Time Series Clustering using Granger Causality to Identify Time Series Applicable to Forecasting Internal Waves in Lake and Marine Environments

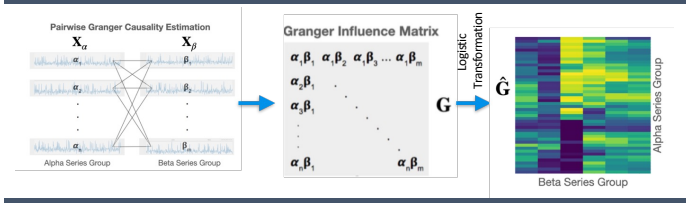
Joshua Sylvester[1], Matthew Lee[1], Mitchell Thornton[1], Eric Larson[1] Sunjoli Aggarwal[1], Matthew J. Hornbach[2] , Michael Manga[3],Bartshe Miller[4]

1) Southern Methodist University, Darwin Deason Institute for Cyber Security, Dallas, Texas, USA 2) Southern Methodist University, Department of Earth Sciences, Dallas, Texas, USA 3) University California at Berkeley, Department of Earth and Planetary Science, Berkeley, California, USA 4) Mono Lake Committee, Lee Vining, California, USA

Abstract

Analysis of water column and surface data collected in Mono Lake California establishes proof of concept that time series clustering based on Granger Causality could be used to identify relationships between atmospheric data and internal waves in lake and marine environments. Time series clustering is typically utilized in application domains where a plethora of time series are available and relationships between the time series are unknown. The method groups time series based on their similarities such that time series clustered together have higher similarities and time series across clusters have less similarity [1]. Most existing clustering methods group time series together based on the correlation between values in each time series. However, for many applications, time series clustering may be more meaningful when time series are grouped based upon how they influence external events, rather than their similarity. Utilizing Granger Causality to carry out time series clustering presents a new method for establishing influence among a group of time series. Granger Causality was originally designed to determine how well one time series forecasts another [2]. Substituting our method for traditional similarity methods in time series clustering provides a new way to identify meaningful relationships in a time series dataset. We have applied Granger-based time series clustering to water column temperature, wave height, and barometric pressure time series data collected from Mono Lake, CA during the spring and summer of 2021. Initial results indicate that Granger-based clustering could be used to identify relationships between barometric pressure and internal waves along thermocline boundaries in lake and potentially ocean environments. This suggests that Granger-based time series clustering might be applied to more complex lake and ocean environments to identify potentially unrecognized relationships between time series data and supplement existing data analysis techniques. Further studies will include water column pressure data, methane gas sensor data, and higher sample rates with the hopes of identifying time series associated with a broader set of environmental factors including internal waves, gas release, and slope stability.

Granger-based Time Series Clustering

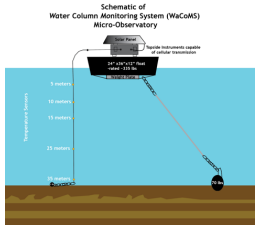


Our Experiment

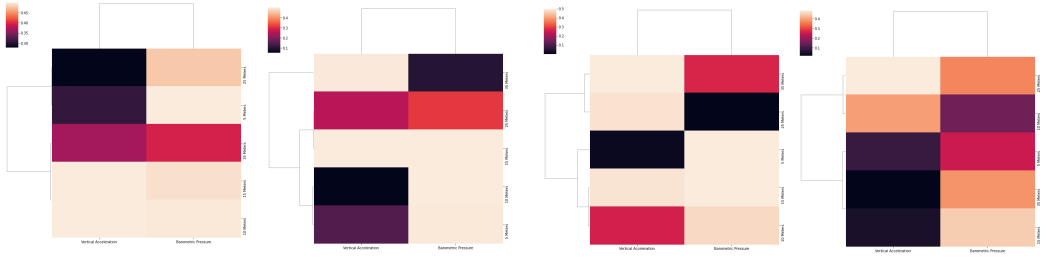
Data analyzed over 4 months in Mono Lake, CA by the probe diagramed to the right

Alpha time series: Barometric pressure and accelerometer data acting as proxy for wave height

Beta time series: Water column temperature data at 5, 10, 15, 25, and 35 meter depths



Results



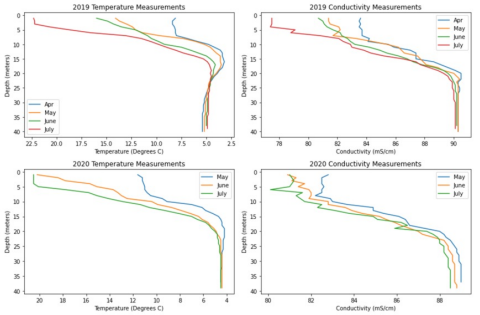
Granger Clustering from April 2021 *Granger Clustering from May 2021* *Granger Clustering from June 2021* *Granger Clustering from July 2021*

High Granger causality between surface barometric pressure and temperature at 15 meters

Inconsistent results across months with accelerometer data

Evaluation

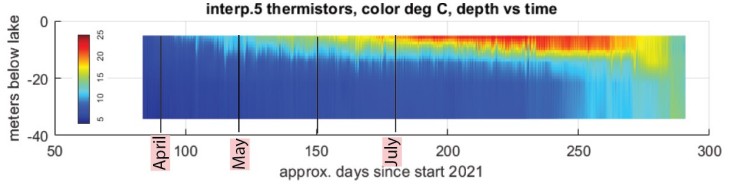
Historical LADWP Water Column Data



Likely thermocline boundary located at ~15 meters

See Magnani et al. V15I-0148 for seismic evidence of thermocline and internal waves in Mono Lake

Mono Lake 2021 Water Column Data
interp.5 thermistors, color deg C, depth vs time



Preliminary Conclusions

- Proof-of-concept established that Granger-Based Time Series Clustering is applicable to complex geophysical problems
- Barometric pressure appears granger-casual with thermocline temperature fluctuations predictor in Mono Lake, CA, suggesting a relationship between barometric pressure and internal waves.
- A relationship between surface accelerometer data and temperature data in the water column is present but is not yet fully understood

Future Work

- Collect additional sensor time series data in the form of water column pressure data and methane gas sensor data
- Analyze a broader set of factors including gas release and slope stability

References

[1] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering – A decade review," *Inf. Syst.*, vol. 53, pp. 16–38, Oct. 2015, doi: 10.1016/j.is.2015.04.007.

[2] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969, doi: 10.2307/1912791.