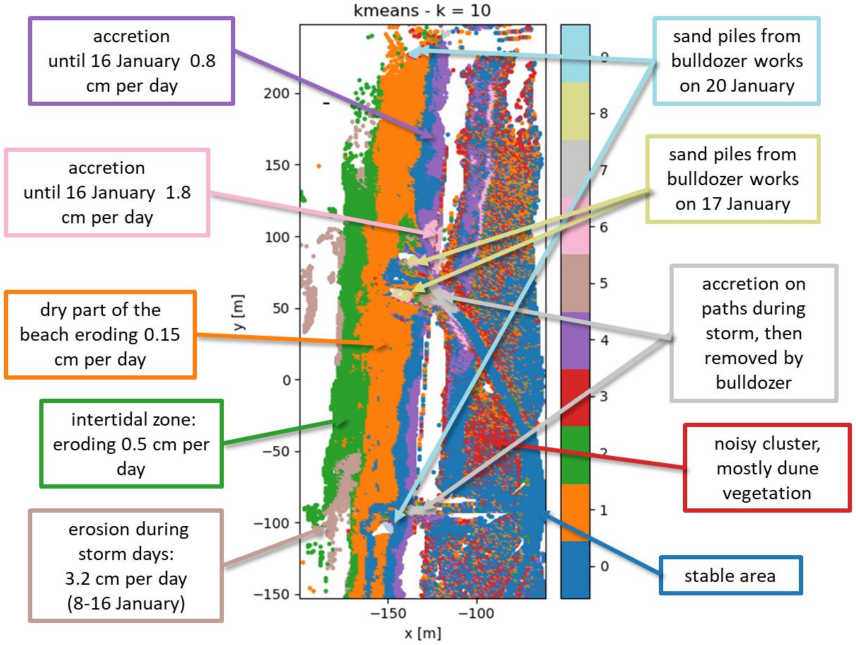
K-Means Clustering for time series from laser scanning data



accretion until 16 January 0.8 cm per day

accretion until 16 January 1.8 cm per day

dry part of the beach eroding 0.15 cm per day

intertidal zone: eroding 0.5 cm per day

erosion during storm days: 3.2 cm per day (8-16 January)

sand piles from bulldozer works on 20 January

sand piles from bulldozer works on 17 January

accretion on paths during storm, then removed by bulldozer

noisy cluster, mostly dune vegetation

stable area

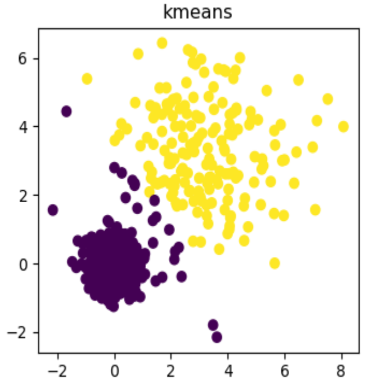
Based on M. Kuschnerus, R. Lindenbergh and S. Vos: **Coastal change patterns from time series clustering of permanent laser scan data**. Earth Surface Dynamics, 9, 89–103, 2021.

Clustering

Goal: Group data with similar properties together

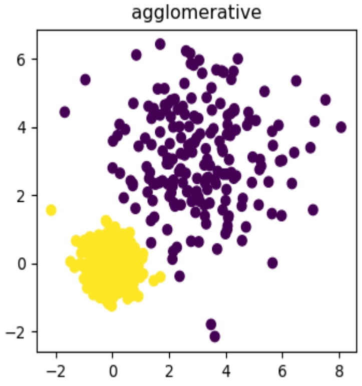
- There are many different methods for example:
  - K-means
  - Agglomerative Clustering
  - DBSCAN
  - Nearest Neighbours
  - Gaussian Mixture
  - …

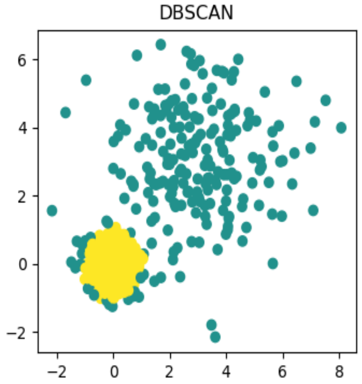
Check the scikit-learn package for more info: https://scikit-learn.org/stable/modules/clustering.html



Rough idea:

- Derive labels of different categories for each data point without providing a 'solution'
- Some algorithms use number of clusters/categories as input, others determine it as part of the output





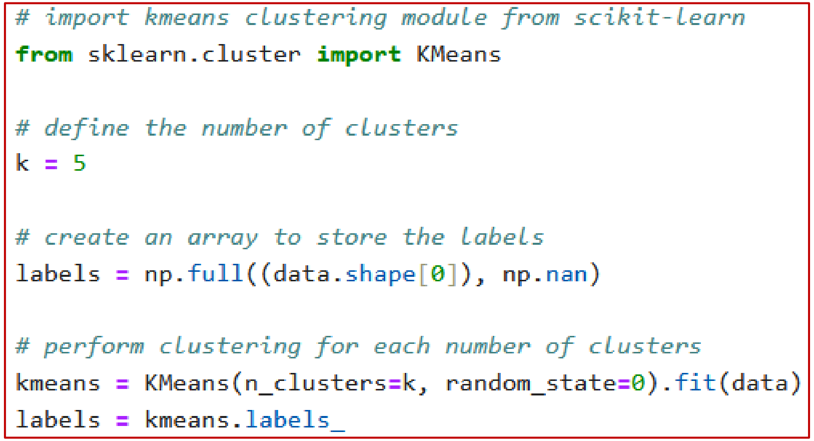
K-Means

Algorithm:

1. Assign each point x_i to the cluster with closest centroid v_i
2. Move centroid to mean of each cluster
3. Calculate sum of distances over all clusters: $\mathrm{Min}_V\, J(V) = \sum_{j=1}^{k} \sum_{x_i \in v_j} ||x_i - v_j||^2,$

Shown convergence to local minimum if Euclidean distance is used. Variations:

- K-medoids (uses actual point as centroid)
- K-means++ (different initialization)
- MiniBatch K-means (speeds up calculation)



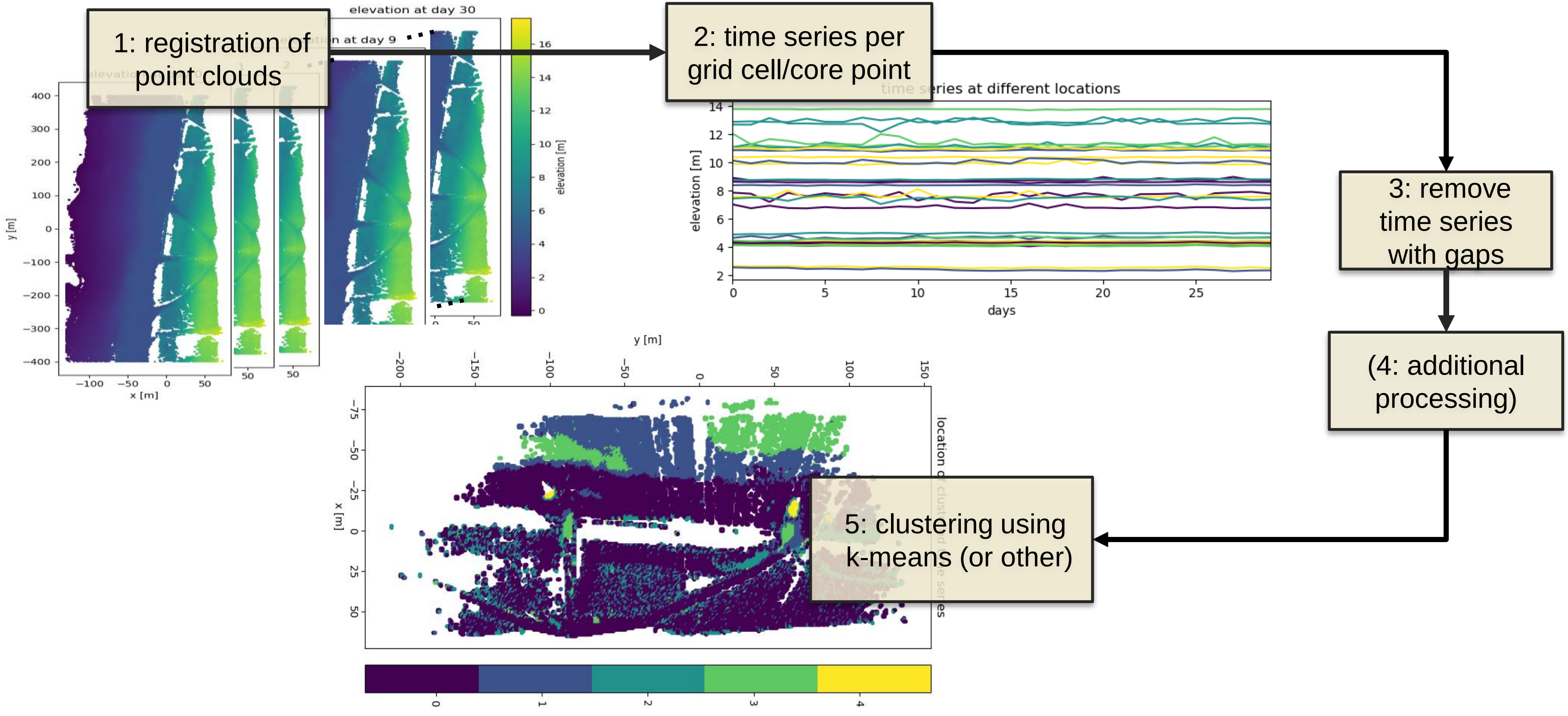
```
# import kmeans clustering module from scikit-learn
from sklearn.cluster import KMeans

# define the number of clusters
k = 5

# create an array to store the labels
labels = np.full((data.shape[0]), np.nan)

# perform clustering for each number of clusters
kmeans = KMeans(n_clusters=k, random_state=0).fit(data)
labels = kmeans.labels_
```

Work-flow



1: registration of point clouds

2: time series per grid cell/core point

3: remove time series with gaps

(4: additional processing)

5: clustering using k-means (or other)

Time Series Clustering

Different ways to cluster time series:

- Using all epochs of the time series as input data
  - Entire time series (elevation)
  - Time series difference values/deviation from reference epoch
  - Normalized time series
  - ...
- Extract features as input to clustering algorithm
  - For example: min, max, mean, std, length, etc.
  - Many other options: PCA, Fourrier, wavelets, …

Define distance measure:

- For different algorithms different distance measures are suitable:
  - Euclidean distance, Correlation, Dynamic time warping (DTW), …

Define evaluation criteria:

- When is a clustering method successful?

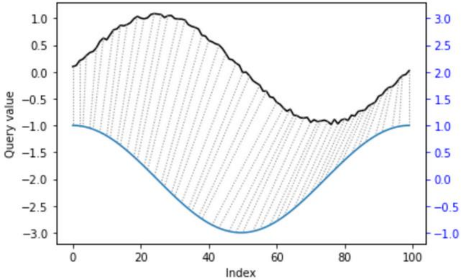
Euclidean distance:

$$d_E(Z_0, Z_1) = ||Z_0 - Z_1|| = \sqrt{\sum_{i=1}^{n} |Z_{0i} - Z_{1i}|^2},$$

Correlation:

$$\mathrm{Cor}(Z_0, Z_1) = 1 - \frac{(Z_0 - \bar{Z}_0) \cdot (Z_1 - \bar{Z}_1)}{||Z_0 - \bar{Z}_0|| \cdot ||Z_1 - \bar{Z}_1||},$$

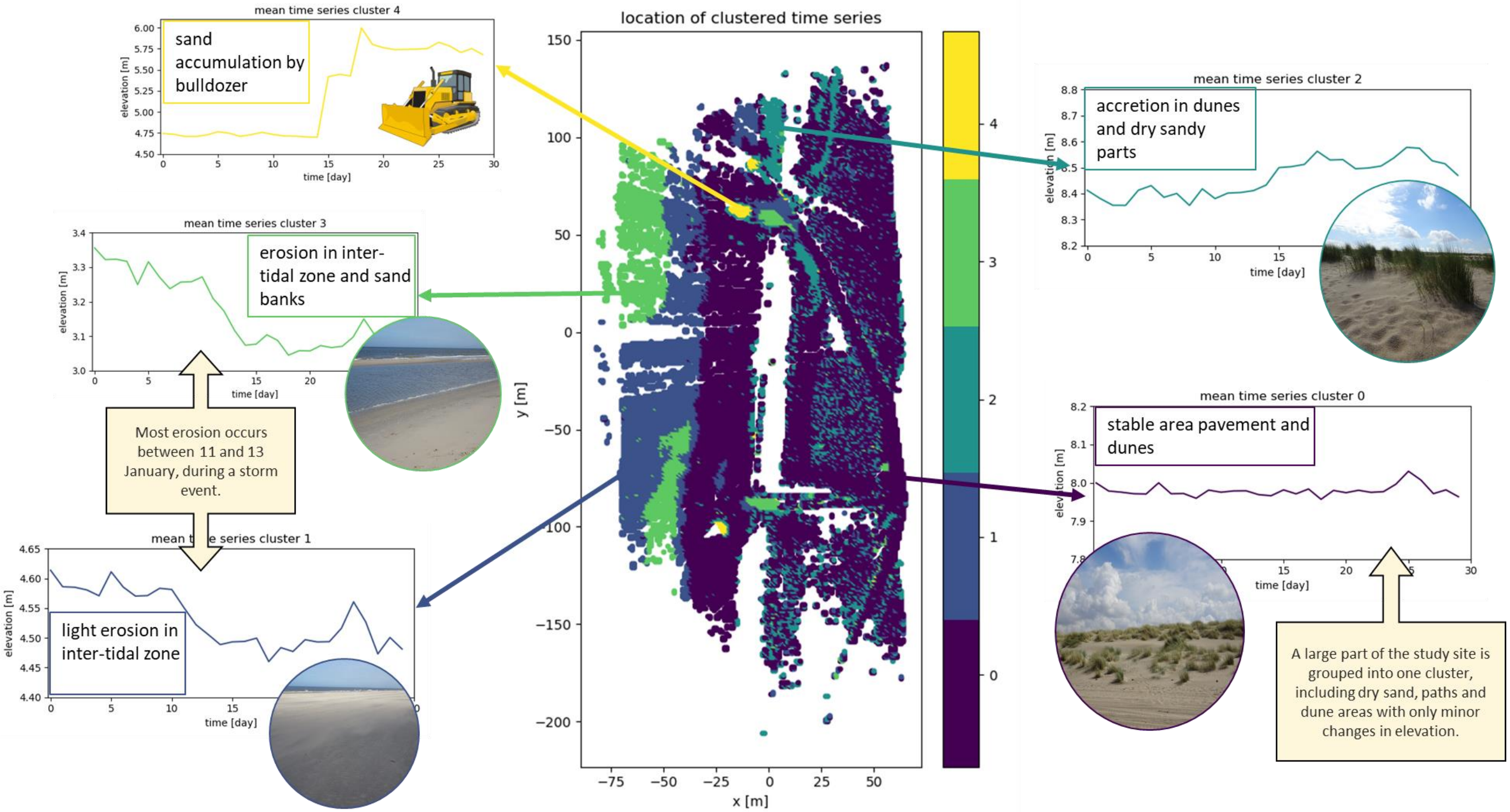
DTW



Berndt and Clifford, 1994

Figure: Wikipedia.org

Example: Clustering time series from January 2017 in Kijkduin (1 epoch per day at low tide)



sand accumulation by bulldozer

accretion in dunes and dry sandy parts

erosion in inter-tidal zone and sand banks

Most erosion occurs between 11 and 13 January, during a storm event.

stable area pavement and dunes

light erosion in inter-tidal zone

A large part of the study site is grouped into one cluster, including dry sand, paths and dune areas with only minor changes in elevation.

References

Mieke Kuschnerus, Roderik Lindenbergh and Sander Vos. 2021. Coastal change patterns from time series clustering of permanent laser scan data. Earth Surface Dynamics, 9, 89–103, https://doi.org/10.5194/esurf-9-89-2021.

Donald J. Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (AAAIWS'94). AAAI Press, 359–370. *(Details on dynamic time warping distance)*

T. Warren Liao. 2005. Clustering of time series data - a survey, Pattern Recognition, 38, 1857–1874, https://doi.org/10.1016/j.patcog.2005.01.025. *(Comparison of different clustering algorithms and distance measures)*

Katharina Anders, Lukas Winiwarter, Hubert Mara, Roderik Lindenbergh, Sander Vos and Bernhard Höfle. 2021. Fully automatic spatiotemporal segmentation of 3D LiDAR time series for the extraction of natural surface changes. IPSRS Journal of Photogrammetry and Remote Sensing, 173, 297-308, https://doi.org/10.1016/j.isprsjprs.2021.01.015. *(Use of LiDAR data to generate time series, use of DTW)*

Lukas Winiwarter, Katharina Anders, Daniel Czerwonka-Schröder and Bernhard Höfle. 2023. Full four-dimensional change analysis of topographic point cloud time series using Kalman filtering. Earth Surface Dynamics, 11, 593-613, https://doi.org/10.5194/esurf-11-593-2023. *(Clustering of time series after feature extraction)*