Process-Oriented Dynamical Models Evaluation for Seasonal Prediction through the Lens of Causal Network

Nachiketa Acharya

Science, Analytics, and Innovation (SAI) Consultants LLC.

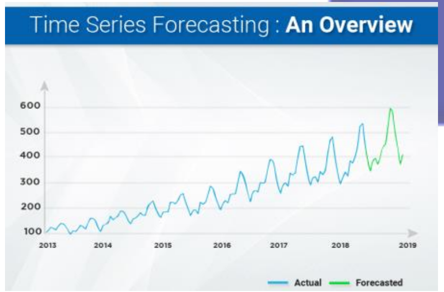
9th International Verification Methods Workshop

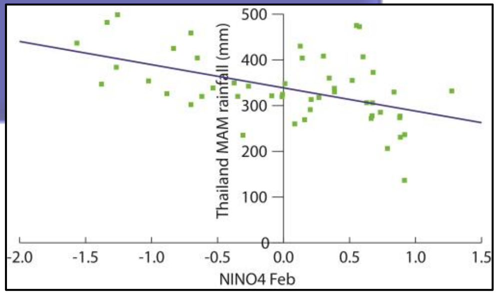
Tools for Climate Forecast

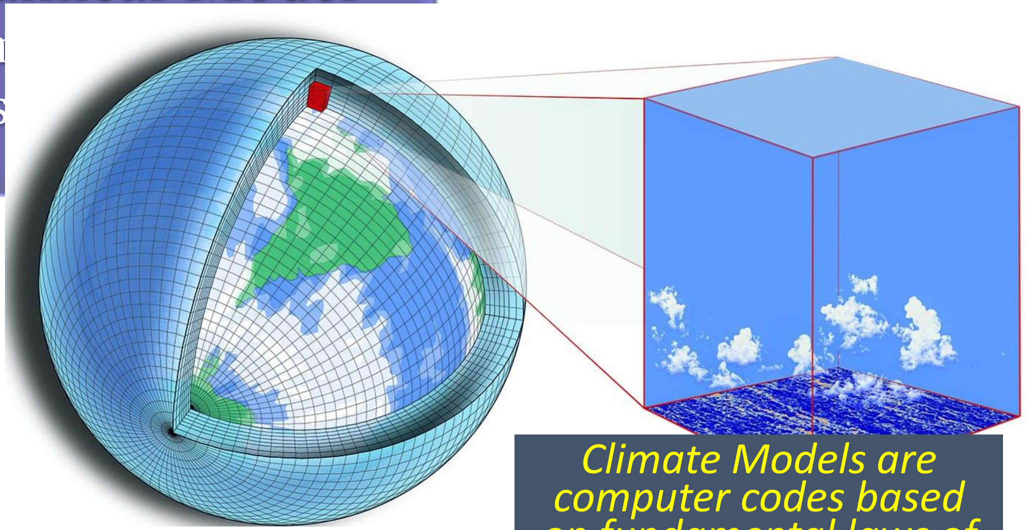
Seasonal Prediction

Statistical Model

Dynamical Model
(Form
phys







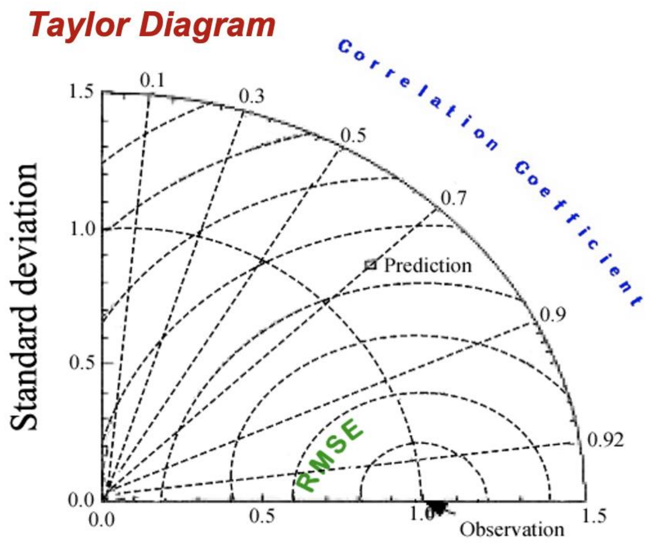
Climate Models are computer codes based on fundamental laws of physics

Limitations

- **The autocorrelation of rainfall is very poor.**
- **The relation between rainfall and predictors go on changing .**

Standard Evaluation process of Dynamical Model

Observed rainfall vs Model forecasted rainfall (mostly descriptive statistics)



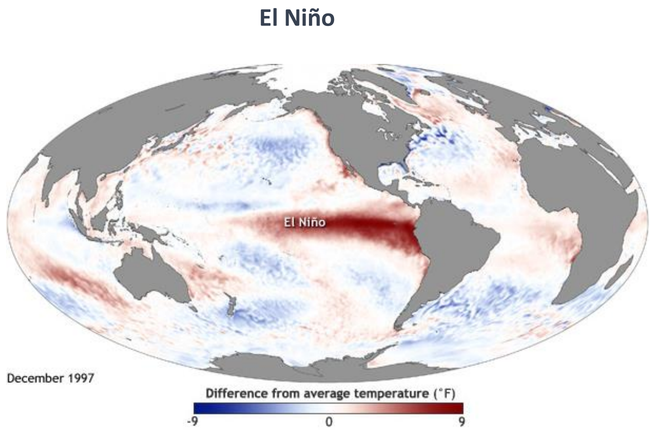
The Taylor diagram communicates three skill scores in one plot due to the cosine properties between :

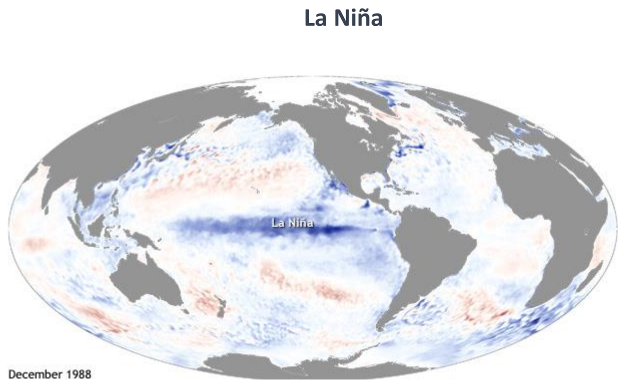
- Correlation.
- Centered RMSE.
- S.D.

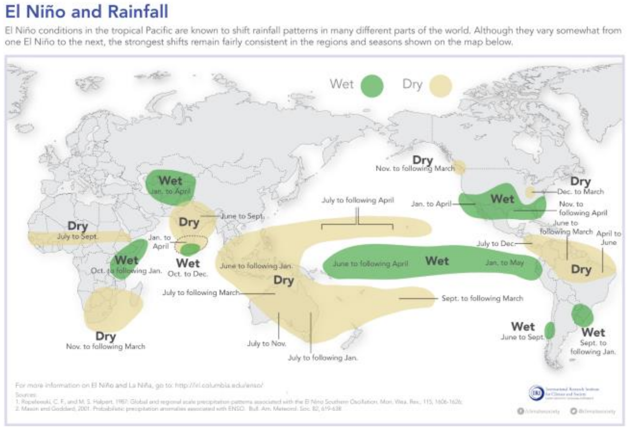
Taylor (2001)

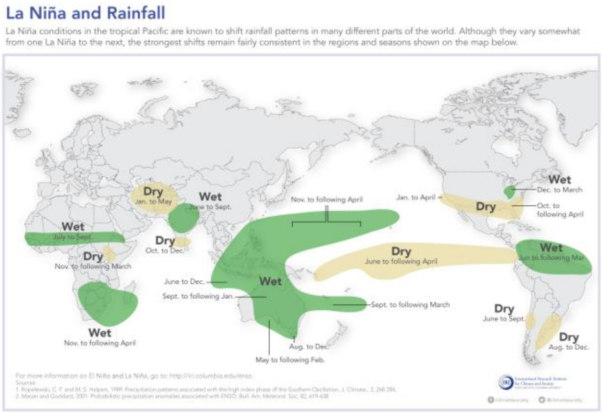
Limitation: Dose not provide “Why” part which will help dynamical modeling community to improve model physics, teleconnection or others factors.

Major Predictability of Seasonal Forecast time scale: ENSO







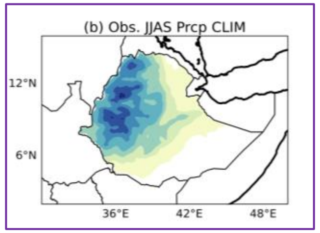


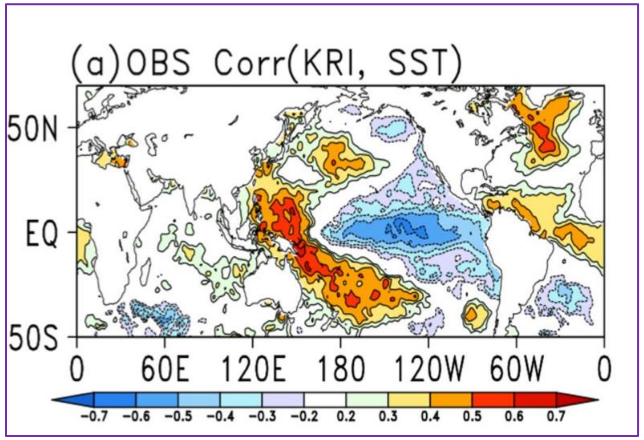
Process-oriented Evaluation based on Teleconnection

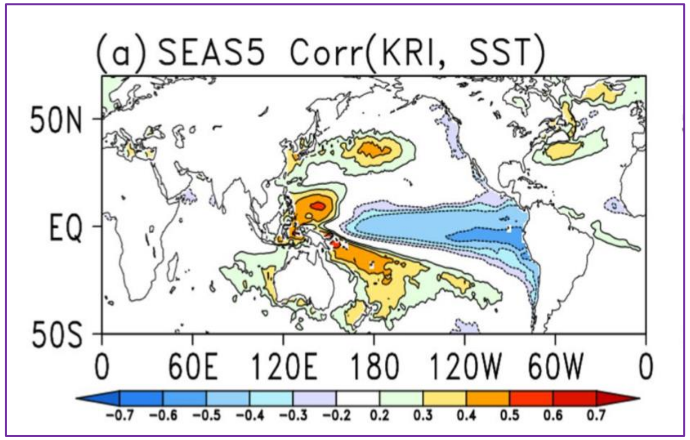
Observed teleconnection (rainfall-ENSO) vs Model forecasted teleconnection (rainfall-ENSO)

Common tool: Pearson’s Correlation Coefficient

Example: Ethiopian “Kiremt” season, Acharya et al.2022, Ehsan et al 2021







Causal Pathways for Teleconnection

BAMS
Article

Quantifying Causal Pathways of Teleconnections

Marlene Kretschmer, Samantha V. Adams, Alberto Arribas, Rachel Prudden, Niall Robinson, Elena Saggioro, and Theodore G. Shepherd

ABSTRACT: Teleconnections are sources of predictability for regional weather and climate, but the relative contributions of different teleconnections to regional anomalies are usually not understood. While physical knowledge about the involved mechanisms is often available, how to quantify a particular causal pathway from data are usually unclear. Here, we argue for adopting a causal inference-based framework in the statistical analysis of teleconnections to overcome this challenge. A causal approach requires explicitly including expert knowledge in the statistical analysis, which allows one to draw quantitative conclusions. We illustrate some of the key concepts of this theory with concrete examples of well-known atmospheric teleconnections. We further discuss the particular challenges and advantages these imply for climate science and argue that a systematic causal approach to statistical inference should become standard practice in the study of teleconnections.

KEYWORDS: Atmospheric circulation; Teleconnections; Statistical techniques; Time series; Interannual variability; Regional effects

https://doi.org/10.1175/BAMS-D-20-0117.1
Corresponding author: Marlene Kretschmer, m.j.a.kretschmer@reading.ac.uk
In final form 27 April 2021

- "…the relative contributions of different teleconnections to regional anomalies are usually not understood. While physical knowledge about the involved mechanisms is often available, how to quantify a particular causal pathway from data are usually unclear".
- However, they used "Partial Correlation", "Multiple Linear Regression" and "Conditional Probability" to quantifying the Causal Pathways using five examples.
- Examples: Common drivers, Mediating pathways, Direct and indirect pathways, Blocking the correct paths in the network, Measuring nonlinear dependencies.

Goal of this study

Causal discovery algorithms go beyond correlation-based measures by systematically excluding common driver effects and indirect links.

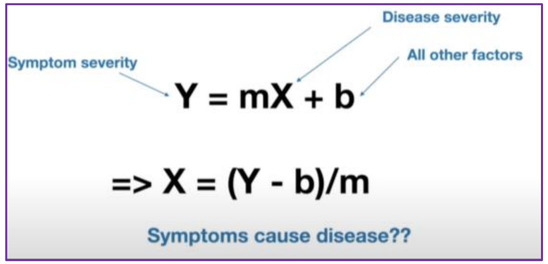
Here, we explore a causal network for model evaluation as a type of process-oriented framework.

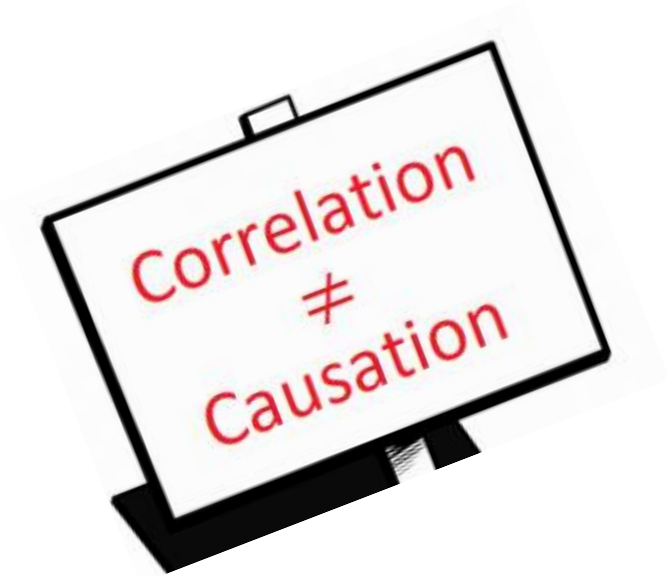
Based on data-driven causal fingerprints, the causal network can understand differences between models and observations based on the physical process which potentially influences model biases in simulating climate variables.

This process based evaluation and informed model development community. to improve the teleconnection within model world.

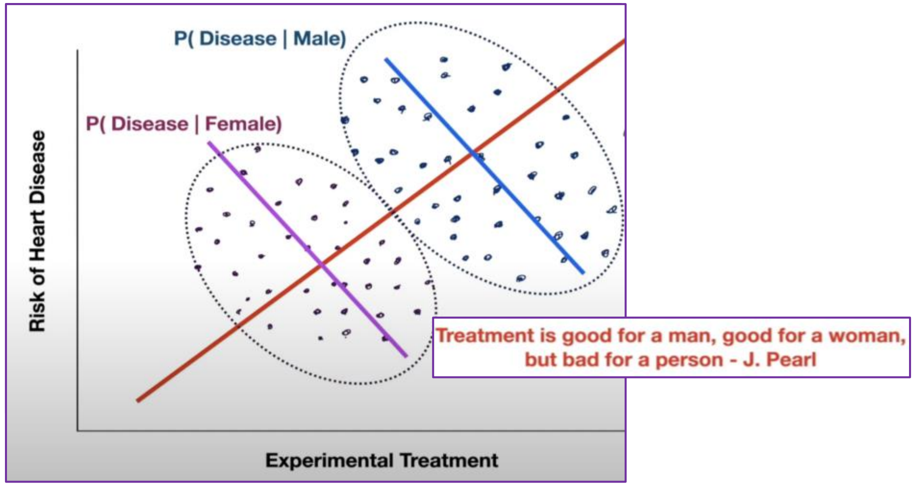
3 Traps of Statistics (3S')

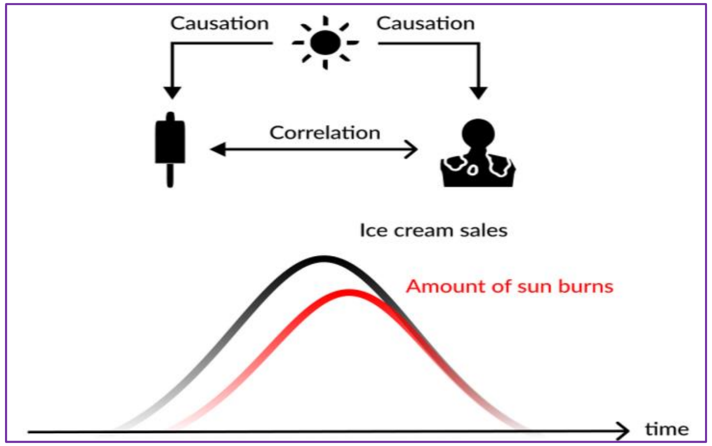
Symmetry



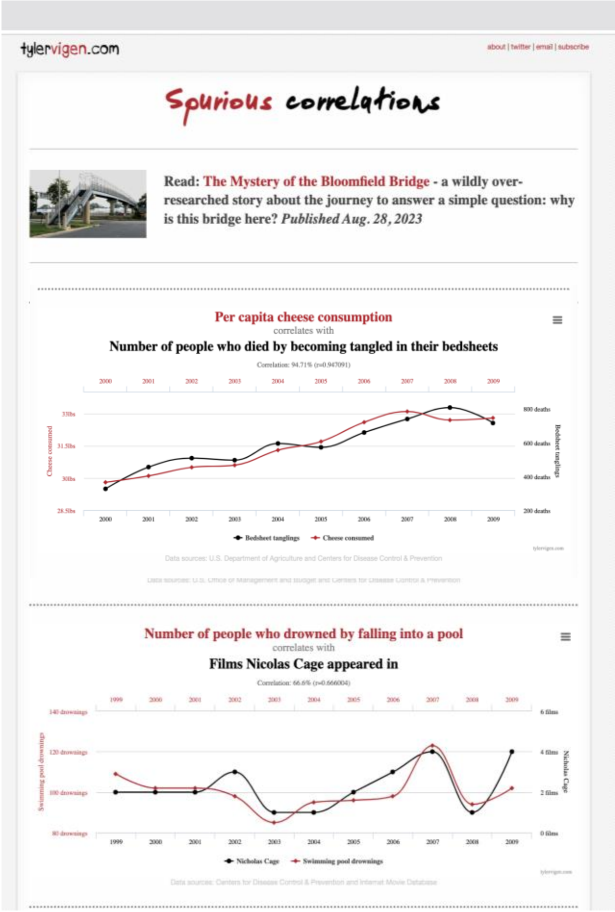


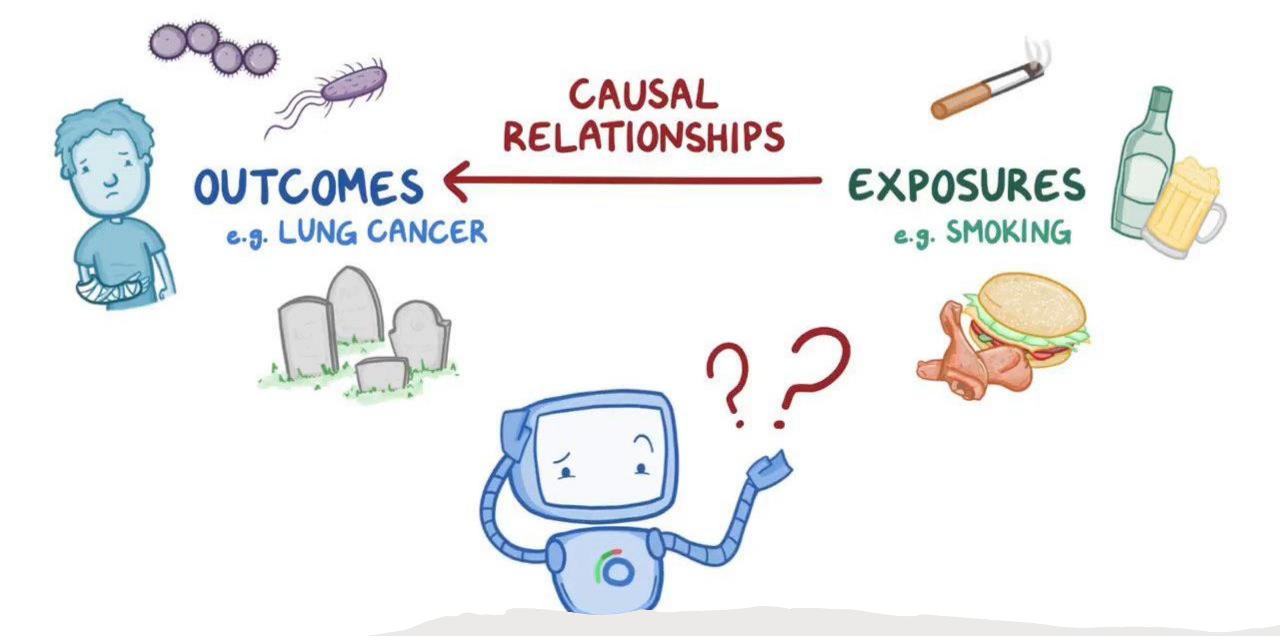
Simpson's Paradox





Spurious Correlation



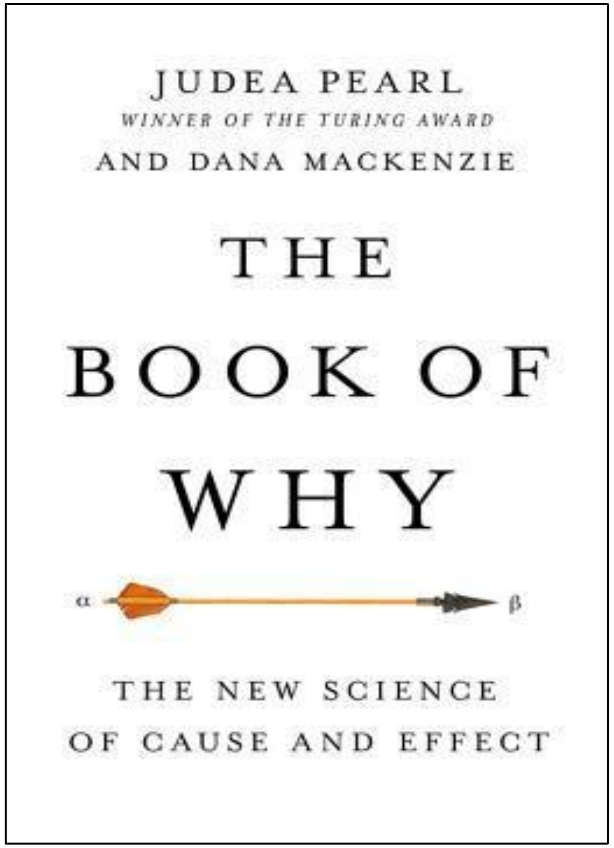


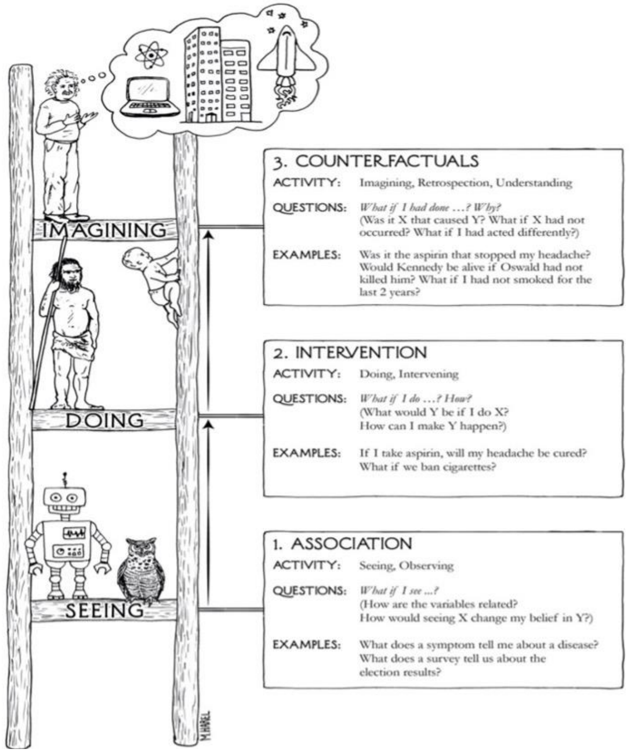
Causality

- Causality is the science of understanding the “cause – and- effect” relationships in the world around us.
- X (new drug) caused Y (patient’s health) if when all confounders (age, severity of illness etc.) are adjusted, an intervention in X results in a change in Y, but intervention in Y does not change X.

Pearl's Ladder of Causation

Causality





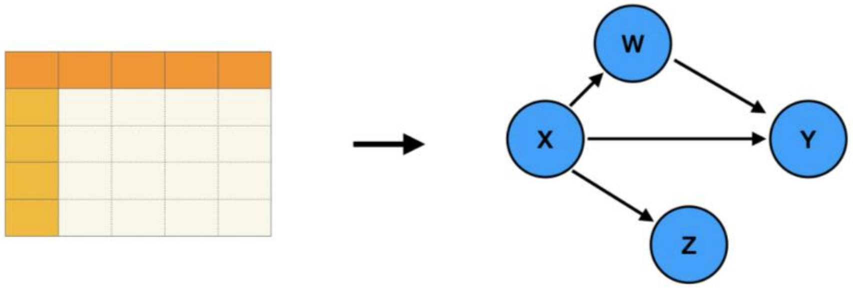
The Three Layer Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Causal Modeling

Causal Discovery

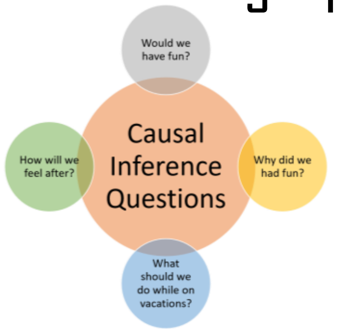
Learn the graph/structure from the data



Build a graphical representation (often a **Directed Acyclic Graph** or DAG) that captures the causal relationships among variables.

Causal Inference

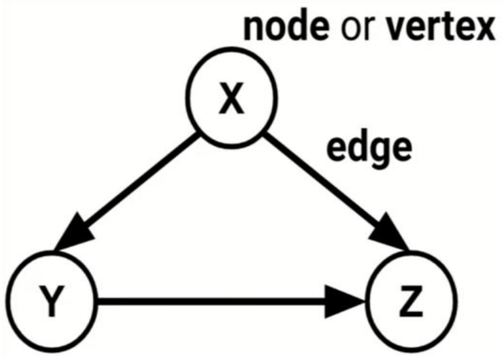
Inferring/answering conditional questions from causal graph



It is about drawing meaningful and well-supported causal conclusions within a known causal framework.

"**Causal discovery** is the process of **building the causal model from data** when the model is unknown, while **causal inference** is the process of using the causal model (whether discovered or assumed) to **make meaningful causal statements and predictions**."

Basics of Causal Graphs



Adjacent nodes: X and Y, Y and Z

Non-adjacent nodes: X and Z

X is **parent** of Y	Y is **parent** of Z
Y is **child** of X	Z is **child** of Y
X is **ancestor** of Y and Z	Y is **ancestor** of Z
Y is **descendant** of X	Z is **descendant** of Y and X

Causal Graphs are Directed Acyclic Graphs (DAGs)

A DAG is a graph that provides a visual representation of causal relationships among a set of variables.

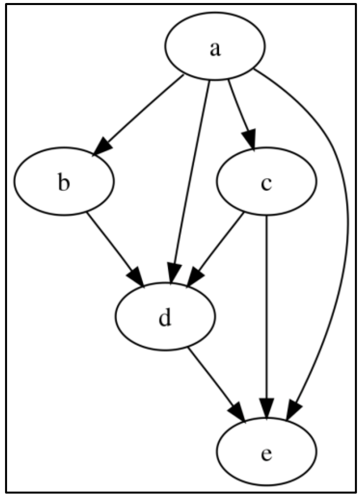
D = directed (all arrows point in only a single direction).

The direction of the arrow is the direction of causation:
A ➡ B means A causes B.

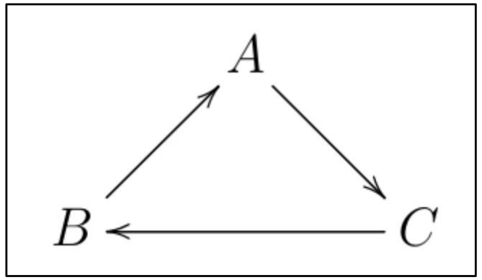
A = acyclic (no sequence of arrows forms a closed loop, which would be backwards causation).Causal Graph should be acyclic.

Several Methods available to find out DAG for Causal Discovery.

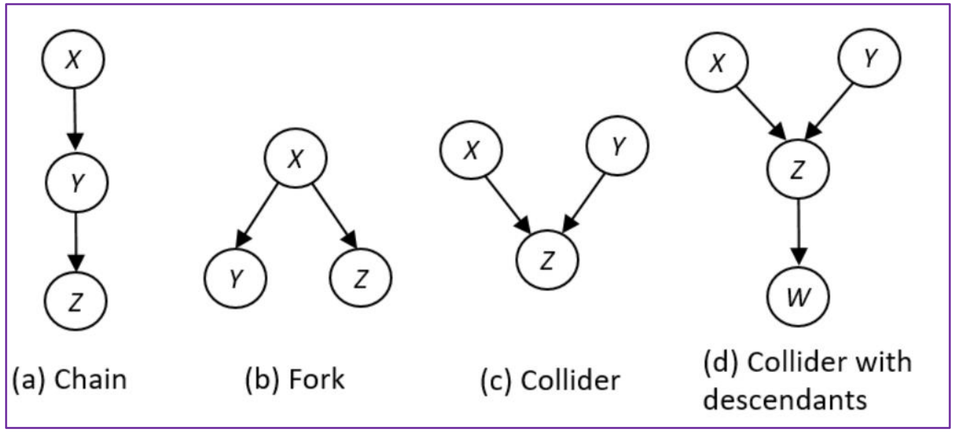
DAG



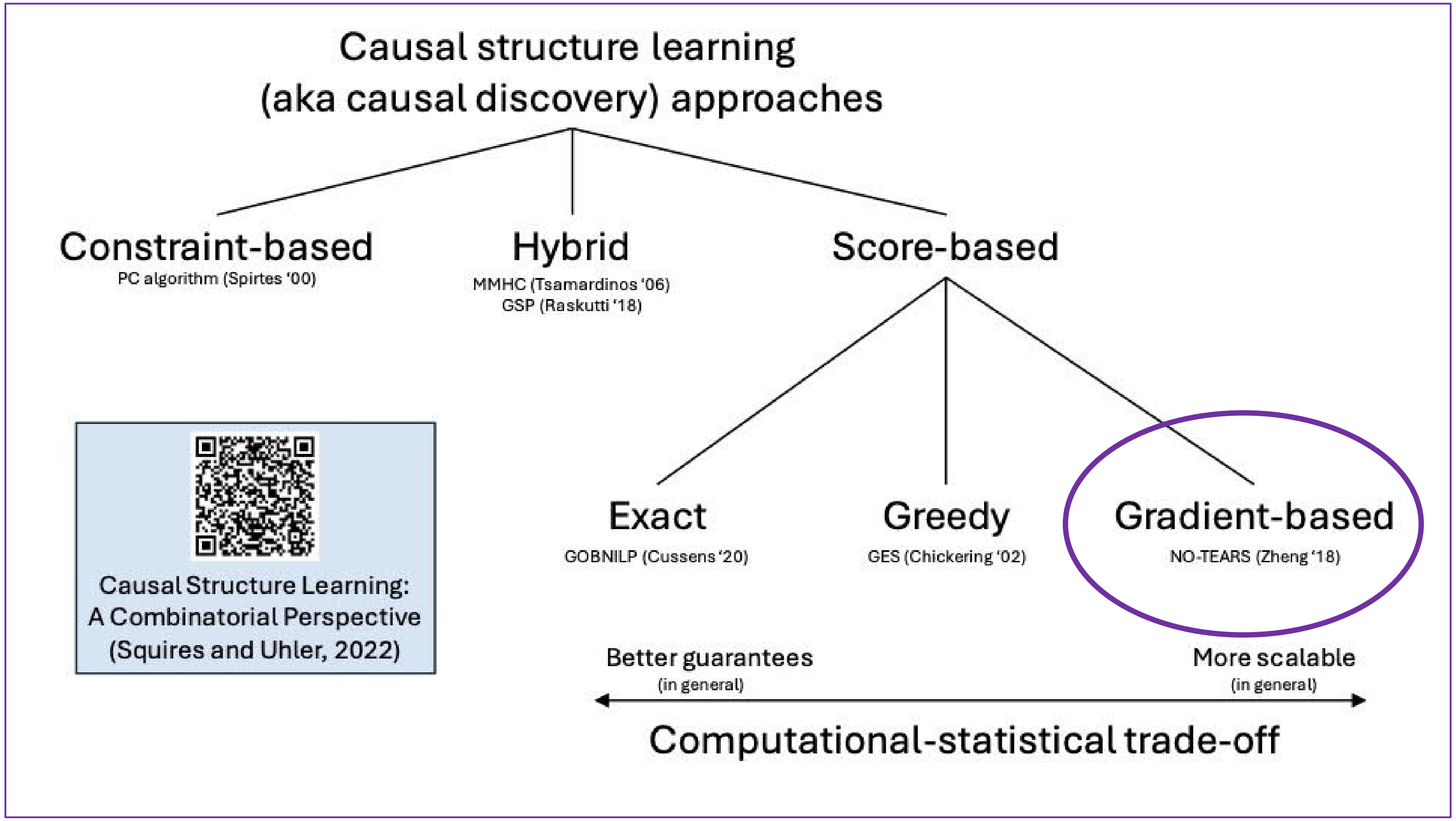
Not a DAG, Correlation

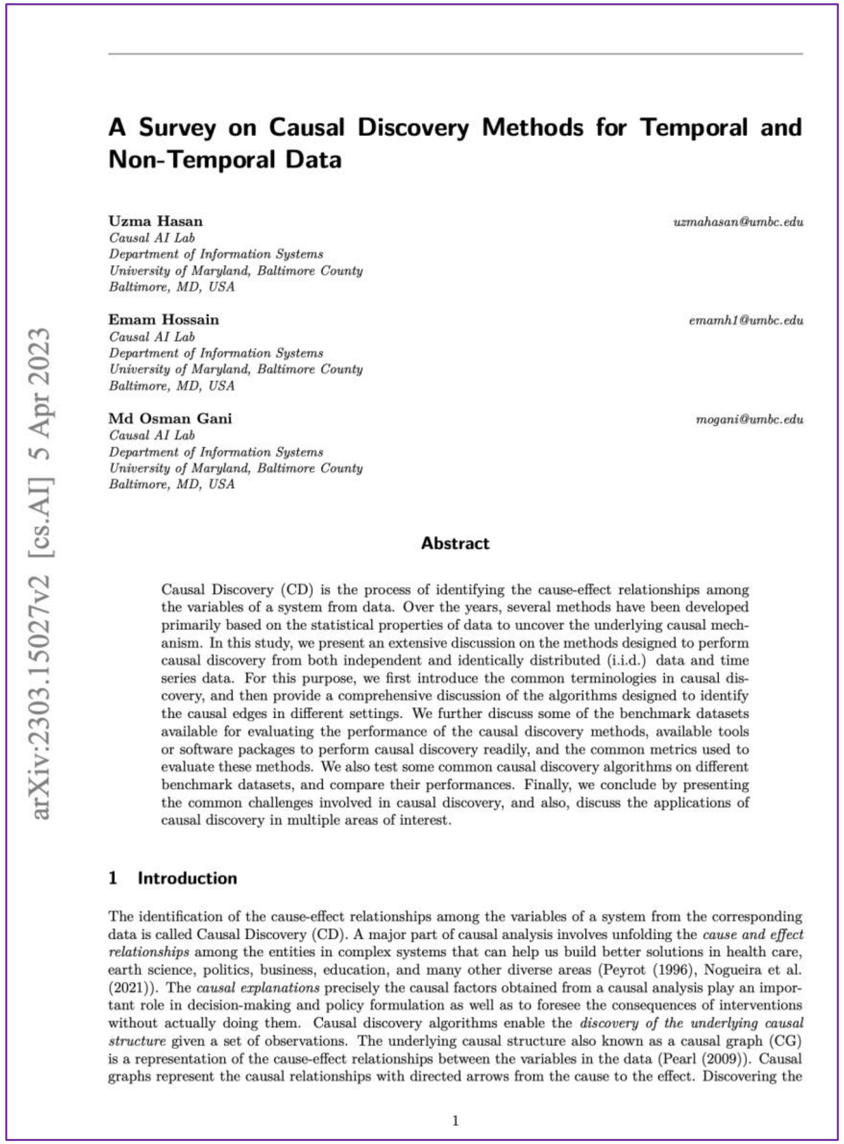


Types



Methods for Causal Discovery





A Survey on Causal Discovery Methods for Temporal and Non-Temporal Data

Uzma Hasan uzmahasan@umbc.edu
Causal AI Lab
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD, USA

Emam Hossain emamh1@umbc.edu
Causal AI Lab
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD, USA

Md Osman Gani mogani@umbc.edu
Causal AI Lab
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD, USA

arXiv:2303.15027v2 [cs.AI] 5 Apr 2023

Abstract

Causal Discovery (CD) is the process of identifying the cause-effect relationships among the variables of a system from data. Over the years, several methods have been developed primarily based on the statistical properties of data to uncover the underlying causal mechanism. In this study, we present an extensive discussion on the methods designed to perform causal discovery from both independent and identically distributed (i.i.d.) data and time series data. For this purpose, we first introduce the common terminologies in causal discovery, and then provide a comprehensive discussion of the algorithms designed to identify the causal edges in different settings. We further discuss some of the benchmark datasets available for evaluating the performance of the causal discovery methods, available tools or software packages to perform causal discovery readily, and the common metrics used to evaluate these methods. We also test some common causal discovery algorithms on different benchmark datasets, and compare their performances. Finally, we conclude by presenting the common challenges involved in causal discovery, and also, discuss the applications of causal discovery in multiple areas of interest.

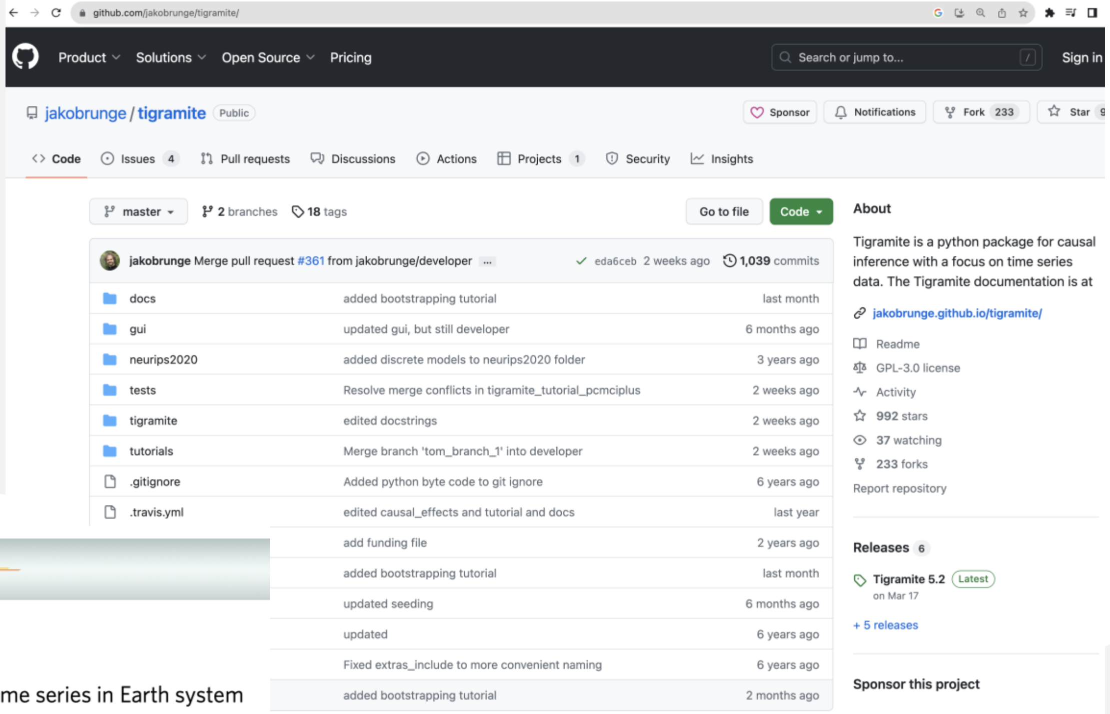
1 Introduction

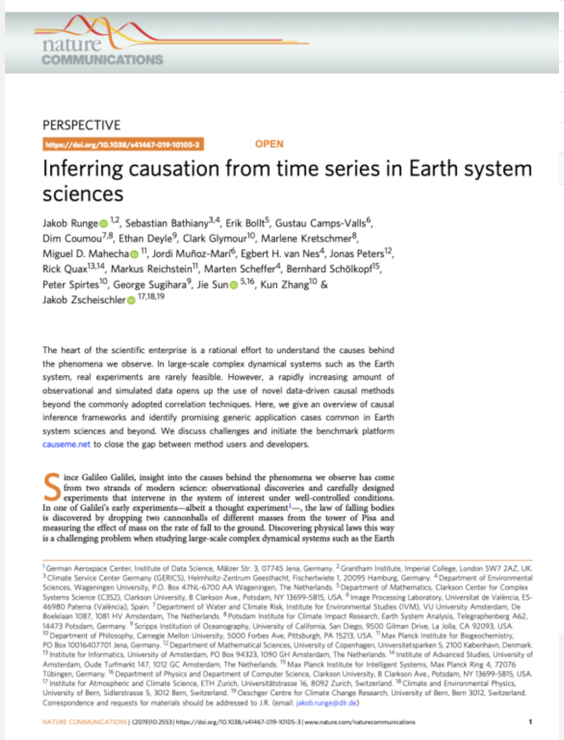
The identification of the cause-effect relationships among the variables of a system from the corresponding data is called Causal Discovery (CD). A major part of causal analysis involves unfolding the *cause and effect relationships* among the entities in complex systems that can help us build better solutions in health care, earth science, politics, business, education, and many other diverse areas (Peyrot (1996), Nogueira et al. (2021)). The *causal explanations* precisely the causal factors obtained from a causal analysis play an important role in decision-making and policy formulation as well as to foresee the consequences of interventions without actually doing them. Causal discovery algorithms enable the *discovery of the underlying causal structure* given a set of observations. The underlying causal structure also known as a causal graph (CG) is a representation of the cause-effect relationships between the variables in the data (Pearl (2009)). Causal graphs represent the causal relationships with directed arrows from the cause to the effect. Discovering the

1

“PCMCI” de facto model Earth Science

- Peter-Clark (PC)+Momentary Conditional Independence (MCI)
- After the paper by Runge (2019) PCMCI become “THE METHOD” for Causal Structure Learning in Earth Science.
- Part of it because of the “tigermite” package by Runge et al.





nature COMMUNICATIONS

PERSPECTIVE

https://doi.org/10.1038/s41467-019-10105-3

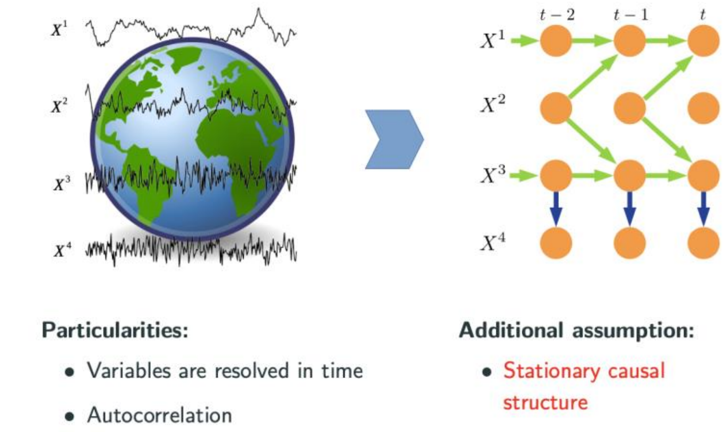
OPEN

Inferring causation from time series in Earth system sciences

Jakob Runge[1,2], Sebastian Bathiany[3,4], Erik Bollt[5], Gustau Camps-Valls[6], Dim Coumou[7,8], Ethan Deyle[9], Clark Glymour[10], Marlene Kretschmer[8], Miguel D. Mahecha[11], Jordi Muñoz-Marí[6], Egbert H. van Nes[4], Jonas Peters[12], Rick Quax[13,14], Markus Reichstein[11], Marten Scheffer[4], Bernhard Schölkopf[15], Peter Spirtes[10], George Sugihara[9], Jie Sun[5,16], Kun Zhang[10] & Jakob Zscheischler[17,18,19]

The heart of the scientific enterprise is a rational effort to understand the causes behind the phenomena we observe. In large-scale complex dynamical systems such as the Earth system, real experiments are rarely feasible. However, a rapidly increasing amount of observational and simulated data opens up the use of novel data-driven causal methods beyond the commonly adopted correlation techniques. Here, we give an overview of causal inference frameworks and identify promising generic application cases common in Earth system sciences and beyond. We discuss challenges and initiate the benchmark platform causeme.net to close the gap between method users and developers.

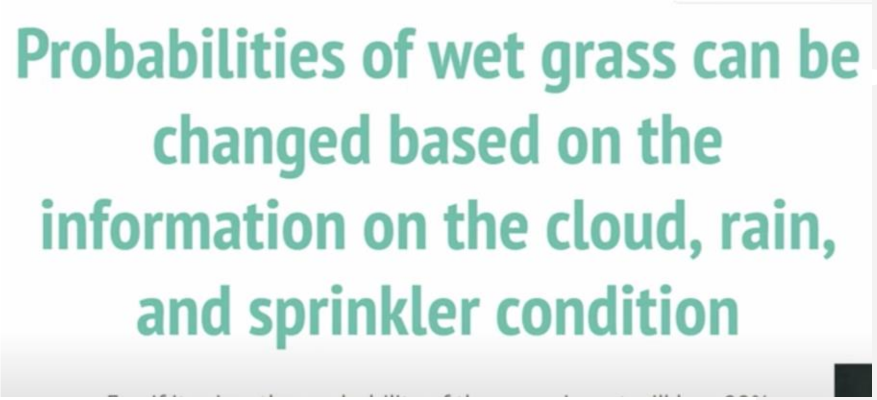
Since Galileo Galilei, insight into the causes behind the phenomena we observe has come from two strands of modern science: observational discoveries and carefully designed experiments that intervene in the system of interest under well-controlled conditions. In one of Galilei's early experiments—albeit a thought experiment[1]—, the law of falling bodies is discovered by dropping two cannonballs of different masses from the tower of Pisa and measuring the effect of mass on the rate of fall to the ground. Discovering physical laws this way is a challenging problem when studying large-scale complex dynamical systems such as the Earth

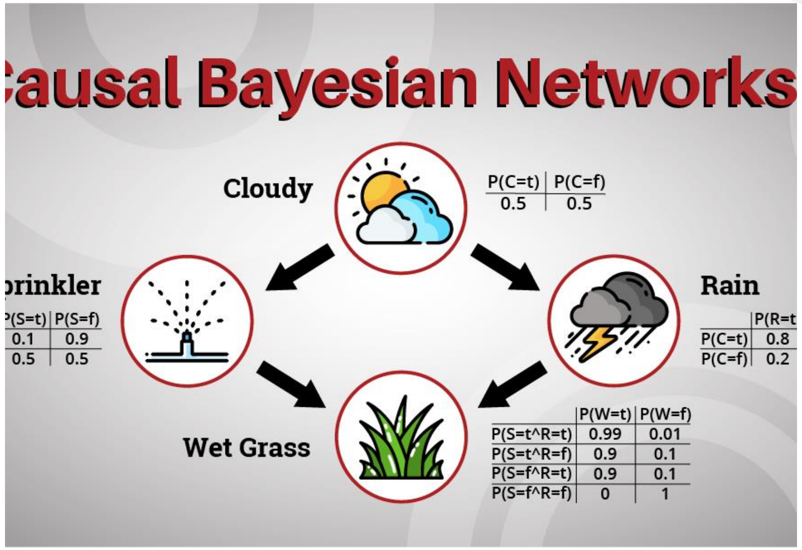


Why we choose DAG with No tears over PCMCI?

PCMCI is tailored for time series data, considering the **the temporal ordering and seeks to identify the time lag between cause and effect**. It infers causal relationships based on partial correlation estimation at different time points.

DAG with NO TEARS is designed for general observational data and focuses on **inferring causal relationships without assuming a specific temporal order**. It estimates partial correlations and optimizes the DAG structure to capture the most significant direct associations between variables.





“DAGs with NO TEARS (Nonlinear Optimization of TEmporal Relationships in Systems) ”

- It is a novel method for Bayesian Network (BN) structure learning based on continuous optimization. BN is probabilistic graphical model consist of two parts: a structure and parameters.

- The structure is a directed acyclic graph (DAG) that expresses conditional independencies and dependencies among random variables associated with nodes. The parameters consist of conditional probability distributions associated with each node.

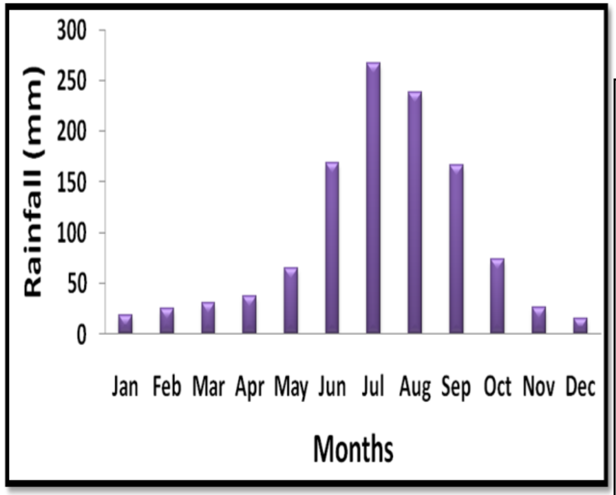
- Estimating the structure of DAGs, is a challenging problem since the search space of DAGs is combinatorial and scales superexponentially with the number of nodes.

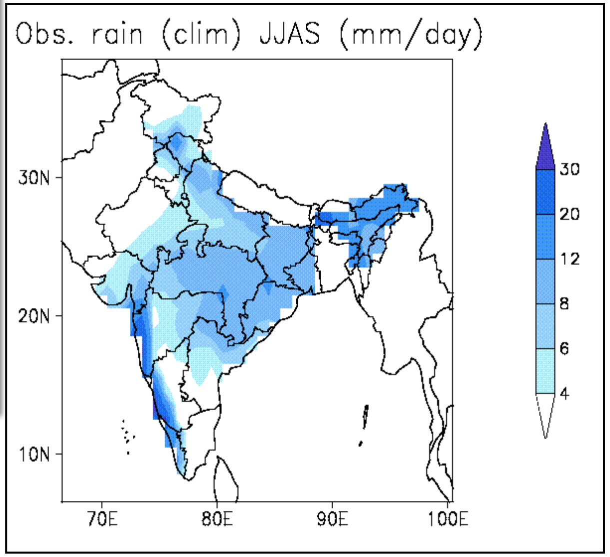
- "DAGs with NO TEARS” introduced a fundamentally different strategy: formulate the structure learning problem as a purely continuous optimization problem over real matrices that avoids this combinatorial constraint entirely (Zheng et al.,2018)

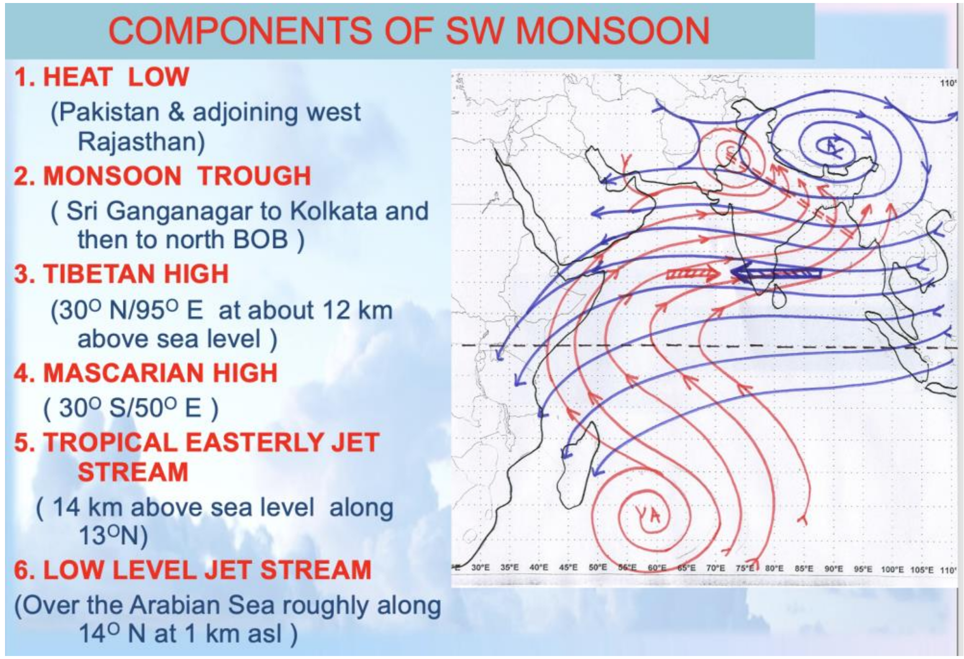
Steps for DAGs with NO TEARS

- **Estimating Partial Correlations:**
  - Calculate the partial correlations between pairs of variables while controlling for the effects of other variables.
  - Estimate the strength of the direct associations between variables.
- **Score Function and Optimization:**
  - Define a score function to evaluate the goodness of fit between the observed partial correlations and the hypothetical set of partial correlations in the DAG.
  - Employ an optimization algorithm to search for the DAG structure that maximizes the score function.
  - Iteratively explore different DAG structures by adjusting the presence or absence of edges between variables.
- **Sparsity Control:**
  - Apply a threshold or criteria to determine the significance of the estimated partial correlations.
  - Remove weaker or less significant edges to create a sparse DAG that focuses on the most important causal relationships.
- **Edge Orientation:**
  - Utilize additional techniques, such as constraint-based methods or local search algorithms, to orient the edges in the DAG and determine the direction of influence between variables.
- **Plotting the DAG:**
  - Visualize the resulting DAG, representing the estimated causal relationships among variables.
  - Use arrows to indicate the direction of influence between variables.

Case study: All India Summer Monsoon (Jun-Jul-Aug-Sep)







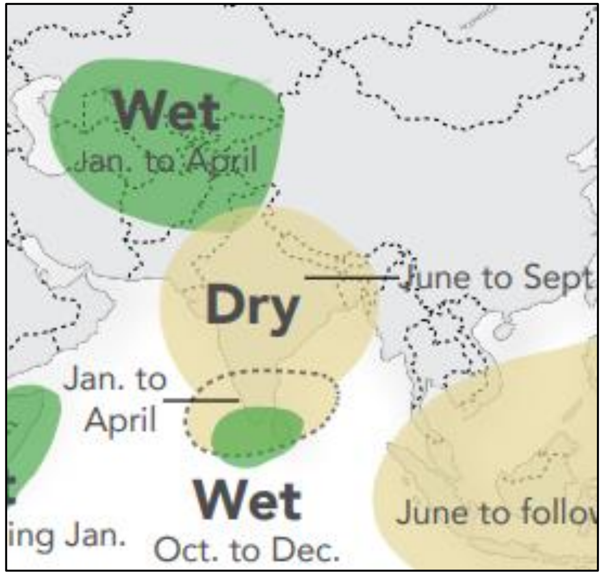
Monsoon is characterized by seasonal wind reversal in tropics

How good is CFSv2 to predict AISMR?

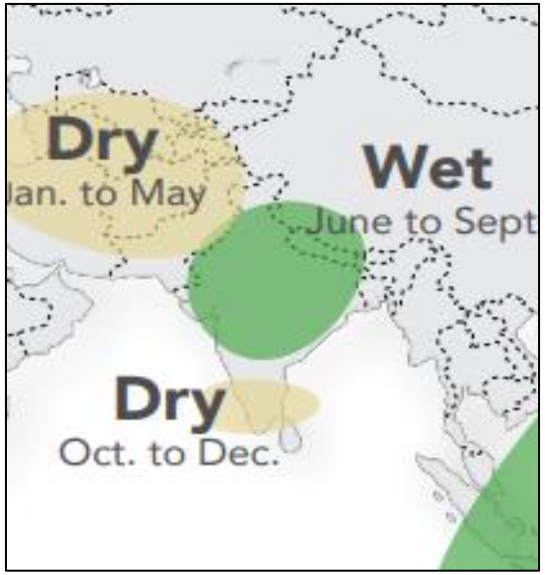
Predictability of AISMR: Process

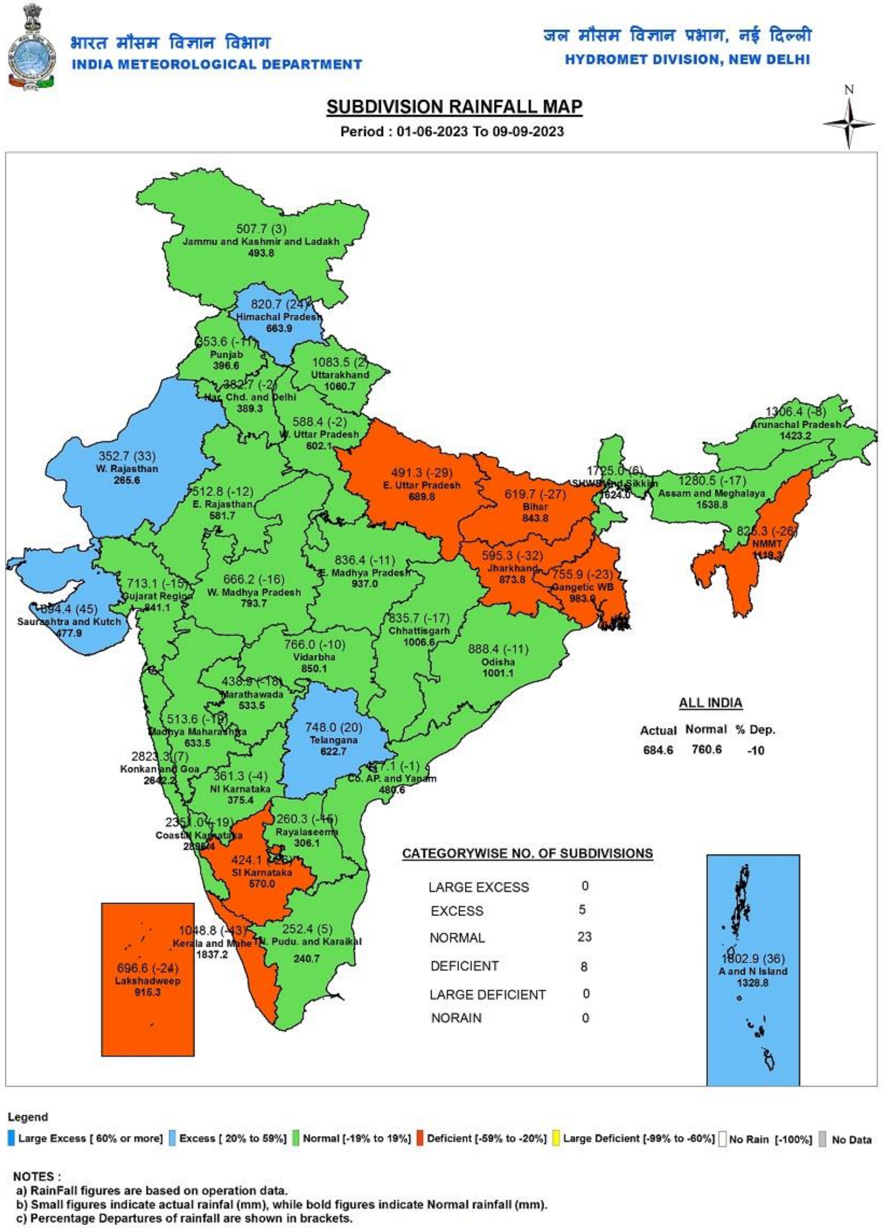
What about last year (El Niño)?

El Niño and Rainfall



La Niña and Rainfall





Predictability of AISMR: Process

GEOPHYSICAL RESEARCH LETTERS, VOL

Impact of the Indian Ocean Dip… between the Indian Monsoon Rainfa…

Karumuri Ashok[1], Zhaoyong Guan[2] and Toshio Y…
Institute for Global Change Research, Frontier Research System f…
Kanazawa-Ku, Yokohama City, Kanagawa, 236-0001, Japan

Unraveling … Monsoon …

Earth Syst. Dynam., 11, 903–923, 2020
https://doi.org/10.5194/esd-11-903-2020

Earth System Dynamics
EGU

The synerg… summer mon… simulation…

Indian Summer Monsoon Variability
El-Nino Teleconnections and Beyond
2021, Pages 157-182
ELSEVIER

Chapter 8 - Indian Ocean Dipole influence on Indian summer monsoon and ENSO: A review

Annalisa Cherchi [a], Pascal Terray [b], Satyaban B. Ratna [c], Syam Sankar [d], K P Sooraj [e], Swadhin Behera [f]

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/B978-0-12-822402-1.00011-9

Get rights and content

Abstract

The Indian Ocean Dipole (IOD) is one of the dominant modes of variability of the tropi… Indian Ocean and it has been suggested to have a crucial role in the teleconnection between the Indian summer monsoon and El Niño Southern Oscillation (ENSO). The main ideas at the base of the influence of the IOD on the ENSO-monsoon teleconnectio… include the possibility that it may strengthen summer rainfall over India, as well as the opposite, and also that it may produce a remote forcing on ENSO itself. In the future, th… IOD is projected to increase in frequency and amplitude with mean conditions mimick… the characteristics of its positive phase. Still, state-of-the-art global climate models hav… large biases in representing the mean state and variability of both IOD and ISM, with potential consequences for their future projections. However, the characteristics of the IOD and ENSO are likely to continue in a future warmer world, with persistence of thei… linkage.

VayuMandal 48(1), 2022

Role of ENSO and IOD in the Indian Summer Monsoon Variability: A Review

Pratap Kumar Mohanty and Sanjukta Rani Padhi
Department of Marine Sciences
Berhampur University, Berhampur-760007, Odisha
Email: pratap_mohanty@yahoo.com

ABSTRACT

The Indian Summer Monsoon Rainfall (ISMR) has contributed almost 75.3% to the annual rainfall during 1901-2020 and is considered as the lifeline of India for a sustainable agriculture and economy. ISMR exhibits significant spatial and temporal variability in the form intra-seasonal, seasonal, interannual and biennial oscillations. In the present study, we have used gridded Indian Meteorological Department (IMD) rainfall data from 1901 to 2020 with 0.250x0.250 resolution and have focused on ISMR variability due to coupled ocean atmosphere processes in the Indian and Pacific oceans. As proxies of these coupled ocean atmosphere processes, we consider the role of ENSO and IOD on ISMR. Although several studies were carried out on these aspects during the last two and half decades, the present study is different from other and aims to examine the ISMR variability during 1901-2020 over All India (AI) and different homogeneous zones (NEI, NWI, CI and SPI) under El Nino, La Nina, +ve and -ve IOD (without any co-occurrence) and with co-occurrence of El Nino with +ve IOD and La Nina with –ve IOD. Considering the changing relationship of ISMR with ENSO and IOD, this study also focuses on regional ISMR variability due to various ENSO-IOD conditions.

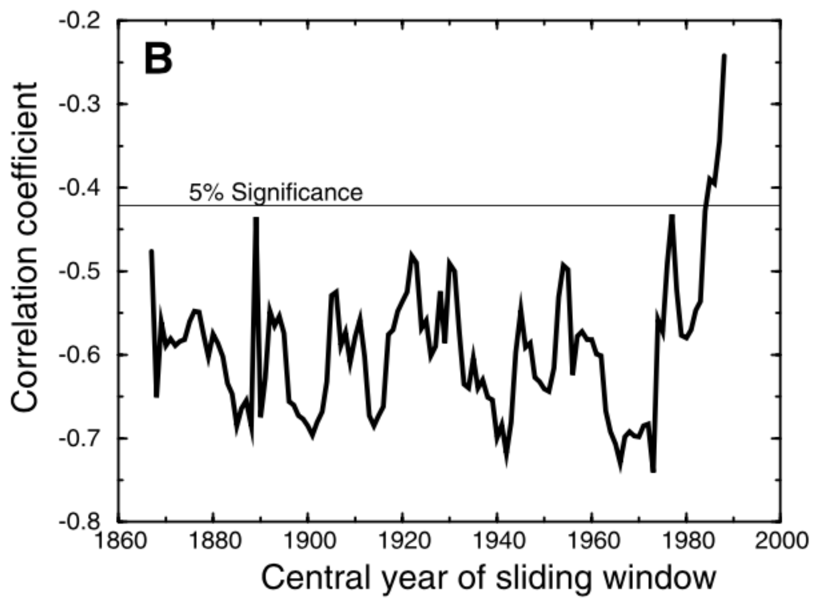
Keywords: ISMR, ENSO, IOD, Homogeneous zones and Precipitation concentration index.

1. Introduction

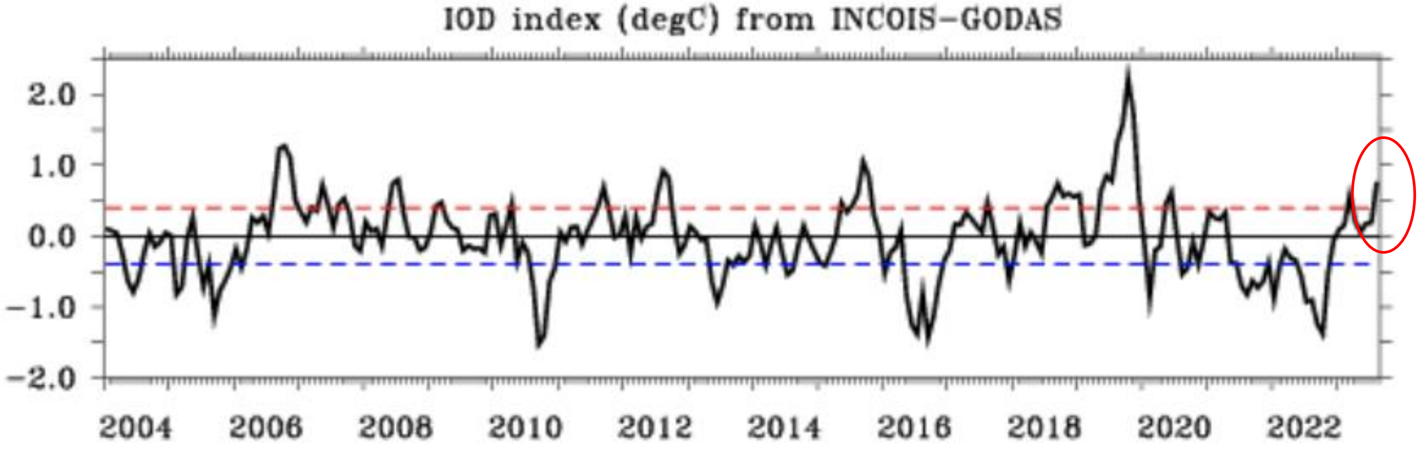
Agricultural practices in India and in many other south Asian countries are intricately linked to the performance of the monsoon, particularly ISMR (Parthasarathy et al., 1988). Considering significant spatial-temporal variabilities of ISMR forced from both internal and external forcing, understanding ISMR variability and its prediction is extremely important. Since the internal forcing, mainly intra-seasonal oscillations set a limit to the predictability, major focus are on understanding external forcing including the coupled ocean-atmosphere interaction, SST, snow cover etc. to improve prediction of ISMR. Besides other external forcing, the El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) are widely considered as the two major climate drivers of ISMR (Ashok et al., 2001; Behera & Ratnam, 2018; Cherchi et al., 2021; Hrudya et al., 2020; Krishnaswamy et al., 2014; Rajeevan & Pai, 2007; Saji et al., 1999; Varikoden et al., 2020; Webster et al., 1999). Past studies have elucidated on the relationship of ISMR with ENSO and IOD. However, weakening relationship between ENSO and ISMR after 1970s was revealed by studies conducted during the last two decades (Ashok et al., 2019; Kawamura et al., 2005; Kumar et al., 1999), which lead to study the impacts of IOD on ENSO-ISMR relationship (Ashok et al., 2001). The weakening relationship was attributed to shift in the spatial correlation pattern over the Indian subcontinent from northwest to north east. The study revealed that when the ENSO-ISMR correlation is low (high), the IOD-ISMR correlation is high (low). Many other studies (Anil et al., 2016; Ashoket al., 2004; Gadgil et al., 2004; Webster et al., 1999) also indicated that frequent emergence of the IOD have weakened the otherwise robust relationship between ENSO and ISMR. Thus, it was made apparent that IOD, which moderates the meridional circulation by inducing anomalous convergence (divergence) pattern over Bay of Bengal during positive (negative) IOD events, leads to excessive (deficit) monsoon rainfall over the monsoon trough region (Ashok et al., 2003). This feature was evident in the typical IOD year of 1994 (Behera et al., 1999) and positive IOD year 1997.Studies indicated the influences of ENSO and IOD on the ISMR as opposite to one another (Ashoket al., 2004; Ashok & Saji, 2007). Years with co-occurrence of +ve IOD with El Nino (1961,1963,1967,1972,1977,1982,1983,1994 and 1997) have positive anomalies of rainfall along the monsoon trough area, the west coast and northwest India while in years with pure +ve IOD events

In summary…

- Weakening relationship between ENSO and ISMR after 1970s was revealed by studies conducted during the last two decades.
- There is impacts of IOD on ENSO-ISMR relationship.
- Frequent emergence of the IOD have weakened the otherwise robust relationship between ENSO and ISMR.
- Years with co-occurrence of +ve IOD with El Nino (1961,1963,1967,1972,1977,1982,1983,1994 and 1997) have positive anomalies of rainfall along the monsoon trough area, the west coast and northwest India.



Current IOD/DMI index



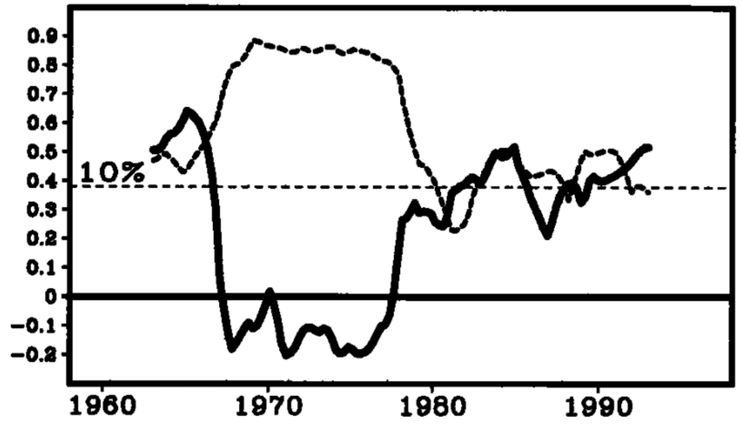
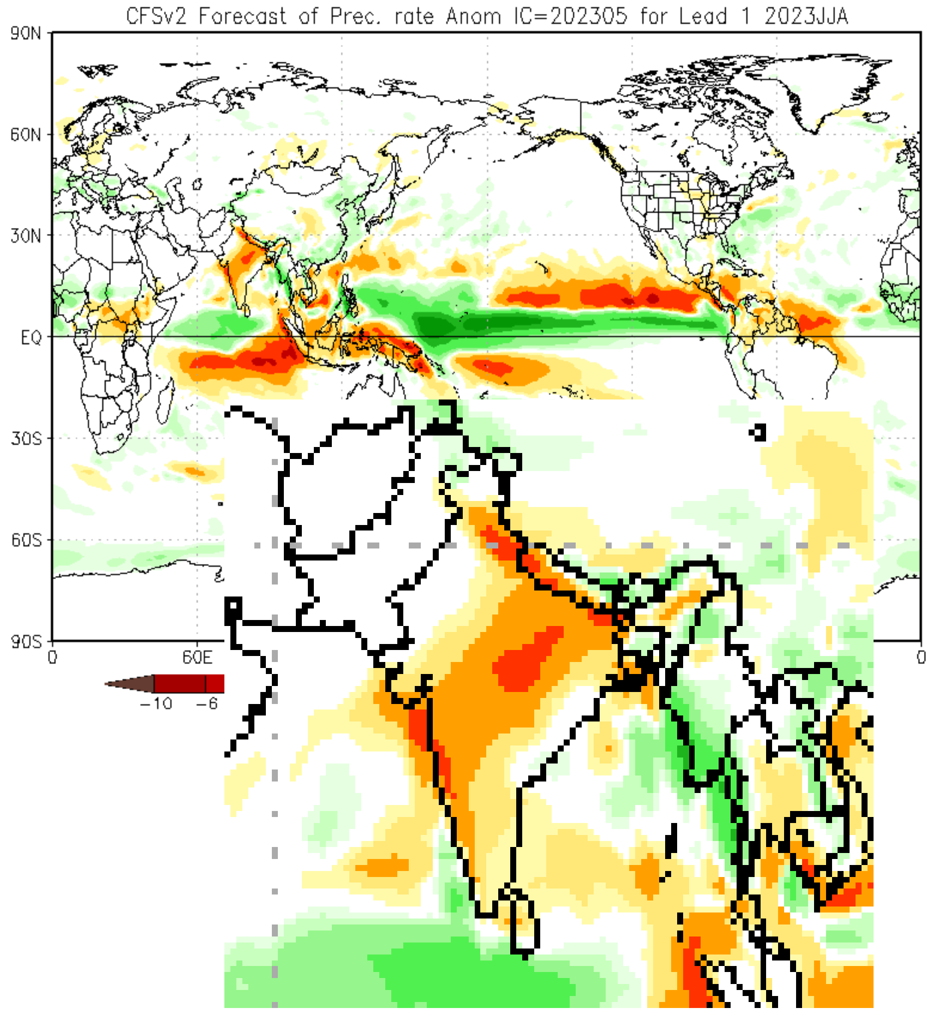
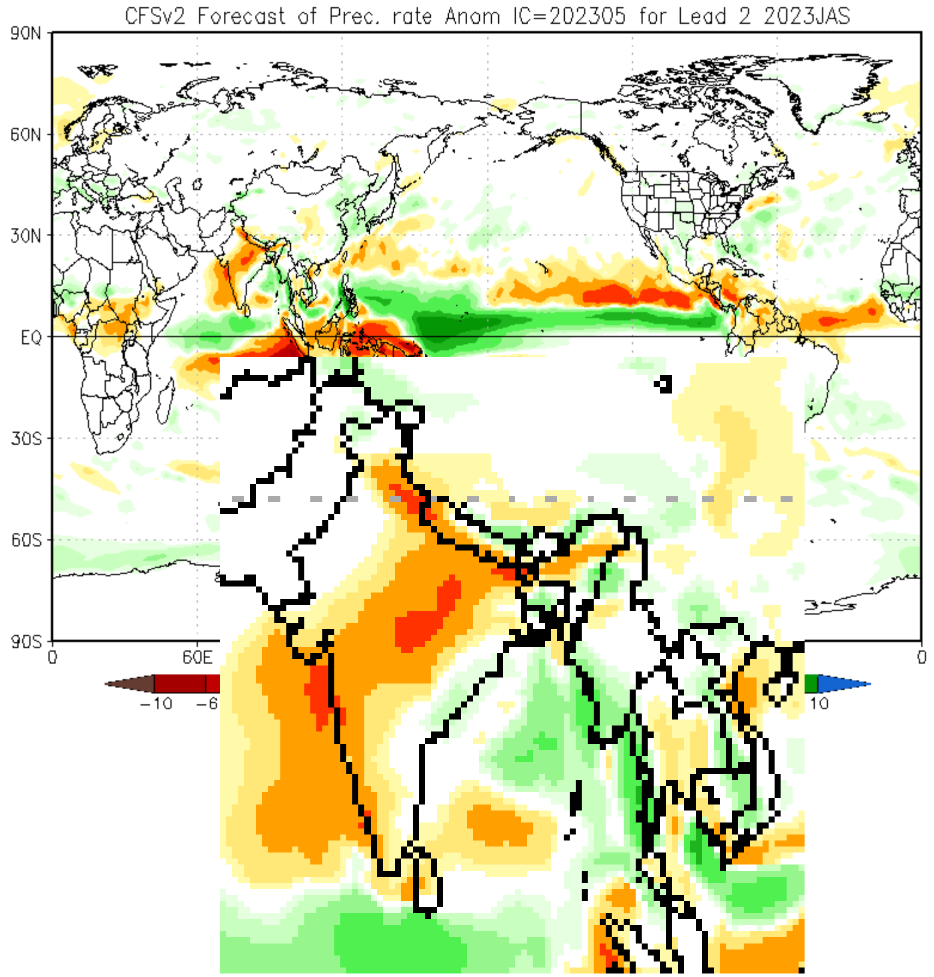


Figure 1. The 41-month sliding correlation coefficients between ISMR and IODMI (solid), and those between monthly ISMR and NINO3 SST (dashed; to be multiplied by −1) during 1958-1997. The significant correlation value at 90% confidence level is 0.38 (verified by 1,000 randomized time series, using the Monte-Carlo simulations)

How good is CFSv2 to predict AISMR?





CFSv2 forecast for 2023 monsoon

How good is CFSv2 to predict AISMR?

Observational References:

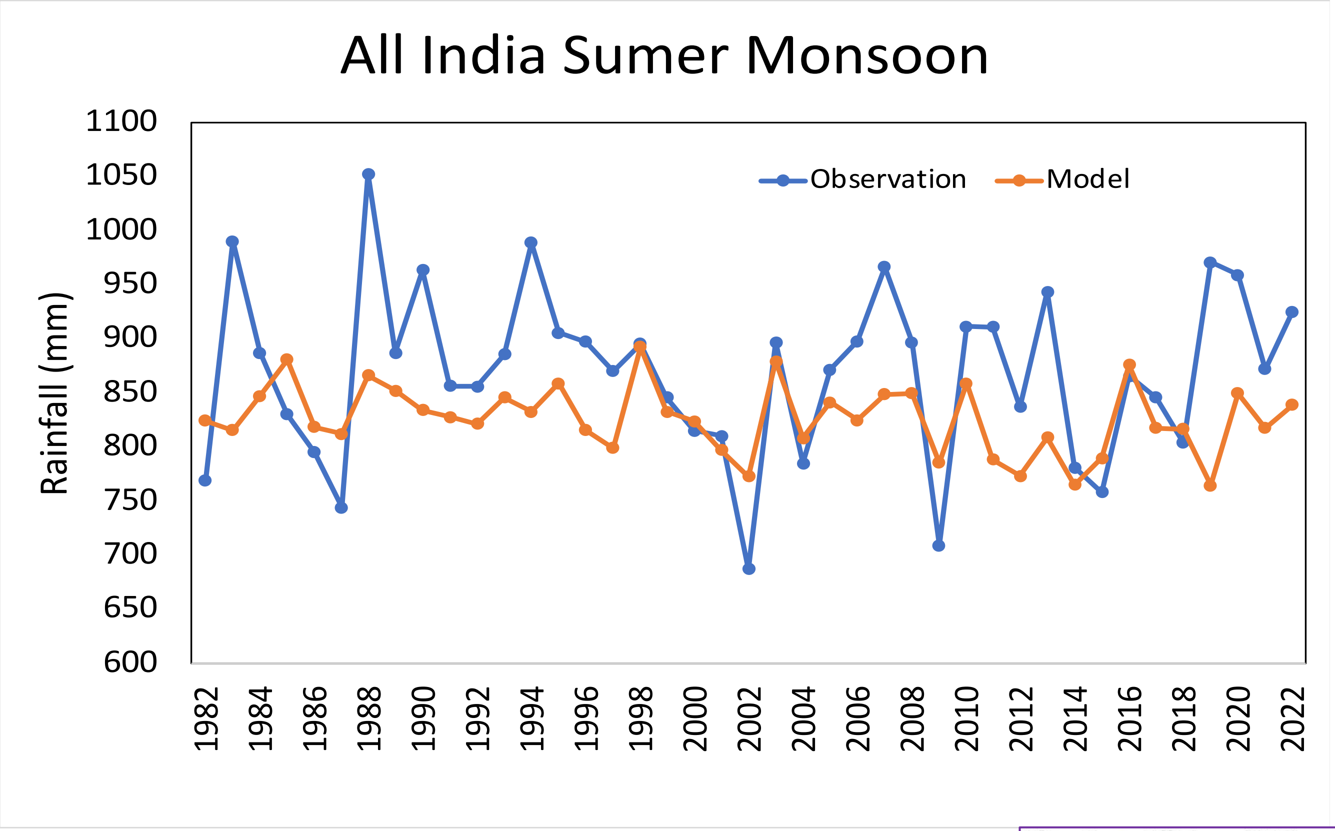
Rainfall data from India Met. Dept.

Niño 3.4 and DMI SST Indices from NOAA's PSL webpage.

Obs_{JJAS} =[O_k], k=1982-2022

Seasonal Climate Model: Climate Forecast system v2 from NCEP-Long (lead -1 hindcast from 1982-2022). Niño3.4 and DMI SST Indices calculated from SST.

$Model_{Jun \rightarrow JJAS}$ =[M_k] ,k=1982-2022



Kling-Gupta Efficiency=0.167

Kling-Gupta Efficiency (KGE)

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}$$

- where r = correlation coefficient

  β = bias ratio $= \frac{\mu_E}{\mu_O}$

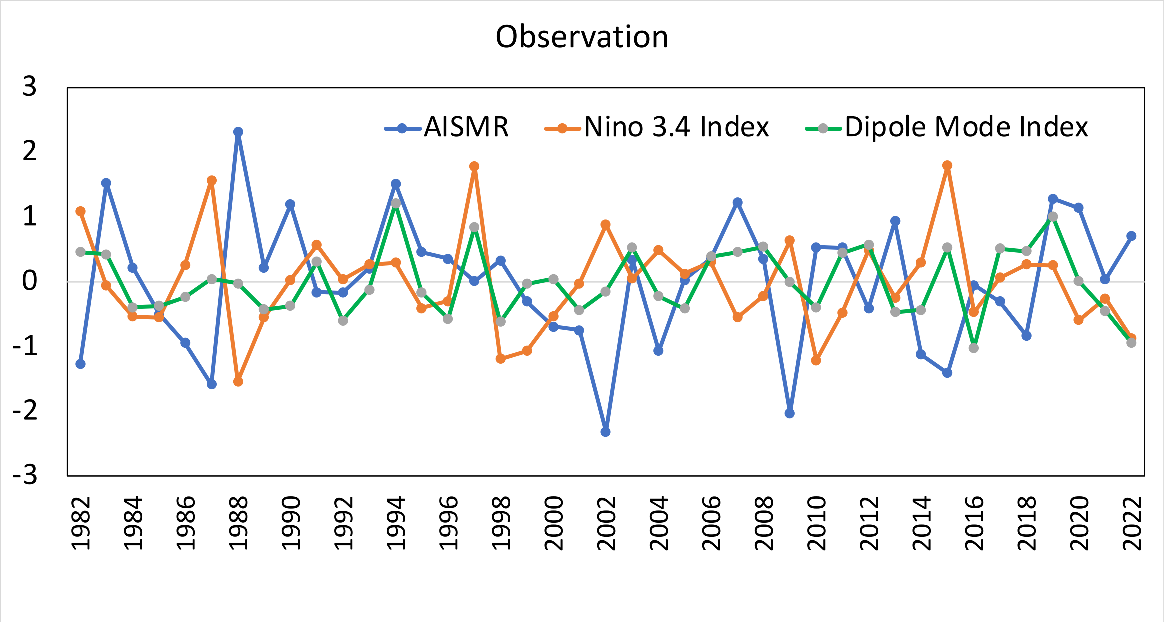
  γ = variability ratio $= \frac{CV_E}{CV_O} = \frac{\sigma_E/\mu_E}{\sigma_O/\mu_O}$

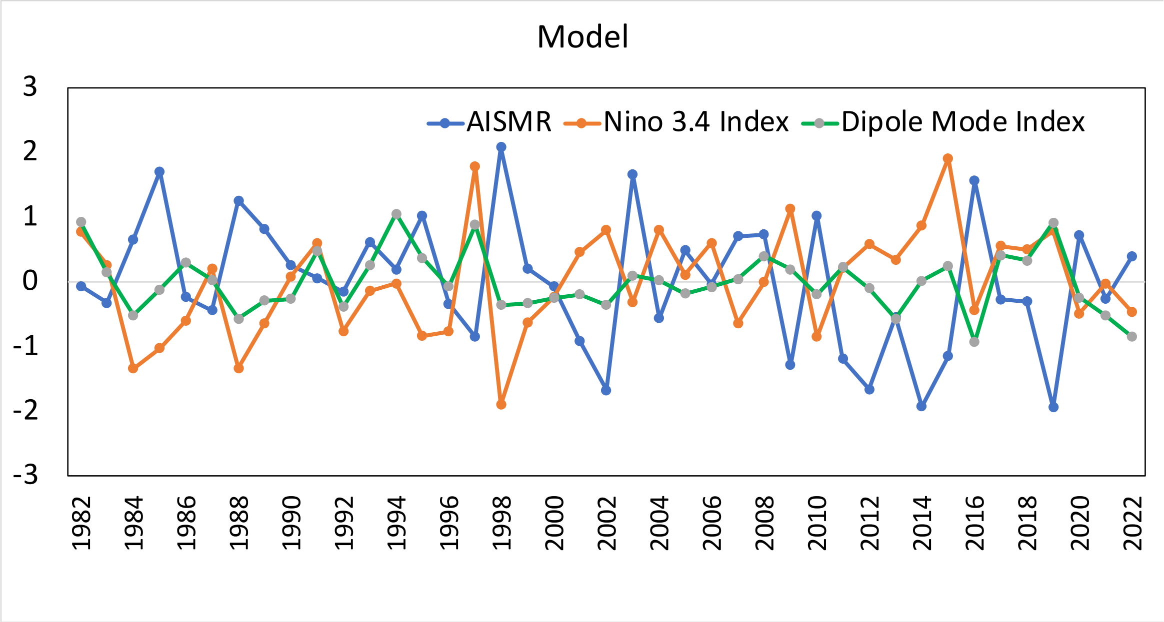
- and CV = coefficient of variation

  μ = mean

  σ = standard deviation

Process-oriented dynamical Model evaluation: Traditionally



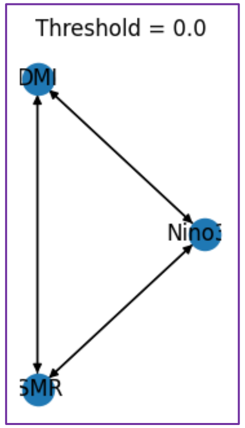


	AISMR	Nino 3.4 Index	Dipole Mode Index
AISMR	1	-0.57	0.11
Nino 3.4 Index		1	0.44
Dipole Mode Index			1

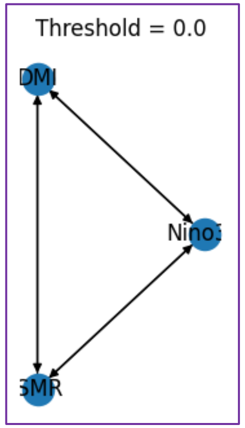
	AISMR	Nino 3.4 Index	Dipole Mode Index
AISMR	1	-0.75	-0.31
Nino 3.4 Index		1	0.48
Dipole Mode Index			1

Process-oriented dynamical Model evaluation: Causal Structure learning

Observation



CFSv2

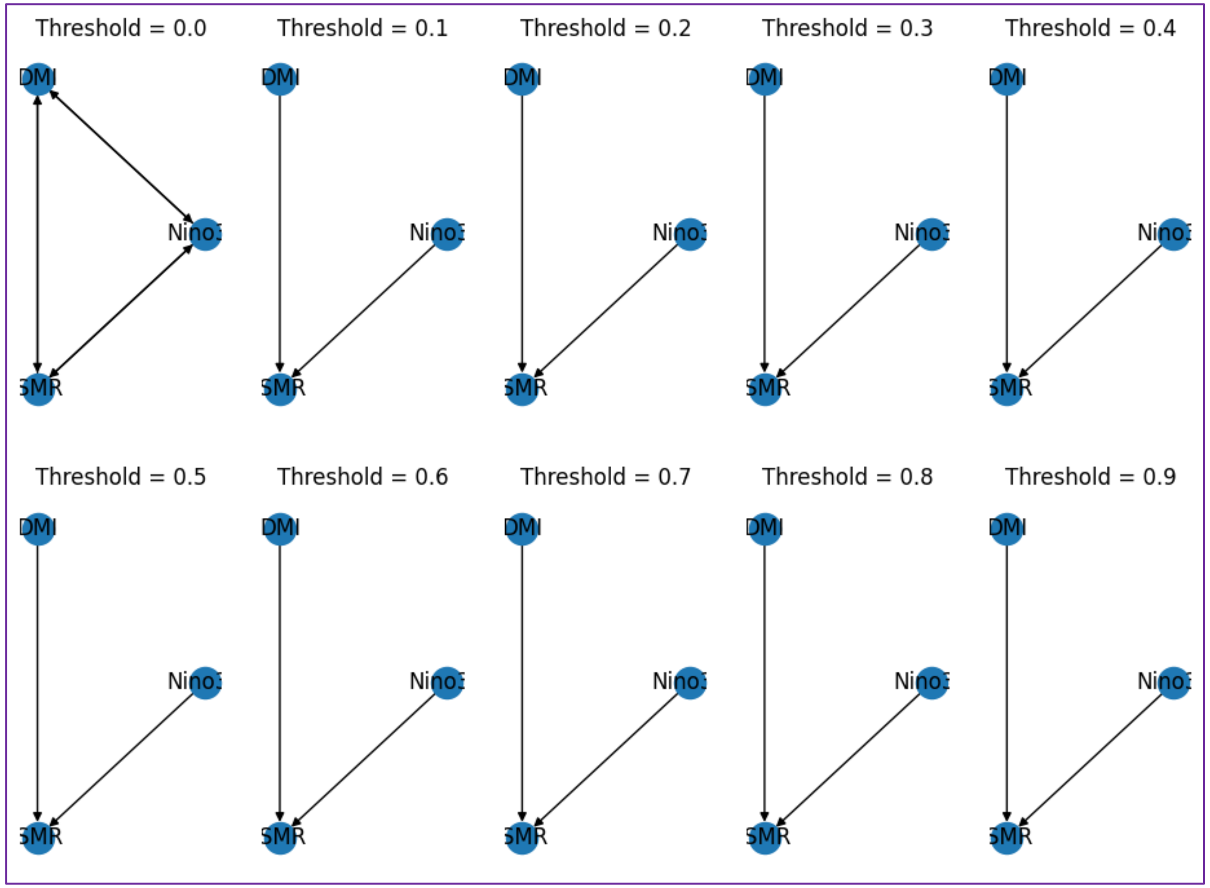


DAG between AISMR,Nino3.4 and DMI

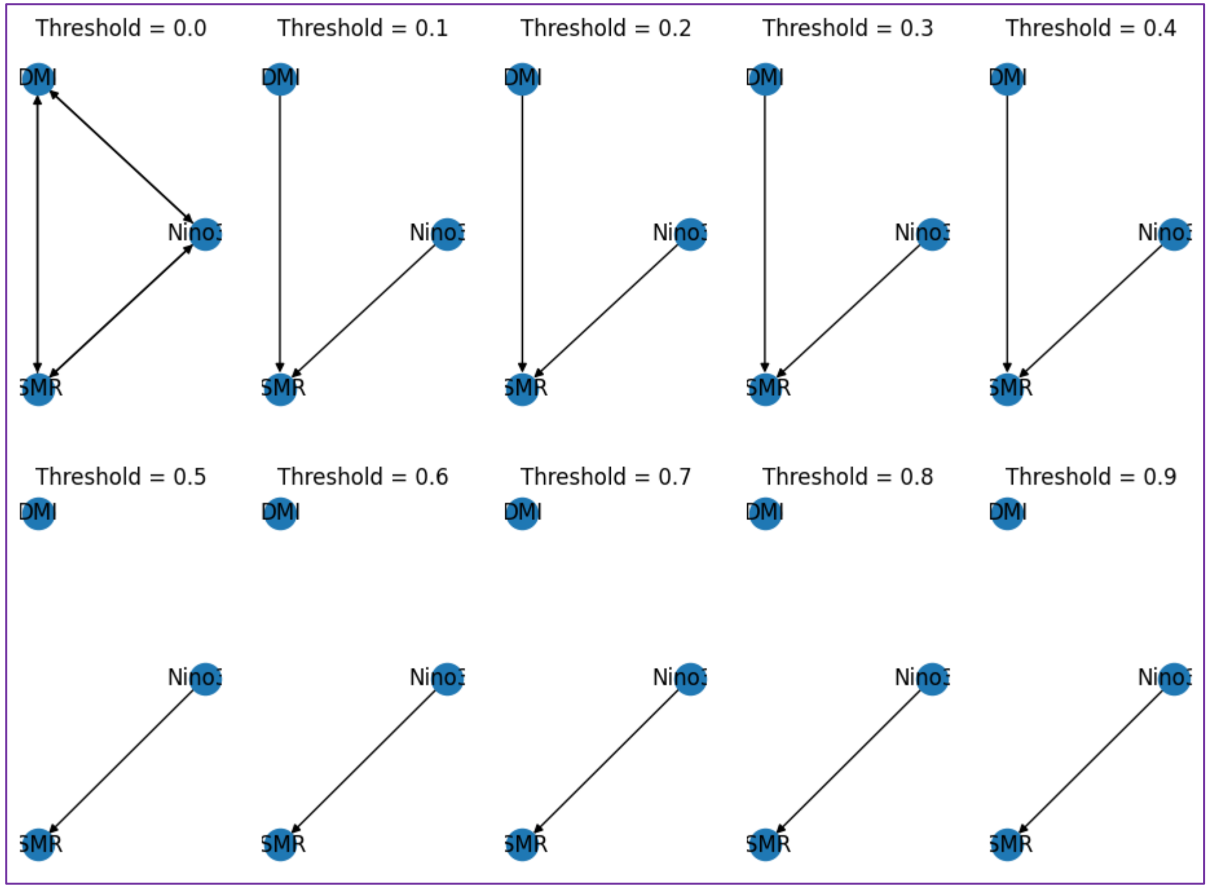
Threshold: Partial Correlation value

Process-oriented dynamical Model evaluation: Causal Structure learning

Observation



CFSv2



DAG between AISMR,Nino3.4 and DMI

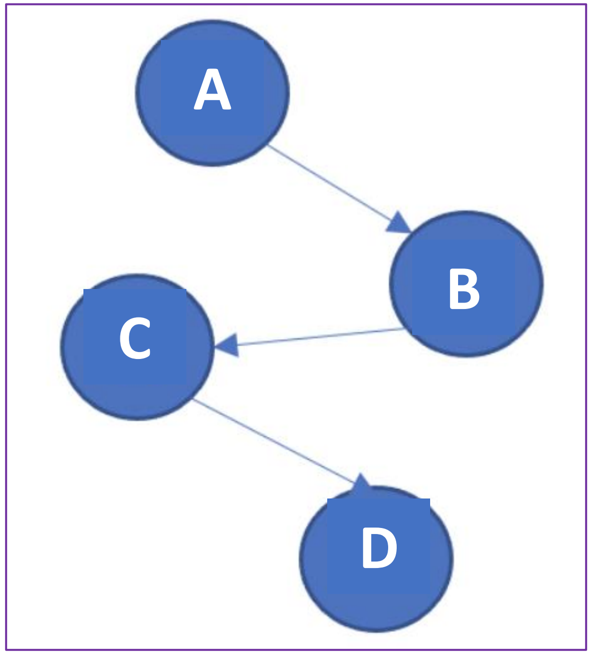
Threshold: Partial Correlation value

How to Quantify the Similarity between Model and observational DAG?

- **Step 1: Convert DAG in Adjacency Matrix for both Obs. and Model.**
- **Step 2: Estimate Confusion Matrix from two Adjacency Matrix.**
- **Step 3: Calculate a score (F1?).**

- **Step 1**

DAG

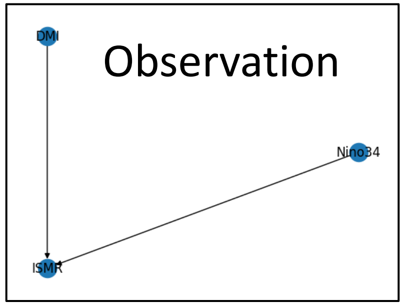


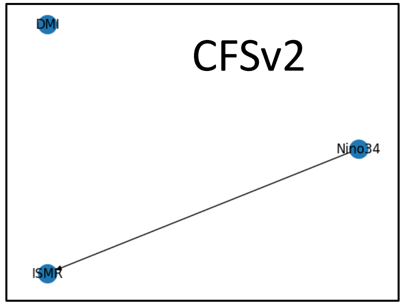
Adjacency Matrix

	A	B	C	D
A	0	1	0	0
B	0	0	1	0
C	0	0	0	1
D	0	0	0	0

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge from j to } \mathbf{i} \\ 0 & \text{otherwise} \end{cases}$$

Quantify the Similarity between Model and observational DAG

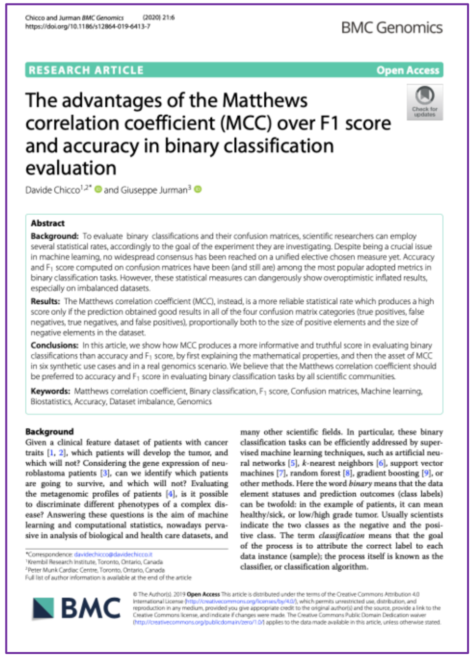




Adjacency Matrix

	Nino3.4 Index	Dipole Mode Index	AISMR
Nino3.4 Index	0	0	1
Dipole Mode Index	0	0	1
AISMR	0	0	0

	Nino3.4 Index	Dipole Mode Index	AISMR
Nino3.4 Index	0	0	1
Dipole Mode Index	0	0	0
AISMR	0	0	0


Chicco and Jurman *BMC Genomics* (2020) 21:6
https://doi.org/10.1186/s12864-019-6413-7
BMC Genomics
RESEARCH ARTICLE Open Access

The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation

Davide Chicco[1,2*] and Giuseppe Jurman[3]

Abstract

Background: To evaluate binary classifications and their confusion matrices, scientific researchers can employ several statistical rates, accordingly to the goal of the experiment they are investigating. Despite being a crucial issue in machine learning, no widespread consensus has been reached on a unified elective chosen measure yet. Accuracy and F_1 score computed on confusion matrices have been (and still are) among the most popular adopted metrics in binary classification tasks. However, these statistical measures can dangerously show overoptimistic inflated results, especially on imbalanced datasets.

Results: The Matthews correlation coefficient (MCC), instead, is a more reliable statistical rate which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), proportionally both to the size of positive elements and the size of negative elements in the dataset.

Conclusions: In this article, we show how MCC produces a more informative and truthful score in evaluating binary classifications than accuracy and F_1 score, by first explaining the mathematical properties, and then the asset of MCC in six synthetic use cases and in a real genomics scenario. We believe that the Matthews correlation coefficient should be preferred to accuracy and F_1 score in evaluating binary classification tasks by all scientific communities.

Keywords: Matthews correlation coefficient, Binary classification, F_1 score, Confusion matrices, Machine learning, Biostatistics, Accuracy, Dataset imbalance, Genomics

Background

Given a clinical feature dataset of patients with cancer traits [1, 2], which patients will develop the tumor, and which will not? Considering the gene expression of neuroblastoma patients [3], can we identify which patients are going to survive, and which will not? Evaluating the metagenomic profiles of patients [4], is it possible to discriminate different phenotypes of a complex disease? Answering these questions is the aim of machine learning and computational statistics, nowadays pervasive in analysis of biological and health care datasets, and many other scientific fields. In particular, these binary classification tasks can be efficiently addressed by supervised machine learning techniques, such as artificial neural networks [5], *k*-nearest neighbors [6], support vector machines [7], random forest [8], gradient boosting [9], or other methods. Here the word *binary* means that the data element statuses and prediction outcomes (class labels) can be twofold: in the example of patients, it can mean healthy/sick, or low/high grade tumor. Usually scientists indicate the two classes as the negative and the positive class. The term *classification* means that the goal of the process is to attribute the correct label to each data instance (sample); the process itself is known as the classifier, or classification algorithm.

*Correspondence: davidechicco@davidechicco.it
[1]Krembil Research Institute, Toronto, Ontario, Canada
[2]Peter Munk Cardiac Centre, Toronto, Ontario, Canada
Full list of author information is available at the end of the article

Confusion Matrix

	Positive	Negative
Positive	1 (TP)	0 (FP)
Negative	1 (FN)	7 (TN)

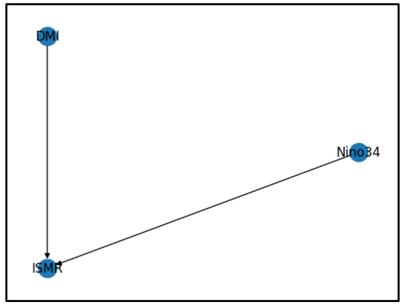
TP: True Positives
TN: True Negatives.
FP: False Positives.
FN: False Negatives.

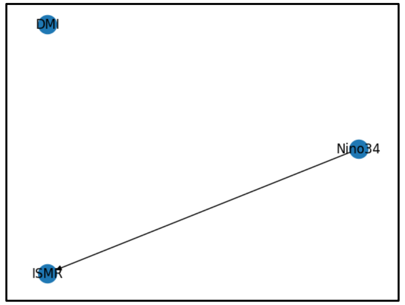
F1 Score: 0.67
MCC Score: 0.67

$$F1\ \text{score} = \frac{2TP}{2TP + FN + FP}$$

$$MCC = \frac{TN \times TP - FP \times FN}{\sqrt{(TN + FN)(FP + TP)(TN + FP)(FN + TP)}}$$

Quantify the Similarity between Model and observational DAG





Adjacency Matrix

	Nino3.4 Index	Dipole Mode Index	AISMR
Nino3.4 Index	0	0	1
Dipole Mode Index	0	0	1
AISMR	0	0	0

	Nino3.4 Index	Dipole Mode Index	AISMR
Nino3.4 Index	0	0	1
Dipole Mode Index	0	0	0
AISMR	0	0	0

Confusion Matrix

	Positive	Negative
Positive	1 (TP)	0 (FP)
Negative	1 (FN)	2 (TN)

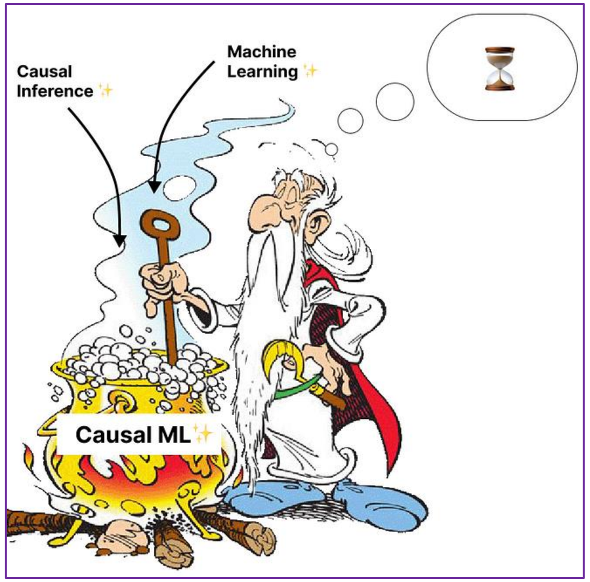
*Remove all irrelevant TN

F1 Score: 0.67
MCC Score: 0.57

F1 ignores the True Negatives and thus is misleading for unbalanced classes

Concluding Remarks

- Teleconnection is the most important factor for the process-oriented model diagnostic for seasonal forecast.
- Linear Correlation is de facto major for such practice.
- Proposing "Causal Structure Learning" for model evaluation as a type of process-oriented framework. Causality deals with understanding the cause and effect between different fields
- Causal discovery graphs from observation and dynamical model shows physical pathways of interaction.
- Quantification of similarity between causal discovery graphs of dynamical model and observations provides a causal-metric to assess the fidelity of dynamical models.



"CausalML combines techniques from machine learning and causal inference to understand and model causal relationships in data."

Thank you!

dr.nachiketaacharya@gmail.com