AUGMENTING THE SENSOR NETWORK AROUND HELGOLAND USING UNSUPERVISED MACHINE LEARNING METHODS

Viktoria Wichert, Holger Brix

Institute of Coastal Research

EGU 2020- Sharing Geosciences Online

Helmholtz-Zentrum Geesthacht

Centre for Materials and Coastal Research

[Slide design and cover photos: HZG]

REMINDER: ABSTRACT

Part 1

A sensor network surrounds the island of Helgoland, supplying marine data centers with autonomous measurements of variables such as temperature, salinity, chlorophyll and oxygen saturation. These measurements yield data collections containing information about the complicated physical and biogeochemical conditions around Helgoland, an island situated at the interface between coastal area and open sea. Spatio-temporal phenomena, such as passing river plumes and pollutant influx through flood events are reflected in this data set. The data provided by the existing measurement network allow for detection and investigation of these events.

Because of its important role in understanding the transition between coastal and sea conditions, plans are made to augment the sensor network around Helgoland with another underwater sensor carrier, an Underwater Node (UWN). The new node is supposed to optimally complement the existing sensor network. Therefore, it should be placed in an area that is not yet represented well by other sensors. The exact spatial and temporal extent of the area of representativity around a sensor is hard to determine, but is assumed to be characterized by similar statistical conditions as measured by the sensor. In the complex system around Helgoland these specifications might change with both, space and time.

ABSTRACT

Part 2

Using an unsupervised machine learning approach, we determine areas of representativity around Helgoland with the goal of finding an ideal placement for a new sensor node. The areas of representativity are identified by clustering a dataset containing time series of the existing sensor network and complementary model data for a period of several years. The computed representativity areas are compared to the existing sensor placements to decide where to deploy the additional UWN to achieve an optimized coverage for further investigations of spatio-temporal phenomena.

A challenge associated with the clustering analysis is to determine whether the spatial areas of representativity remain stable enough over time to base the decision of long-term sensor placement on its results. We compare results across different periods of time and investigate how fast areas of representativity change spatially with time and if there are areas that remain stable over the course of several years. This also allows insights on the occurrence and behavior of spatio-temporal events around Helgoland in the long-term.

Future network design decisions depend on the spatio-temporal stability of the representativity areas as only the stability of observational representativity that is captured by an extended sensor network will allow to assess the variety of spatio-temporal phenomena as well as provide an overview of the long-term behavior of the marine system.

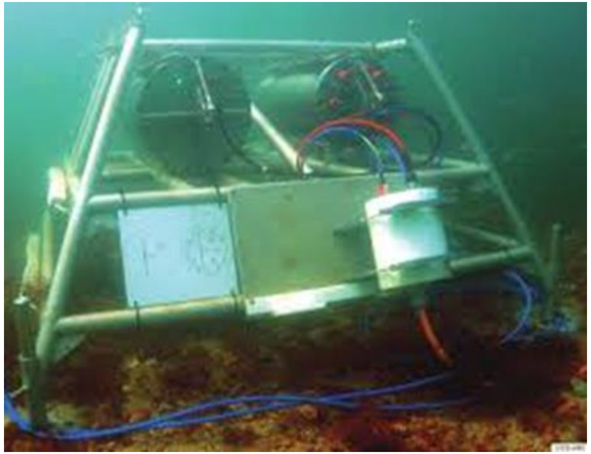
SENSORS AROUND HELGOLAND

Current network and plans

- Current Sensor Network:
  - Autonomous measurement devices & measuring
  - Campaigns
  - UWN Helgoland: underwater node providing infrastructure for sensors (project by HZG & AWI)
  - FerryBoxes
  - Part of the COSYNA project (Coastal Observing Systems for Northern and Arctic Seas, for more information visit https://www.hzg.de/institutes_platforms/cosyna/index.php.en)

- Plans:

Augment the sensor network by deploying another UWN near Helgoland

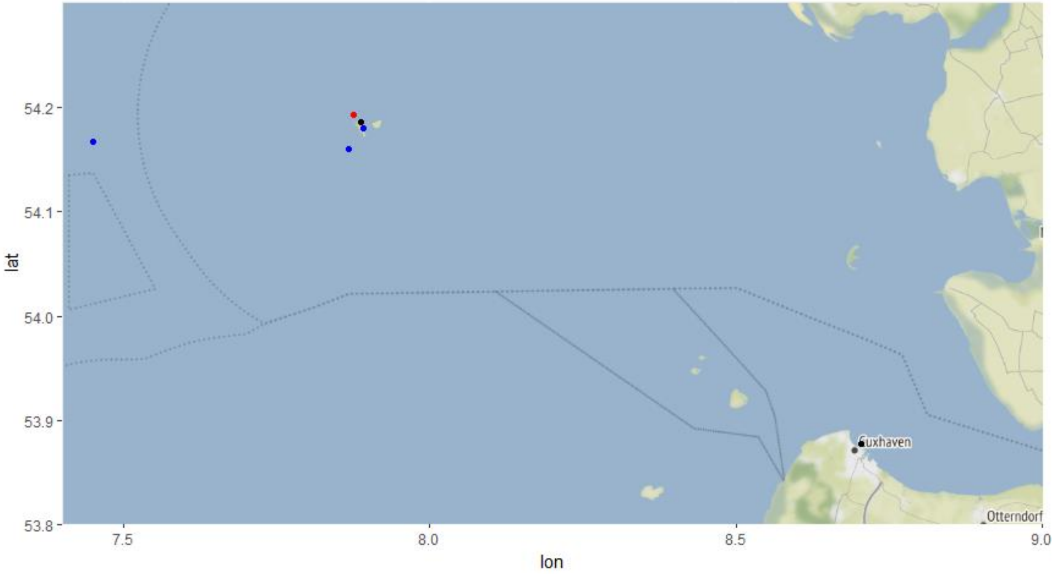


Underwaternode near Helgoland [photo: P. Fischer, AWI]

SENSORS AROUND HELGOLAND

Study region in the German Bight

60
58
56
54
52
lat
-5
0
5
10
lon



[Maps: Stamen Design LLC, OpenStreetMap]

- Map of sensor stations around Helgoland:
  - The existing Underwaternode in red,
  - Stationary FerryBoxes in **black**
  - and MARNET stations in blue.

AUGMENTING THE SENSOR NETWORK

Where to deploy the new UWN?

Helmholtz-Zentrum Geesthacht
Centre for Materials and Coastal Research

- There are plans to augment the existing sensor network by a new Underwaternode (UWN) near Helgoland
- Where should it be placed to yield as much additional information on the complex conditions around Helgoland as possible?
- One approach is to define areas of representativity, where conditions are similar
- To ensure a good coverage of the conditions in the study area, it would be advantageous to deploy the UWN in an area that is not yet represented well in existing measurements

AREAS OF REPRESENTATIVITY

Concept and Applications

- Representativity of Observations
  - Area represented by a measurement
  - Similar conditions
  - Not necessarily temporally stable

- Applications:
  - Optimization of measurement campaigns
  - Data validation
  - Exploit concept to inform measurement strategies

- Related Concepts:
  - SMART monitoring (**s**pecific, **m**easurable, **a**ccepted, **r**elevant and **t**rackable)
  - Decide what data is relevant for a specific use

METHODS

Datasets and algorithm

Helmholtz-Zentrum Geesthacht
Centre for Materials and Coastal Research

- Data:
  - 147 multivariate time series from in-situ measurements and model-data
  - Two time periods: 2013-2015 and 2017-2019 (monthly means)
  - Parameters: Sea surface temperature (SST) and Salinity
  - In-situ data have undergone automated quality control
  - Pre-processing: for each time series and parameter, the data is scaled to an [0,1] interval to minimize the impact of systematic errors

- Algorithm:
  - Partitional clustering
  - Distance measure: Dynamic Time Warping (dtw)

METHODS

Validation: Cluster validity indices

- The clustering approach is an unsupervised machine learning method, therefore no „ground truth“ exists
- The number of clusters **k** generated by the algorithm must be provided by the user
- There are no clues inherent to the problem that point to the „true“ number of areas of representativity in the data
- Cluster Validity Indices assess how good a number of clusters matches the underlying data and can be used to determine the probable number of areas of representativity
- For more information on the indices used, see Arbelaitz et al (2013)
- Depending on the validation method, the index needs to be minimal or maximal to indicate a good match btw. data and number of clusters. This will be indicated by (min/max) next to the method‘s name in the next slide.

VALIDATION

Results

- Validation of cluster number **k** based on 2017-2019 data
- Set of internal cluster validity indices

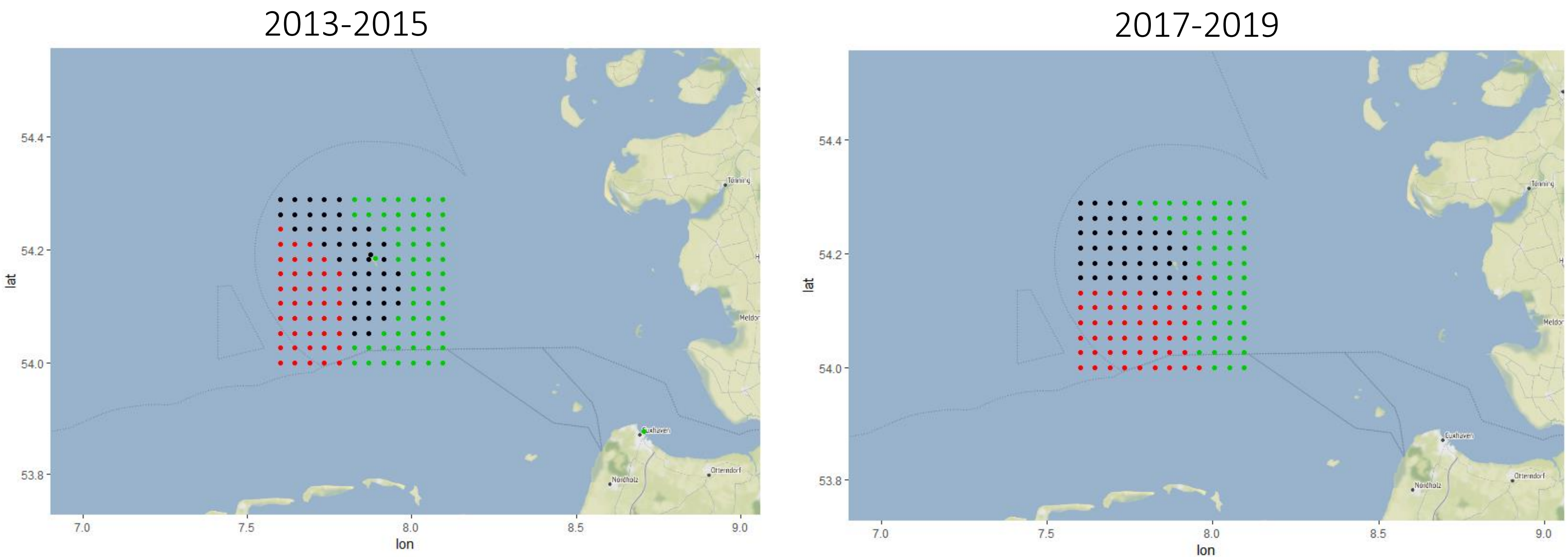
Methods	k=2	k=3	k=4
Sil (max)	2.360624e-01	**3.433842e-01**	3.303423e-01
SF (max)	**5.906701e-05**	1.934212e-06	1.403409e-07
CH (max)	**9.012715e+01**	8.238793e+01	6.989779e+01
DB (min)	1.502164e+00	**9.849641e-01**	1.073846e+00
DB* (min)	1.502164e+00	**1.131937e+00**	1.315200e+00
D (max)	5.456693e-02	**1.182699e-01**	7.128583e-02
COP (min)	5.441190e-01	3.592888e-01	**2.921868e-01**

Result:

Most validity indices favor k=3 clusters,
i.e. three representative regions

PRELIMINARY RESULTS

Areas of representativity and their spatio-temporal stability



2013-2015

2017-2019

[Maps: Stamen Design LLC, OpenStreetMap]

- 3 areas of representativity identified through clustering algorithm
- Not temporally stable though
- Close to Helgoland, more than one area of representativity is present (left panel)

SUMMARY

And open questions

- A partitional clustering algorithm with DTW distance measure was applied to a dataset of multivariate time series from the German Bight
- Three areas of representativity were identified for the study-area around Helgoland
- Representativity patterns change over the investigation period
- Preliminary results indicate that conditions near Helgoland need to be investigated on a smaller scale (temporally and spatially)

Open Questions:

- What are the crucial properties of the areas of representativity?
- Is this approach a valid method to assess sensor-placement?
- Investigate the mismatch btw. spatial data resolution and local topography around Helgoland