Unsupervised Deep Clustering on Spatiotemporal Objects Extracted from 4D Point Clouds for Automatic Identification of Topographic Processes in Natural Environments

Jiapan Wang[1], Katharina Anders[1]

[1] Professorship of Remote Sensing Applications, TUM School of Engineering and Design, Technical University of Munich, Germany - (jiapan.wang, k.anders)@tum.de

Keywords: Time Series, Autoencoder, Deep Clustering, Topographic Change Analysis, 4D Objects-by-change

Abstract

Topographic processes, such as sediment erosion, accumulation, and transport are crucial for understanding the evolution of natural landscapes. Current developments in permanent laser scanning (PLS) technology and 4D change detection methods have made it possible to extract spatiotemporal change objects from near-continuous 3D observations, e.g., 4D objects-by-change. However, the automatic characterization and identification of these processes remain challenging due to the complex spatiotemporal data and unpredictable types of topographic processes in natural environments. In this paper, we present a time series-based unsupervised deep clustering framework for identifying topographic processes without manual feature engineering and annotations. By leveraging the representation learning capability of autoencoders, especially using convolutional neural networks (CNNs) as feature extractors, our approach implements the dimensionality reduction of the original inputs to uniform low-dimensional vectors in latent space. Subsequently, after jointly optimizing the reconstruction and clustering loss, our model generates unique clusters with high intra-cluster similarity and inter-cluster variability. We validated the proposed method on a six-month 4D dataset, acquired at Kijkduin sandy beach (The Netherlands), yielding distinctive clusters that correspond to sediment change phenomena. Our results demonstrate that the deep learning-based method successfully identifies topographic processes, providing an efficient and scalable alternative to traditional feature engineering-based approaches. This work highlights the potential for automating topographic process identification and supporting long-term environmental monitoring.

1. Introduction

Topographic changes are countlessly and constantly occurring across spatial and temporal scales on Earth's surface. These changes shape diverse landscapes with complex morphology from high-mountain to coastal environments (O'Dea et al., 2019). The factors that induce topographic changes can be attributed to different environmental forces (e.g., wind, water, tectonic activities) or anthropogenic influences (e.g., infrastructure and different forms of land use). Understanding the impact of these activities on surface evolution poses significant challenges, such as the complex composition of activities, unpredictability of natural hazards, and long-timescale monitoring. However, through the systematic observation and analysis of surface changes, it is possible to detect and infer patterns that cause topographic dynamics (Kromer et al., 2017). This is essential for better understanding the underlying environmental processes and their interactions with human activities and provides valuable information to mitigate, for example, potential climate change consequences (Eitel et al., 2016).

1.1 From Change Detection to Characterization

Multi-temporal 3D observation acquired by LiDAR sensors or photogrammetry has shown the potential to increase our insight into the mechanisms of topographic dynamics. Surface changes can be detected and quantified by magnitude or volume between multiple epochs of 3D point clouds (Qin et al., 2016). As high-frequency (sub-hourly to weekly time intervals) and high spatial resolution (sub-centimeter to meter) 4D point cloud datasets are becoming increasingly available through permanent laser scanning (PLS) (Eitel et al., 2016), the exploitation of the dense point cloud time series that has richer spatiotemporal information is actively encouraged in the detection of subtle processes of surface dynamics (Anders et al., 2020; Vos et al., 2022; Kuschnerus et al., 2024). Leveraging the spatial and temporal properties of 4D datasets, many studies demonstrate a research focus shifting from few epochs to near-continuous monitoring of topographic dynamics (Woodcock et al., 2020; Kromer et al., 2017; Williams et al., 2018; O'Dea et al., 2019).

To automatically extract surface changes over a long period, a time series-based topographic change analysis advances the pairwise change detection to '4D objects-by-change' (4D-OBCs) (Anders et al., 2021). Spatial neighborhoods that experience similar topographic processes within identified subperiods are detected based on their time series similarity. The resulting object defined by spatial extent and timespan represents an individual topographic change process (Anders et al., 2020, 2021). While this enables the automatic extraction of numerous topographic changes from thousands of epochs of point clouds, their interpretation is difficult without further grouping. However, methods for the characterization of these changes are still lacking in research. An accurate way to categorize these changes is expert labeling, but it is time- and human-resource-consuming. Therefore, it is highly relevant to develop a pipeline with high degree of automation, i.e. low supervision and manual inputs, to assign meaningful labels to change processes. This is a key step in pushing the paradigm from topographic change detection to change characterization and bridging 4D change analysis and environmental monitoring.

Recent research directly applies clustering methods (e.g., K-Means, DBSCAN, and agglomeration clustering) on full point cloud time series, which enable grouping of full-length surface time series (Kuschnerus et al., 2021). Furthermore, for object-wise topographic changes, specific spatial and temporal handcrafted features derived from 4D-OBCs are used as the inputs of self-organizing map (SOM) for clustering change objects into different groups (Hulskemper et al., 2022). Beyond this unsupervised grouping, a supervised classification solution can be considered: a classifier is trained with labeled data, and the target class of each change object from real-world data is inferred. Zahs et al. (2024) propose to leverage synthetic 4D datasets and pre-defined labels for this. However, these methods highly depend on feature engineering and require a good understanding of the dataset and preferably prior knowledge of types of changes in scenarios.

In general, topographic monitoring contains a series of tasks like detection, classification, identification, and prediction of changes. Taking sequential 3D observations as an example, change detection determines "when" and "where" a change occurs. Given the location and timestamps of the change information, change classification decides on the "what" the type of change is (i.e., erosion, deposition, and transportation of surface materials) and may include information relating to the drivers (e.g., environmental forces or anthropogenic influences). The "when", "where", and "what" together contribute to the change identification. As automatic change detection methods have recently been developed for 3D time series analysis (Anders et al., 2021; Winiwarter et al., 2023; Kuschnerus et al., 2024), this paper aims to enhance change identification to understand the extracted change objects by using deep learning strategies.

1.2 Deep Learning for Topographic Change Analysis

Typical feature engineering and machine learning algorithms are commonly used to find intrinsic patterns among geospatial or time series data (Sarker, 2021). However, defining the features to be used for classification is subjective, depending on expert knowledge. The reproducibility of feature engineering strategies across different geographic settings is even more challenging due to the different spatiotemporal properties and unpredictable types of changes, especially when spatial and temporal features need to be considered simultaneously. Recent advances in representation learning research are driven by deep learning models using neural networks. These models automate feature engineering and extract general deep features that represent input data in a latent feature space. The extracted features can then be used for downstream tasks, such as being connected to classifiers. Learning good representations from the data makes it easier to extract information (Bengio et al., 2013). Many studies have shown the promising representation learning capability of deep learning models on spatiotemporal datasets (Wang et al., 2022), which primarily focus on understanding of image series. There is less research about deep learning for topographic change analysis, especially using point cloud time series data. This paper presents the use of deep learning models to extract comprehensive representations (i.e., deep features) instead of manual feature engineering from topographic changes.

So far, research focuses on deep learning for change analysis on urban objects (Stilla and Xu, 2023; de Gélis et al., 2023). Insights on applying deep learning techniques to automatically monitor topographic changes in natural environments are still worth investigating. The performance of deep learning models usually relies on extensive annotated training data, i.e., supervision. While handling point cloud time series datasets, especially acquired in natural environments (e.g., beach, mountain, glacier, landslide), there are some apparent challenges in annotating the datasets: 1) the boundaries of topographic events are not clearly defined in contrast to the distinct boundaries of urban objects; 2) the type of changes are not all known in advance; 3) the spatial and temporal properties of different processes are highly variable; (4) a parameter adaptation of methods is always needed for different geographic settings. Therefore, it is important to develop unsupervised data-driven methods that require few or even no annotations to perform topographic change identification. Inspired by the deep embedded clustering (DEC) principle that jointly optimizes deep representations learned by autoencoders and performs clustering with latent representations (Caron et al., 2018; Xie et al., 2016), we propose an unsupervised learning strategy to enhance the identification of topographic change processes in this paper. The contribution of this work can be summarised as follows:

- Conversion of 4D objects-by-change (4D-OBCs) into a topographic change process dataset.
- Time series-based DEC for unsupervised grouping of topographic change processes.
- Derivation of geometric properties from clusters to characterize topographic processes.

2. Data and Study Area

This study was conducted on a point cloud time series acquired hourly by a permanent terrestrial laser scanner installed at the sandy beach of Kijkduin, The Netherlands ($52°04'14''N, 4°13'10''E$) (Vos et al., 2022). The beach was monitored during the winter of 2016 - 2017 using a Riegl VZ-2000 laser scanner mounted on a stable frame overlooking the beach, generating hourly point clouds with densities of 2 - 20 points/m^2. The 4D point cloud time series of around six months duration (2,942 epochs) is published in the PANGAEA data repository (Vos et al., 2022). The area of interest, shown in Fig. 1, has a spatial extent of around 300 x 600 m.

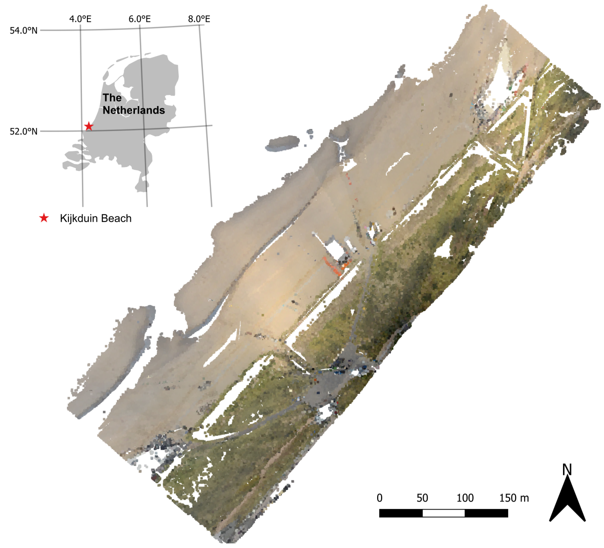

Figure 1. RGB colored 3D point cloud of the sandy beach in Kijkduin. The star marks the location of the study area. World Borders © thematicmapping.org 2017.

The extraction of 4D-OBCs (Anders et al., 2020) from the full dataset and all required preprocessing is detailed in Anders et al. (2021). The resulting 4D change objects, defined by spatial extent and timespan, depict the individual topographic change process of the surface. Four examples are shown in Fig. 2.

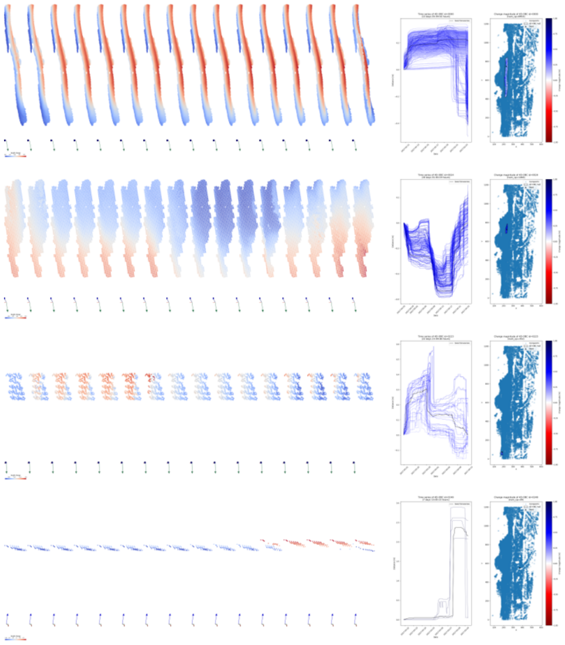

Figure 2. Different types of 4D-OBCs are shown from top to bottom: sand accumulation, sand erosion, unknown activity, and beach hut in the summer. Left plots show change maps of 16 epochs (subsampled) and right plots show the time series and locations of 4D-OBCs.

In total, 2,021 4D objects were extracted from the Kijkduin sandy beach dataset, waiting for characterization and summarizing. An overview of the spatiotemporal occurrence map of 4D-OBCs is shown in Fig. 3.

3. Methods

This paper proposes a three-stage method to implement topographic change process identification in an automatic, unsupervised manner (Fig. 4). Firstly, a spatiotemporal segmentation method is applied to extract change objects (so-called 4D-OBCs) from dense point cloud series (presented by Anders et al. (2021)). From this, we build the topographic change process dataset by selecting multiple time series from each change object. Secondly, we use a DEC-based method to obtain clustering centers and label assignments of time series. In the final stage, we summarize the geometric properties of clustering results to provide information for the characterization of the dataset. The method was validated on a 4D dataset collected by PLS over the Kijkduin sandy beach (Vos et al., 2022).

3.1 Topographic Change Process Dataset

As a first step, we bring the 4D objects as extracted from the 3D time series into a data format that can be effortlessly used for training and inference by deep neural models. Commonly used digital media are images, text, and signals, predominantly deployed in computer vision, natural language processing, and sequence analysis, respectively. To investigate the changes in the topographic surface, we use time series as major representations of change objects. All spatial locations, so-called core points, belonging to an object, ranging from ten to tens of thousands, have similar time series during the timespan of the change object occurrence. Considering a balance of samples, we select 10 core points (the lowest number constituting one object in this dataset) from each object as representatives to contribute to the time series dataset. In the end, a total of 20,210 data points are present in the topographic change process dataset, which is separated into 12,050 positive and 8,160 negative changes.

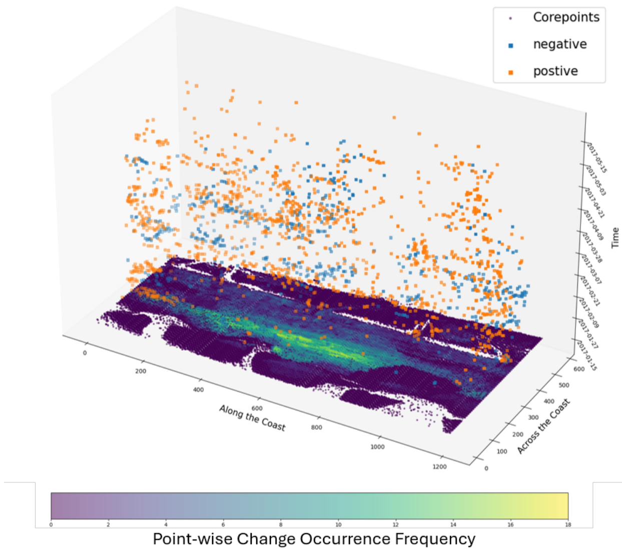

Figure 3. Spatiotemporal occurrence map of 4D-OBCs with occuring time as z axis, and location that along and across coast as x and y axis. The "positive" and "negative" points correspond to surface accumulation and erosion.

3.2 Deep Embedded Clustering

The concept of deep clustering has been proven to tackle the problem of clustering complicated high-dimensional data in a latent space (i.e., deep feature space) leveraging the huge success of representation learning techniques with deep neural networks (Caron et al., 2018; Xie et al., 2016; Guo et al., 2017). It typically includes two major steps: 1) learning comprehensive representations from original data; 2) clustering data instances in the latent vectorized space. The major objective is to jointly optimize deep representation learning and clustering within one iterable framework. In contrast to traditional machine learning methods (i.e., feature engineering and clustering), deep clustering not only automates the extraction of deep features from high-dimensional data but also optimizes the clustering itself during the learning process.

Common network structures used for representation learning are autoencoders, generative learning networks, or contrastive learning networks. We use a time series-based autoencoder to implement the dimensional reduction of inputs of different lengths into deep features of uniform low-dimensional vectors and to initialize the weights of the model. In the clustering learning phase, there are two common types: hard assignment clustering (Caron et al., 2018) and soft assignment clustering (Xie et al., 2016). Hard assignment produces discrete one-hot cluster labels $\tilde{y}_i$ for each data point, while soft assignment generates a set of continuous cluster assignment probabilistic vectors $z_i \in \mathcal{R}^K$, which are the outputs of softmax activated K-dimensional logits of deep features. For the final clustering assignment, the label can be obtained by selecting the dimension with the maximum probability.

Considering our problem of clustering a set of n time series data points $\{x_i \in X\}_{i=1}^{n}$ into k clusters, we propose to use an

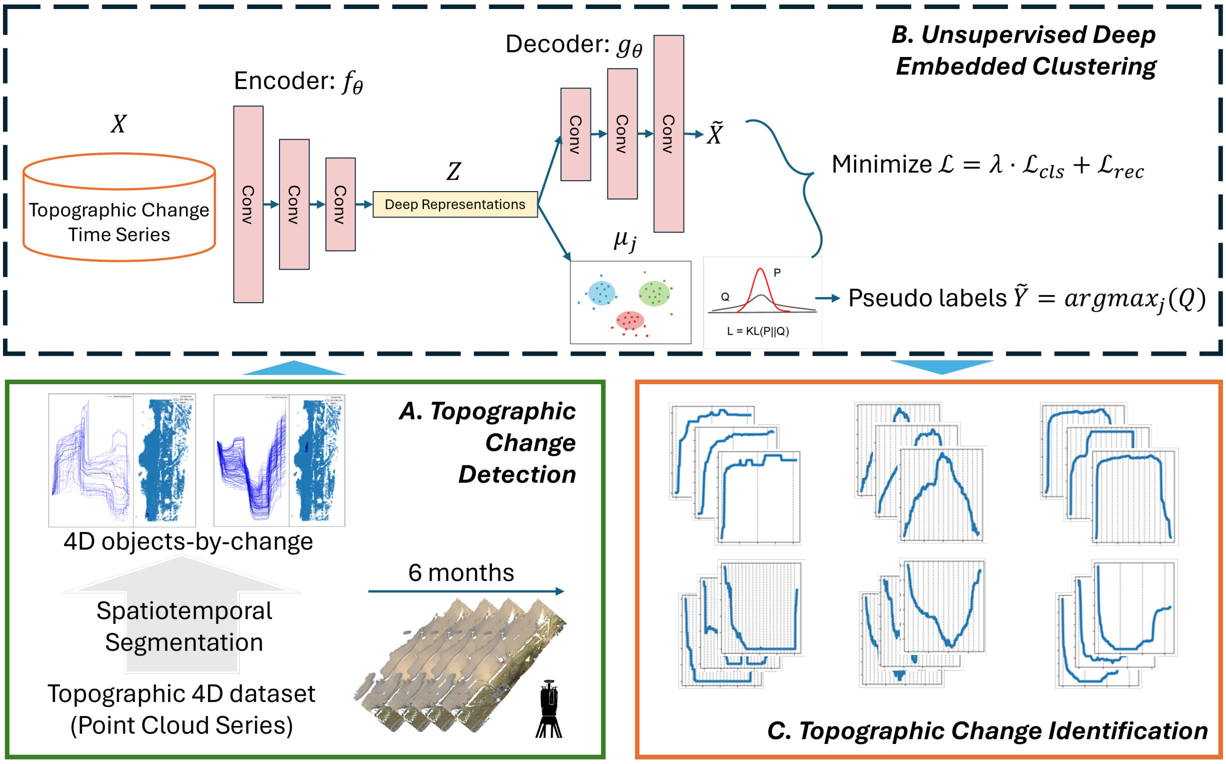

Figure 4. A three-stage identification pipeline for topographic changes. A. Extracting spatiotemporal change objects from 4D point cloud dataset. B. Implementing two phases (reconstruction and clustering) of unsupervised deep embedded clustering (DEC) by considering time series representatives of 4D-OBCs as inputs. C. Identify the cluster of individual topographic change processes. The first and second rows are different types of positive changes and negative changes (i.e., accumulation and erosion), respectively.

autoencoder-based architecture to extract latent representations, i.e., embeddings $\{z_i \in Z\}_{i=1}^n$, and cluster them into different groups $\{\tilde{y}_i \in \tilde{Y}\}_{i=1}^n$. We follow the technique details in Guo et al. (2017), which jointly optimizes the reconstruction loss $\mathcal{L}_{rec}$ and clustering loss $\mathcal{L}_{cls}$. The training can be split into two phases: the pretraining phase and the clustering phase.

3.2.1 Latent Representations by Autoencoders We apply a time series-based autoencoder to automate the feature extraction process from the topographic change process dataset. The autoencoder consists of an encoder-decoder structure, where the encoder is implemented using three-layer convolutional neural networks (CNNs), and the decoder is implemented using the mirror structure with transposed CNNs. The encoder f_θ, where θ are learnable parameters, compresses the input time series X into a lower-dimensional latent space Z. The decoder g_θ then tries to reconstruct the input from the latent space through transposed CNNs, resulting in the reconstructed time series $\tilde{X} = g_\theta(f_\theta(X))$, as shown in Fig. 5. The training objective of the first phase is to minimize the reconstruction loss $\mathcal{L}_{rec}$, such as the Mean Squared Error (MSE) between the input and reconstructed time series. This can be described as:

$$\mathcal{L}_{rec}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x}_i)^2 \tag{1}$$

By minimizing $\mathcal{L}_{rec}$ loss, the autoencoder is guided to extract meaningful features from the topographic change process dataset. After the pretraining phase, the weights θ of the encoder f_θ are initialized for deep clustering, as the deep representations $Z = f_\theta(X)$ generated by the encoder will later be used as input for the clustering phase.

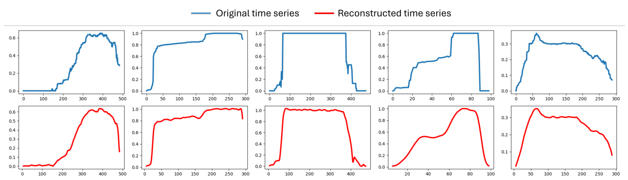

Figure 5. Original and reconstructed time series.

3.2.2 Clustering and Optimizing Following the pretraining of the autoencoder, we add a soft clustering layer after the encoder to cluster the deep representations in latent space. We first inherit the model weights from the encoder to extract deep representations of each topographic change process ($f_\theta : x_i \rightarrow z_i$), and then initialize k cluster centers $\{\mu_j\}_{j=1}^k$ in latent space using k-means. In the clustering phase, the similarity between embedded point z_i and cluster μ_j is calculated by Student’s t-distribution (van der Maaten and Hinton, 2008):

$$q_{ij} = \frac{(1 + \|z_i - \mu_j\|^2)^{-1}}{\sum_j (1 + \|z_i - \mu_j\|^2)^{-1}} \tag{2}$$

where $\{q_{ij} \in Q\}$ can be interpreted as the soft assignment probability of data point x_i to cluster μ_j. In order to iteratively refine the clusters by learning from their high-confidence assignments, we use an auxiliary target distribution $\{p_{ij} \in P\}$, same as used in Xie et al. (2016). The target distribution is:

$$p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_j (q_{ij}^2 / \sum_i q_{ij})} \tag{3}$$

where the target distribution $\{p_{ij} \in P\}$ is actually defined by

q_{ij}. Thus, this can be seen as a form of self-training. Specifically, the target distribution can give more attention to data points with high confidence (i.e., close to the cluster centers μ_j) and it is iteratively updated based on the current soft assignments Q. During the clustering phase, we employ the Kullback-Leibler (KL) divergence between distributions P an Q as the clustering loss, defined as:

$$\mathcal{L}_{cls} = KL(P||Q) = \sum_{i}^{n} \sum_{j}^{k} p_{ij} log \frac{p_{ij}}{q_{ij}} \tag{4}$$

where $\mathcal{L}_{cls}$ is the KL clustering loss. To fine-tune the encoder f_θ, we minimize a combined loss function that consists of the reconstruction loss and the clustering loss. The final learning objective is defined as:

$$\mathcal{L}(\theta) = \mathcal{L}_{rec} + \lambda \mathcal{L}_{cls} \tag{5}$$

where $\mathcal{L}_{rec}$ is the reconstruction loss, shown as Eq. 1 (i.e., measured by MSE), and $\mathcal{L}_{cls}$ is the clustering loss, shown as Eq. 4. $\lambda > 0$ is a hyperparameter that controls the degree of distorting embedded space by clustering (Guo et al., 2017). The clustering phase of the training process jointly optimizes this combined loss $\mathcal{L}(\theta)$ while continuously updating the target distribution to guide the clustering process. Training is stopped once the cluster assignments become stable, with insignificant changes between consecutive iterations. Finally, pseudo-labels $\tilde{y}_i$ can be derived for each topographic time series x_i using softmax activation of the soft assignments: $\tilde{y}_i = argmax_j(q_{ij})$. The maximum probability is used to indicate the certainty of the pseudo labels, referred to as the confidence score ϕ. The pseudo-code of the proposed method is shown in Algorithm 1.

Algorithm 1 DEC for Topographic Processes

Input: Topographic Change Process Dataset: X; Number of clusters: k; Autoencoder's model: f_θ, g_θ; Pretraining and clustering epochs: E_1, E_2
Step 1: Autoencoder's pretraining
Initialize encoder f_θ and decoder g_θ with random weights
for epoch = 1 to E_1 **do**
  # Compress the input data into deep representations
  $Z = f_\theta(X)$
  # Reconstruct the input data from deep representations
  $\tilde{X} = g_\theta(Z)$
  # Compute reconstruction loss by MSE
  $\mathcal{L}_{rec} = MSE(X, \tilde{X})$
  # Update model's weights by back-propagation
  Update f_θ, g_θ
end for
Step 2: Clustering and optimizing
Initialize cluster centers μ_j by k-means
$\mu_j = KMeans(f_\theta(X), k)$
for epoch = 1 to E_2 **do**
  # Compute deep representations and reconstruction
  $Z = f_\theta(X), \tilde{X} = g_\theta(Z)$
  # Update soft assignments and target distribution
  Update Q and P using Eq. 2 and Eq. 3
  # Compute reconstruction loss and clustering loss
  $\mathcal{L} = \mathcal{L}_{rec} + \lambda \mathcal{L}_{cls}$
  # Update model's weights by back-propagation
  Update $f_\theta, g_\theta, \mu_j$
  if clustering assignments become stable **then**
    Early stop training.
  end if
end for
Step 3: Pseudo label assignment
Get pseudo labels from soft assignments
$\tilde{Y} = argmax_j(Q), \phi = max(Q)$
Output: Predicted pseudo-labels: $\tilde{Y}$; Confidence score: ϕ; Trained models: f_θ, g_θ; Clustering centers: μ_j

3.3 Experimental Settings

All experiments in this paper were conducted on a machine equipped with an Intel® Xeon(R) w7-3455 CPU and an NVIDIA RTX A4500 GPU with 20 GB of memory. This setup was used for both training and evaluation of the models.

Autoencoder model: The autoencoder consists of a 3-layer 1D CNN for the encoder and mirrored transposed CNN for the decoder. The encoder takes time series data as input and reduces the dimensionality to a deep representation of size 32, while the decoder reconstructs the original input from the deep representations. We use MSE as the reconstruction loss function and employ an Adam optimizer to improve training efficiency.

Deep clustering layer: A soft assignment clustering layer is attached to the encoder to assign data points to cluster distributions. The objective is to jointly minimize the reconstruction loss and the KL clustering loss (Eq. 5). During the clustering phase, following the settings of Guo et al. (2017), we use a Stochastic Gradient Descent optimizer to jointly optimize the combined loss of reconstruction and clustering.

Hyperparameters: Several hyperparameters were carefully chosen in this paper to ensure effective training and model performance. The clustering loss weight is set to $\lambda = 0.01$ to balance the trade-off between reconstruction accuracy and clustering quality, ensuring that the clustering does not overly distort the latent space. An update interval of 10 is selected to avoid instability of the clustering layer. A learning rate of 1e-3, together with the weights decay of 1e-5, ensures smooth convergence and prevents overfitting of the model. The batch size is 512 to optimize computational efficiency. The data are normalized from the 0th to the 95th percentile to reduce outliers and ensure the input is between 0 and 1. The time series length was set to 1400 time steps, which corresponds to the longest 4D objects timespan, ensuring to capture sufficient temporal patterns of all 4D objects. The shorter time series are extended by zero padding. Finally, the number of clusters ranges from 2 to 20 at increments of 1 to explore different clustering structures and assess if the model can adapt to different levels of data granularity. These hyperparameters are chosen based on grid search experiments as well as best practices in deep embedded clustering and time series modeling with the goal of jointly optimizing reconstruction and clustering performance.

4. Results

We test the proposed pipeline on 4D objects extracted from the Kijkduin dataset of hourly PLS observations over a 6-month period. 2,021 4D-OBCs are detected from this 3D time series, from which we derive 20,210 topographic change time series as input for our method. The full dataset was split into positive (n=12,050) and negative (n=8,160) changes, indicating accumulation and erosion activities, respectively. We then train the autoencoder and deep clustering layer for each subset to obtain the soft cluster assignments and corresponding probability distributions for each data point (i.e., time series x_i). The cluster assignment with the maximum probability is selected as the pseudo label $\tilde{y}_i$, and the probability can be considered as the level of confidence that the data point belongs to the target group. Thus, we call the maximum probability as the confidence score ϕ. The clusters and confidence scores can provide us with patterns of time series to help us characterize topographic change processes.

We implement deep clustering with k values from 2 to 20 and use the silhouette score to assess the clustering performance (Rousseeuw, 1987). The silhouette score is an internal metric that measures how similar a data point is to its cluster assignment compared to other clusters, between -1 and 1, with higher values indicating better clustering performance.

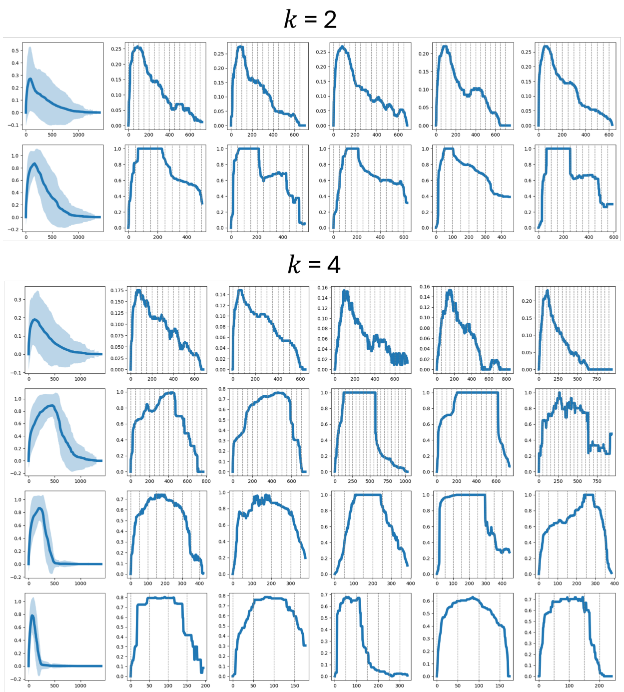

Figure 6. The clustering results of $k = 2$ and $k = 4$. The first column is the cluster-wise mean and standard deviation, and the other five columns are time series closest to cluster centers.

Clustering interpretation: We show the clustering results of positive changes (i.e., accumulation activity) with $k = 2$ and $k = 4$ in Fig. 6 and the clustering results of both positive and negative changes with $k = 9$ in Fig. 7 (which is the highest k while maintaining silhouette scores higher than 0.5). From the clustering results, we observe several interesting aspects. First, the binary clustering categorizes the time series into small-magnitude (first row) and large-magnitude (second row) changes. When increasing k, new distinct patterns appear, such as the third and fourth row of the quad-cluster (Fig. 6), reflected in the duration of the changes. This also demonstrates the ability of CNN-based models to extract localized features from active data periods. Second, although increasing k enables mining more varied patterns of change, this behavior does not persist with infinite growth. When k is increased to a certain level, different clusters emerge with apparent similarities. For instance, C_3 and C_5 of positive changes in Fig. 7 both show a pattern of gradual starting and sudden ending. Most importantly, during the clustering training phase, the clustering layer constantly keeps similar data points close and different points far away in latent space, which ensures consistency within every single category and heterogeneity between different categories to some extent. As shown in Fig. 7, the time series in each row presents a very high degree of similarity, whereas the time series in each column presents differences in some aspects, such

	Positive			Negative		
K	**duration (h) mean (±std)**	**magnitude (m) mean (±std)**	**shape**	**duration (h) mean (±std)**	**magnitude (m) mean (±std)**	**shape**
K_1	458 (± 187)	0.56 (± 0.24)	GG	388 (± 140)	0.53 (± 0.21)	GG
K_2	329 (± 258)	0.37 (± 0.24)	SG	229 (± 186)	0.51 (± 0.24)	SG
K_3	669 (± 93)	0.98 (± 0.07)	GS	480 (± 132)	0.93 (± 0.11)	SS
K_4	344 (± 127)	0.97 (± 0.08)	SS	635 (± 114)	0.99 (± 0.05)	SS
K_5	495 (± 145)	0.97 (± 0.07)	GS	919 (± 131)	0.54 (± 0.23)	GG
K_6	875 (± 120)	0.95 (± 0.10)	GS	673 (± 190)	0.43 (± 0.18)	GG
K_7	222 (± 145)	0.91 (± 0.15)	SS	1042 (± 62)	0.99 (± 0.02)	GS
K_8	735 (± 211)	0.43 (± 0.22)	GG	823 (± 103)	0.99 (± 0.04)	GS
K_9	1136 (± 96)	0.99 (± 0.27)	SS	253 (± 134)	0.95 (± 0.11)	SS

Table 1. Characteristics of clusters.

as active duration, shape, and magnitude.

Characterization of change objects: To provide information for the characterization and classification of 4D objects, we summarize spatiotemporal properties of each cluster shown in Fig. 7 in Table 1. Duration refers to the timespan of the change objects. Magnitude represents the peak value of the change, and the numbers shown are normalized by the max-min normalization. Each time series is then clipped into two phases (i.e., the beginning and the end of the change) based on its peak, with G representing a gradual change and S representing a sudden change. The combination of letters represents the shape of topographic change processes. The properties exhibited in our results show clear differences across clusters, providing valuable characteristics for describing the change objects.

5. Discussion

The proposed DEC method for topographic change processes demonstrates notable performance in automatically grouping topographic processes into distinct clusters. This is a good indication that data-driven self-supervised learning shows the potential to provide meaningful features and reduce traditional feature engineering. The autoencoder effectively extracts deep representations from the inputs. It furthermore exhibits denoising capabilities that keep the significant change information, such as trend, shape, and volume of the change (Fig. 5). The clustering layer assigns pseudo labels depending on the distance between each data point and the cluster centers in the latent space. The final clustering results demonstrate evident intra-cluster similarity and inter-cluster variability.

Compared to traditional machine learning methods used for topographic change process characterization, e.g., raw time series clustering (Kuschnerus et al., 2021) or SOMs (Hulskemper et al., 2022), our DEC-based approach has several strengths. First, it eliminates the manual feature engineering, allowing the model to automatically learn features from raw data without compression. This preserves the full temporal information and increases the inference efficiency. Second, the clusters produced by our model reflect variability in the spatial and temporal properties of the topographic change process, enabling the capture of unknown patterns. Third, the confidence scores of each data point provide us with information on how certain a data point belongs to its cluster. In contrast, data with a low confidence score shows the uncertainty or ambiguity of changes. More potential remains to explore self-supervised learning methods for topographic change analysis and the model's transferability across different datasets.

A key limitation of our approach is the lack of prior knowledge about the number of clusters, which requires us to perform a grid search of an appropriate confidence threshold and to use silhouette scores to determine the optimal number of

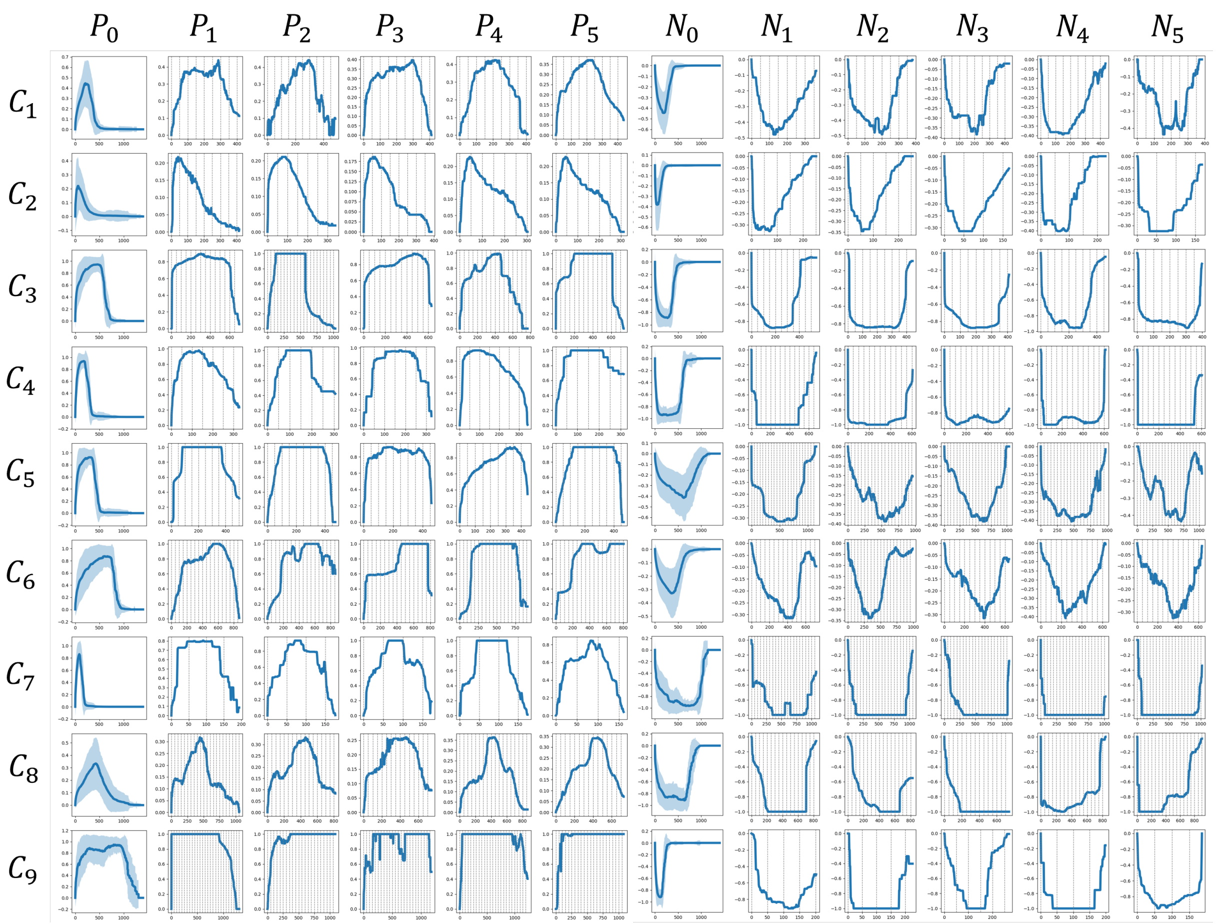

Figure 7. The clustering results of $k = 9$. C_1 to C_9 are cluster id, P_0 and N_0 are the cluster-wise mean and standard deviation, and P_0 to P_5 and N_0 to N_5 are top 5 time series of each cluster. P and N mean positive and negative, respectively.

clusters (k). While this approach successfully identified meaningful clusters, it requires more non-parametric computation (e.g., density-based clustering) and transferability to new datasets that need investigation. However, the approach can provide valuable information for understanding whether new inputs belong to existing clusters, helping in real-time classification of topographic change processes. Another limitation could be that we select the spatially closest points of interest for the input dataset, but we do not incorporate spatial features into model training. We are investigating incorporating handcrafted spatial features of 4D objects as auxiliary features in the training (e.g., size, height, width, and volume of change objects), but so far, this has not notably improved the clustering results. The next phase will focus on taking the spatial dependencies between points or 4D objects into account. Additionally, the CNN module requires fixed-length inputs, for which we use zero padding. This achieves evident results due to CNN’s ability to focus on localized features, but it is less effective for handling variable-length time series data. A more dynamic module, such as a recurrent or transformer-based neural network that could process varying input shapes, should be further investigated.

Future research could focus on explicitly incorporating spatial information into the training process, which would likely enhance the model’s ability to capture geospatial dependencies of topographic processes. Integrating multi-modal data sources, such as environmental sensor data and optical imagery, could provide a more comprehensive understanding of topographic dynamics also to a deep clustering model. Furthermore, we may convert our clustering results into preliminary semantic labels to accelerate the annotation process of topographic changes, potentially by leveraging semi-supervised learning to support the community in deep learning with topographic change analysis.

6. Conclusion

This study develops a deep embedded clustering (DEC) approach for the characterization of topographic change processes, focusing on automatically clustering and identifying distinct patterns of 4D objects that are automatically extracted from time series of 3D point clouds. The results show that our approach effectively clusters topographic change processes into notable clusters without requiring any annotations or manual feature engineering. The identified clusters show evident intra-cluster similarity and reflect the inter-cluster variability in meaningful properties, such as duration, magnitude and shape, which provide discrepancy characteristics of topographic change processes. Our approach offers several advantages, including eliminating manual feature engineering, preserving the full temporal information, and providing confidence scores for cluster assignments. However, limitations remain, such as the determination of cluster numbers requiring hyperparameter tuning and the need for fixed-length time series as input. In future research, we will explore explicitly incorporating spatial features and improving the automation of hyperparameter tuning. With this, our research presents the first step towards using deep learning to empower automated characterization and annotation of topographic change processes. Overall, this study highlights the potential of deep learning-based methods to advance the understanding of complex topographic changes and improve automation in geoscience applications.

Acknowledgements

This research was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 535733258 (Extract4D). K. Anders is grateful to the Baden-Württemberg Stiftung for the financial support of the research project CharAct4D through the Postdoctoral fellowship for leading early career researchers.

References

Anders, K., Winiwarter, L., Lindenbergh, R., Williams, J. G., Vos, S. E., Höfle, B., 2020. 4D Objects-by-Change: Spatiotemporal Segmentation of Geomorphic Surface Change from LiDAR Time Series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 352–363.

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.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8), 1798–1828.

Caron, M., Bojanowski, P., Joulin, A., Douze, M., 2018. Deep Clustering for Unsupervised Learning of Visual Features. *Proceedings of the European Conference on Computer Vision (ECCV)*, 132–149.

de Gélis, I., Lefèvre, S., Corpetti, T., 2023. DC3DCD: Unsupervised Learning for Multiclass 3D Point Cloud Change Detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 206, 168–183.

Eitel, J. U. H., 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.

Guo, X., Gao, L., Liu, X., Yin, J., 2017. Improved deep embedded clustering with local structure preservation. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, AAAI Press, Melbourne, Australia, 1753–1759.

Hulskemper, D., Anders, K., Antolínez, J. a. Á., Kuschnerus, M., Höfle, B., Lindenbergh, R., 2022. Characterization of Morphological Surface Activities Derived from Near-Continuous Terrestrial LiDAR Time Series. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2-W2-2022, 53–60.

Kromer, R. A., Abellán, A., Hutchinson, D. J., Lato, M., Chanut, M.-A., Dubois, L., Jaboyedoff, M., 2017. Automated Terrestrial Laser Scanning with Near-Real-Time Change Detection – Monitoring of the Séchilienne Landslide. *Earth Surface Dynamics*, 5(2), 293–310.

Kuschnerus, M., Lindenbergh, R., Vos, S., 2021. Coastal Change Patterns from Time Series Clustering of Permanent Laser Scan Data. *Earth Surface Dynamics*, 9(1), 89–103.

Kuschnerus, M., Lindenbergh, R., Vos, S., Hanssen, R., 2024. Statistically Assessing Vertical Change on a Sandy Beach from Permanent Laser Scanning Time Series. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 11, 100055.

O'Dea, A., Brodie, K. L., Hartzell, P., 2019. Continuous Coastal Monitoring with an Automated Terrestrial Lidar Scanner. *Journal of Marine Science and Engineering*, 7(2), 37.

Qin, R., Tian, J., Reinartz, P., 2016. 3D Change Detection – Approaches and Applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, 41–56.

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.

Sarker, I. H., 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.*, 2(3), 160.

Stilla, U., Xu, Y., 2023. Change Detection of Urban Objects Using 3D Point Clouds: A Review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 228–255.

van der Maaten, L., Hinton, G., 2008. Visualizing Data Using T-SNE. *Journal of Machine Learning Research*, 9(86), 2579–2605.

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. *Sci Data*, 9(1), 191.

Wang, S., Cao, J., Yu, P. S., 2022. Deep Learning for Spatio-Temporal Data Mining: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 34(8), 3681–3700.

Williams, J. G., Rosser, N. J., Hardy, R. J., Brain, M. J., Afana, A. A., 2018. Optimising 4-D Surface Change Detection: An Approach for Capturing Rockfall Magnitude-Frequency. *Earth Surface Dynamics*, 6, 101–119.

Winiwarter, L., Anders, K., Czerwonka-Schröder, D., Höfle, B., 2023. Full Four-Dimensional Change Analysis of Topographic Point Cloud Time Series Using Kalman Filtering. *Earth Surface Dynamics*, 11(4), 593–613.

Woodcock, C. E., Loveland, T. R., Herold, M., Bauer, M. E., 2020. Transitioning from Change Detection to Monitoring with Remote Sensing: A Paradigm Shift. *Remote Sensing of Environment*, 238, 111558.

Xie, J., Girshick, R., Farhadi, A., 2016. Unsupervised Deep Embedding for Clustering Analysis. *Proceedings of The 33rd International Conference on Machine Learning*, PMLR, 478–487.

Zahs, V., Höfle, B., Federer, M., Weiser, H., Tabernig, R., Anders, K., 2024. Automatic Classification of Surface Activity Types from Geographic 4D Monitoring Combining Virtual Laser Scanning, Change Analysis and Machine Learning. *(No. EGU24-1640). Copernicus Meetings.*