12-10-2022

# On the Use of Machine Learning for Causal Inference in Extreme Weather Events

Yuzhe Wang
*Purdue University*, wang4380@purdue.edu

# Causal Inference Report

Yuzhe Wang

December 16, 2022

## 1 Background

Machine learning become useful tools in analyzing data. Causal Inference is the method that can be used in machine learning to determine the causal relationships in data. This technology can be applied in Climate study and predicting relationships among weather events.

## 2 Granger Causality

Granger causality is a statistical test for identifying whether one time series is useful in forecasting the other time series. This section is about Granger Causality's theory, algorithm, application and discussion

### 2.1 Theory and Algorithm

Firstly, here are definitions in Granger Causality:
Causality: $\sigma^2(X|U) < \sigma^2(X|\overline{U-Y})$ This equation means that by using all information without Y, the variance in predicted X is larger than the variance in predicted X by using all information included Y. In other words, by using Y, it is better to predict X. In this situation, Y causes X ($Y_t \Rightarrow X_t$)
Feedback: Feedback in Granger Causality is when X cause Y and Y cause X($Y_t \Leftrightarrow X_t$)
Instantaneous Causality: $\sigma^2(X|\overline{U}, \overline{\overline{Y}}) < \sigma^2(X|\overline{U})$ This equation means that the current X is better predicted by using information that current Y value is included
Causality Lag: If $Y_t \Rightarrow X_t$, casualty lag m is the least value of k that $\sigma^2(X|U - Y(k)) < \sigma^2(X|U - Y(k+1))$

Secondly, here is a simple Granger causality model:
Let $X_t$ and $Y_t$ be two stationary time series: Then in a causal model, for $X_t$, include $Y_t$, for $Y_t$, include $X_t$:

$$X_t = \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon 1$$

$$Y_t = \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \varepsilon 2$$

This equation implies that $Y_t$ cause $X_t$ provided some $b_j$ is not zero.
In terms of the time shift operator U, U$X_t$=$X_{t-1}$, the equation can be written as:

$$X_t = a(U)X_t + b(U)Y_t + \varepsilon 1$$

$$Y_t = c(U)X_t + d(U)Y_t + \varepsilon 2$$

Apply Cramer representation of the series, the equation can be written as:

$$\int_{-\pi}^{\pi} e^{it\omega}[(1 - a(e^{-it\omega}))dZ_x(\omega) - b(e^{-it\omega})dZ_y\omega - dZ_{\varepsilon 1}\omega] \; = 0$$

$$\int_{-\pi}^{\pi} e^{it\omega}[-c(e^{-it\omega})dZ_x(\omega) - (1 - d(e^{-it\omega}))dZ_y\omega - dZ_{\varepsilon 2}\omega] \; = 0$$

Then,

$$\begin{bmatrix} 1-a & -b \\ -c & 1-d \end{bmatrix} \begin{bmatrix} dZ_x \\ dZ_y \end{bmatrix} = \begin{bmatrix} dZ_{\varepsilon 1} \\ dZ_{\varepsilon 2} \end{bmatrix}$$

,

$$\begin{bmatrix} dZ_x \\ dZ_y \end{bmatrix} = \begin{bmatrix} 1-a & -b \\ -c & 1-d \end{bmatrix}^{-1} \begin{bmatrix} dZ_{\varepsilon 1} \\ dZ_{\varepsilon 2} \end{bmatrix}$$

from the above equation, these equations can be retrieved using properties of $dZ_{\varepsilon 1}$ and $dZ_{\varepsilon 2}$ :

$$f_x(\omega) = \frac{1}{2\pi\Delta} + (|1-d|^2\sigma_{\varepsilon 1}^2 + |b|^2\sigma_{\varepsilon 2}^2)$$

$$f_y(\omega) = \frac{1}{2\pi\Delta} + (|c|^2\sigma_{\varepsilon 1}^2 + |1-a|^2\sigma_{\varepsilon 2}^2)$$

,

$$\Delta = |(1-a)(1-d) - bc|^2$$

The cross spectrum has form:

$$C_r(\omega) = \frac{1}{2\pi\Delta} + (|1-d|\bar{c}\sigma_{\varepsilon 1}^2 + |1-\bar{a}|b\sigma_{\varepsilon 2}^2)$$

$$C_r(\omega) = C_1(\omega) + C_2(\omega)$$

$$C_1(\omega) : \frac{\sigma_{\varepsilon 1}^2}{2\pi\Delta} + (|1-d|\bar{c})$$

$$C_2(\omega) : \frac{\sigma_{\varepsilon 2}^2}{2\pi\Delta} + (|1-\bar{a}|b)$$

If Y cannot cause X, $C_2(\omega) = 0$ since b = 0, then $C_r(\omega) = C_1(\omega)$. On the other side, if X cannot cause Y, $C_1(\omega)$ = 0 since c = 0, then $C_r(\omega) = C_2(\omega)$. Therefore, the causality coherence can be defined as:

$$X \Rightarrow Y : C_{x\to y} = \frac{|C_1(\omega)|^2}{f_x\omega f_y\omega}$$

$$Y \Rightarrow X : C_{y\to x} = \frac{|C_2(\omega)|^2}{f_x\omega f_y\omega}$$

These causality coherences can be the measure for the strength of causal inference between X and Y.

## 2.2 Application and Discussion

In the project, Granger Causality is applied to determine the causal relationship between Nino-3.4 index and surface temperature.

Nino 3.4 index refers to the Sea Surface Temperature month running means in the region from the dateline to the South American coast. From the sea surface temperatures dataset from NOAA, the Nino 3.4 index is calculated in Python and as a time series form. The correlation coefficient between the calculated Nino 3.4 index and NOAA's actual Nino 3.4 data is 0.95.

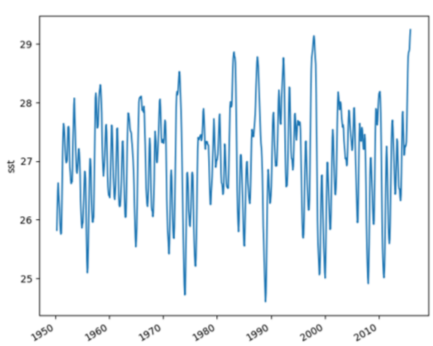



**Figure 1: Nino 3.4 index.**

Based on the surface temperature dataset from NOAA(NOAA/CIRES/DOE 20th Century Reanalysis), surface temperature data in grid points can be retrieved through longitude and latitude, and transformed into time series form. Then, granger causality function take surface temperature and Nino 3.4 index time series as parameters and output the result of causal relationship between them. The program will read the outputs, count the number of grid points that have causal relationship and generate the graph of result. Here, both direction's relationships are tested and the graphs of causal relationship between Nino 3.4 index and Surface Temperature T are shown:

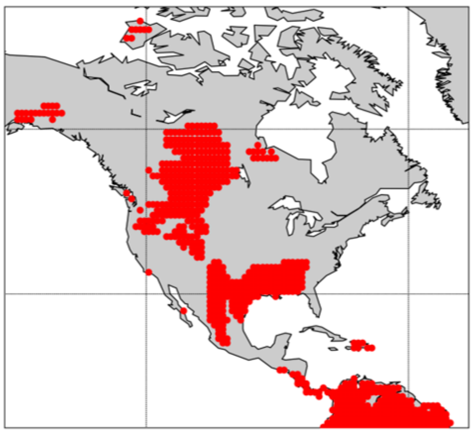


**Figure 2: Nino 3.4 index to T**
Number of red marks: 743

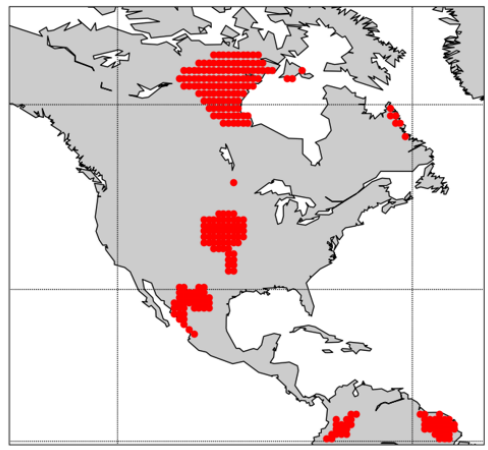


**Figure 3: T to Nino 3.4 index**
Number of red marks: 272

The Figure 2 is to test whether Nino 3.4 index causing surface temperature T in North America($Nino3.4 \Rightarrow T$). The Figure 3 is to test whether temperature T in North America causing Nino 3.4 index surface ($T \Rightarrow Nino3.4$). The grid points marked in red color mean that the causality exists for the causal relationship in that coordinate point. For example, in Figure 2, the red mark means that Nino 3.4 index can granger-cause Surface Temperature T in that coordinate.
As shown above, $Nino3.4 \Rightarrow T$ shows much more red marks than $T \Rightarrow Nino3.4$. This difference points out the strength of causality $Nino3.4 \Rightarrow T$, is stronger than $T \Rightarrow Nino3.4$ and the strength of $T \Rightarrow Nino3.4$ is weak. In other words, the result indicates that Nino 3.4 index is causing Surface Temperature T, but Surface Temperature T is less likely to cause Nino 3.4 index.

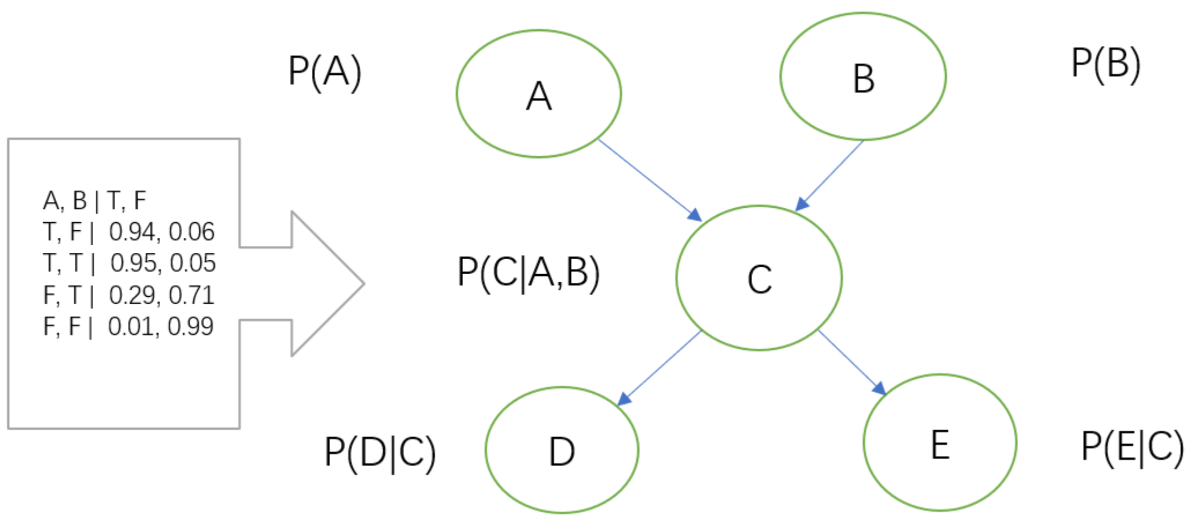
## 3 Causal Inference Methodologies

Besides Granger Causality, there are also other methodologies used in causal inferenece.

### 3.1 probabilistic graphical causal models

Probabilistic graphical model (PGM) is a probabilistic model for which a graph expresses the conditional dependence structure between random variables. Graphical causal model is based on Bayesian networks. Bayesian network can take advantage of conditional and marginal independences among random variables, so it can get the probability of one event happening given another event happens.

Here is an example of probabilistic graphical causal models based on Bayesian Network:



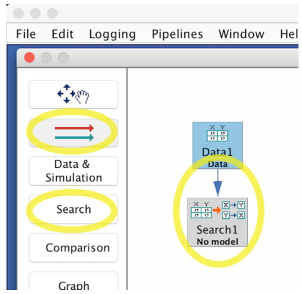

There are two major components in this model. The first one is Directed Acyclic graph, which is also called DAG. It provide a compact visual representation of the interactions between a set of random variables by representing the variables as nodes of a Directed Acyclic Graph (DAG). The direct causal relationships between variables are also shown as arrows/directed edges. The second one is Parameters, which are local conditional probability distributions for variable-parent configuration. In this example, the table of probability distribution of $P(C|A, B)$ is one of parameters.
Based on these parameters and directed causal relation in variables, through Bayesian network and variable elimination, the distribution of unknown conditional probability, such as $P(D|A)$ in the example, can be calculated.
There are two major inference tasks in causal model: The first one is Diagnostic Inference, from effect to cause: $P(A = T|D = T)$ The second one is Causal Inference, from cause to effect $P(D = T|A = T)$

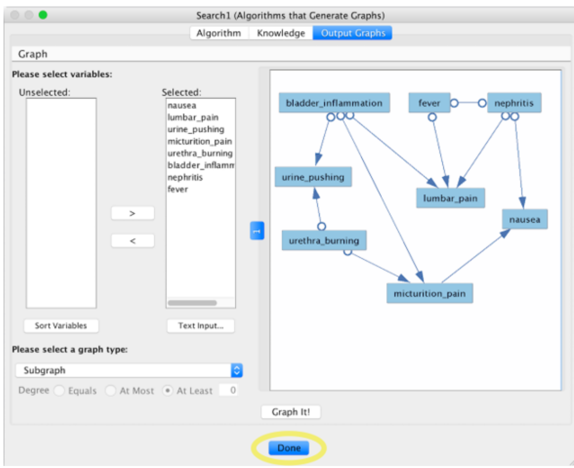
Advantage: It uses Bayesian belief networks and take advantage of conditional and marginal independences among random variables. Therefore, it can get a comprehensive analysis of a causal relationship based on variables' distributions.
Disadvantage: The calculation needs variable elimination, which can lead to high complexity of calculation. The methodology is based on probability distributions, so its application in time series is unknown

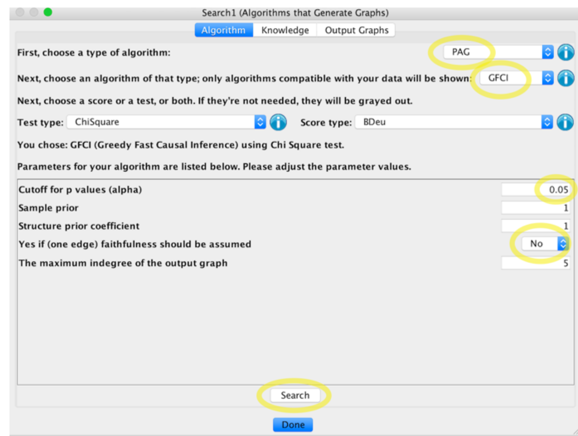
### 3.2 TETRAD - Program

TETRAD is a program that simplifies the process of calculating causal inference. It is a friendly program for users to apply various causal methods on their datasets and getting results just in this program. TETRAD is open-source and it is available on github.
There have been researches that test TETRAD on real and simulated data and see its application on Climate and Earth research. By using UI, TETRAD can read the preprocessed data into TETRAD. It can also perform additional data manipulation and causal model searches. For troubleshooting. TETRAD can provide a first graphical

representation that helps user. It can also export either the graph image or the graph edges as a text file from TETRAD.




TETRAD source code: https://github.com/cmu-phil/tetrad
TETRAD Tutorial: https://bd2kccd.github.io/docs/tetrad/
Advantage: As a program with user-friendly UI, TETRAD is an easy and convenient tool to use.
Disadvantage: TETRAD is not complete version yet, and it is still in development.

### 3.3 Causal Chain

A causal chain describes a process where a variable A causes a variable B, which again causes a variable C. The graphical form of such a causal chain is $A \rightarrow B \rightarrow C$. C is conditionally independent of A, given B. This means the change in C due to a change in A is mediated by B. B is a mediator between A and C.

Example: Fire and Alarm are not independent of each other without giving a mediator. The situation that Fire is correlated with Alarm is only reasonable if we check the underlying mechanism. That is Fire is correlated with Smoke, and Smoke can cause Alarm. In other words, Fire and Alarm are conditionally independent given Smoke.

$$P(Fire) \neq P(Alarm) \tag{1}$$

$$P(Fire|Smoke) = P(Alarm|Smoke) \tag{2}$$

For causality, a causal chain is more like a concept that is often used in probabilistic inference. It helps calculate the target probability from other known probabilities.

Advantage: Causal Chain can be used in probabilistic inference and find mediator between two events
Disadvantage: It is used in probability distribution and it is unknown for its application in data like time series

### 3.4 PCMCI

PCMCI is PC(Peter and Clark algorithm) combined with MCI (Momentary Conditional approach). Therefore, there are two major steps in the algorithm, that are PC and MCI.

Algorithm:
PC-step: it is given a set of univariate time series, which is called "actors". Then, PC algorithm calculates plain correlations between first elements with lag 0 and remaining elements in P at lag $\tau$. If the first element is significantly correlated with three other actors, it will form the set of parents. For each element in parents set, the partial correlations are calculated, like $\rho(x, y|z)$. If the partial correlation is significant at a confidence level $\alpha$, x and y are conditionally dependent given z. Otherwise, x and y are conditionally independent and y is kept in the parent set. Based on the comparisons between partial correlations and confidence level $\alpha$, the parent set will be update. Then, using the updated parent set, repeat the previous partial correlations and update process until the number of elements in parent set is less than the requirement of the process.
Then, the parent set for first element in actor set converges, it begin the same previous process on second element(third element,..., and so on) in actor set.

MCI-step: When the process for all elements in actor sets are done, the selected parent sets will pass in MCI-step. In MCI-step, the partial correlation between actor and the corresponding parent set is calculated. Different from calculation in PC-step, it is conditioning on the parent set of parent set of the current actor in calculation. Then, compare the results with confidence level $\alpha$ and update the parent set. The kept parents will be the final parent set of the actor. So, the causal relationships between variables are determined.

Advantage: PCMCI takes time series as input, so it is helpful in calculating causal inference in time series. It also avoids conditioning on irrelevant variables during calculation.
Disadvantage: Time complexity of calculation can be high

### 3.5 Causal Effect Network

A Causal Effect Network, called CEN, can detect and visualizes the causal relationships among a set of univariate time series of variables. It is useful in the application of PCMCI algorithm.
For CEN based on PCMCI algorithm: After getting causal links detected by PCMCI, CEN visualizes the causal links. Each CEN is composed of circles representing the various actors and of arrows, with the color indicating the strength and the arrow the direction of the detected causal links.

Advantage: CEN can be a useful tool in the application of algorithm, such as PCMCI. It gives the direction and the strength of causal relationships among variables clearly.
Disadvantage: CEN requires the algorithm, such as PCMCI.

## References

Di Capua, G., Kretschmer, M., Donner, R., Hurk, B., Vellore, R., Raghavan, K., and Coumou, D. (2019). Tropical and mid-latitude teleconnections interacting with the indian summer monsoon rainfall: A theory-guided causal effect network approach. *Earth System Dynamics Discussions*, pages 1–27.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods.

Ramsey, J., Zhang, K., Glymour, M., Romero, R. S., Huang, B., Immé, Ebert-Uphoff, Samarasinghe, S. M., Barnes, E. A., and Glymour, C. (2018). Tetrad - a toolbox for causal discovery.

Samarasinghe, S. M., Connolly, C., Barnes, E. A., Ebert-Uphoff, I., and Sun, L. (2021). Strengthened causal connections between the mjo and the north atlantic with climate warming. *Geophysical Research Letters*, 48:e2020GL091168. e2020GL091168 e2020GL091168.
Granger (1969) Samarasinghe et al. (2021) Ramsey et al. (2018) Di Capua et al. (2019)