Time Series Clustering of Arctic Water Column and Benthic Community Data with Dynamic Time Warping (CC44B-1358)

Spencer Blackwood, *Yale University, Maryland Sea Grant REU Fellow 2023 (spencer.blackwood@yale.edu)*
Jacqueline M. Grebmeier, Lee W. Cooper, and Vyacheslav Lyubchich
Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science

University of Maryland CENTER FOR ENVIRONMENTAL SCIENCE

Yale University

Sea Grant Maryland

DBO DISTRIBUTED BIOLOGICAL OBSERVATORY

NSF

Introduction

Figure 1: DBO regions 1-4 and the stations of study.

- As the Arctic warms 2-3 times faster than the global average, benthic biomass is shifting poleward, suggesting both spatial and temporal dimensions of changing conditions (Wassmann et al. 2011).
- Biomass and abundance data — as well as other chemical, biological, and physical data — are available from time-series observations at "hotspots" of biological activity identified within the Distributed Biological Observatory (DBO; Fig. 1; Grebmeier et al. 2018).
- However, traditional Euclidean distance measures are insufficient to detect similar dynamics which occur with a time lag across locations (Maharaj et al. 2019).
- Dynamic time warping (DTW) finds the optimal alignment between two time series across time points, accounting for similarities among processes which vary in starting time or duration (Maharaj et al. 2019).

Through a cluster analysis of these time series, we aim to:
1) identify redundant information recorded at these research stations, and
2) develop new understanding of how Arctic benthic communities are responding to rapidly changing conditions.

Methods

Data: Of the 42 stations sampled in July between 2010 and 2019, we selected 21 stations with at least five years of data during the study period for analysis in each of the following variables:
- bottom water temperature
- bottom water salinity
- surface sediment chlorophyll content
- benthic biomass
- percent of silt and clay in surface sediment.

Missing data were imputed using maximum likelihood with Kalman smoothing.

DTW: For each pair of (demeaned) time series, time indices were realigned to the nearest value within two time steps (years) of the original (Fig. 2). We then calculated Euclidean distances for each pair.

Clustering: Using the DTW distances, we conducted a cluster analysis of the stations. For each of the five variables of study, we compared the following globular clustering algorithms to identify the optimal method:
- k-medoids
- single-linkage
- complete-linkage
- and Ward's method (Hastie et al. 2009).

For each variable, we then produced two clusterings using the selected globular clustering algorithm and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm.

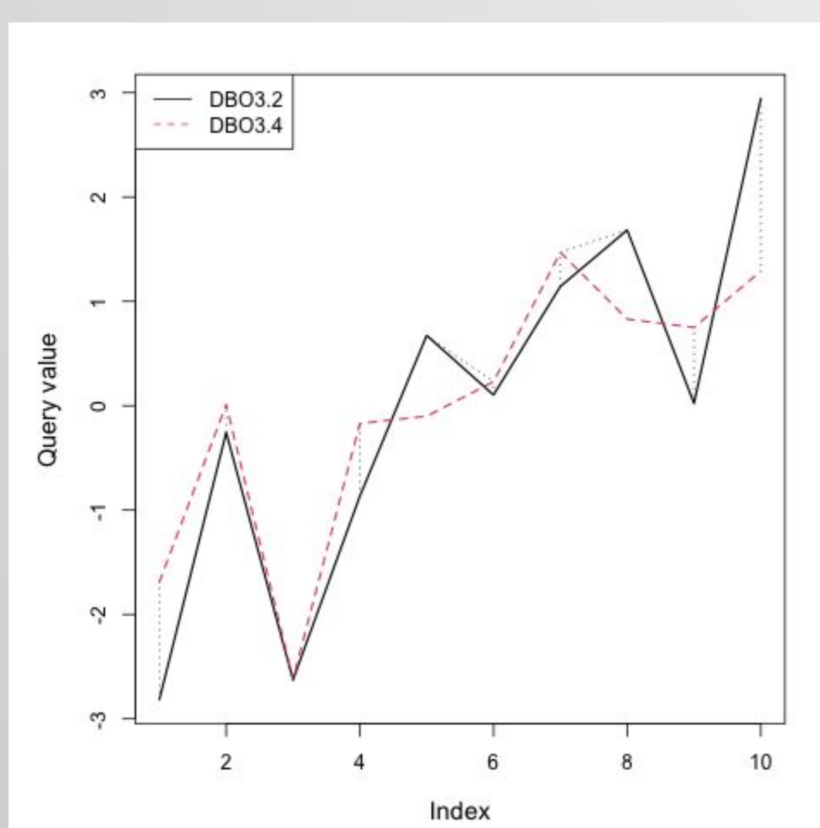

Figure 2: DTW plot of demeaned bottom water temperature time series from stations DBO3.2 and DBO3.4 where the gray dotted lines indicate the DTW alignment of the time indices.

Results

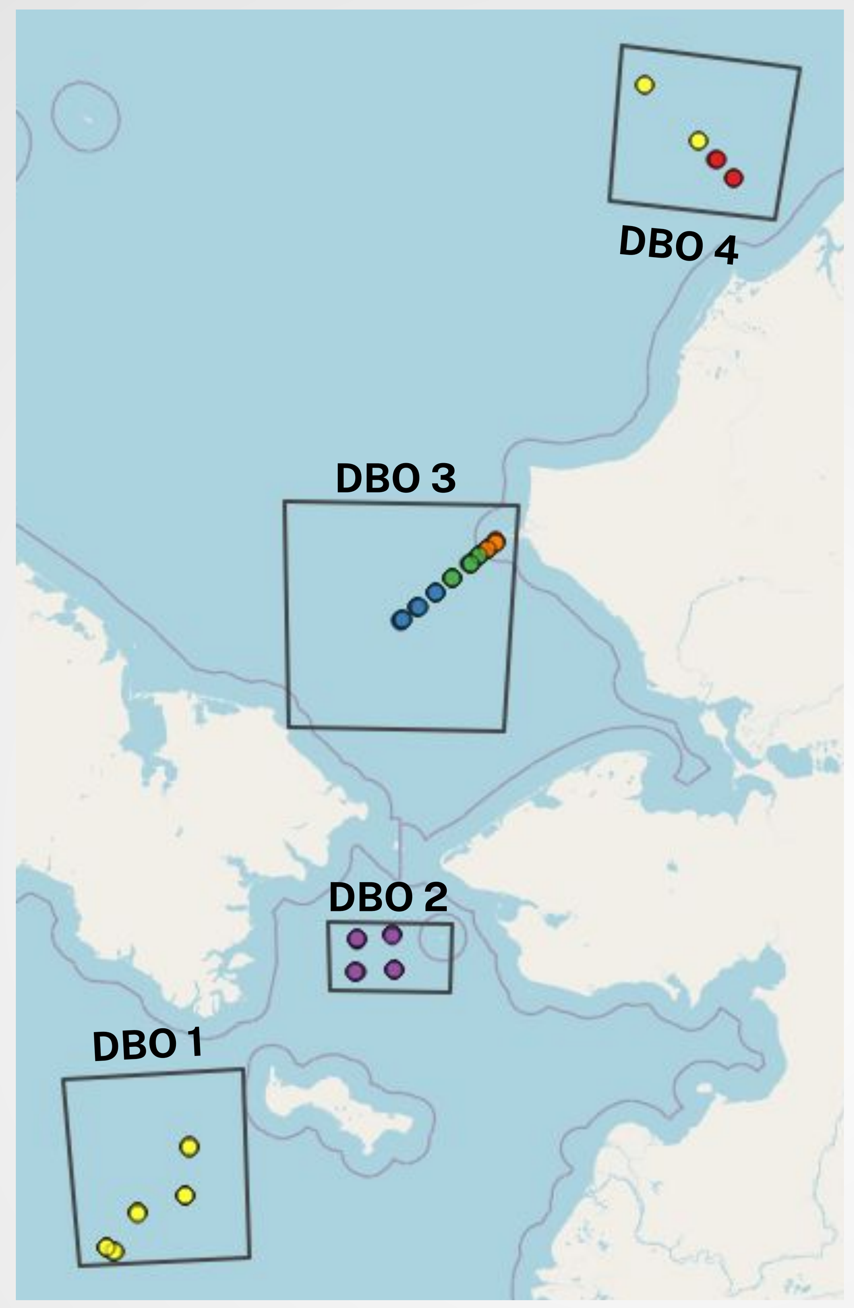

Figure 3: Map of stations clustered by their bottom water temperature using the complete-linkage algorithm. Each color corresponds to one of six clusters. The maximum station depth (in meters) in each region is as follows: 86 (DBO 1), 49 (DBO 2), 60 (DBO 3), 49 (DBO 4).

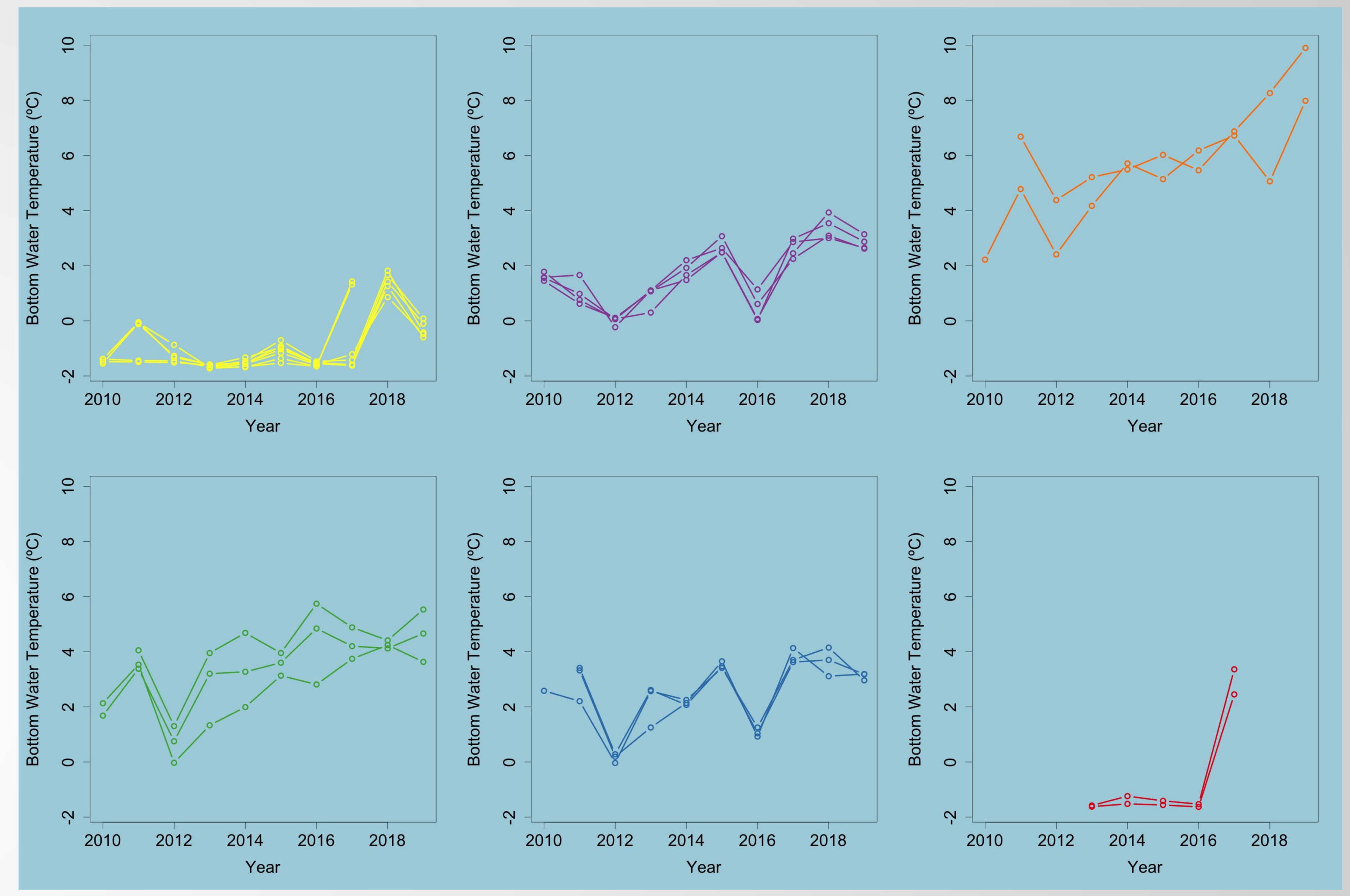

Figure 4: Bottom water temperature time series for each station of study. Each plot corresponds to a cluster identified by color in Fig. 3.

For the three sedimentary variables (surface sediment chlorophyll content, biomass, and percent of silt and clay in surface sediment), all but three of the stations were grouped into one large cluster with the remaining stations classified as "noise."

When grouped by bottom water salinity, an additional cluster formed within DBO region 3.

Bottom water temperature yielded the most clusters (see Figs. 3 and 4). Each cluster is specific to one DBO region with the exception of the yellow cluster, which spans regions 1 and 4.

Conclusions

- Water column variables demonstrated clearer spatial patterns than sedimentary variables, likely because sediment moves less quickly across the sea floor.
- In cases where stations were grouped into multiple clusters, they were roughly clustered by region, indicating similar dynamics within each region. There were two notable exceptions:
  - Variability within region 3 (due to the region spanning two water masses with different characteristics; Grebmeier et al. 2018);
  - Similarities between regions 1 and 4 (due to winter sea ice formation and cooling of bottom water that remains at depth in the summer period; Grebmeier et al. 2018).

Further work could explore the extent and significance of the similarities between regions 1 and 4. More detailed, region-specific analyses (e.g., a formal test for trend synchrony) could more definitively assess information redundancy as well as how changing bottom water temperatures impact benthic community composition.

Acknowledgements

We thank Alynne R. Bayard for contributing to data preparation for this project. This work was made possible thanks to the Maryland Sea Grant Research Experiences for Undergraduates program, funded by the National Science Foundation, and the University of Maryland Center for Environmental Science. This analysis is based upon work supported by the National Science Foundation under Grant No. OPP 2232596 (VL, JG, LC) and OPP-1917469 (JG and LC) and NOAA CINAR (Grant #25984.02).

References

Grebmeier, J. M., K. E. Frey, L. W. Cooper, and M. Kędra. 2018. Trends in Benthic Macrofaunal Populations, Seasonal Sea Ice Persistence, and Bottom Water Temperatures in the Bering Strait Region. *Oceanography* 31. Oceanography Society: 136–151. doi: 10.5670/oceanog.2018.224

Hastie, T., R. Tibshirani, and J. Friedman. 2009. Unsupervised Learning. In The Elements of Statistical Learning: Data Mining, Inference, and Prediction, ed. T. Hastie, R. Tibshirani, and J. Friedman, 485–585. Springer Series in Statistics. New York, NY: Springer. doi: 10.1007/978-0-387-84858-7_14

Maharaj, E. A., P. D'Urso, and J. Caiado. 2019. *Time series clustering and classification*. Chapman & Hall/CRC Computer Science and Data Analysis Series. Boca Raton, Florida: CRC Press.

Wassmann, P., C. M. Duarte, S. Agustí, and M. K. Sejr. 2011. Footprints of climate change in the Arctic marine ecosystem. *Global Change Biology* 17: 1235–1249. doi: 10.1111/j.1365-2486.2010.02311.x