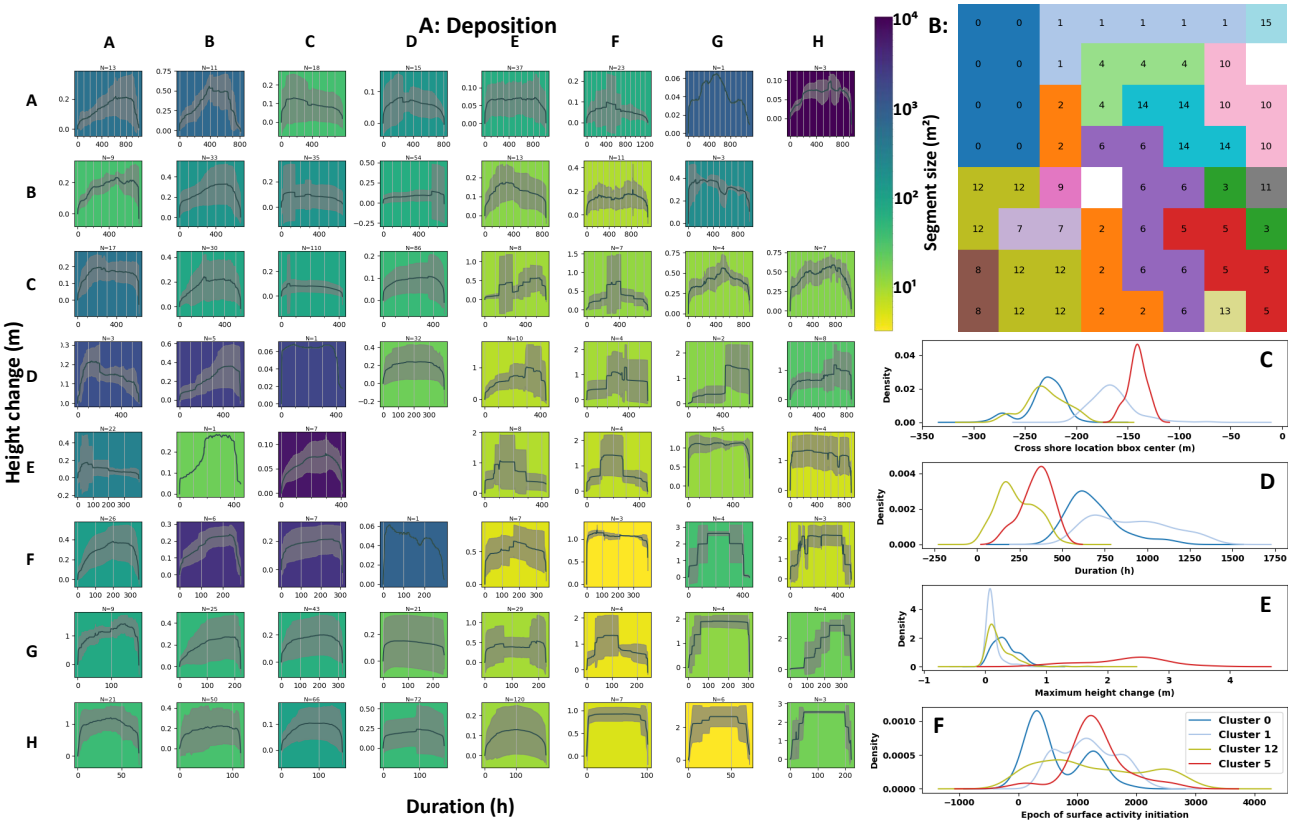

Figure 5. A) Visualization of the deposition Self-organizing Map (i.e., SOM D). For detailed explanation refer to the caption of Figure 4. B) The clusters found with a distance threshold of 230. Each grid point refers to a node in A, each color and number represents one cluster of nodes. C-F) Density 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 surface 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.