IS 800 - Causal AI and Machine Learning

Midterm Progress Report

Causal Discovery with Convergent Cross Mapping

Francis Ndikum Nji - RN49536

Seraj Mostafa - HC15573

submitted to

Prof. Dr. Jianwu Wang

Department of Information Systems

UMBC
AN HONORS UNIVERSITY

2022-10-30

Contents

1 Abstract

In geoscience, data are typically generated from different dynamic and thermodynamic atmospheric processes. Complex interactions between measured time series are involved, and nothing or only little is known about the underlying dynamic system. Recent developments in complex systems analysis have led to new techniques for detecting causal relationship using time series. In this project we use Convergent Cross Mapping (CCM) to both investigate nonlinear causal interactions between time series by using nonlinear state space reconstruction. We further investigate the general applicability of CCM and present potentials over Granger causality and TCDF causality. Using examples from simulated and real-world artic sea-ice dataset, we test the ability of CCM to detect causal relationships between variables were our baseline models couldn't detect.

Keywords— convergent cross mapping, Dynamical systems, causality, non-linearity, correlation, time series

2 Problem definition:

The ability to identify truly causal relationships is fundamental to developing impactful interventions in climate science, medicine, policy and business. While controlled randomized experiments remain the gold standard for causal discovery and constitute the primary tool for identifying causal relationships, such experiments are in many cases either unethical, too expensive, or technically impossible(1). The development of causal discovery methods to infer causal relationships from uncontrolled data constitutes an important current research topic. For continuous-valued observed data, linear models are often used not because their true causal relationships are believed to be linear, but because they are well understood and easy to work with. For linear models, non-Gaussianity in the data can actually help in distinguishing the causal directions and allow one to uniquely identify the generating graph under favourable conditions (2). When causal relationships are nonlinear it typically helps break the symmetry between the observed variables and allows the identification of causal directions. Non-invertible functional relationships between the observed variables can provide clues to the generating causal model (3). The problem of inferring causal interactions from data has challenged scientists and philosophers for centuries (4) and it is often seen as an NP-hard problem. In reality, the observations are often naturally ordered in time, and data are usually stored in the form of time series. Among the many definitions of causality, it is generally agreed that the "cause" occurs before the "effect" (5) and the Granger causality still is the most prominent and widely used concept. The Granger causality built on a vector autoregression underlying model is used to measure whether one time series could be used to predict another time series and works on two principles: first, the effect does not precede its cause in time, and second, the causing series contains unique information about the series being caused that is not available otherwise. Although the definition of Granger causality is intuitive and elegant, its applicability is limited when the time series are coupled, e.g., in dynamical systems. The intuition behind Granger causality is that a variable A causes B if we have a model that improves prediction of B if we include A. This framework assumes that we can “separate” the information about A from the rest of the system. However, in many systems that have interacting parts, information about a variable may not necessarily be removable from the rest of the system. They are integrated or in other words, not separable. the assumption of separability does not hold when the time series are coupled (6). In orther words, Granger causality framework does not apply to cases where information about variables are not separable from the rest of the system especially for those whose causalities are weak to moderate. Finally, Granger Causality is dependent on the performance of a linear predictive model that assumes linear combination of variables. This model may not perform well for some systems that need nonlinear models to improve predictions.

Convergent Cross Mapping (CCM) is designed to identify and quantify causalities in systems whose variables are not separable. Beyond this, it is designed to identify and quantify weak to moderate causalities, which Granger Causality may miss. Finally, CCM is independent of a predictive model so we're not at the mercy of model performance. This study observes an unsupervised learning approach were we apply CCM on simulated ground truth data with known causal effects, and artic sea ice data with unknown causal effects with our goal to discover bi-directional causal relations between variables that other approaches like the Granger Causality and Temporal Causal Discovery Framework(TCDF) may fail. We further compare the causal relationship strengths between variables discovered by our CCM model against the baseline models GC and TCDF. Unlike GC, CCM works on the principle of Convergence and Cross Mapping and states that If A causes B, then we say some information about A gets stored in B. If this is the case, then we can predict the values of X given values from Y, and this contradicts GC. The results of our project will be a table of values representing correlation coefficient and P-value of each connection between two variables. We run the models to discover causal relationship, e.g X -> Y and X <- Y. In cases were there is no reverse causal relationship, we apply "strong forcing"., a technique applied to CCM when X has very strong effects on Y yet Y has 0 effect on X. Finally, we plan to use the “accuracy” (MAE, MSE, or correlation) of this prediction as our metric for causality.

3 Related Work

CCM is mostly used for time series analysis however it is hard to find contribution in CCM improvement where there are quite a few work related to experiment and application. Clark et. al. came up with a novel multispatial CCM by combining the CMM and dewdrop regression to detect causal relationships in fewer steps (7). The idea is to work with short time series and and discovering the relationships in a shorter time which is truely a great work. Some experimental work such as, Ye et. al. checked on the ability of CCM for variable time delayed interactions using undirectional and bidirectional causality (8). The main idea is to resolve the transitive issues in the causal chain. This work is an experiment using Veilleux's Paramecium-Didinium Experiment to see whether the extented CCM works with resolving transitive problems. In another work, Wang et. al. applied CCM to explore the causality between soil moisture and precipitation over low- and mid- latitude regions (9).

Comparing other works, we are looking at a little bit different direction followed by the similar approach of experimenting. The dataset we are using is related to climate change where we want to see whether the CCM is able to detect the feedbacks in variables in any way. If we able to find any feedback from even using negatively correlated variables we would like to move forward with discovering the same relationships using other methods such as Granger Causality.

4 Approach

Our objective is to show that CCM can identify causal relationships that other causal discovery approaches like GC and TCDF may not identify. We split our process into four phases.

4.1 Data selection

In this project we use both simulated as well as real world dataset. Simulated spatio-temporal data is generated using ‘Python’ programming language. Our real world data set consist of 15,584 spatio temporal records of data collected over 41 years precisely from Jan-1979 to Jun-2021 spanning across the artic. Furthermore embedded in the data are mean values of both Sea-ice concentration values from Nimbus-7 SSMR and DMSP SSM/I-SSMIS passive and

Date	wind_10m	specific_humidity	LW_down	SW_down	rainfall	snowfall	sst	t2m	surface_pressure	sea_ice_extent
1979-01	5.531398	0.811961	186.687054	3.127880	1.009872	0.892319	273.355237	250.388101	984.633032	15604191
1979-02	5.328020	0.688896	174.794571	18.541594	0.920831	0.781347	273.121885	247.071202	983.980418	16378929
1979-03	5.432511	0.916124	190.741933	67.690429	0.983327	0.855266	273.088099	252.954138	985.140468	16521089
1979-04	4.792836	1.272056	212.937925	156.223673	0.890723	0.705203	273.126062	259.557456	989.314698	15561238
1979-05	4.819028	2.239776	253.690478	230.950833	1.201308	0.688723	273.393551	269.375118	984.483658	14085613

Figure 1: subset of the Sea Ice dataset

Num	Variable Name	Unit	Records	Variable Type
1	wind_10m	m\s	15584	Continuous
2	specific_humidity	kg\kg	15584	Continuous
3	LW_down	W/m2	15584	Continuous
4	SW_down	W/m2	15584	Continuous
5	rainfall	mm/day	15584	Continuous
6	snowfall	mm/day	15584	Continuous
7	sst	K	15584	Continuous
8	t2m	K	15584	Continuous
9	surface_pressure	Pa	15584	Continuous
10	sea_ice_extent	%	15584	Continuous

Figure 2: Existing correlations between variables

4.2 Data preprocessing

We observe the general trend of the data and answer answer questions like; is the trend rising or falling. Out time series consists of four components, three of which we need to get rid of. These components are the Cyclical effect (C_t), seasonal effect (S_t), random variation (R_t) and long-term trend (T_t). We are most interested in the long-term trend so we decompose using the seasonal_decompose library from statsmodels with the additive model parameter Series where Series = T_t + C_t + S_t + R_t. To deseasonalize our time series or measure the seasonal and cyclical effect, we compute the seasonal indexes which is the degree of variation of seasons in relation to global average. Finally to deal with random variables, we apply Exponential Smoothing which smoothes our data and eliminates random movements. Figure 4 is an additive decomposition of all the components present in the time series variable sea-ice extent.

After deseasoning our time series, we observe that it has input values with differing scales, hence the need for standardization. We standardize our dataset by rescaling the distribution of values so that the mean of observed values is 0 and the standard deviation is 1. This process assumes that your observations fit a Gaussian distribution (bell curve) with a well behaved mean and standard deviation. For a data point X, we apply the following: y = (x - mean) / standard_deviation. For simplicity we used the Object StandardScaler() from scikit-learn.

4.3 The application of Granger causality

The Granger causality test is a statistical hypothesis test for determining whether one time series is a factor and offer useful information in forecasting another time series. In particular, whether a variable has an effect on another variable. The cause and effect relationships are found using directed acyclic graph (DAG).

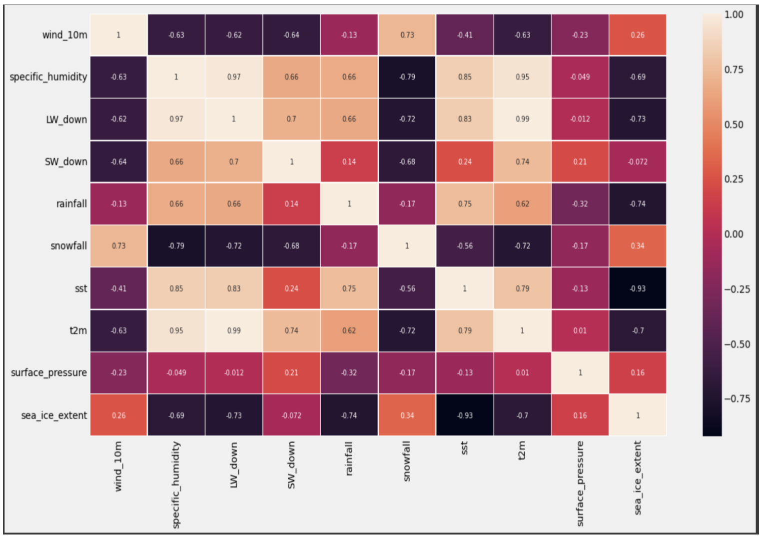

Figure 3: Existing correlations between variables

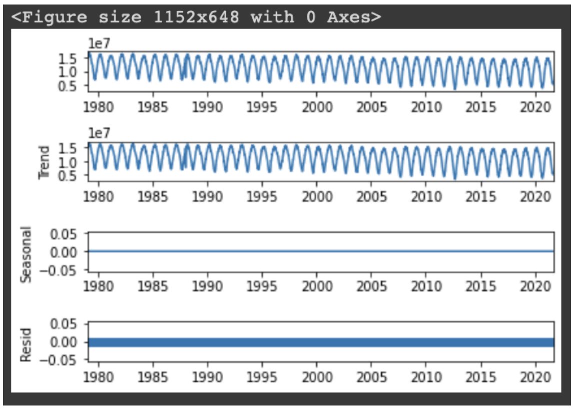

Figure 4: Additive Decomposition of sea-ice extent

4.4 The application of convergent cross mapping

CCM tests for causation between time series X and Y by looking at the correspondence between so-called shadow manifolds M_x and M_y constructed from lagged coordinates (nonlinear state space reconstruction) of the time series values of X and Y (10). Running contrary to intuition (and Granger causality) the basic concept of CCM is that when causation is unilateral (X drives Y), then it is possible to estimate X from Y, but not Y from X. Therefore, CCM is measuring the extent to which the historical record of Y values can estimate the states of X (cross mapping of X by using M_y:X/M_y) or vice versa (cross mapping of Y by using M_x:Y/M_x). Practically, correlation coefficient X and X/M_y or Y and Y/M_x are examined using increasing data length L. The basic algorithm for cross mapping of X by using M_y for each L is:

- generate 'shadow manifold M_y
- find defined number of nearest neighbors at each time point t in M_y
- generate weight matrix by use of nearest neighbors
- estimate $\hat{x}$(t)/M_y by use of these weights
- calculate correlation coefficient ρ (or other error metric) between x(t) and $\hat{x}$(t)/M_y.

Estimation is performed accordingly for cross mapping of Y by using M_x.

In a bi-directional case (X drives Y stronger than vice versa) correlation between X and X/M_y converges faster/reaches a higher plateau than correlation between Y and Y/MX respectively.

4.5 We also planned to apply some techniques as follows for the rest of the project

- The application of TCDF
- The application of Sims causality
- The merging of their corresponding causal graphs
- The "strong forcing" of reverse edges where reverse edges are absent

5 Progress:

So far we have applied the Granger Causality *(discussed in chapter 3.3)* and the Convergence cross mapping *(discussed in chapter 3.8)* to the Sea-Ice dataset. The goal of the Granger Causality is to observe whether one time series (a variable) is affecting another time series (another variable). Among 10 variables we consider 3 variables from which ‘snowfall’ and ‘wind-10m’ is highly correlated and ‘sst’ negatively highly correlated with ‘sea-ice-extent’. The main purpose of this analysis is to see what granger causes what and then from then we can move forward for further analysis. To do so we first check whether the above mentioned variables are stationary, if not we make them stationary by shifting method. Stationarity means that the statistical properties of a time series i.e. mean, variance and covariance do not change over time. Shifting helps us to understand the changes in the shifted variable and then comparing with the target variable. Shifting will not always help but we need our data to be stationary to run the Granger test. KPSS and ADF are two well known tests that are used to check stationarity. KPSS (Kwiatkowski–Phillips–Schmidt–Shin) is used for testing a null hypothesis that an observable time series is stationary around a deterministic trend (i.e. trend-stationary) against the alternative of a unit root. On the other hand, ADF (augmented Dickey–Fuller) tests the null hypothesis that a unit root is present in a time series sample. So, we apply these two methods to test whether they finally become stationary, if not we redo the shifting with higher values. For example, first we try with one shift to check the stationarity, if still we see that there is no stationarity then increase by 1 and keep checking. We did a maximum of 6 shifting for a pair of variables in this analysis. Following is an example of how it looks before shifting and after shifting. The details are in the granger causality test with sea ice data notebook (11).

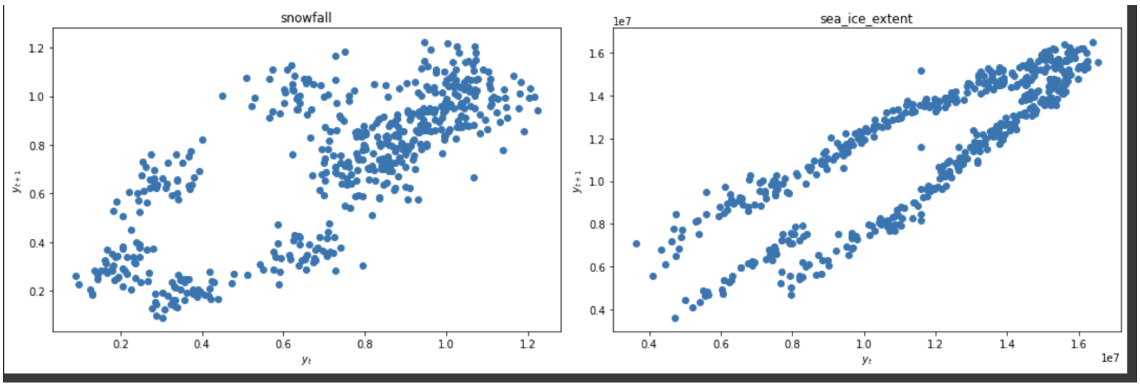

Figure 5: Snowfall and sea-ice-extent with No shifting

The KPSS and ADF test gives us the ‘test statistic’ value against ‘p-value’ and ‘critical values’ where if the test-statistic is lower than the p-value and critical values, then we can say that they are now stationary.

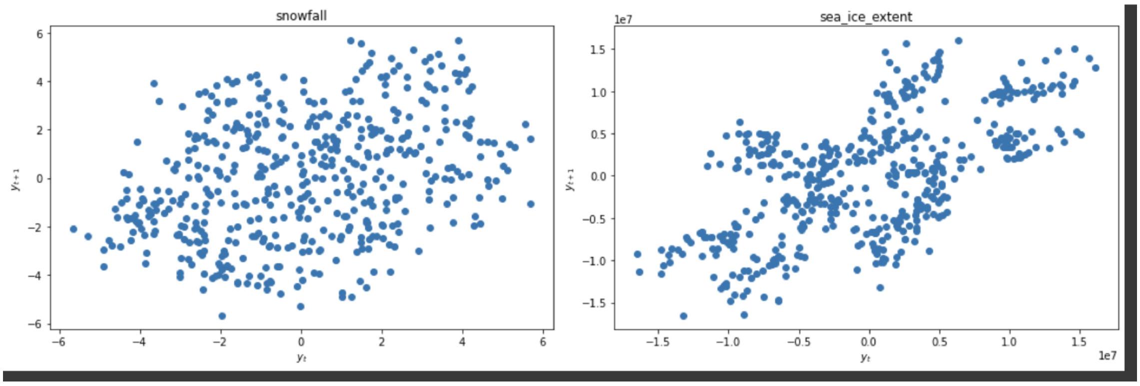

Figure 6: Snowfall and sea-ice-extent with 2 shifting

	snowfall	sea_ice_extent
Test statistic	0.0186	0.0201
p-value	0.1000	0.1000
Critical value - 1%	0.2160	0.2160
Critical value - 2.5%	0.1760	0.1760
Critical value - 5%	0.1460	0.1460
Critical value - 10%	0.1190	0.1190

Figure 7: Test statistic vs p-value and critical values before applying stationarity

	snowfall	sea_ice_extent
Test statistic	-6.9219	-12.1748
p-value	0.0000	0.0000
Critical value - 1%	-3.4438	-3.4435
Critical value - 5%	-2.8675	-2.8674
Critical value - 10%	-2.5699	-2.5699

Figure 8: Test statistic vs p-value and critical values after applying stationarity

Following the processes we then move on to AIC, BIC, HQIC, and FPE tests which are indicated through an 'elbow' shape as the following plots. The purpose is to choose the perfect elbow value to train the model to fit into the Granger causality test.

AIC, BIC, HQIC, and FPE tests. Both AIC (Akaike information criterion) and BIC (Bayesian information criterion) are both Maximum Likelihood estimate driven and penalize free parameters in an effort to combat overfitting. AIC tries to select the model that most adequately describes an unknown, high dimensional reality. This means that reality is never in the set of candidate models that are being considered. BIC tries to find the TRUE model among the set of candidates (12). Hannan–Quinn information criterion (HQIC) is also a likelihood estimator but it uses log likelihood instead of maximum likelihood for model selection (13). Final Prediction Error (FPE) criterion provides a measure of model quality by simulating the situation where the model is tested on a different data set. After computing several different models, you can compare them using this criterion. According to Akaike's theory, the most accurate model has the smallest FPE (14).

We equally applied convergence cross mapping on our real world dataset. To reduce time complexity for our algorithm, we selected a subset of variables from our entire dataset based on the correlation matrix in Figure 3. Sea ice extend is our variable of interest. To avoid a bias selection as well as cover the entire sample space, we selected two highly positively correlated, two highly negatively correlated and one with the least correlation with sea ice extent. From the correlation, we observed that snowfall and wind_10m were highly positively correlated with sea ice extent while SW_down was close to zero and rainfall and sst were highly negatively correlated. We tested for pairwise causal relationships between sea ice extent and all shortlisted variables. We further reduced our dataset and selected just dates bewteen 2015 to 2021 in order to reduce runtime. Below is a summary of the results.

sea ice extent -> snowfall r 0.69 p value 0.0 snowfall -> sea ice extent r 0.34 p value 0.0

sea_ice_extent -> wind_10m r 0.12 p value 0.0 wind_10m -> sea_ice_extent r 0.4 p value 0.0

sea_ice_extent -> SW_down r 0.6 p value 0.0 SW_down -> sea_ice_extent r 0.7 p value 0.0

sea_ice_extent -> sst r 0.98 p value 0.0 sst -> sea_ice_extent r 0.97 p value 0.0

The full results and visualization can be found on (15)

6 Discussions

In this study we considered Granger and CCM based causality tests within the Sea Ice dataset. For the granger causality test we saw that the highly correlated factors, such as, Snowfall Sea-Ice-extent show the granger causes, but negative correlation such as, sst Sea-Ice-extent does not show an effective granger cause. We are convinced that this is due to the natural fact that snowfall causes sea ice extent and sst does not cause sea ice extent. On the other hand, while testing with CCM, we are unable to find yet whether there is an opposite effect such as, sea ice extent cases snowfall. At this point we would like to highlight that, our knowledge on CCM is limited and still we are trying to find a way on how to apply the CCM to determine effective opposite effects for some variables. To be more specific, we would like to force some directed edges and check the results of KPSS and ADF if they produce any convincing evidence. Otherwise we might look for some other scopes to find a CCM based solution that needs further discussion with the course instructor.

7 Plan:

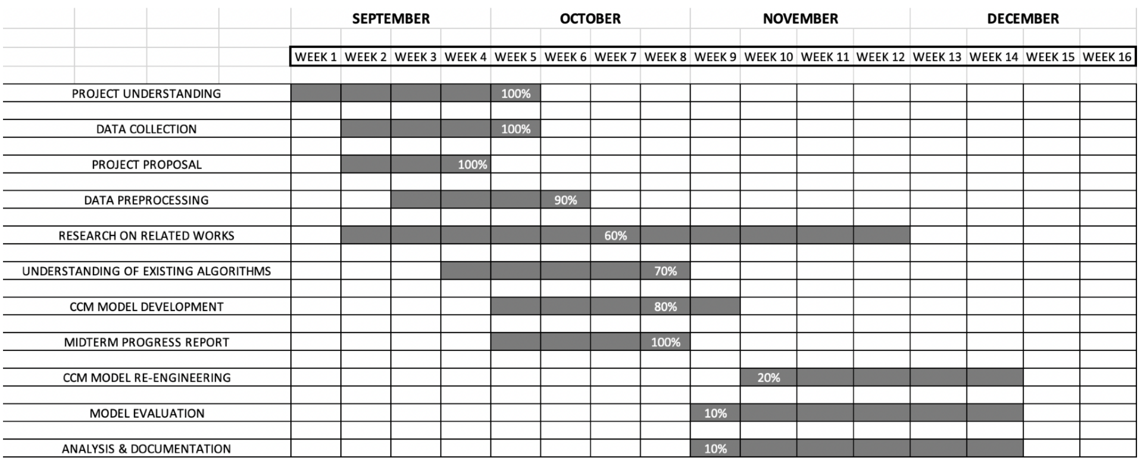

Figure 9: Project Timeline

Figure 9 is a timeline for our project completion. It details the entire life cycle of of our project from project understanding to project delivery. The phases preceding midterm progress report are explained in previous sections above. While developing our CCM model we encountered the following problems and hope to reach ground-breaking solutions after seeking further guidance from our Professor. How can we force a reverse causal relation for our CCM model to prove if there actually exist a reverse causal relation or not? From current literature, forcing a reverse causal relation on 2 variables with very strong forward causal relations will result to the CCM model to identify an existing backward causal relation.

At the moment we have successfully run our CCM model on the arctic sea ice dataset. We have equally run the Granger test on the same dataset and initial results are summarized. We plan to evaluate our model by comparing the results with Granger Causality and Temporal Causal Discovery Framework. Our project follows an agile methodology and we will continue to re-engineer each phase for better performance and more accurate results.

References

[1] P. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf, "Nonlinear causal discovery with additive noise models," *Advances in neural information processing systems*, vol. 21, 2008.

[2] S. Shimizu, P. O. Hoyer, A. Hyvärinen, A. Kerminen, and M. Jordan, "A linear non-gaussian acyclic model for causal discovery." *Journal of Machine Learning Research*, vol. 7, no. 10, 2006.

[3] N. Friedman and I. Nachman, "Gaussian process networks," *arXiv preprint arXiv:1301.3857*, 2013.

[4] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for gaussian variables," *Physical review letters*, vol. 103, no. 23, p. 238701, 2009.

[5] P. W. Cheng and M. J. Buehner, "Causal learning," *The Oxford handbook of thinking and reasoning*, pp. 210–233, 2012.

[6] A. A. Tsonis, E. R. Deyle, H. Ye, and G. Sugihara, "Convergent cross mapping: theory and an example," *Advances in nonlinear geosciences*, pp. 587–600, 2018.

[7] A. T. Clark, H. Ye, F. Isbell, E. R. Deyle, J. Cowles, G. D. Tilman, and G. Sugihara, "Spatial convergent cross mapping to detect causal relationships from short time series," *Ecology*, vol. 96, no. 5, pp. 1174–1181, 2015.

[8] H. Ye, E. R. Deyle, L. J. Gilarranz, and G. Sugihara, "Distinguishing time-delayed causal interactions using convergent cross mapping," *Scientific reports*, vol. 5, no. 1, pp. 1–9, 2015.

[9] Y. Wang, J. Yang, Y. Chen, P. De Maeyer, Z. Li, and W. Duan, "Detecting the causal effect of soil moisture on precipitation using convergent cross mapping," *Scientific reports*, vol. 8, no. 1, pp. 1–8, 2018.

[10] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *science*, vol. 338, no. 6106, pp. 496–500, 2012.

[11] [Online]. Available: https://github.com/sammgithub/Cusal-AI-and-ML-course/blob/main/Mid-report-and-notebooks/GC-test_sea_ice_data-Seraj.ipynb

[12]

[13] [Online]. Available: https://en.wikipedia.org/wiki/Hannan%E2%80%93Quinn_information_criterion

[14] [Online]. Available: https://www.mathworks.com/help/ident/ref/idmodel.fpe.html

[15] Francis. [Online]. Available: https://github.com/Frankie0609/Causal_AI/blob/main/CCM_Team_Project.ipynb