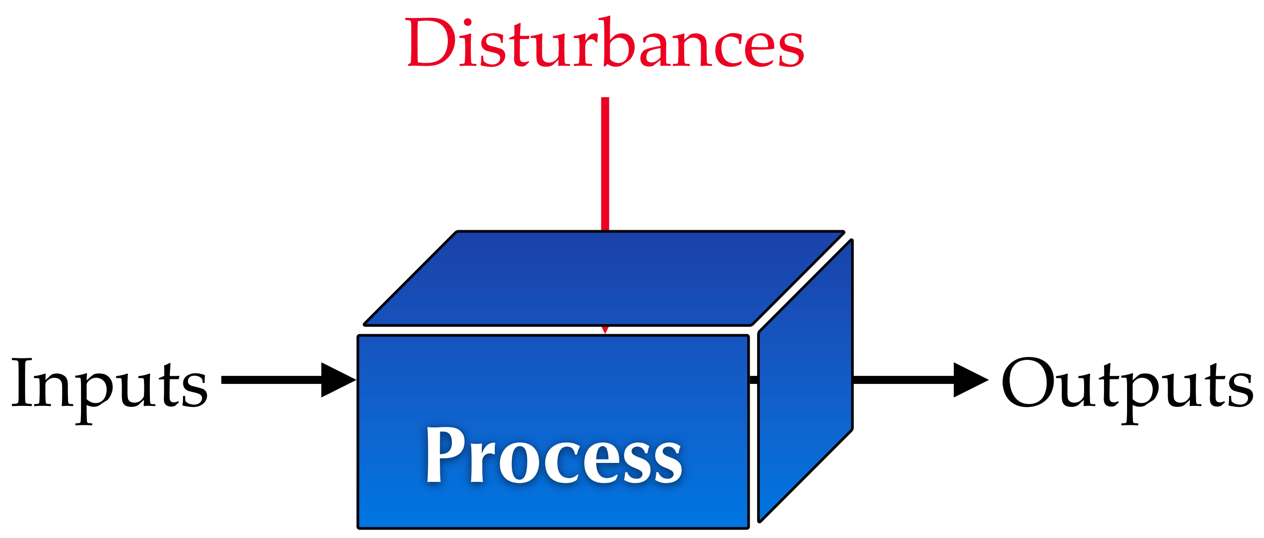
Causal (Climate) Networks - Overview, Reconstruction & Applications

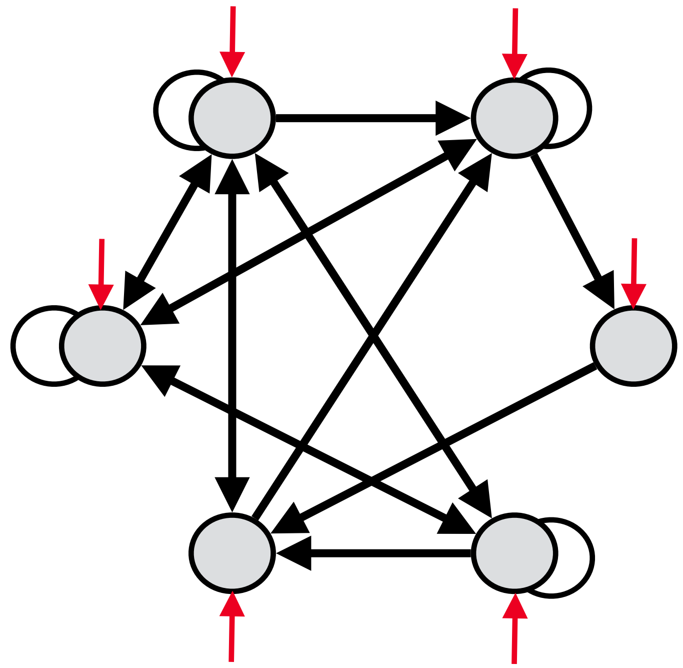
Arun K. Tangirala
Department of Chemical Engineering, IIT Tirupati
Systemic Intelligent Data-Driven and Network-Theoretic Analysis (SIDḌANTA) Lab

On-Lien
Department of DS & AI (Wadhwani School of DS & AI) & Dept. of Chemical Engineering, IITM

A Paradigm Shift

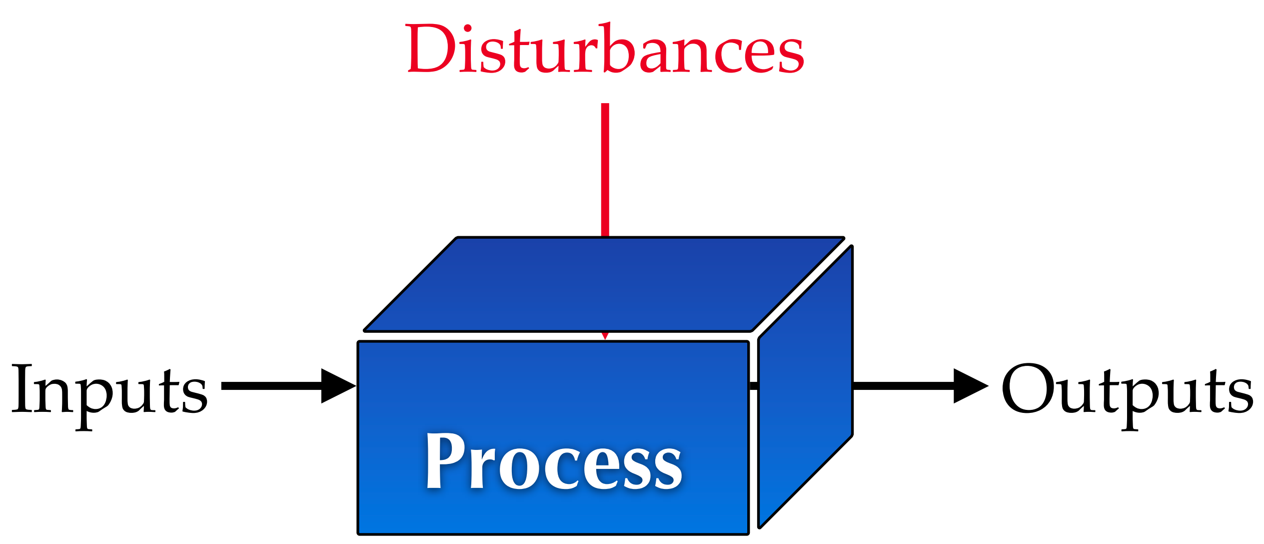


Classical systems-theory representation

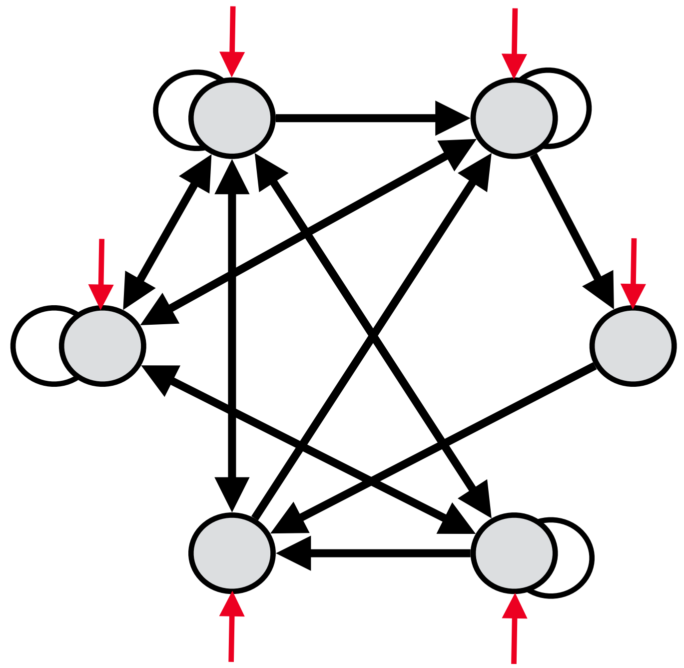


Network representation

A Paradigm Shift



Classical systems-theory representation

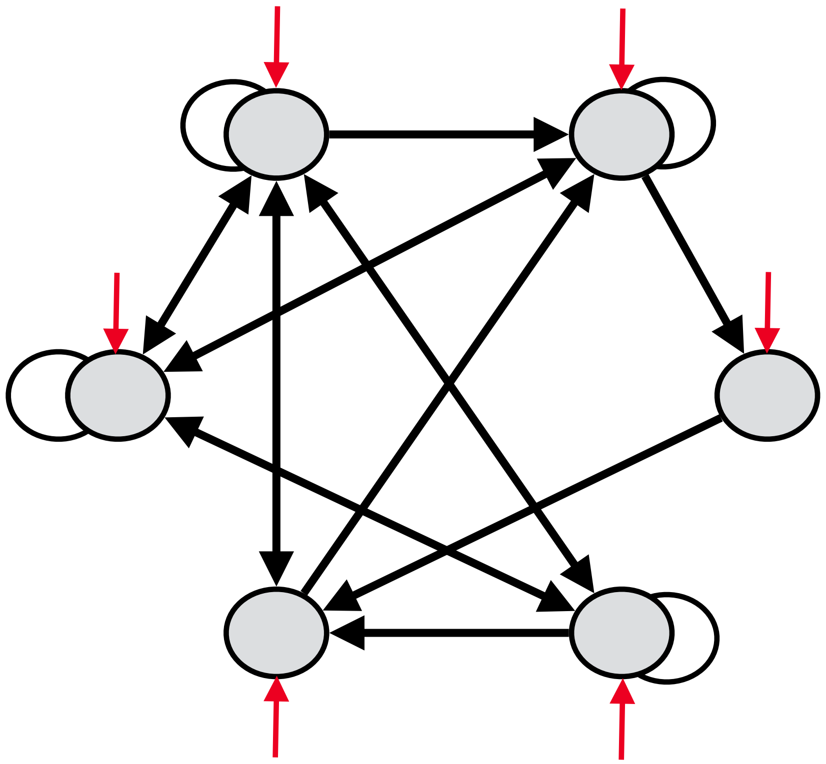


Network representation

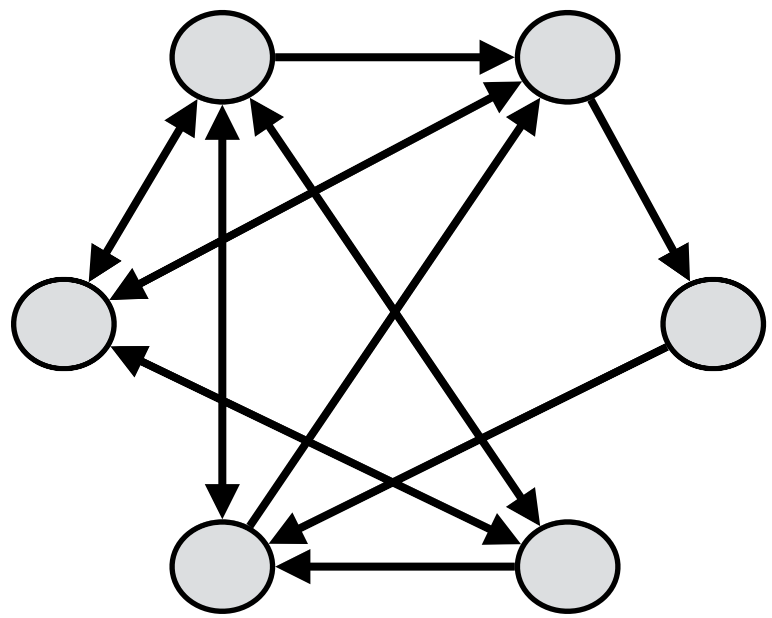
Benefits of graphical models:

- Can handle broader class of systems and situations
- Offers new perspectives and paradigms for analysis of processes
- Many applications require “qualitative” characterisation rather than precise quantitative analysis
- Complex systems are not necessarily equation-friendly - graphical models encode process characteristics in an operator-friendly and effective manner.
- Network models reveal certain universality properties across processes in different domains

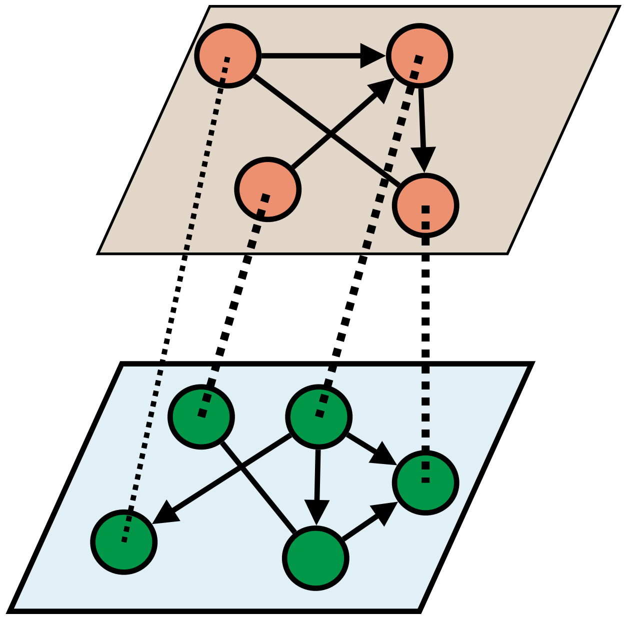
Three Broad Classes of N/W Representations



Complex Network
(on Time-Series / Signals)



Networked System



Multilayer Networks

- Nodes are signals / variables
- Edges encode directed / undirected associations
- Self-loops (past effects) and endogenous excitation

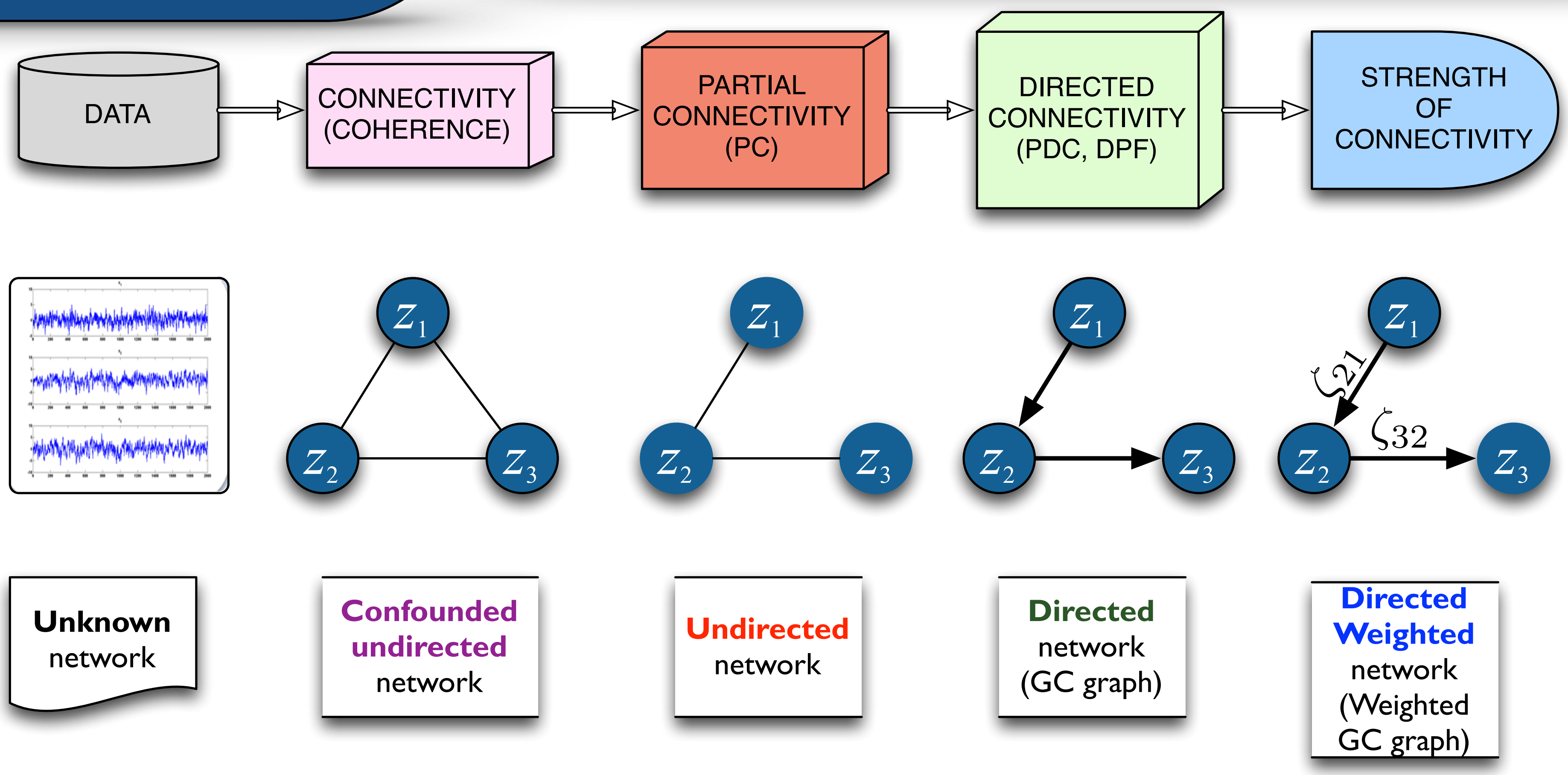
- Nodes are systems
- Edges encode directed / undirected connections
- Network representation of signal-flow graphs

- Several processes have layers of cause & effects
- Each layer can interact with nodes in another
- Ex: Climate Network with Supply Network

Complex Networks in Dynamical Systems

Complex Networks

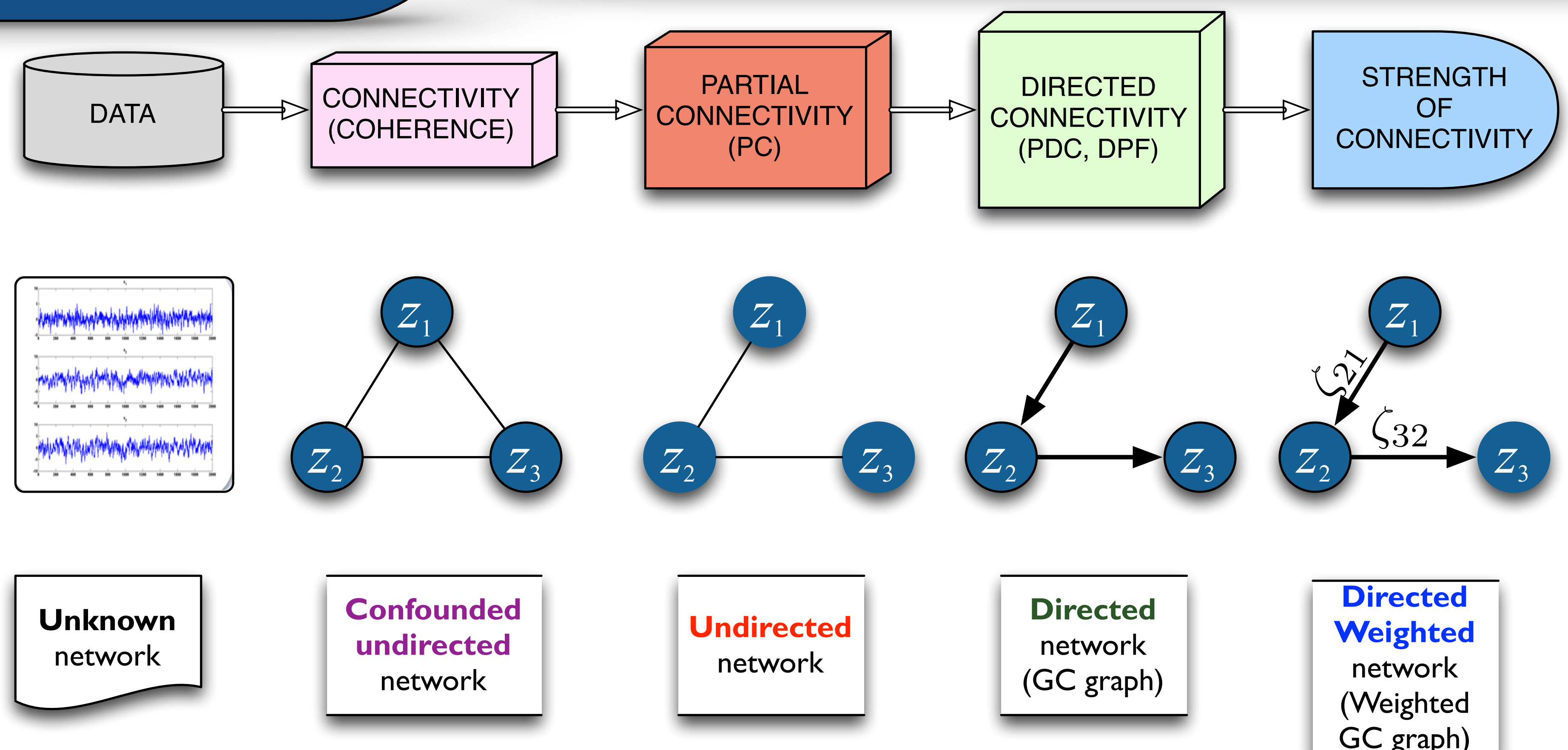
- Reconstruction of causal dynamical networks
- Quantification of interactions along pathways
- Identifying deep causal connections, hubs, etc.

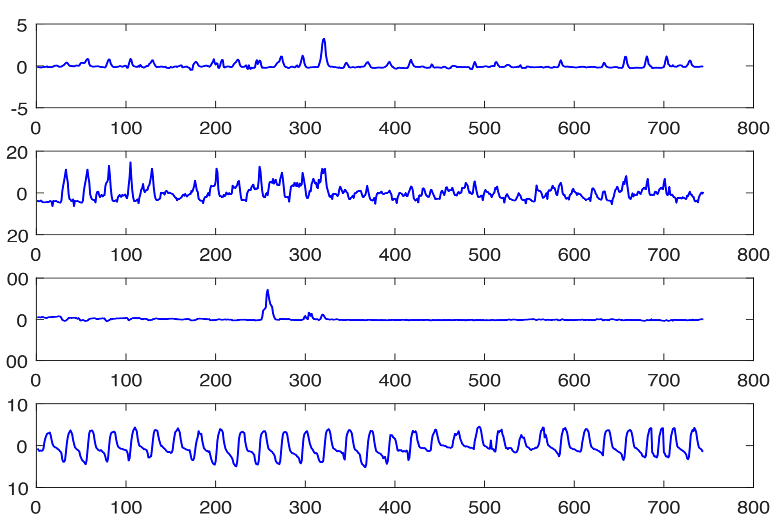


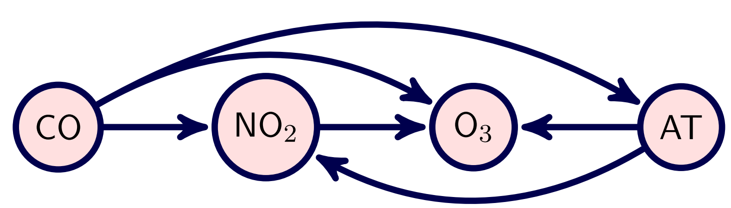
Complex Networks in Dynamical Systems

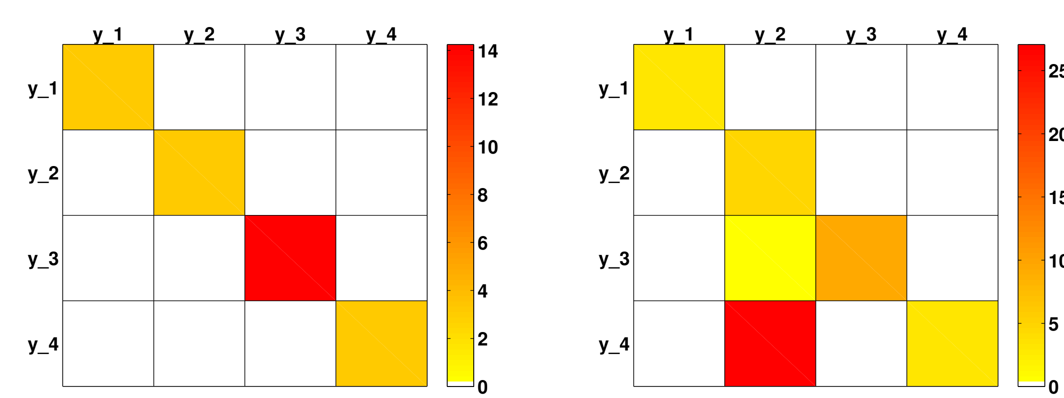
Complex Networks

- Reconstruction of causal dynamical networks
- Quantification of interactions along pathways
- Identifying deep causal connections, hubs, etc.









Comparing interaction strengths in control loops

Data-Driven Networks

All network reconstruction methods essentially
(i) define **vertex attributes** and propose **rules for edges**
(ii) **estimate Adjacency or Laplacian matrix** using suitable algorithms

Determining structure from data has three important advantages:

1. Can calculate **strengths of connectivity:**
   1. Knowing the intensity and characteristics of a coupling can only be determined experimentally
   2. A data-driven approach can give us the **effective connectivity.**
2. Can account for **stochastic effects and uncertainties**
3. Facilitates **inclusion of partial knowledge**

Process and Data Aspects

Development of graphs (networks) from data requires

1. *A choice of **measure** (correlation, causal, recurrence, visibility, etc.)*
2. ***Framework** and scope: predictive, probabilistic, bivariate, multivariate, etc.*
3. *A method of **inference** (non-parametric, parametric)*
4. *A set of **rules for constructing** the associated network / graph*

Process and Data Aspects

Development of graphs (networks) from data requires

1. *A choice of **measure** (correlation, causal, recurrence, visibility, etc.)*
2. ***Framework** and scope: predictive, probabilistic, bivariate, multivariate, etc.*
3. *A method of **inference** (non-parametric, parametric)*
4. *A set of **rules for constructing** the associated network / graph*

Challenges

Process

(P_1) linear / non-linear

(P_2) time-invariant / time-variant

(P_3) deterministic / stochastic

(P_4) confounding, direct / indirect

(P_5) uniqueness of network representation

Data

(D_1) measurement errors

(D_2) small sample sizes

(D_3) "minimal" excitation

(D_4) missing observations

NETWORKS ON TIME-SERIES

Networks differ primarily with respect to the way nodes and edges are characterised.

Network	Vertex	Edge	Directionali	Uni / Multi
Dynamic association n/ws				
Correlation networks	Signal / state	Correlation coefficient	Undirected	Multivariable
Causal networks	Signal / state	Causal measure	Directed	Multivariable
Proximity networks				
Cycle networks	Cycle	Corr. between cycles	Undirected	
Recurrence networks		*Recurrence measures*		
k-NN networks	State (vector)	Distance b/w states	Directed	Univariate
adaptive NN networks	State (vector)	Distance b/w states	Undirected	Univariate
ε-recurrence networks	State (vector)	Distance b/w states	Undirected	Univariate
Visibility graphs				
Natural / Horizontal VG	Time instant	Visibility	Undirected	Univariate
Directed / Weighted VG	Time instant	Directed visibility	Directed	Univariate
Transition networks				
Threshold-based n/ws	Phase-space ...	Temporal succession	Directed	Univariate
Ordinal pattern n/ws	Ordinal patterns	Temporal succession	Directed	Univariate

Y. Zou et al. (2019), Physics Reports 787, 1–97; V.F. Silva et al (2021), WIREs Data Mining & Knowledge Discovery, doi: 10.1002/widm.1404; A.K. Tangirala (2021), Internal Report, IITM

APPLICATIONS

Visibility Graphs

- Detecting transitions to chaos
- Disease onset detection from biomedical data
- Fractality of energy dissipation in turbulent flows
- Analysis of seismic and volatility behaviour, etc.
- …

Recurrence Networks

- Detecting transitions to chaos.
- Characterising phase-space trajectories in NL systems
- Analysis of financial time-series
- …

Causal Networks

- Process topology reconstruction
- Root-cause diagnosis
- Interaction quantification and disturbance propagation
- Process network decomposition
- Effective connectivity in neuronal networks
- Reconstruction of spatio-temporal climate networks
- …

- Other applications include forecasting of non-linear / symbolic time-series
- Graphical representations of signals have given birth to graph signal processing that extends traditional signal processing to graphs including convolution, filtering, generating low-dimensional approximations, etc.

Causality

Causality is a physical concept based on the cause-effect relationship

Three ***temporal*** forms of causality

1. **Strict / Lagged:** $z_2[k] = a_{21}^{(l)} z_1[k-l]$
2. **Instantaneous:** $z_2[k] = a_{21}^{(0)} z_1[k] + a_{21}^{(l)} z_1[k-l]$
3. **Feedback / Bidirectional:**

   $z_2[k] = a_{21}^{(l_1)} z_1[k-l_1]$

   $z_1[k] = a_{12}^{(l_2)} z_2[k-l_2]$

CAUSALITY

Causality is a physical concept based on the cause-effect relationship

Three ***temporal*** forms of causality

1. **Strict / Lagged:** $z_2[k] = a_{21}^{(l)} z_1[k-l]$
2. **Instantaneous:** $z_2[k] = a_{21}^{(0)} z_1[k] + a_{21}^{(l)} z_1[k-l]$
3. **Feedback / Bidirectional:**

   $z_2[k] = a_{21}^{(l_1)} z_1[k-l_1]$

   $z_1[k] = a_{12}^{(l_2)} z_2[k-l_2]$

Two ***variate*** forms of causality

1. **Bivariate (pairwise) causality:** Causal relation between a ***pair of variables*** $z_i[k]$ and $z_j[k]$
2. **Multivariate causality:** Causal relation between a pair of variables $z_i[k]$ and $z_j[k]$ **after discounting** for effects of all other variables

CAUSALITY

Causality is a physical concept based on the cause-effect relationship

Three ***temporal*** forms of causality

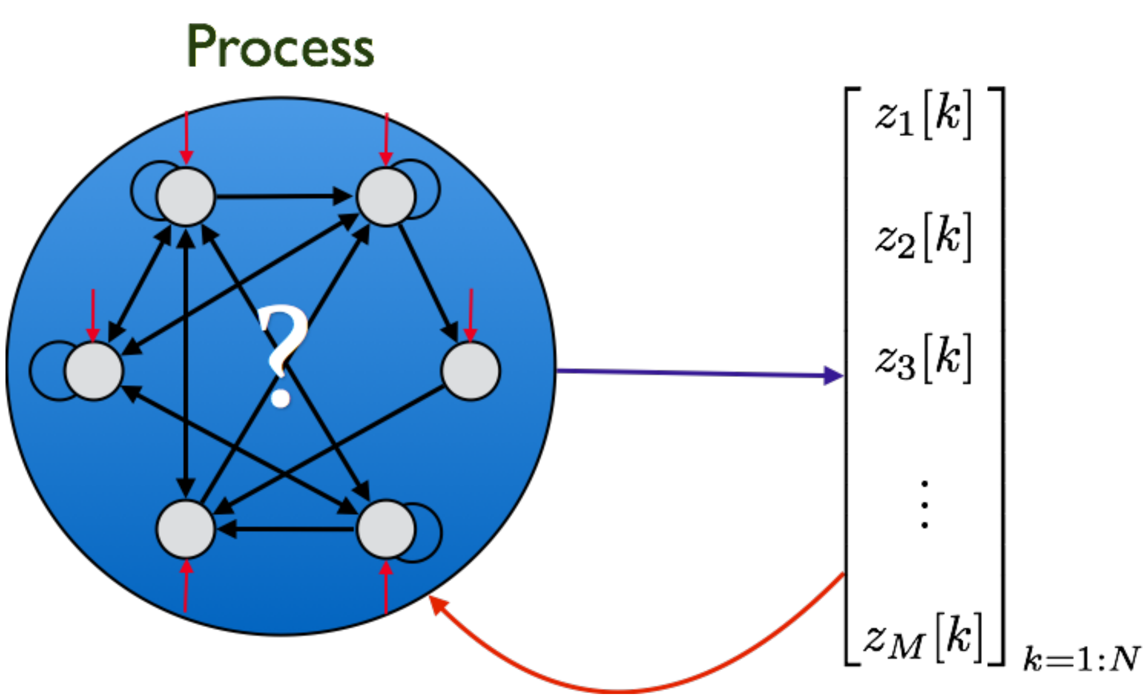
1. **Strict / Lagged:** $z_2[k] = a_{21}^{(l)} z_1[k-l]$
2. **Instantaneous:** $z_2[k] = a_{21}^{(0)} z_1[k] + a_{21}^{(l)} z_1[k-l]$
3. **Feedback / Bidirectional:**

$$z_2[k] = a_{21}^{(l_1)} z_1[k-l_1]$$
$$z_1[k] = a_{12}^{(l_2)} z_2[k-l_2]$$

Two ***variate*** forms of causality

1. **Bivariate (pairwise) causality:** Causal relation between a ***pair of variables*** $z_i[k]$ and $z_j[k]$
2. **Multivariate causality:** Causal relation between a pair of variables $z_i[k]$ and $z_j[k]$ **after discounting** for effects of all other variables

How do we detect all forms of causality from data?



Need a mathematical definition of causality and a statistical inferencing method

Prominent Definitions of Causality - I

Type of Causality	Description	Features / Issues
Granger (wide-sense)	*Signal X G-causes another signal Y if its past aids in "improving" **prediction** of Y*	*Probabilistic form, but predictive causality operationally; stochastic framework; temporal precedence; detects direct causality popular; has several limitations and issues*
Sims	Signal X does not S-cause another signal Y if Y is independent of the past of X.	Stochastic framework; detects *total* causality; similar issues as that of GC
Interventional	A random variable X I-causes Y if $\Pr(Y \mid do(X=x)) \neq \Pr(Y)$	Probabilistic definition; based on physical influence; no extensions to signals are available; difficult to implement
Counterfactual	Quantify the influence of X on Y if X had taken a different value from the observed	Probabilistic definition; based on physical influence; highly difficult to implement

Prominent Definitions of Causality - II

Type of Causality	Description	Features / Issues
Cross-Mapping	*Signal X CCM-causes another signal Y if it can be* ***recovered*** *from Y "consistently" using different data sizes*	*Introduced for non-linear processes; deterministic signals; bivariate; cannot resolve confounding; recent extensions to multivariable processes*
Probabilistic Constraint	Signal / RV X-constraint causes another signal / RV Y if Y is conditionally independent of X	Information-theoretic / probabilistic definition; suitable for big data
Dynamical Constraint	*Signal X-constraint causes another signal Y if X and Y are tied together in a deterministic (dynamical) constraint*	*Introduced for linear deterministic systems; suitable for pair-wise and multivariable processes; PCA- and delay-based inferencing*
Functional-plus Noise-based	Signal / RV X-noise causes another signal / RV Y if Y is related to X with some additional noise	Essential to have noise with specific distributional properties along with the functional relations
Dynamic Bayesian network-based	Signal / RV X causes another signal / RV Y if the node for past of X is connected to the node for present Y	Data is fit onto a DBN using a score-based approach; suitable for stochastic signals; high on computational costs

Causal Discovery Methods - I

Description	Reference	Causality type
Geweke's Spectral Tests; GC Index; Granger-Wald Tests	Granger (1984), Geweke (1979, 1982); Lutkepohl (2005)	Bivariate GC in time-domain
Partial Directed Coherence (PDC) and its variants, extensions	Baccala & Sameshima (2001); Faes and Nollo (2010); Baccala et al (2014); Kannan and Tangirala (2014);	Multivariate GC in frequency domain
Direct Pathway Function (DPF) and its extensions	Gigi and Tangirala (2010); Kathari and Tangirala (2019); Agarwal and Tangirala (2016)	Multivariate GC in freq. domain with extensions to small and missing data
Partial Directed Correlation	Eichler (2005); Lutkephol (2005)	Multivariate GC in time-domain
Directed Transfer Entropy (DTE)	Duan et al (2013)	Multivariate Transfer Entropy-based Causality
Directed Transfer Function (DTF)	Kaminski and Blinowska (1991)	Sims Causality in frequency-domain

Causal Discovery Methods - II

Description	Reference	Type of causality
Embedding-based CCM method	Sugihara (2013); Nithya and Tangirala (2021)	CCM-based
Delay-based and DIPCA-based causal discovery	Kathari and Tangirala (2019); Kathari and Tangirala (2022)	Dynamical Constraint-based
Temporal Causal Discovery Framework (TCDF) using DL	Nauta et al (2019)	Instantaneous and Lagged causality
PC-Momentary Conditional Independence Algorithm (PCMCI)	Spirtes et al (2001); Runge et al (2019)	Probabilistic constraint-based lagged causality
Fast Causal Inference Algo. + Additive NL TS Model / SVAR	Chu and Glymour (2009); Malinsky and Spirtes (2018)	Probabilistic constraint-based inst. and lagged causality
Functional Causal / Structural Eqn. Models + TSM with independent noise (TiMINo)	Peters et al (2013)	Noise-based
DYNOTEARS	Pena et al (2005); Pamfil et al (2020); Sánchez-Romaro et al (2014) and others	DBN-based

A few reconstruct *summary causal graphs* while others reconstruct *window causal graphs*

Causality & Climate (Networks)

Granger Causality of Coupled Climate Processes: Ocean Feedback on the North Atlantic Oscillation

Timothy J. Mosedale, David B. Stephenson, Matthew Collins, and Terence C. Mills

Print Publication: 01 Apr 2006

DOI: https://doi.org/10.1175/JCLI3653.1

Page(s): 1182–1194

Causal feedbacks in climate change

Egbert H. van Nes, Marten Scheffer, Victor Brovkin, Timothy M. Lenton, Hao Ye, Ethan Deyle & George Sugihara

Nature Climate Change **5**, 445–448 (2015) | Cite this article

Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation

Marlene Kretschmer, Dim Coumou, Jonathan F. Donges, and Jakob Runge

Print Publication: 01 Jun 2016

DOI: https://doi.org/10.1175/JCLI-D-15-0654.1

Page(s): 4069–4081

Inferring causation from time series in Earth system sciences

Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Muñoz-Marí, Egbert H. van Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Schölkopf, Peter Spirtes, George Sugihara, Jie Sun, Kun Zhang & Jakob Zscheischler

Nature Communications **10**, Article number: 2553 (2019) | Cite this article

Geophysical Research Letters

Research Letter | Free Access

Detecting Climate Teleconnections With Granger Causality

Filipi N. Silva, Didier A. Vega-Oliveros, Xiaoran Yan, Alessandro Flammini, Filippo Menczer, Filippo Radicchi, Ben Kravitz, Santo Fortunato

First published: 31 August 2021 | https://doi.org/10.1029/2021GL094707 | Citations: 13

Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: a theory-guided causal effect network approach

Di Capua, G., Kretschmer, M., Donner, R. V., van den Hurk, B., Vellore, R., Krishnan, R., and Coumou, D.: Earth Syst. Dynam., 11, 17–34, https://doi.org/10.5194/esd-11-17-2020, 2020.

Networks of climate change: connecting causes and consequences

Petter Holme & Juan C. Rocha

Applied Network Science **8**, Article number: 10 (2023) | Cite this article

Nonlin. Processes Geophys., 31, 115–136, 2024
https://doi.org/10.5194/npg-31-115-2024

A comparison of two causal methods in the context of climate analyses

David Docquier[1], Giorgia Di Capua[2,3], Reik V. Donner[2,3], Carlos A. L. Pires[4], Amélie Simon[4,5], and Stéphane Vannitsem[1]

Wiener-Granger Causality

Wiener-Granger-Causality

Three axioms:

1. cause precedes its effect
2. cause contains unique information about the future values of its effect
3. All causal relationships remain constant in direction throughout time.

Wiener-Granger-Causality

Three axioms:

1. cause precedes its effect
2. cause contains unique information about the future values of its effect
3. All causal relationships remain constant in direction throughout time.

Introduce, $I^*[k]$: information set **up to** 'k', $I^\star_{-z_i}[k]$: information **excluding** z_i.

The series $\{z_i[k]\}$ does not (Granger) cause $\{z_j[k]\}$ if

$$z_j[k+1] \perp\!\!\!\perp \mathcal{I}^\star[k] \mid \mathcal{I}^\star_{-z_i}[k]$$

for all $k \in \mathbb{Z}$, else $\{z_i[k]\}$ is said to cause the series $\{z_j[k]\}$

WIENER-GRANGER-CAUSALITY

Three axioms:

1. cause precedes its effect
2. cause contains unique information about the future values of its effect
3. All causal relationships remain constant in direction throughout time.

Introduce, $I^*[k]$: information set **up to** 'k', $I^{\star}_{-z_i}[k]$: information **excluding** z_i.

The series $\{z_i[k]\}$ does not (Granger) cause $\{z_j[k]\}$ if

$$z_j[k+1] \perp\!\!\!\perp \mathcal{I}^{\star}[k] \mid \mathcal{I}^{\star}_{-z_i}[k]$$

for all $k \in \mathbb{Z}$, else $\{z_i[k]\}$ is said to cause the series $\{z_j[k]\}$

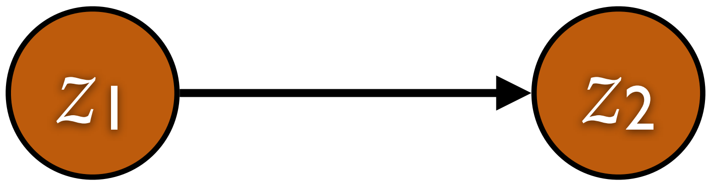
Alternative definition / viewpoint:

- Granger non-causal if and only if **conditional distribution** of $z_j[k]$ is unchanged whether $z_i[k]$ is included or not in the information set.
- Granger non-causal if both sets contain ``same" information about $z_j[k]$.

WGC: Prediction-based Definition

(Bivariate) Granger causality [Granger, 1969]

Suppose we wish to test if a random signal $z_1[k]$ Granger causes another random signal $z_2[k]$. Predict $z_2[k]$ using its own past and construct another forecast by incorporating past of $z_1[k]$. If the forecast is improved then $z_1[k]$ is said to "Granger cause" $z_2[k]$.



WGC: Prediction-based Definition

(Bivariate) Granger causality [Granger, 1969]

Suppose we wish to test if a random signal $z_1[k]$ Granger causes another random signal $z_2[k]$. Predict $z_2[k]$ using its own past and construct another forecast by incorporating past of $z_1[k]$. If the forecast is improved then $z_1[k]$ is said to "Granger cause" $z_2[k]$.

Example:

Construct two models

$$z_2[k] = az_2[k-1] + e_1[k]$$

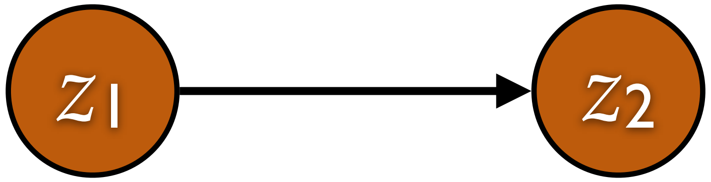
$$z_2[k] = bz_2[k-1] + cz_1[k-1] + e_2[k]$$

Compute corresponding prediction errors

$$\varepsilon_1 = z_2[k] - \hat{z}_2[k|z_2[k-1]]$$

$$\varepsilon_2 = z_2[k] - \hat{z}_2[k|\{z_2[k-1], z_1[k-1]\}]$$

If $\sigma^2_{\varepsilon_2} < \sigma^2_{\varepsilon_1}$, then $z_1[k]$ is said to "Granger cause" $z_2[k]$



WGC: Prediction-based Definition

(Bivariate) Granger causality [Granger, 1969]

Suppose we wish to test if a random signal $z_1[k]$ Granger causes another random signal $z_2[k]$. Predict $z_2[k]$ using its own past and construct another forecast by incorporating past of $z_1[k]$. If the forecast is improved then $z_1[k]$ is said to "Granger cause" $z_2[k]$.

Example:

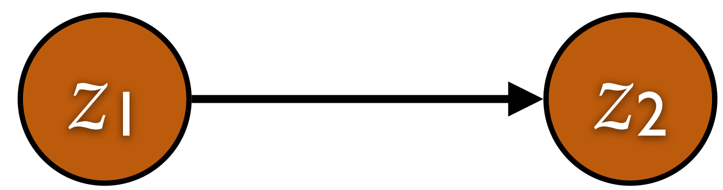
Construct two models

$$z_2[k] = az_2[k-1] + e_1[k]$$
$$z_2[k] = bz_2[k-1] + cz_1[k-1] + e_2[k]$$

Compute corresponding prediction errors

$$\varepsilon_1 = z_2[k] - \hat{z}_2[k|z_2[k-1]]$$
$$\varepsilon_2 = z_2[k] - \hat{z}_2[k|\{z_2[k-1], z_1[k-1]\}]$$

If $\sigma^2_{\varepsilon_2} < \sigma^2_{\varepsilon_1}$, then $z_1[k]$ is said to "Granger cause" $z_2[k]$



Extensions to multivariate, feedback and instantaneous causality are on similar lines.

WGC: Prediction-based Definition

(Bivariate) Granger causality [Granger, 1969]

Suppose we wish to test if a random signal $z_1[k]$ Granger causes another random signal $z_2[k]$. Predict $z_2[k]$ using its own past and construct another forecast by incorporating past of $z_1[k]$. If the forecast is improved then $z_1[k]$ is said to "Granger cause" $z_2[k]$.

Example:

Construct two models

$$z_2[k] = az_2[k-1] + e_1[k]$$

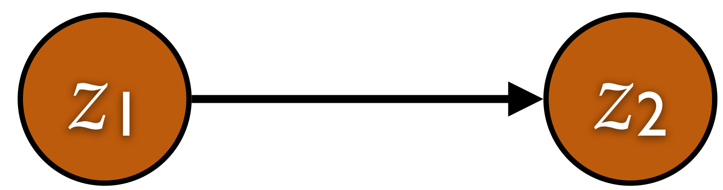
$$z_2[k] = bz_2[k-1] + cz_1[k-1] + e_2[k]$$

Compute corresponding prediction errors

$$\varepsilon_1 = z_2[k] - \hat{z}_2[k|z_2[k-1]]$$

$$\varepsilon_2 = z_2[k] - \hat{z}_2[k|\{z_2[k-1], z_1[k-1]\}]$$

If $\sigma^2_{\varepsilon_2} < \sigma^2_{\varepsilon_1}$, then $z_1[k]$ is said to "Granger cause" $z_2[k]$



Extensions to multivariate, feedback and instantaneous causality are on similar lines.

*The definition is **not tied** to a "method" of prediction!*

Parametric Implementation: VAR Models

- **VAR(p)** model for an M-dimensional process:

$$\mathbf{z}[k] = \sum_{r=1}^{P} \mathbf{A}_r \mathbf{z}[k-r] + \mathbf{e}[k], \qquad \mathbf{e}[k] \sim \text{GWN}(\mathbf{0}, \Sigma_{\mathbf{e}})$$

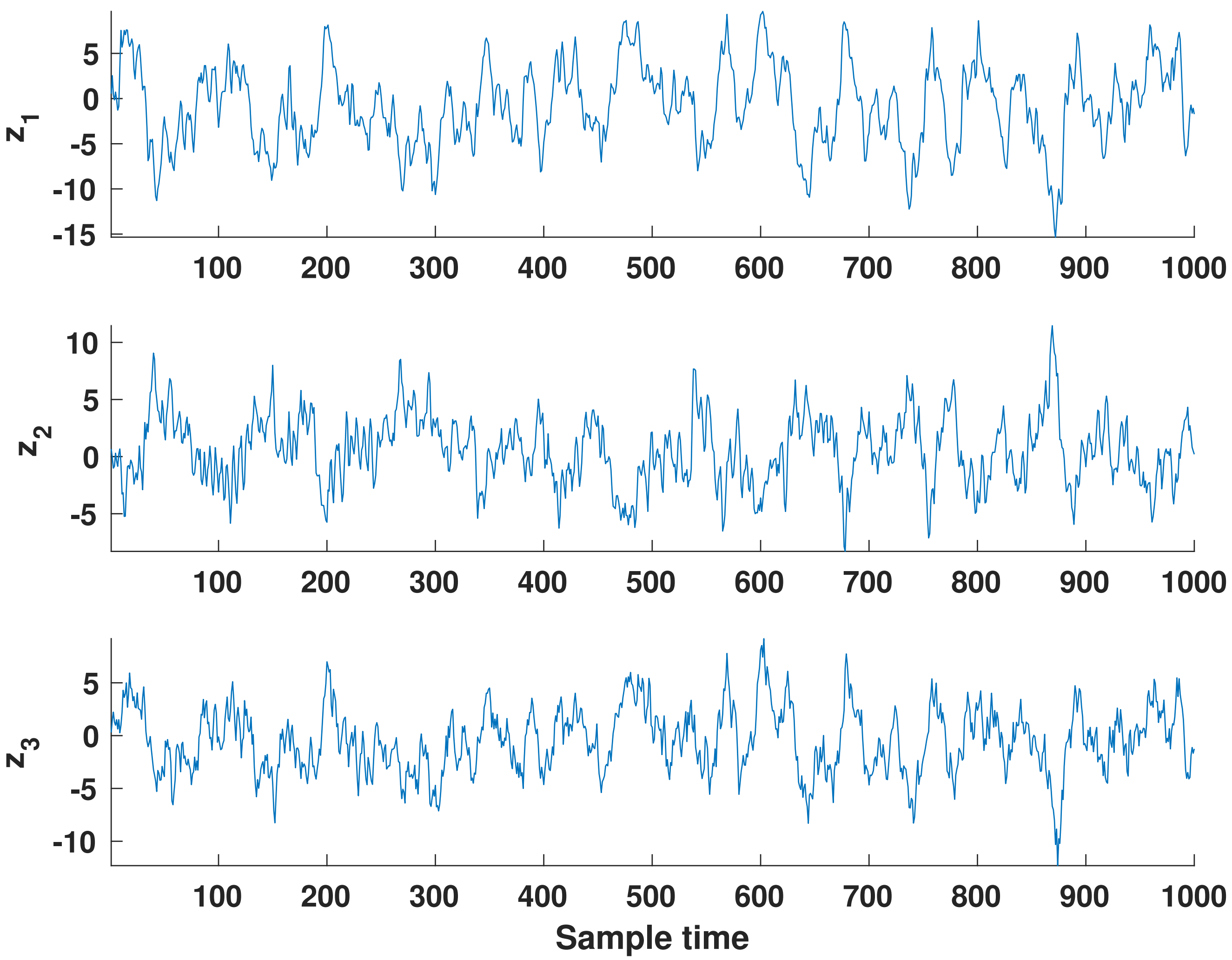
- Define a frequency-domain function: $\bar{\mathbf{A}}(\omega) = \mathbf{I} - \mathbf{A}(\omega) = \mathbf{I} - \sum_{r=1}^{p} \mathbf{A}_r e^{-j\omega r}$

Note: A test on the diagonal structure of $\Sigma_{\mathbf{e}}$ is necessary before using any GC tests!

- With the VAR model in hand, different tests for GC can be set up:
  - **Using matrices $\mathbf{A}_r$:** z_i does not GC z_j if and only if $a_{ji,r} = 0, \forall r = 1, \cdots, P$
  - **Partial directed correlation** (time-domain)
  - **Partial directed coherence** (freq. dom.): $\pi_{ij}(\omega) = \dfrac{\bar{a}_{ij}(\omega)}{\sqrt{\sum_{i=1}^{m} |\bar{a}_{ij}(\omega)|^2}}$
  - **Direct pathway function** (freq. dom.): $h_{D,ij}(\omega) = \dfrac{-\bar{a}_{ij}(\omega)\det(\bar{M}_{ji}(\omega))}{\det(\bar{A}(\omega))}; \quad \psi_{ij}(\omega) \triangleq \dfrac{h_{D,ij}(\omega)}{\sqrt{\sum_{i=1}^{M} |h_{D,ij(\omega)}|^2}}$

Simulation Case Study

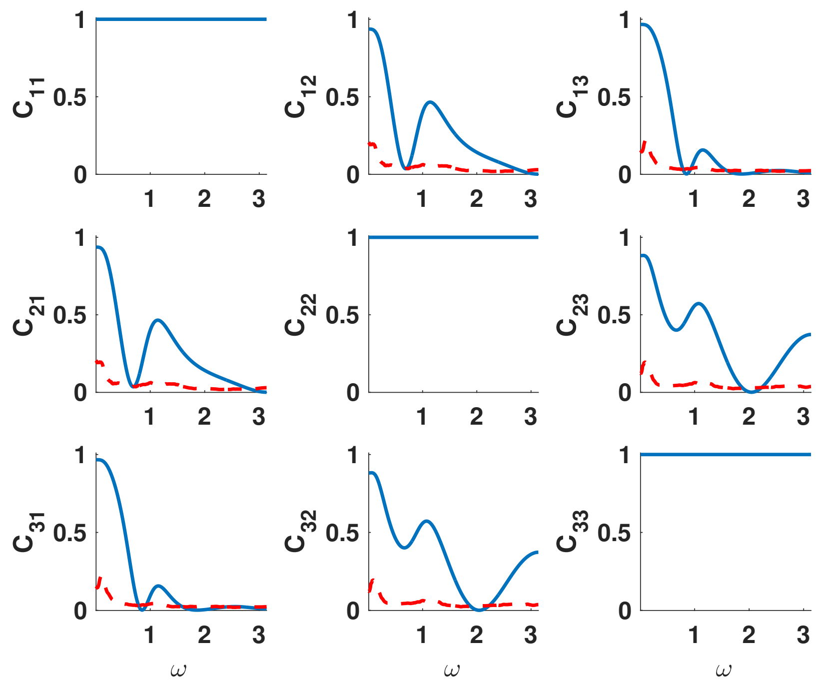
- N = 1000 observations of a 3-D stationary process



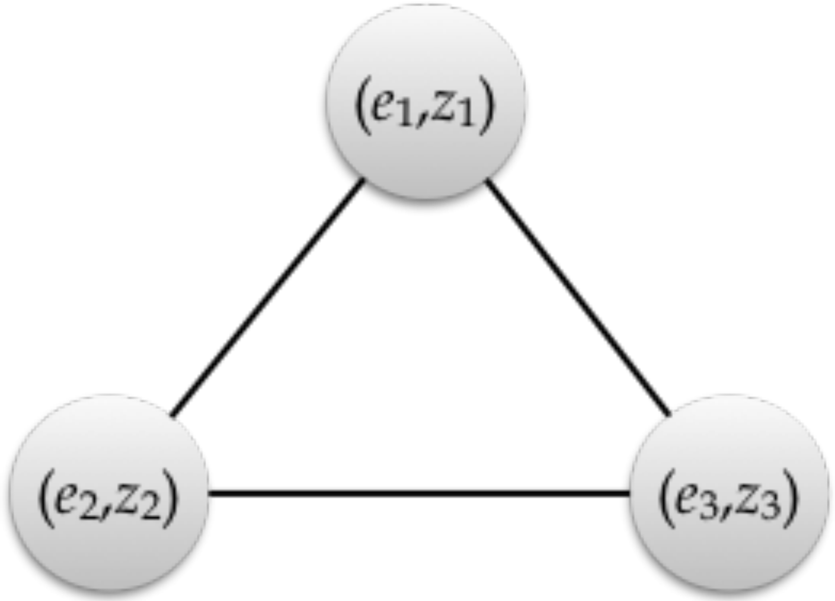
- **Goal:** Reconstruct a class of networks from correlation to the weighted causal network

Coherence and Conditioned (Partial) Networks

Coherence

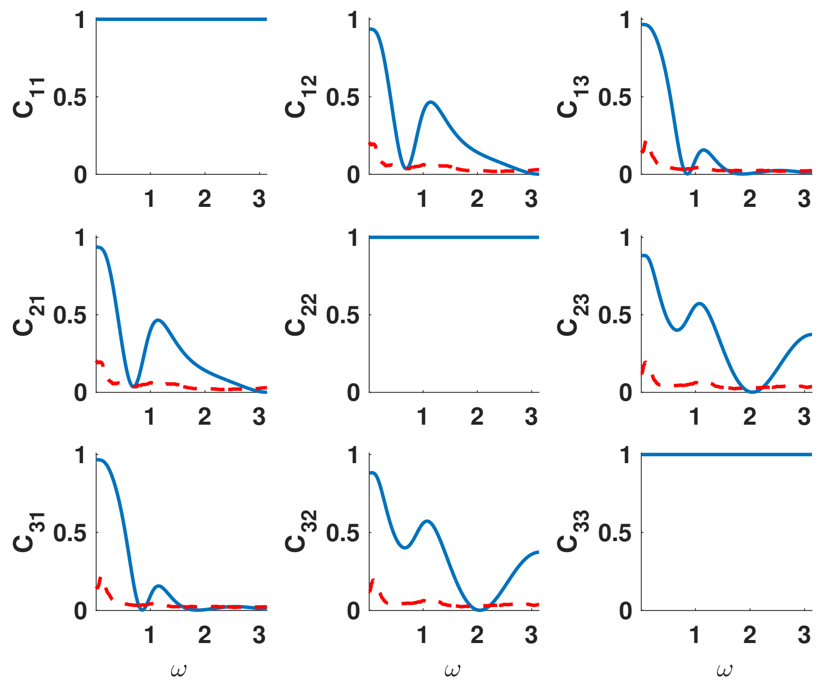


Undirected, confounded network

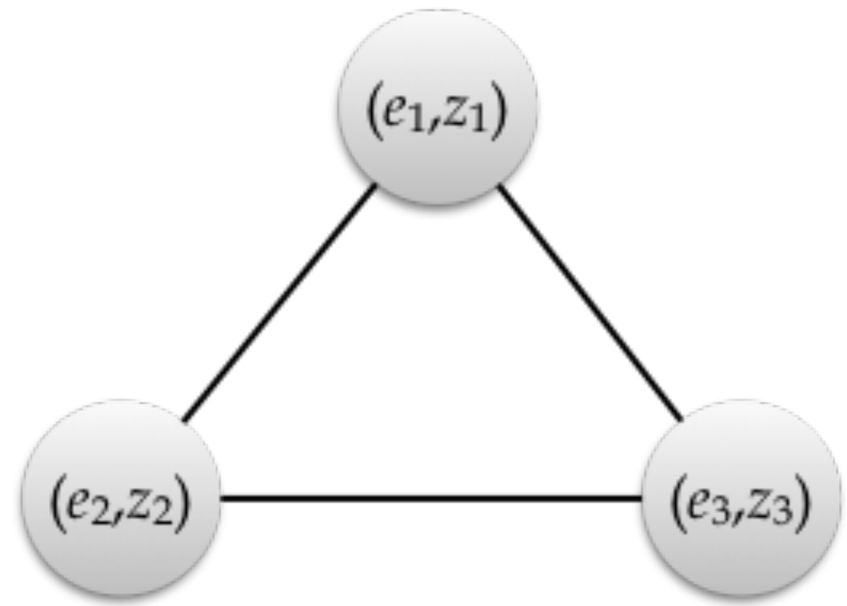


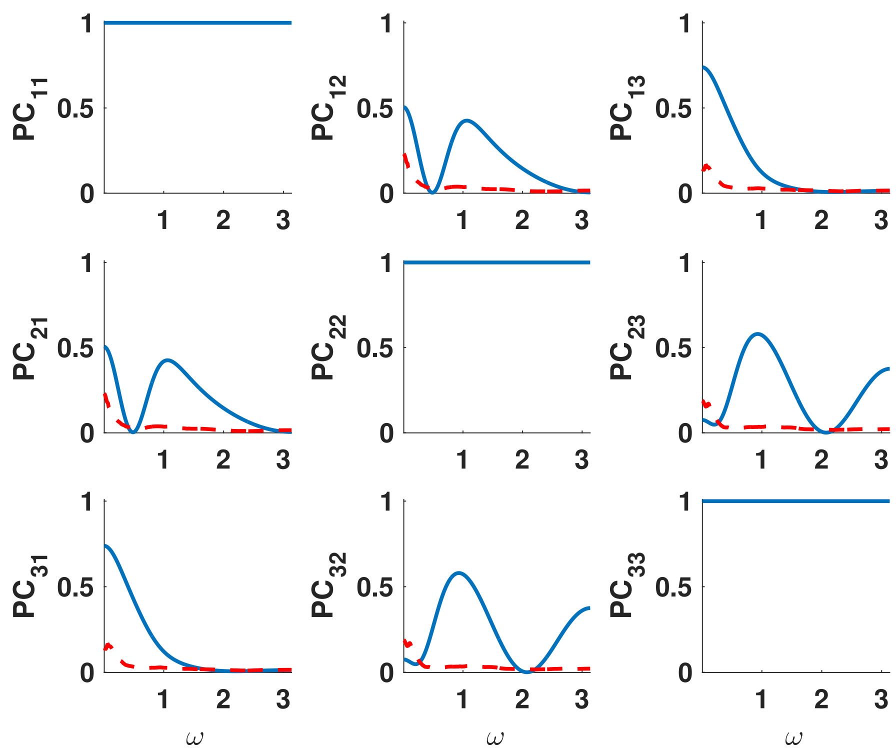
Coherence and Conditioned (Partial) Networks

Coherence

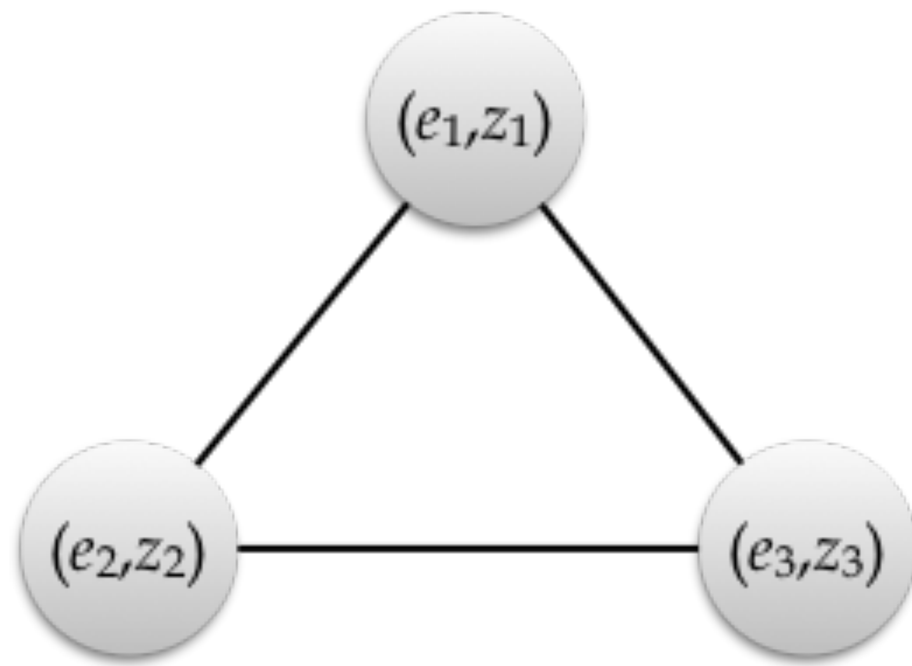


Undirected, confounded network





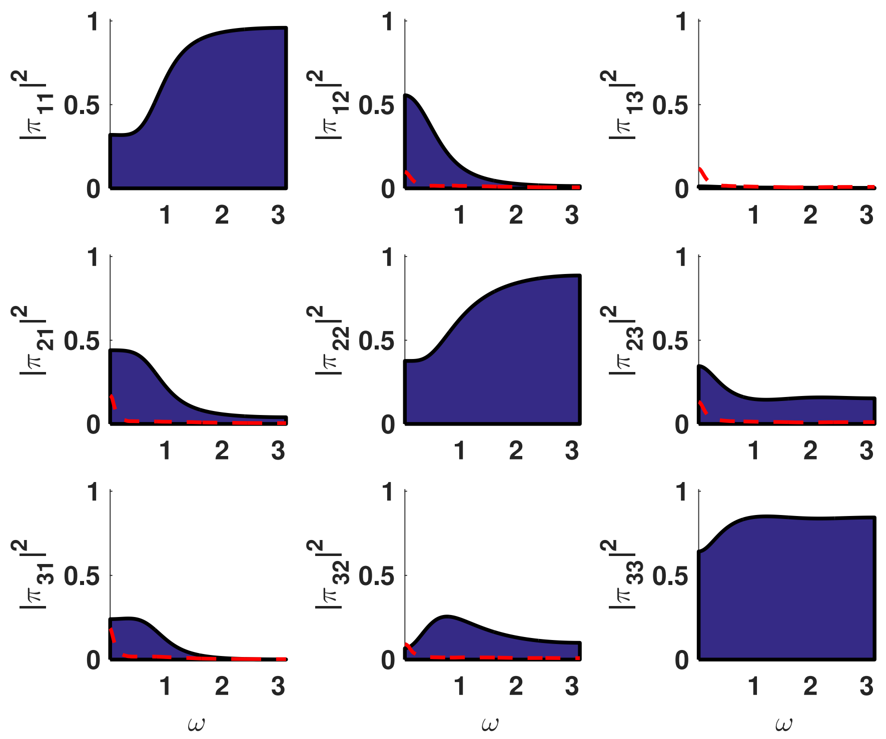
Partial Coherence

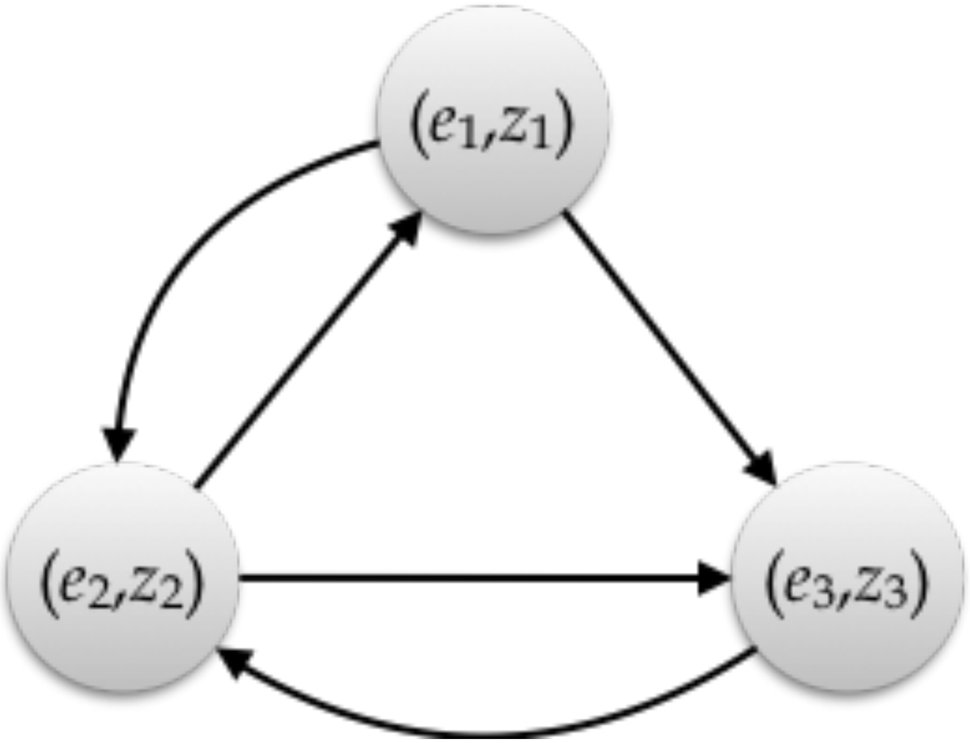


Undirected, direct network

Directed (Causal) Weighted Networks

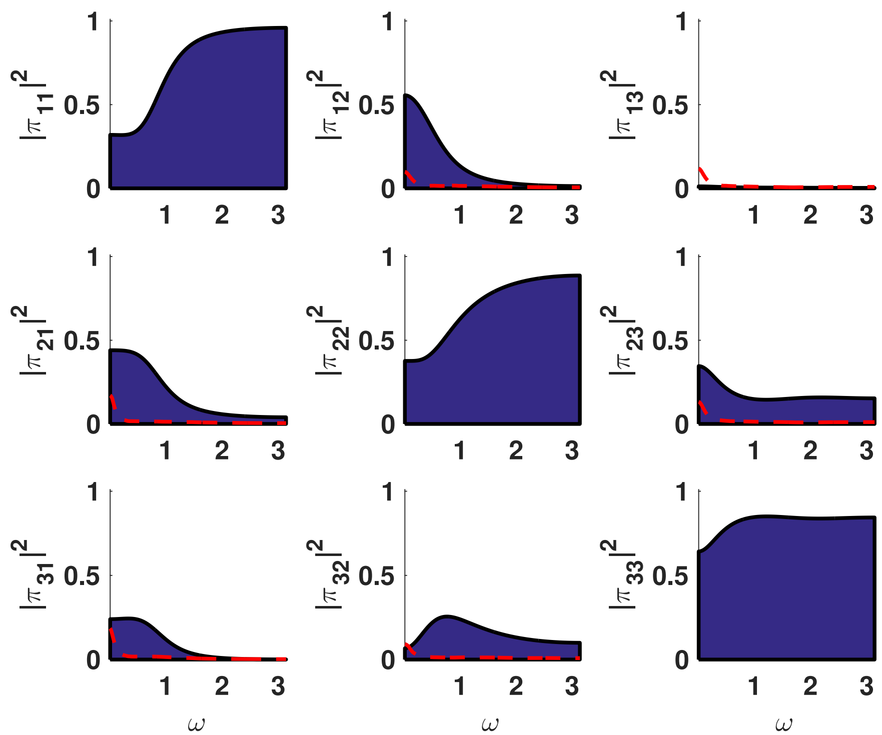
Partial directed coherence

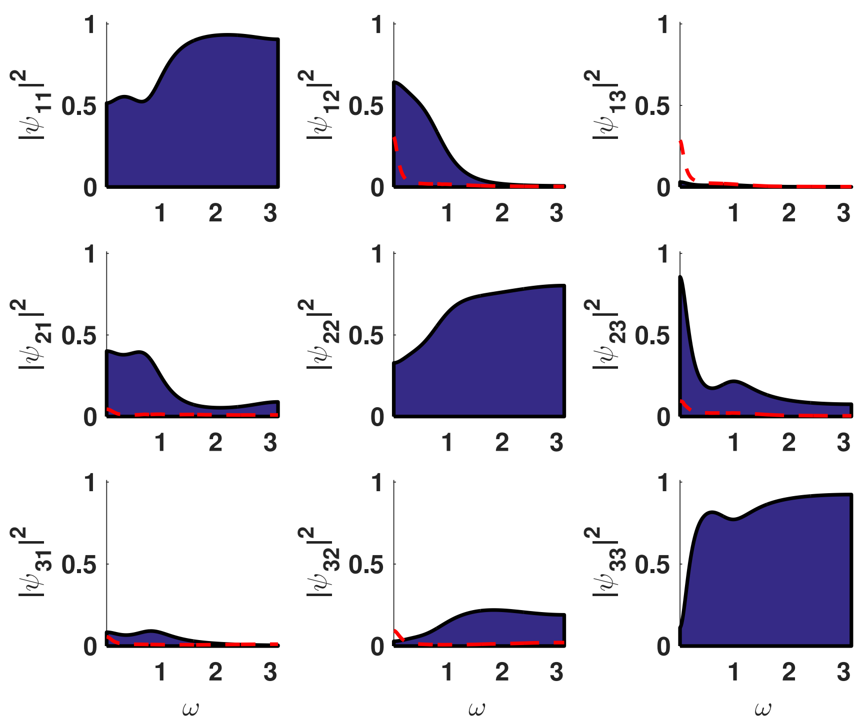




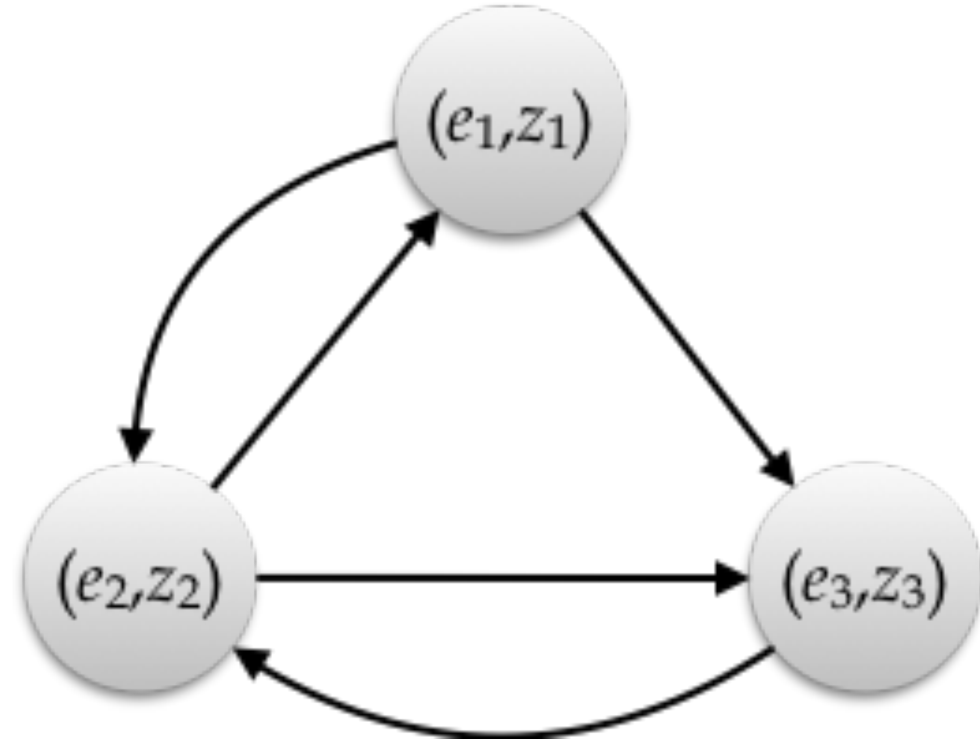
Directed (Causal) Weighted Networks

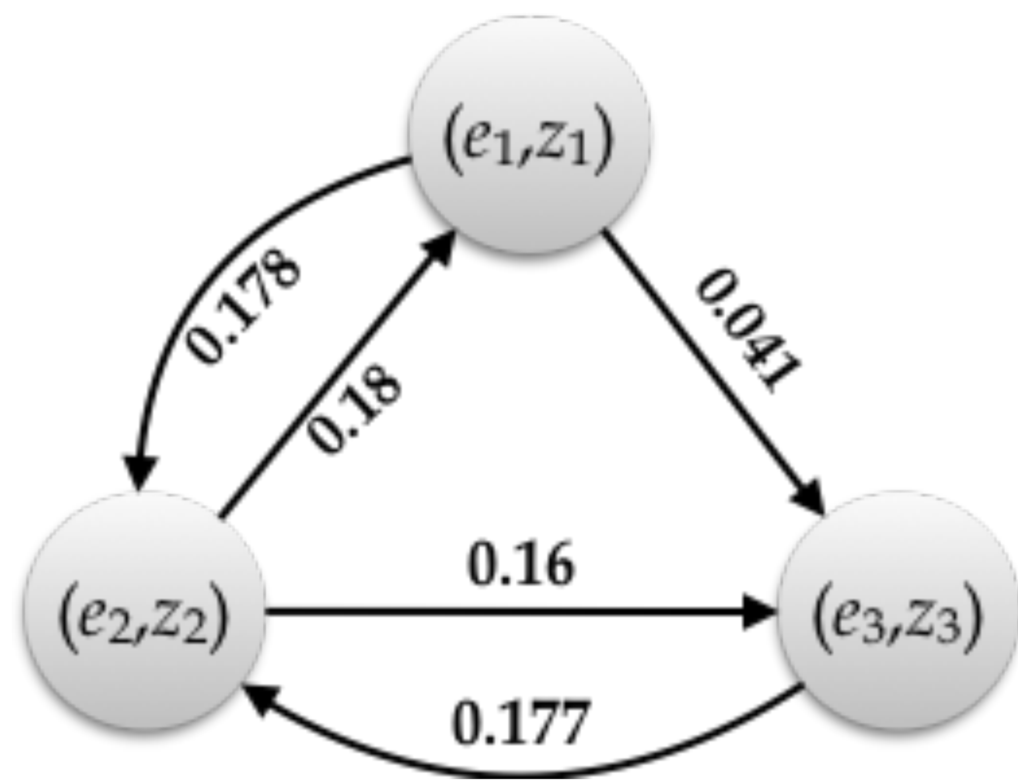
Partial directed coherence





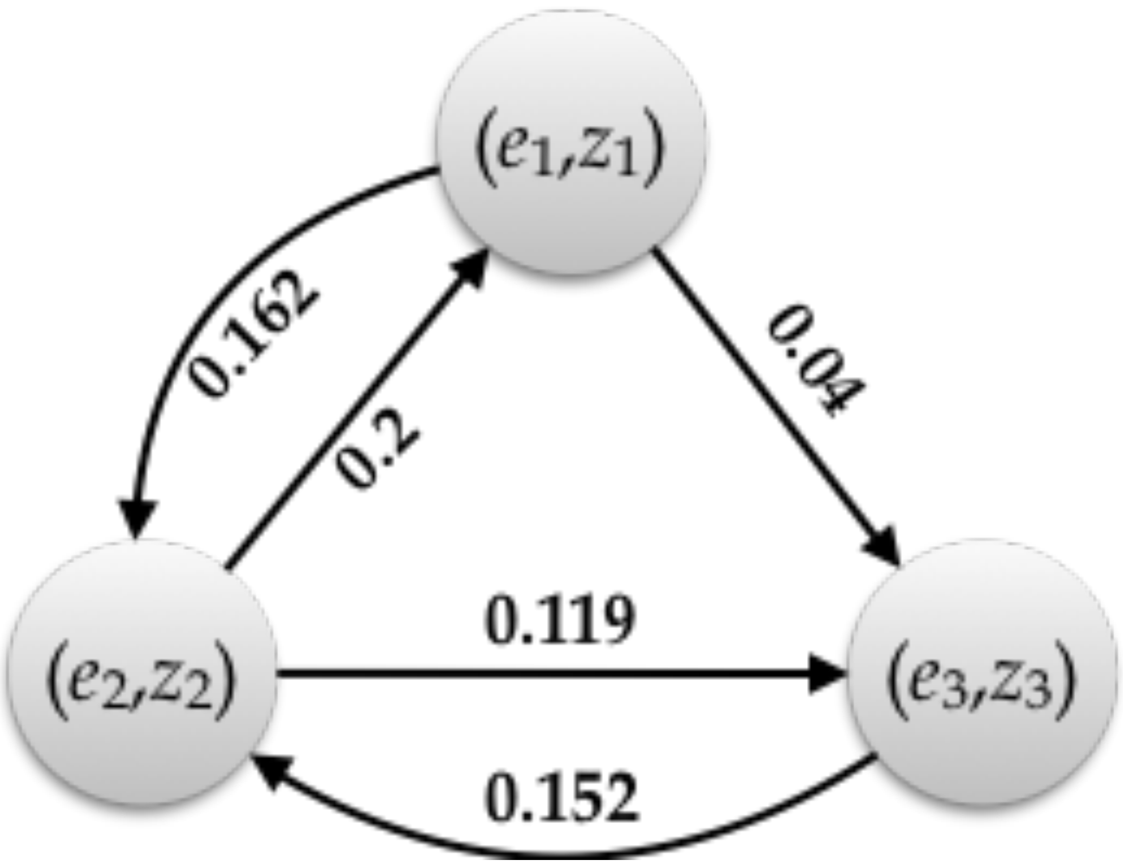
Direct Pathway function



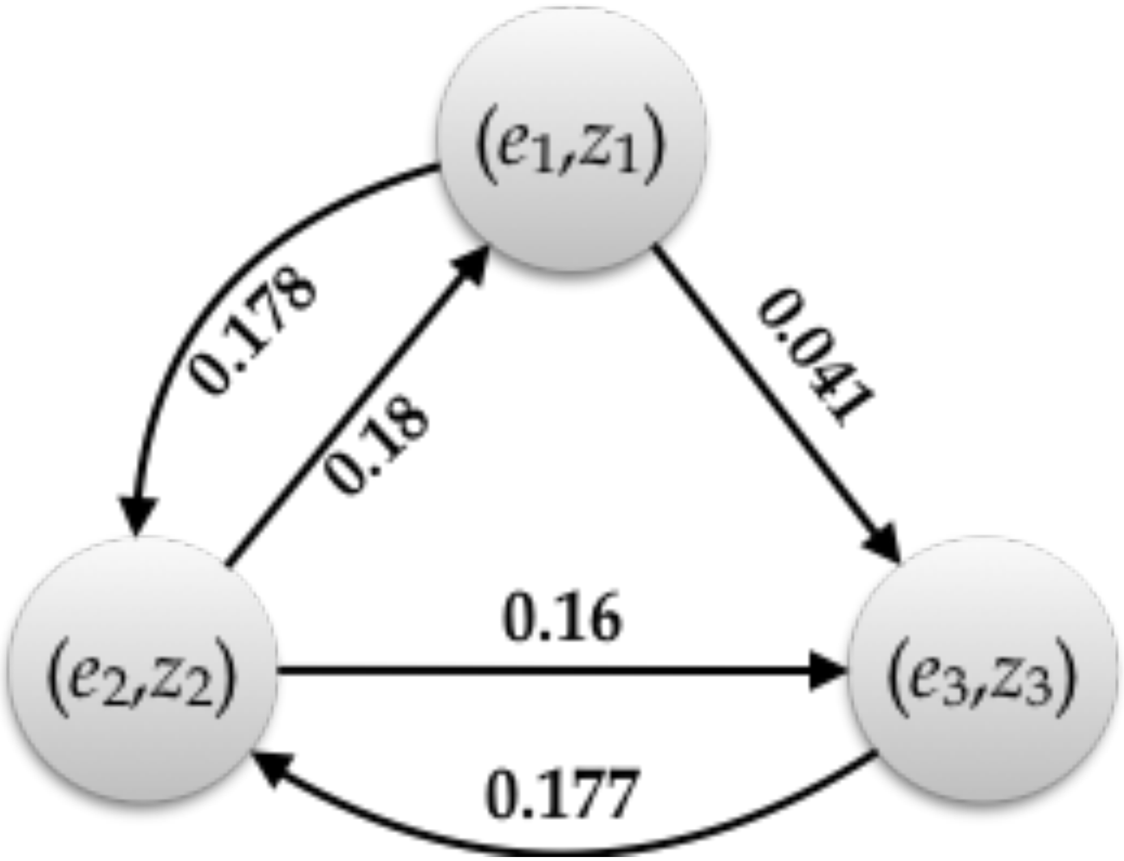


True Process

$$z_1[k] = z_1[k-1] - 0.3z_1[k-2] + 0.4z_2[k-2] + e_1[k] + 0.3e_1[k-1]$$
$$z_2[k] = -0.4z_1[k-1] + 0.6z_2[k-1] + 0.3z_3[k-1] + e_2[k] + 0.4e_2[k-1] + 0.2e_2[k-2]$$
$$z_3[k] = 0.2z_1[k-1] - 0.4z_2[k-2] + 0.3z_2[k-3] + 0.6z_3[k-1] + e_3[k]$$



True causal network



Identified network

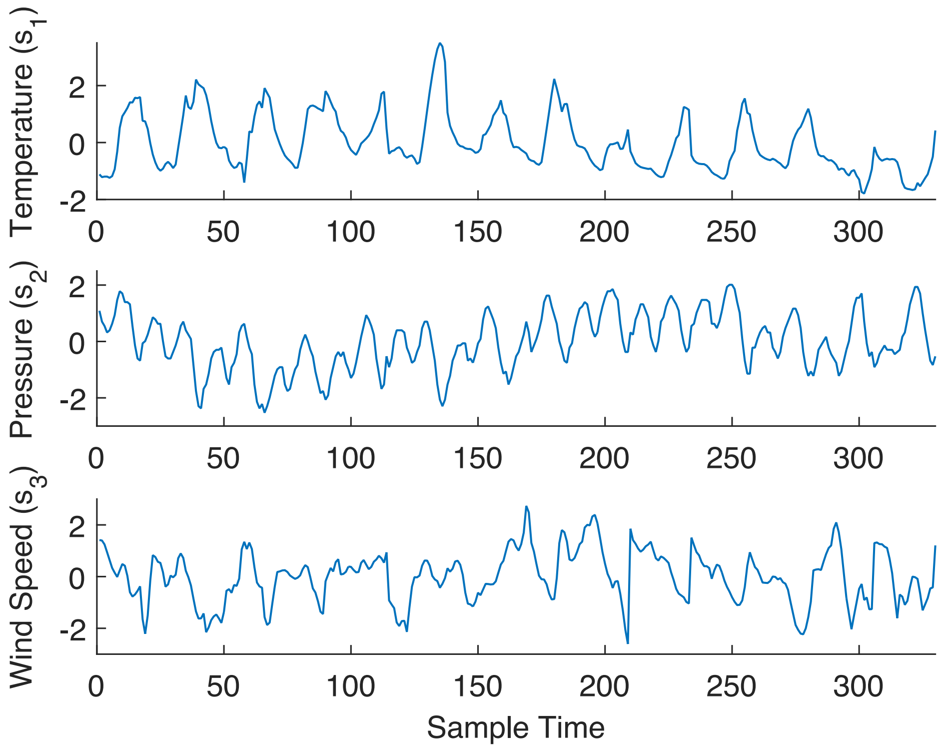
Known Limitations and Issues

- **Assumes stochastic** (stationary signals) framework
- **No scope for handling deterministic only or hybrid** set of signals
- **Sensitive to measurement errors**
- **Filtering** / downsampling can result in **spurious connections**
- **No direct provision** for detecting **instantaneous causality**
- **Can be sensitive to model mis-specification** in parametric testing
- Commonly **misunderstood limitations:**
  - Applicable to linear processes only
  - Only based in the predictive framework
  - No room for handling non-stationarities

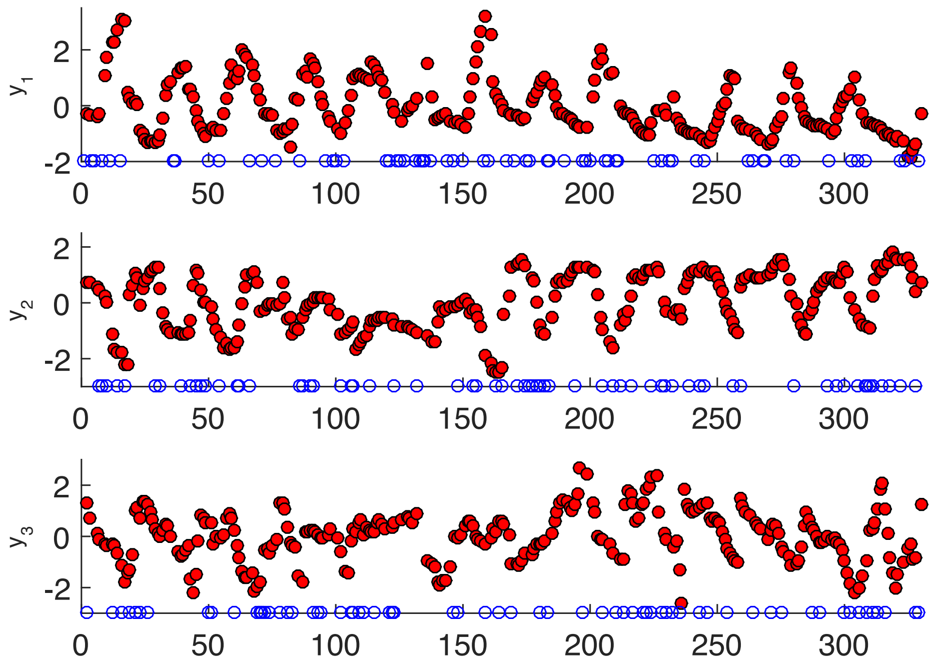
Read: *B.D.O. Anderson, M. Deistler and J-M Defour (2019). On the Sensitivity of Granger Causality to Errors-in-Variables, Linear Transformations and Subsampling. Journal of Time-Series Analysis, 40, 102-123.*

Missing Data

- Missing data can be handled by first reconstructing the series using compressive sensing (CS) techniques



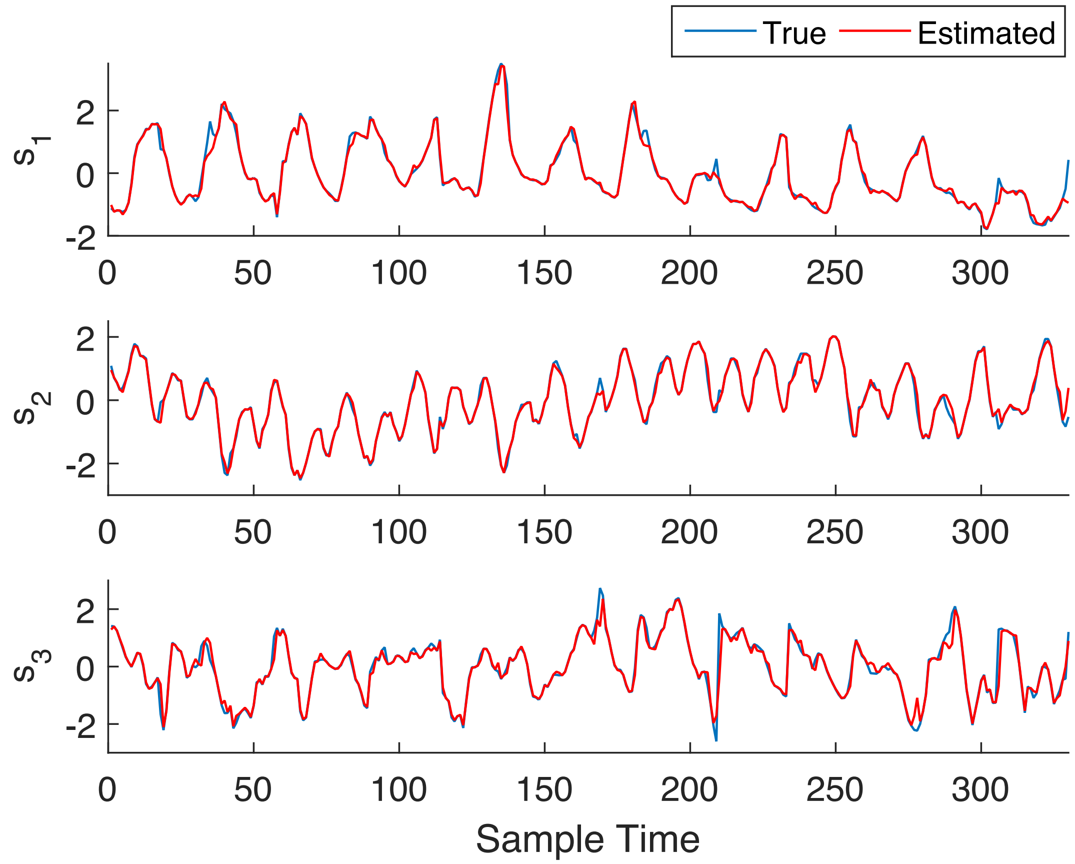
(a) Meteorological Time-series



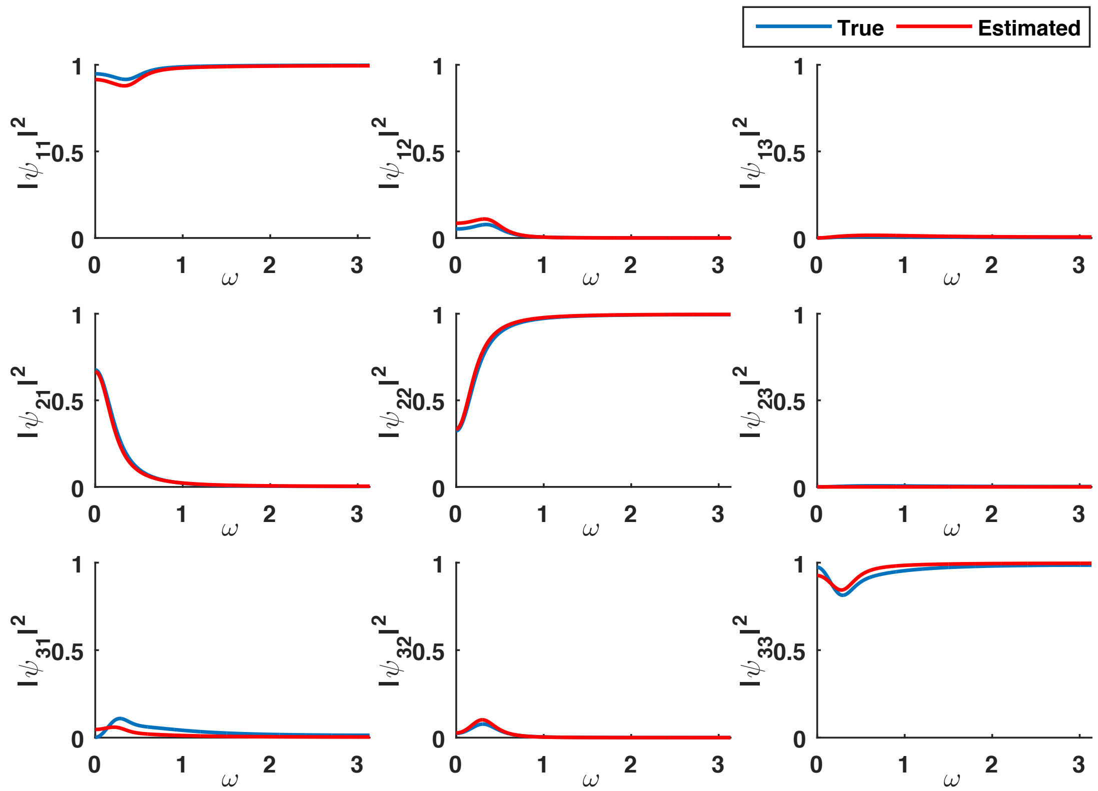
(b) Time-series with missing observations

Complete and incomplete time-series (30% missing data)

Note: ***Missing data located at different locations for each signal***

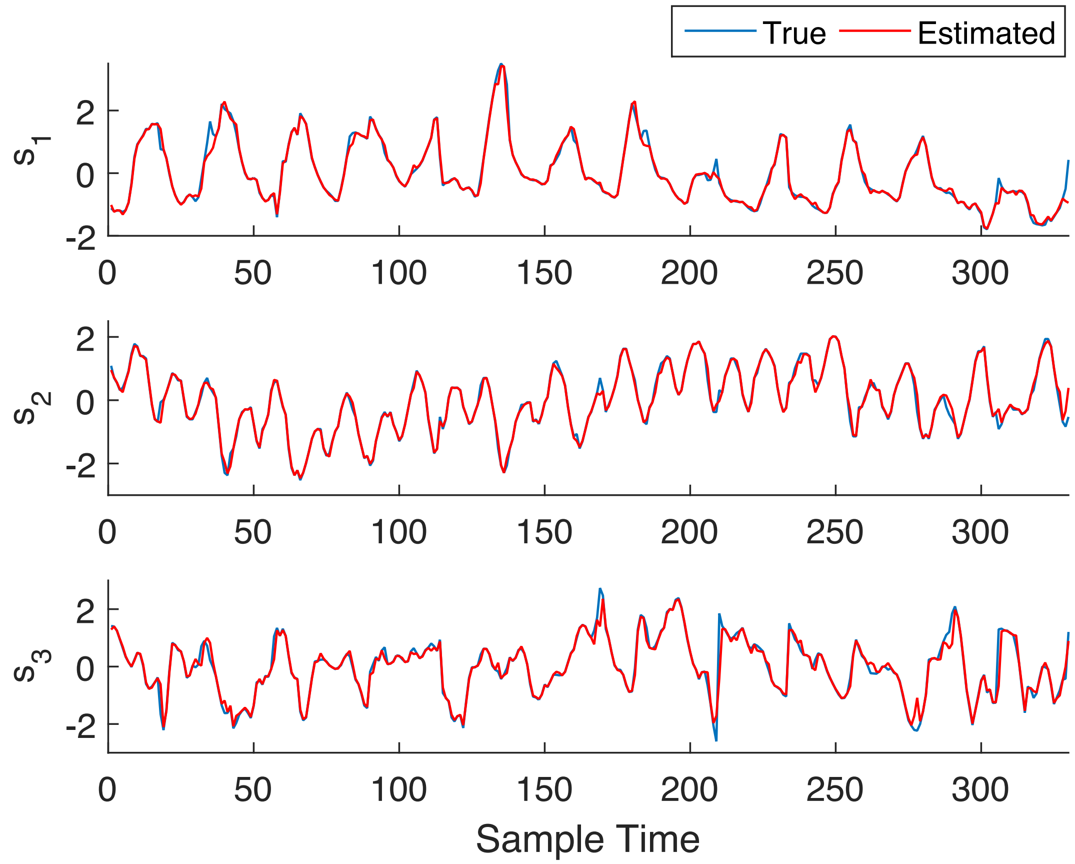


(a) Reconstructed Signal

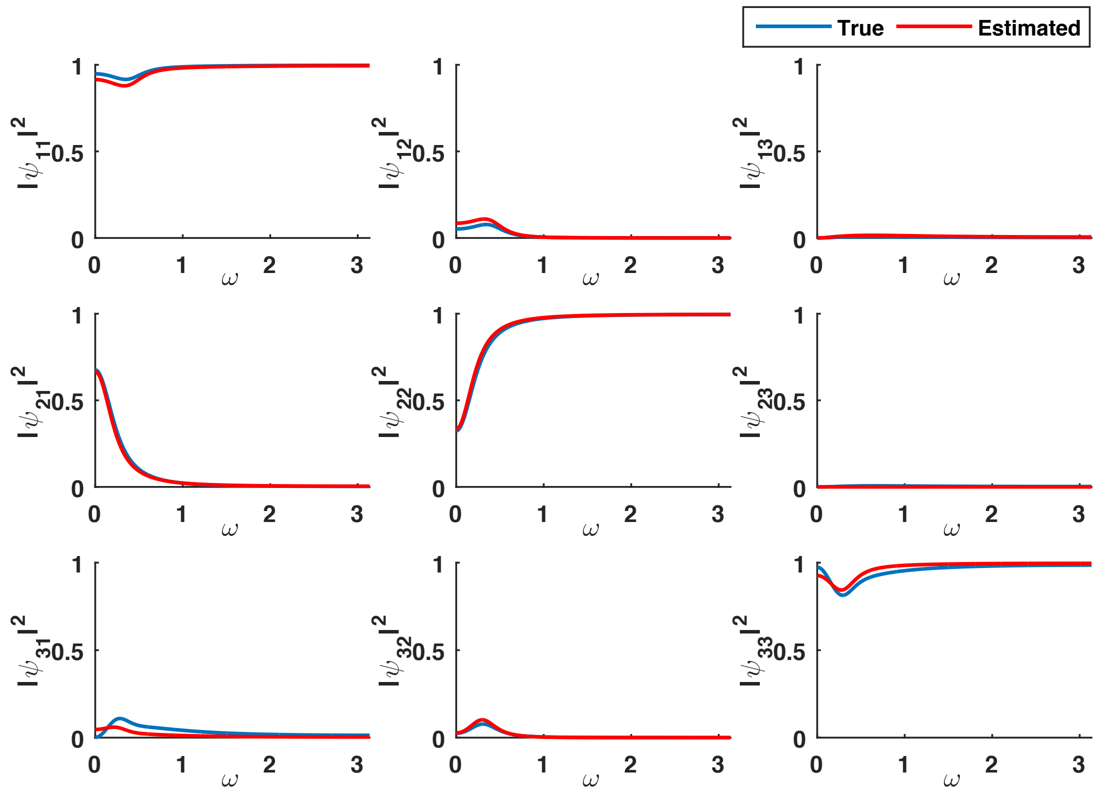


(b) Normalized DPT

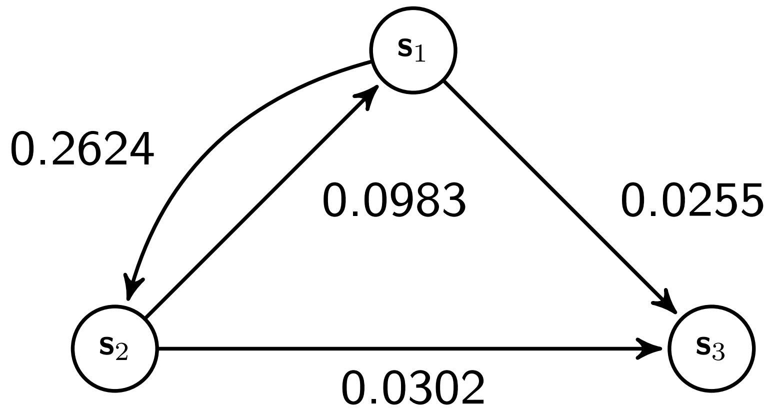
Results



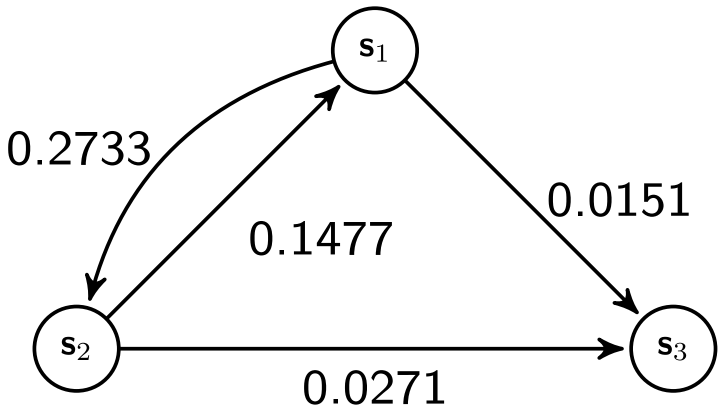
(a) Reconstructed Signal



(b) Normalized DPT



(a) Complete data case



(b) Incomplete data case

NonLinear Systems: Kernel-Granger Causal Inference

- Transform data into feature space and implement linear Granger causal inference methods
- Gaussian kernels were used in conjunction with partial directed coherence

NonLinear Systems: Kernel-Granger Causal Inference

- Transform data into feature space and implement linear Granger causal inference methods
- Gaussian kernels were used in conjunction with partial directed coherence

A three-state bioreactor model [Enszer et al., 2008, Lin and Stadtherr, 2007a] with concentration of the cells (x_1), the substrate (x_2) and the product (x_3) as states:

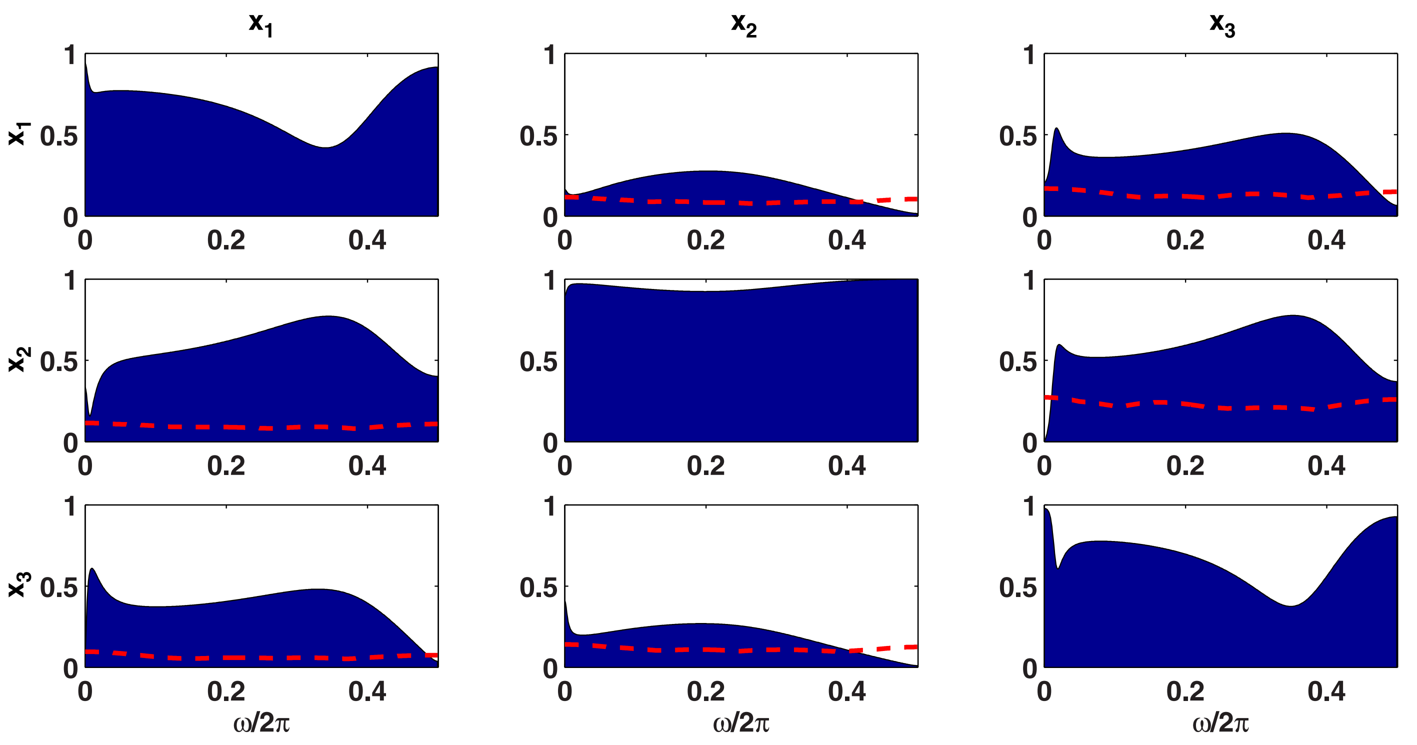
$$\dot{x}_1 = (\mu - D)x_1$$
$$\dot{x}_2 = D(x_{2f} - x_2) - \frac{\mu x_1}{Y}$$
$$\dot{x}_3 = -Dx_3 + (\alpha\mu + \beta)x_1$$

where the growth rate μ is

$$\mu = \frac{\mu_{max}\left[1 - \frac{x_3}{x_{3m}}\right]x_2}{k_s + x_2}$$

with $x_1(0) = 6.50$ g/L, $\mu_{max} = 0.46$, $k_s = 1.1$, $x_2(0) = 5$ g/L, $x_3(0) = 15$ g/L, $Y = 0.4$ g/g, $\beta = 0.2\text{hr}^{-1}$, $D = 0.202\text{hr}^{-1}$, $\alpha = 2.2$ g/g, $x_{3m} = 50$ g/L and $x_{2f} = 20$ g/L.

1. S. Gigi and A. K. Tangirala (2010). Biological Cybernetics. 103(2): 119-133.
2. S. Gigi and A.K. Tangirala (2013). Automatica, 49(5), 1174-1183.
3. R. Kannan and A.K. Tangirala (2014). Physical Reviews E, 89, 062144.
4. P. Agarwal and A.K. Tangirala (2017). Special Issue of Int. J. of Adv. in Engg. Sciences and Applied Math., 9(4):196–213.
5. A. Garg and A.K. Tangirala (2018). Industrial & Engineering Chemistry Research, 57 (3), 967-979.
6. S. Kathari and A.K. Tangirala (2019). Industrial & Engineering Chemistry Research, 58 (26), 11275-11294.
7. S. Kathari and A.K. Tangirala (2019). SICE Conf. Japan, 199-204.
8. S. Nithya and A.K. Tangirala (2021). 7th ICC Proceedings, 436–441



Non-linear Dynamical Systems

Autonomous
Deterministic
Non-linear
Dynamical System

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{F}(\mathbf{x}, \theta)$$
$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n, \theta)$$

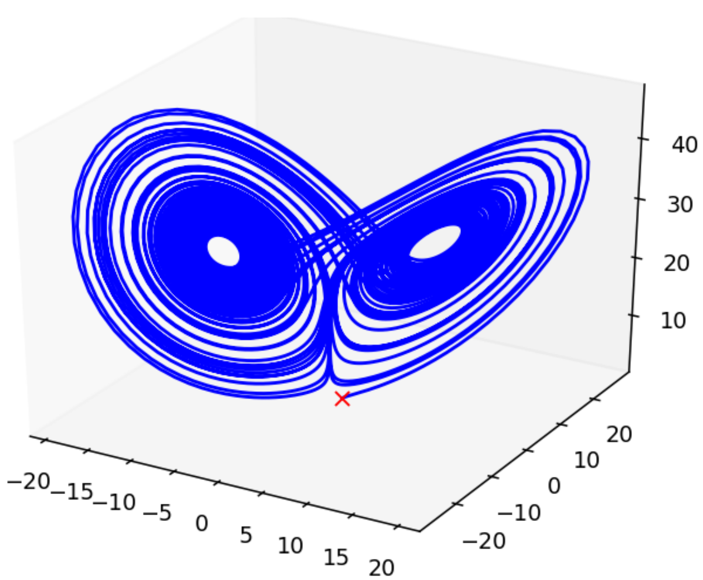
Non-separability

Cannot be broken down into parts that can be analysed (or solved for) individually

State space & trajectories

Trajectory: the time evolution of the system given for **x(0)** or $\mathbf{x}_0$

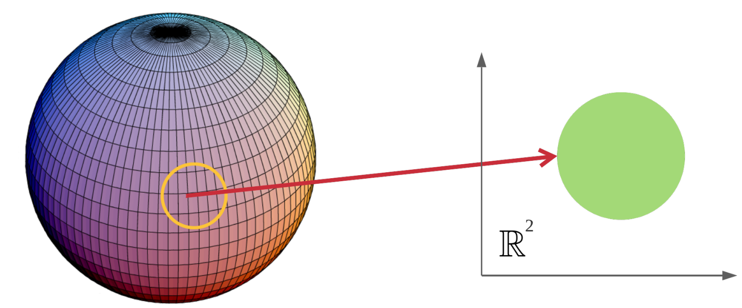
State space: Space filled with trajectories

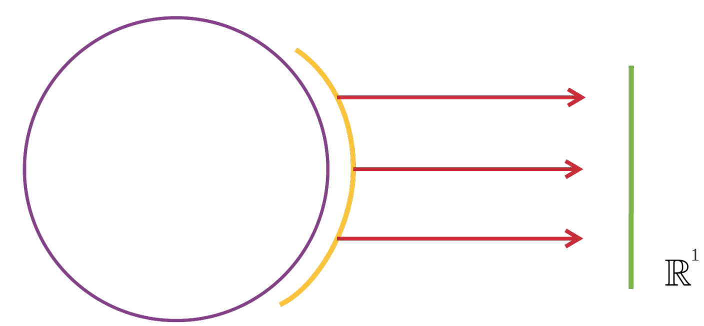


Topological Dynamics

State space: Manifold (M)

A topological space that locally resembles a Euclidean space (in $\mathbb{R}^d$)





- Not all states are measurable. Only a projection of states in a lower-dimensional space can be measured.
- Cannot recover states from measurements, especially when the model is unknown.
- **One can only recover the topology of the phase-space upto a diffeomorphism!**

CONVERGENT CROSS MAPPING

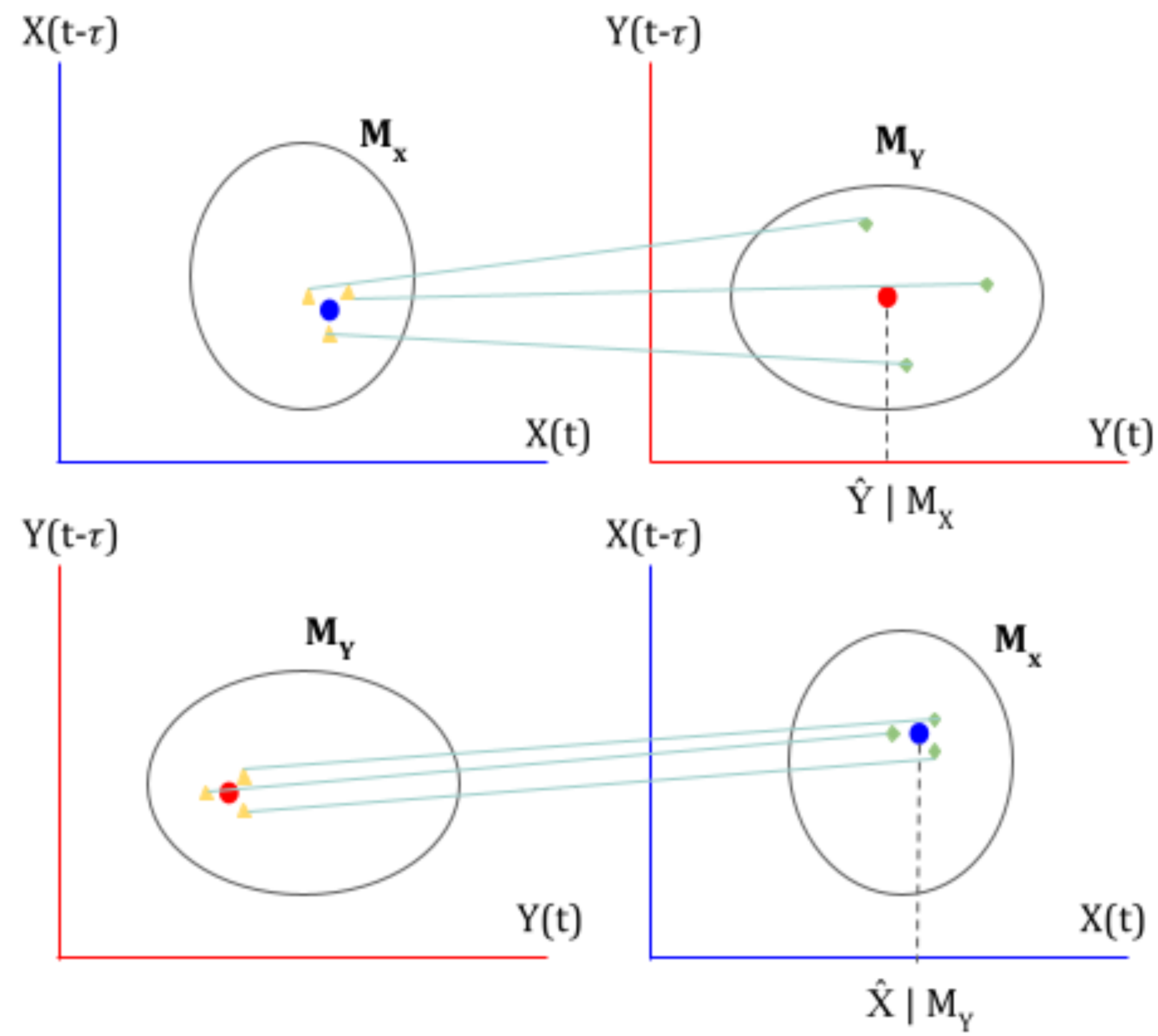
Assumption: The data is generated by a *deterministic nonlinear dynamical system.*

If X causes Y, enough information about X is encoded in Y, but not the other way round.

Hence X can be recovered from Y, i.e., M_Y

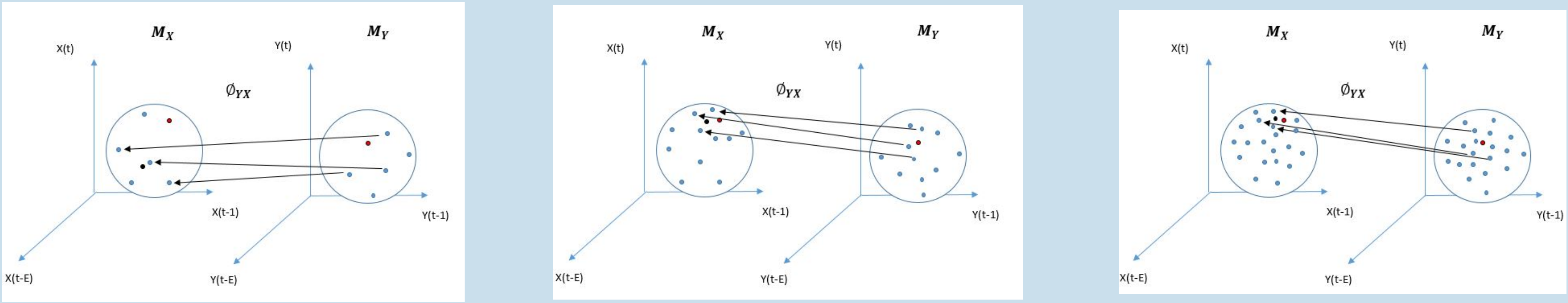
Cross Map Skill (CMS) is computed for $\hat{X} \mid M_Y$ and $\hat{Y} \mid M_X$ as $\varrho(\hat{X}, X)$ and $\varrho(\hat{Y}, Y)$.

The CMS converges to a non-zero value for
$\hat{X} \mid M_Y$ and to 0 for $\hat{Y} \mid M_X$.

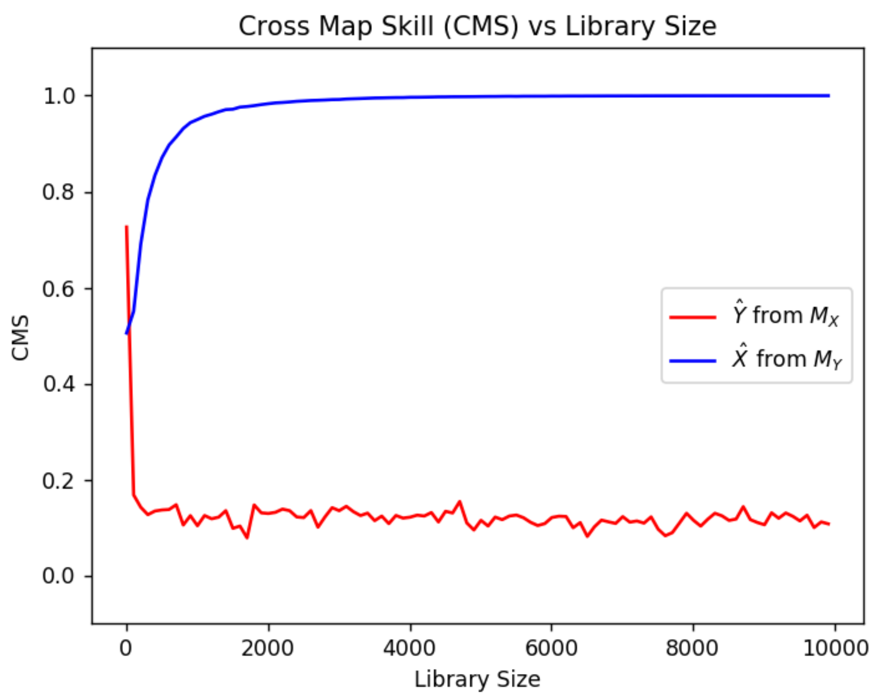


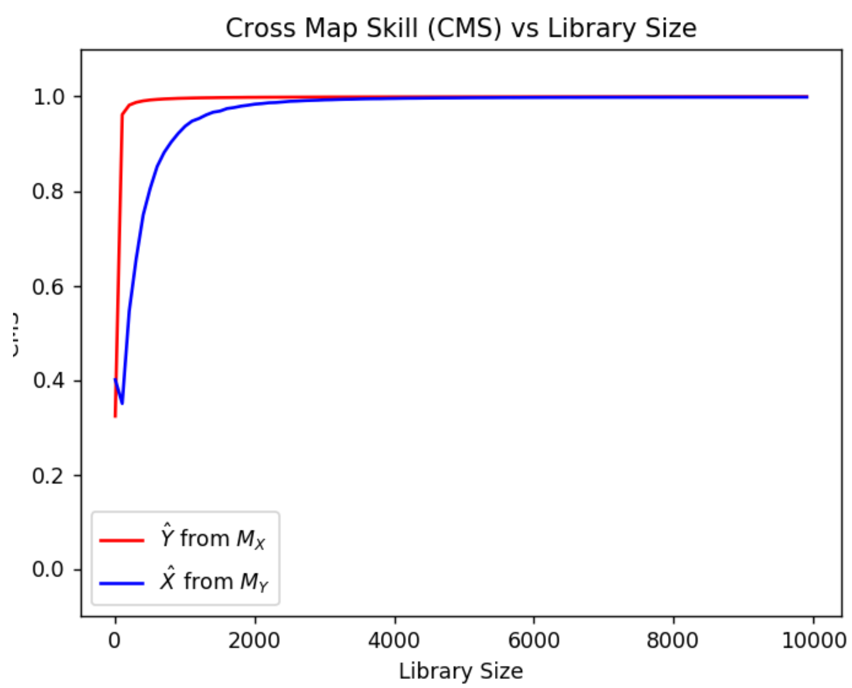
Cross Map Skill

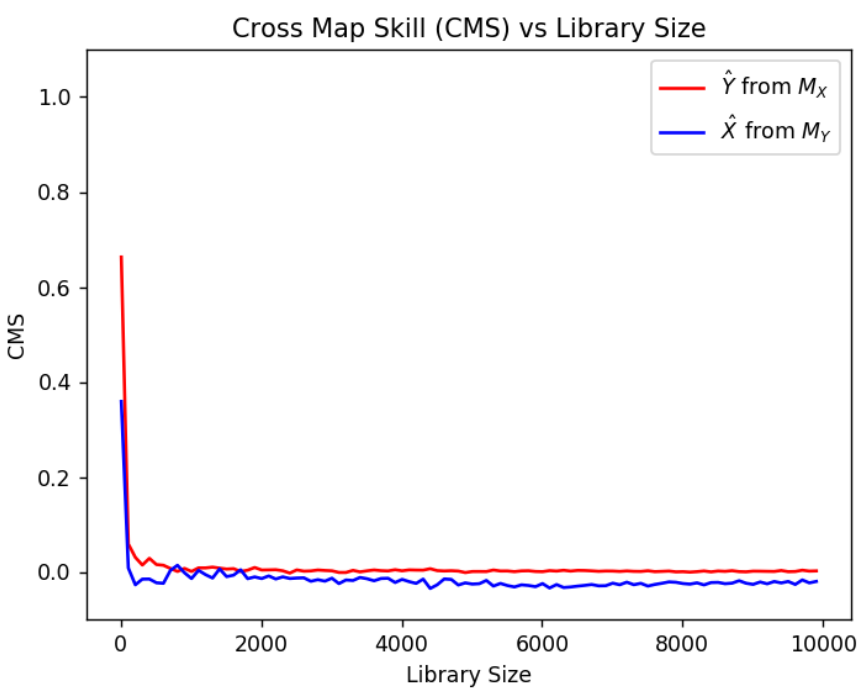
For $x(t)$ causing $y(t)$



Low sample sizes | Moderate sample sizes | High sample sizes



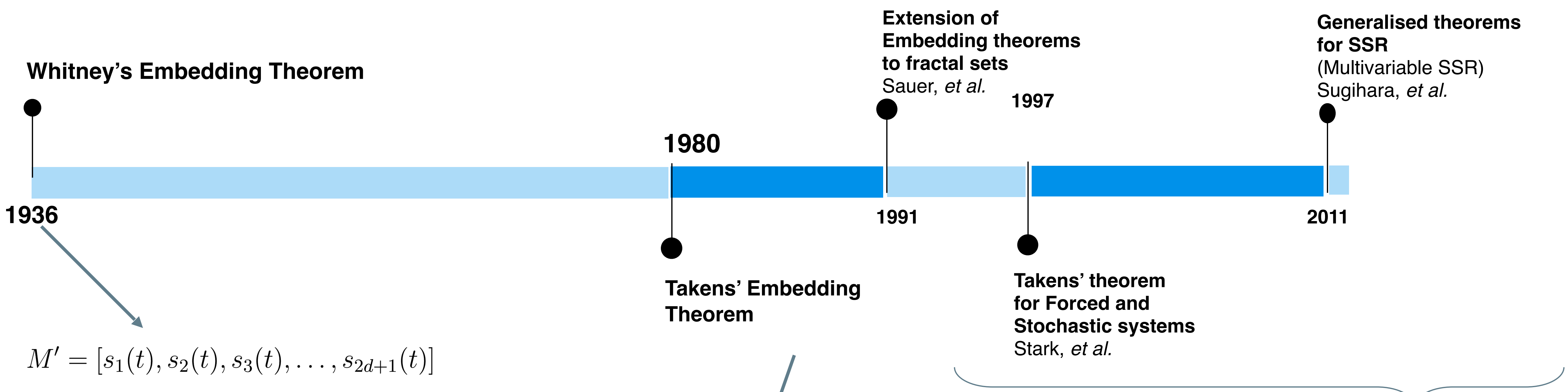




$X \longrightarrow Y$ | $X \longleftrightarrow Y$ | $X \not\leftrightarrow Y$

Embedding

An embedding (of a manifold M) is a one-to-one immersion (smooth diffeomorphism) of M



$$M' = [s_1(t), s_2(t), s_3(t), \ldots, s_{2d+1}(t)]$$
$$s_i(t) = H_i(\mathbf{x})$$
where $\mathbf{x} \in M, H : \mathbb{R}^n \to \mathbb{R}$

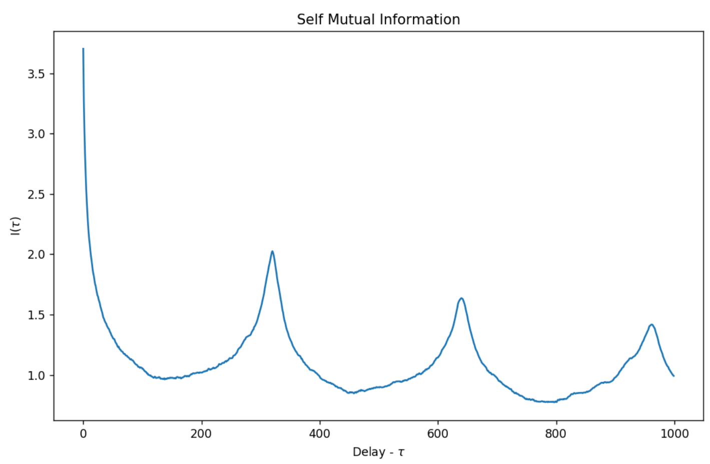
$$M' = [s(t), s(t-\tau), s(t-2\tau), \ldots, s(t-2d\tau)]$$
$$s(t) = H(\mathbf{x})$$
where $\mathbf{x} \in M, H : \mathbb{R}^n \to \mathbb{R}$

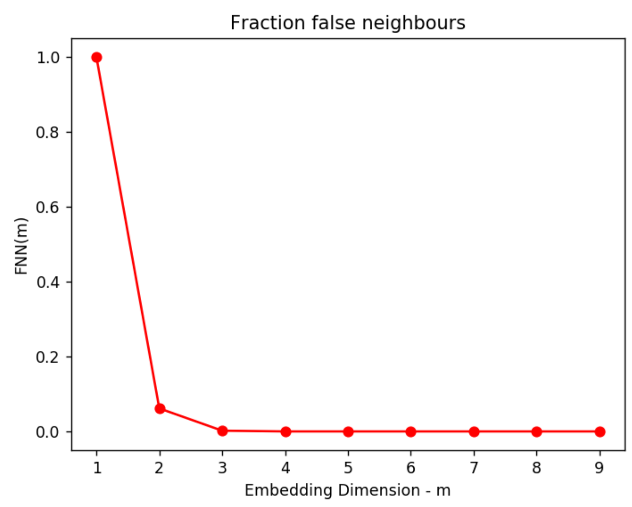
Extensions & generalisations of existing theorems

- Whitney's embedding theorem requires measurements of multiple variables
- Taken's theorem constructs an embedding from a single measurement (by stacking time-lagged measurements judiciously).

Example: Rossler Attractor

Univariable State-Space Reconstruction





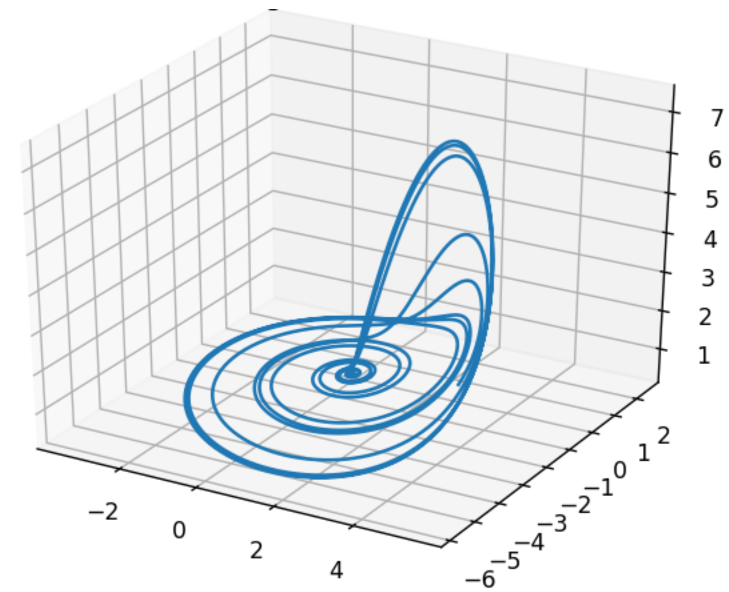
Ideal τ : First minimum of the self mutual information function

Ideal m: Smallest ‘m’ for which the fraction of False Nearest Neighbors reaches zero

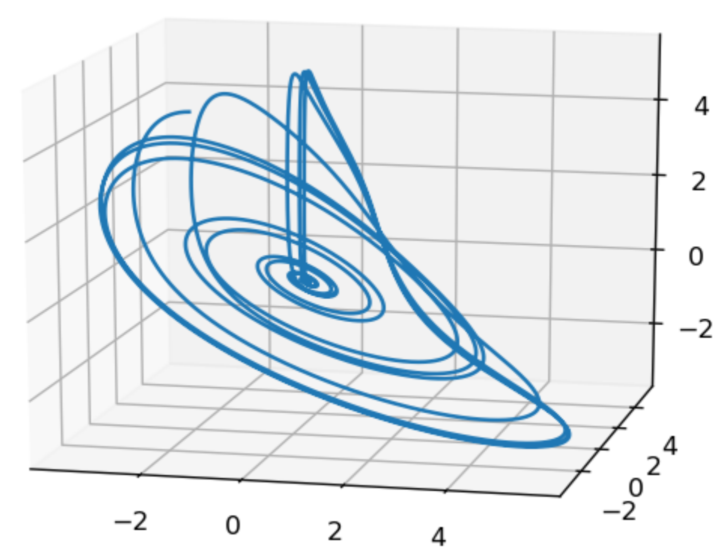
Choice: $\tau = 138$

Choice: $m = 3$

Original attractor



Reconstructed attractor



Shortcomings of Original CCM

Shortcomings of Original CCM

- **Lack of scalability for large gridded datasets**
  - For a dataset consisting of N grid points, since there is no multi-variate version of CCM, a pair-wise analysis is required, (N choose 2) CCM runs
  - CCM implementation in R (Sugihara et al., 2012) requires 2 minutes per run equal to millions of minutes of computation.

Shortcomings of Original CCM

- **Lack of scalability for large gridded datasets**
  - For a dataset consisting of N grid points, since there is no multi-variate version of CCM, a pair-wise analysis is required, (N choose 2) CCM runs
  - CCM implementation in R (Sugihara et al., 2012) requires 2 minutes per run equal to millions of minutes of computation.
- No quantification of coupling strengths over which CCM is applicable

SHORTCOMINGS OF ORIGINAL CCM

- **Lack of scalability for large gridded datasets**
  - For a dataset consisting of N grid points, since there is no multi-variate version of CCM, a pair-wise analysis is required, (N choose 2) CCM runs
  - CCM implementation in R (Sugihara et al., 2012) requires 2 minutes per run equal to millions of minutes of computation.
- No quantification of coupling strengths over which CCM is applicable
- No formal study or discussion on suitability of CCM for stochastic processes

SHORTCOMINGS OF ORIGINAL CCM

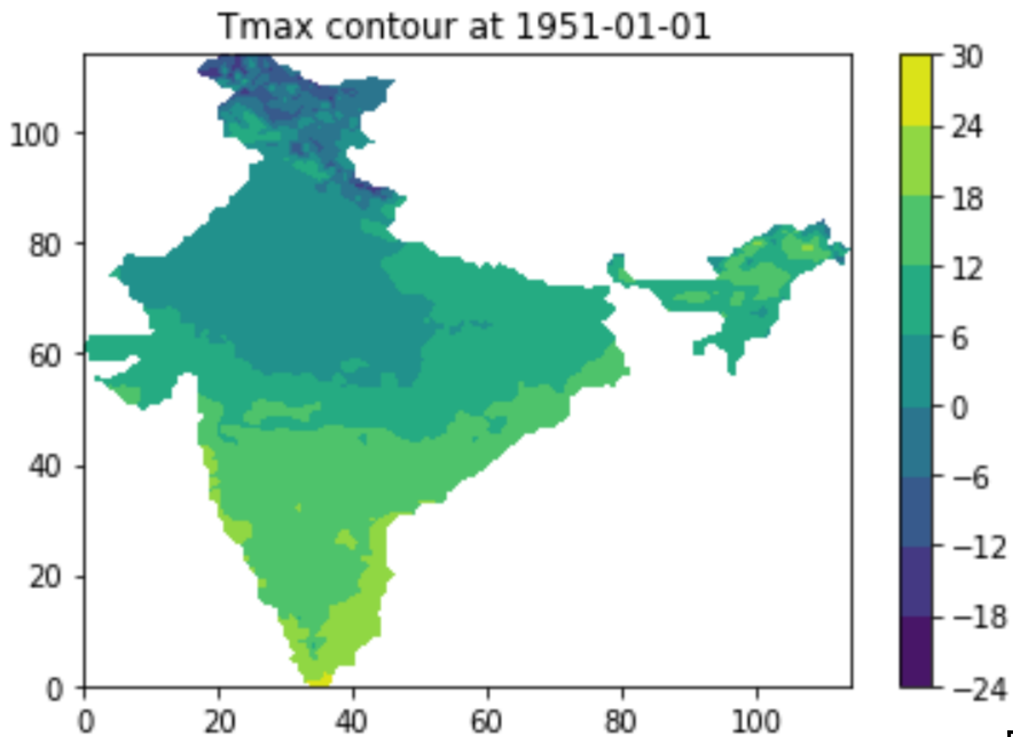
- **Lack of scalability for large gridded datasets**
  - For a dataset consisting of N grid points, since there is no multi-variate version of CCM, a pair-wise analysis is required, (N choose 2) CCM runs
  - CCM implementation in R (Sugihara et al., 2012) requires 2 minutes per run equal to millions of minutes of computation.
- No quantification of coupling strengths over which CCM is applicable
- No formal study or discussion on suitability of CCM for stochastic processes
- Unable to detect direct causality - multivariable version is required

Shortcomings of Original CCM

- **Lack of scalability for large gridded datasets**
  - For a dataset consisting of N grid points, since there is no multi-variate version of CCM, a pair-wise analysis is required, (N choose 2) CCM runs
  - CCM implementation in R (Sugihara et al., 2012) requires 2 minutes per run equal to millions of minutes of computation.
- No quantification of coupling strengths over which CCM is applicable
- No formal study or discussion on suitability of CCM for stochastic processes
- Unable to detect direct causality - multivariable version is required
- Requirement of manual inspection for drawing inference from CMS

Dataset

Global Circulation Model (GCM) output data comprising of spatio-temporal gridded (25 x 25 km^2) precipitation and surface temperature data across about 4000 grid points from 1970 to 2000

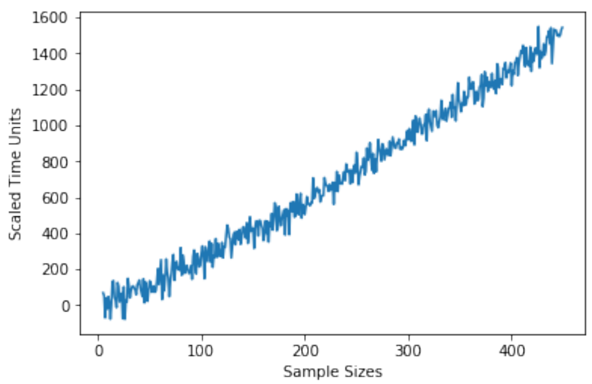


Variable	*Latitudinal Range (0.25 deg res.)*	*Longitudinal Range (0.25 deg res.)*	*Time Range (daily)*	*Datetime Type*
Min. Temp.	8.375-36.875	68.625-97.125	1951-2019	datetime64
Max. Temp.	8.375-36.875	68.625-97.125	1951-2019	datetime64
Precipitation Rate	6.5-38.5	66.5-100	1920-2005	cftime.NoLeap

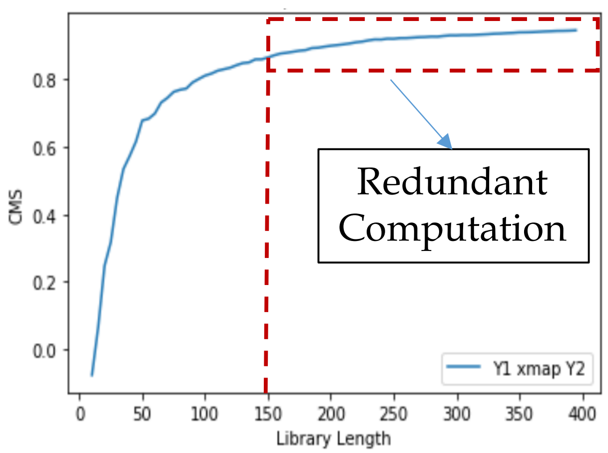
Summary Statistics

COMPONENT ANALYSIS OF CCM

Library sizes for CMS computation

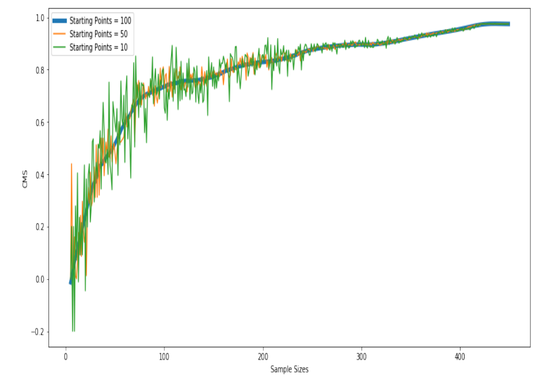


Computational time scales exponentially with Library sizes



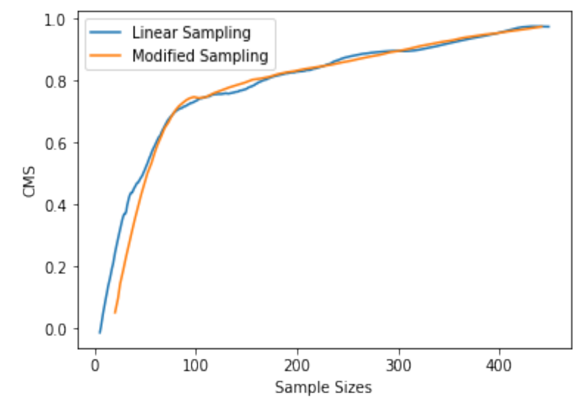
Large library sizes are not a requirement to detect causality

No. of repetitions per library size



CMS curve is smoothed out with increasing number of repetitions

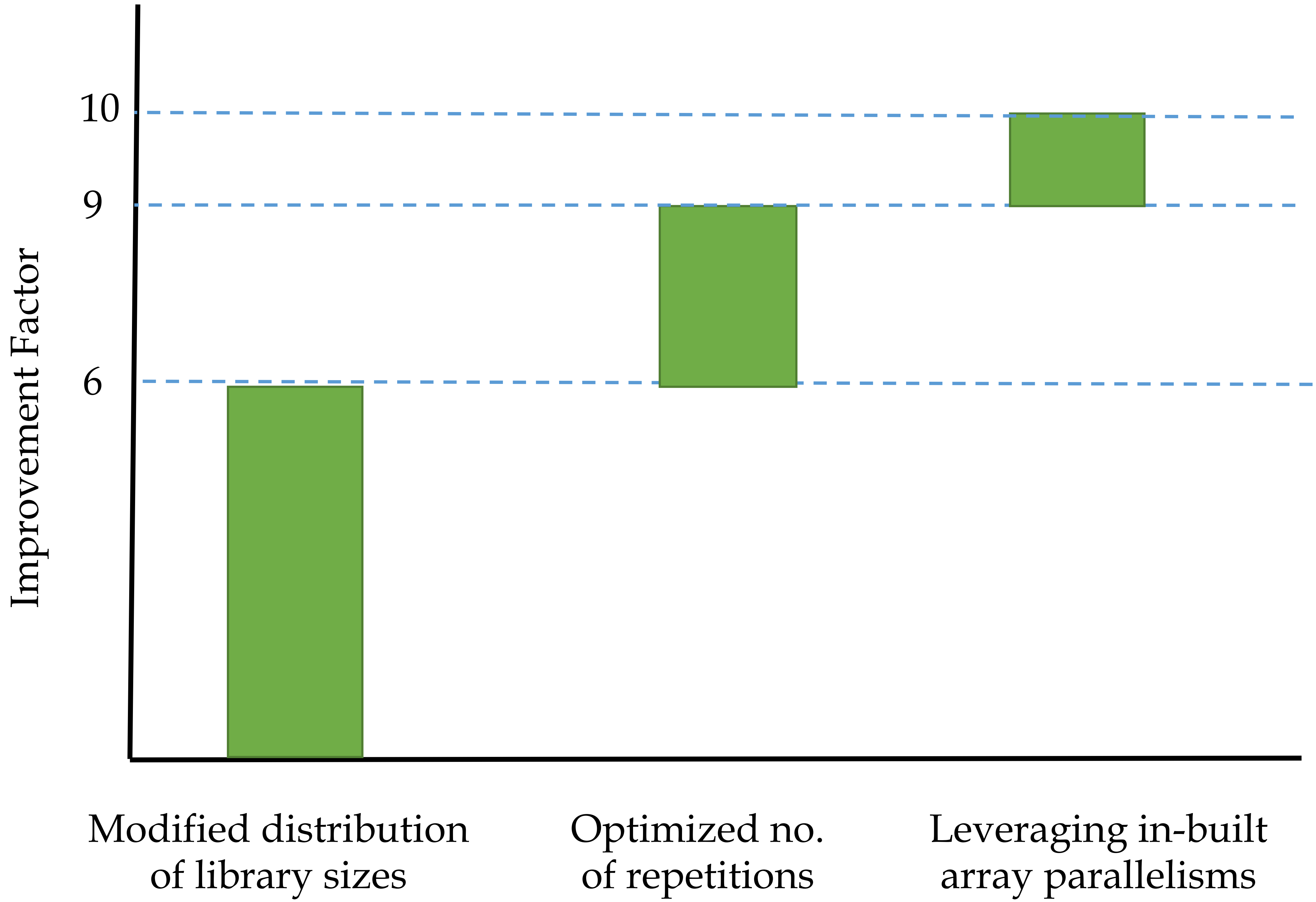
More discontinuities exist for smaller sample sizes



A modified sampling method is proposed with higher density of smaller sample sizes and vice versa.

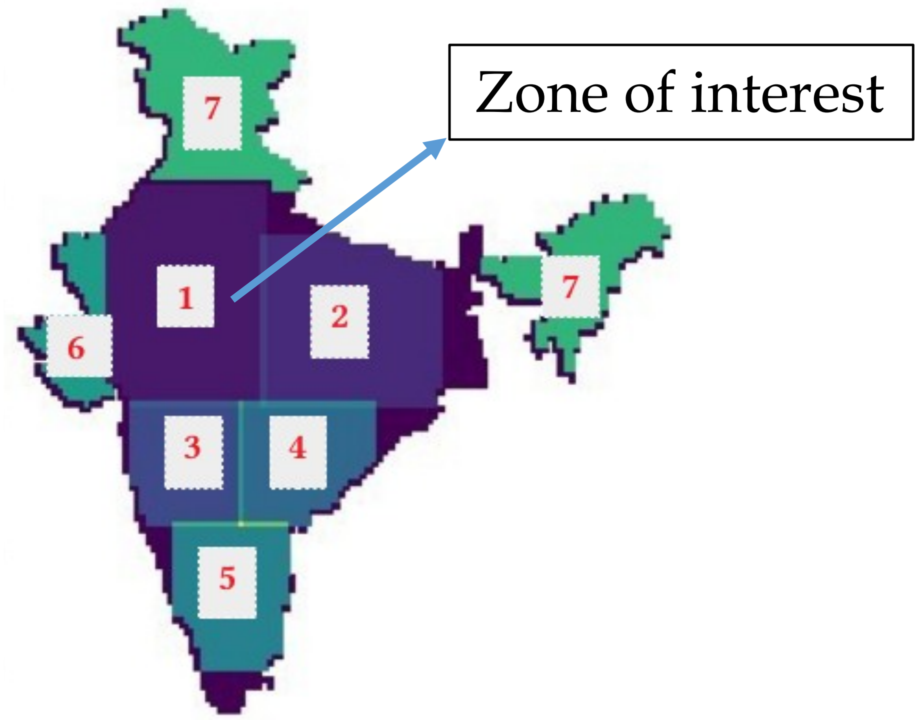
- Built-in NumPy parallelisms were leveraged while computing neighbour distances and eventually the predictions on the target variable. This was observed to reduce compute time to half the time before.

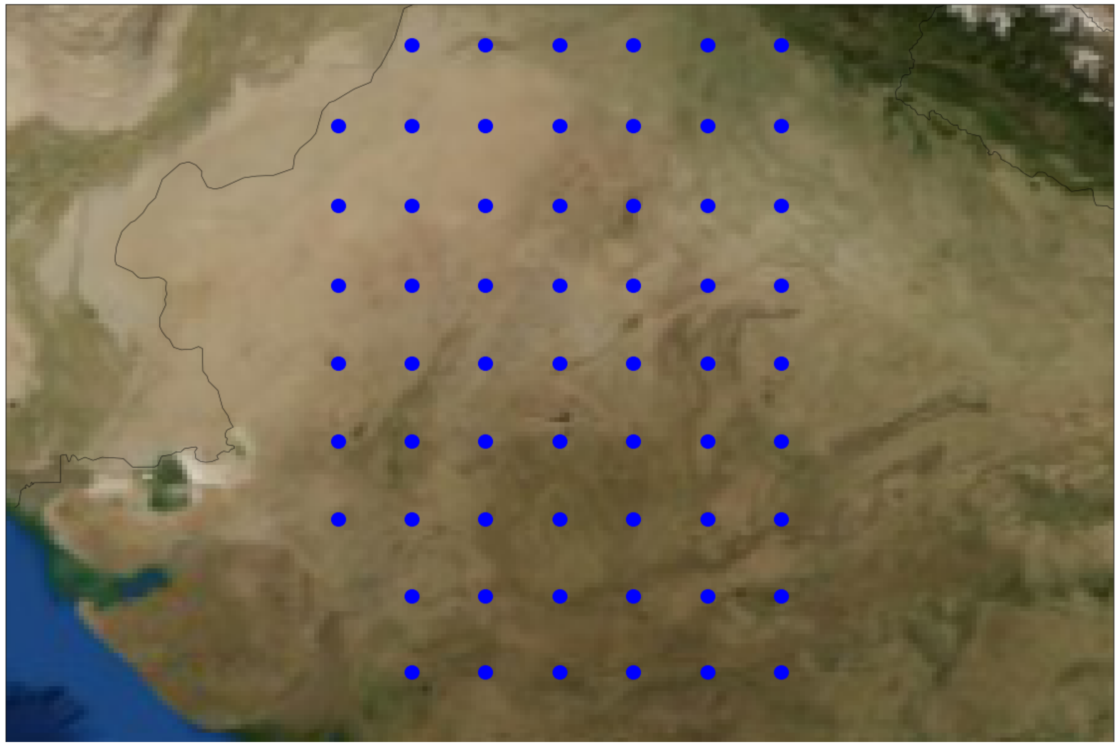
Improvements in Computational Time



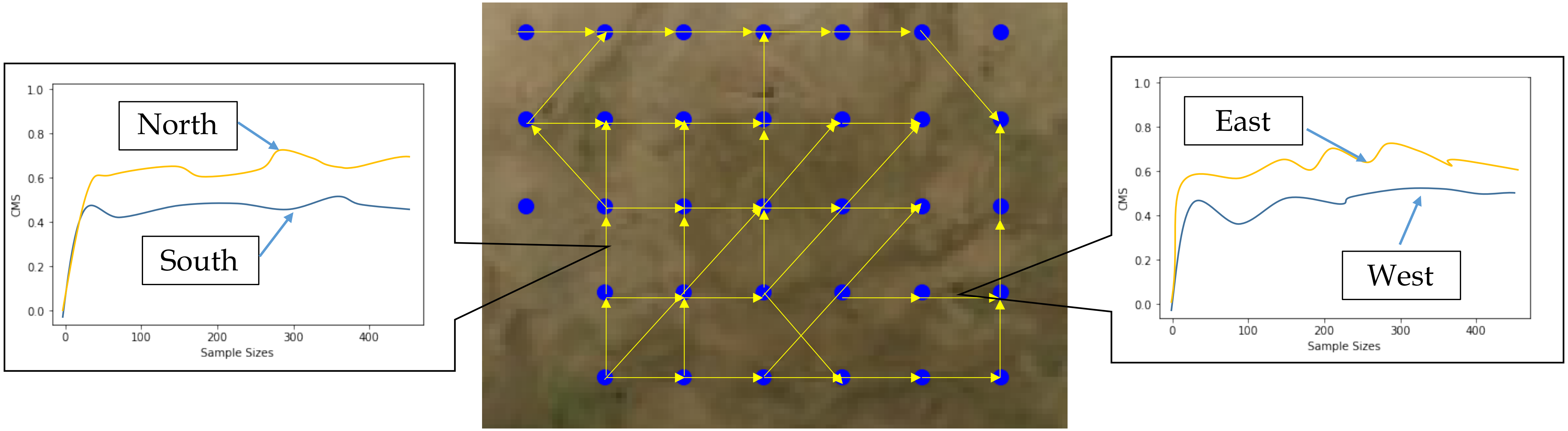
Zonal Dataset

- The whole dataset was broken down into 7 climatologically contiguous zones which had similar surface temperature values
- This division ensures that the networks for individual zones were constructed on a sound climatological bedrock
- Zone 1 (figure below) was chosen as zone of interest with a resolution of 100 x 100 km^2 with about 61 grid points



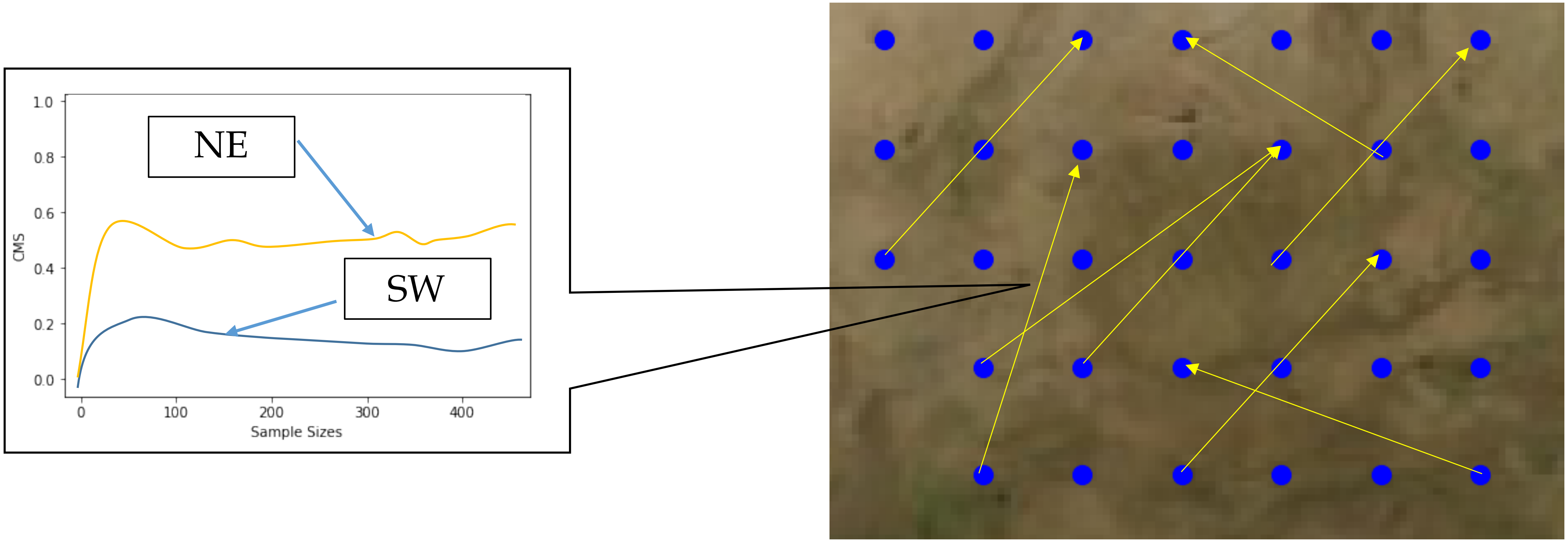


Climatological Insights



- For low lags, bi-directional causality exists between precipitation rates of spatially nearby points
- Stronger causal relationships occur almost consistently in the NE direction

INSIGHTS ... CONTD.



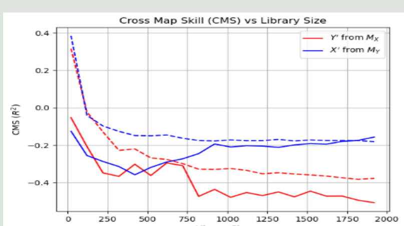
- For greater lag values, nearby causalities seem to vanish while long-distance ones in the NE direction become more prominent but sparse.
- The difference in convergence values for bi-directional cases is observed to increase between far-off points.

Synchronization Likelihood (SL) & CCM

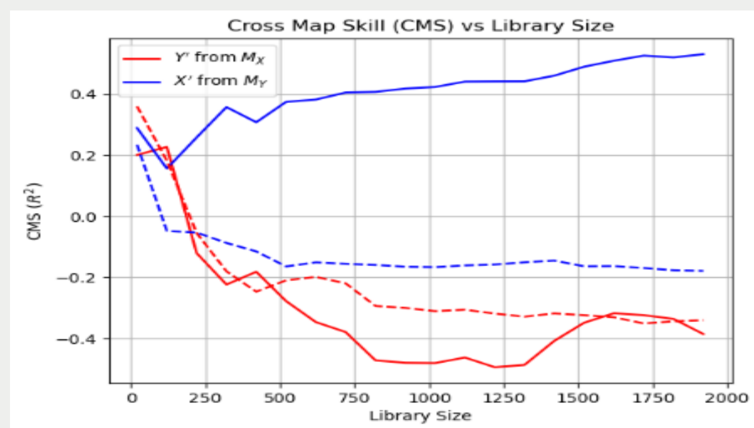
- Generalised synchronisation => exists a functional dependence between cause and effect
- SL (Star and van Dijk, 2002) metric calculates the likelihood of simultaneous auto-recurrence of embedding vectors in bivariate time series data.
- Computation carried out using the “eeglib” Python library

- **SL values divide its range of 0 to 1 into three regions based on the applicability of CCM**

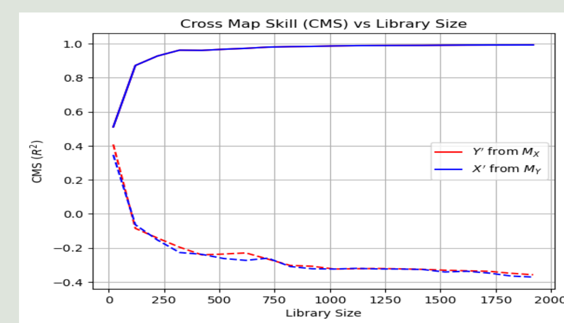
Region I



Region II



Region III



0.15 | 0.6 | 1.0

Causality not detected, CMS is not significant

CCM infers correct Causality

Very high synchrony, CCM doesn't work

Stam, C. and B. van Dijk (2002). *Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. Physica D: Nonlinear Phenomena, 163(3), 236–251. ISSN 0167-2789.*

CCM Detects only Direct links: Logistic Map

- CCM identifies only the presence of **total causal connections**

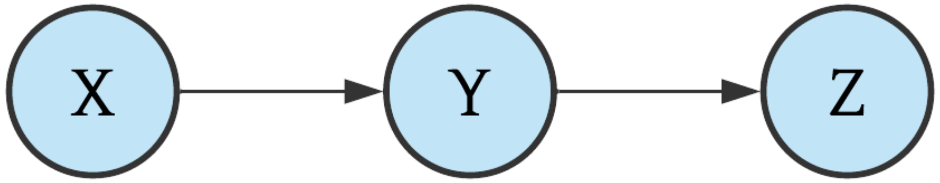
$$X[k+1] = X[k][3.9 - 3.9X[k]]$$

$$Y[k+1] = Y[k][3.3 - 3.3Y[k] - 0.15X[k]]$$

$$Z[k+1] = Z[k][3.5 - 3.5Z[k] - 0.12Y[k]]$$

$$X[0] = 0.8, Y[0] = 0.2, Z[0] = 0.4$$

True network



CCM Detects only Direct links: Logistic Map

- CCM identifies only the presence of **total causal connections**

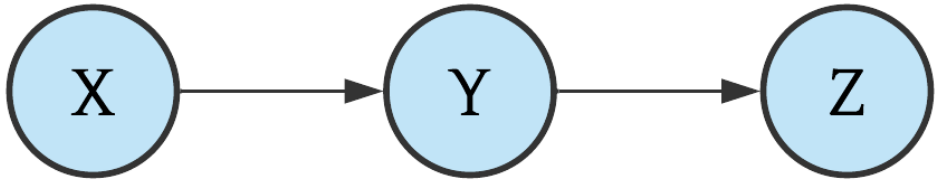
$$X[k+1] = X[k][3.9 - 3.9X[k]]$$

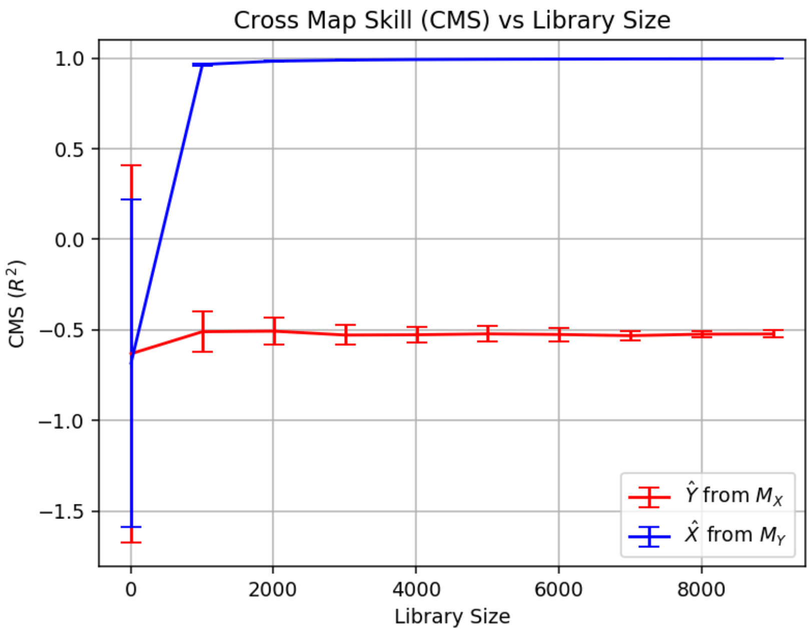
$$Y[k+1] = Y[k][3.3 - 3.3Y[k] - 0.15X[k]]$$

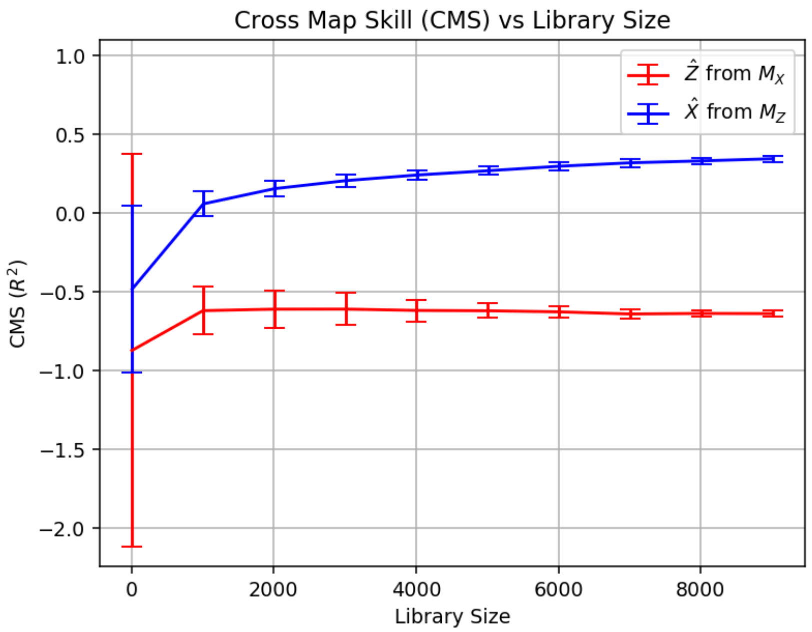
$$Z[k+1] = Z[k][3.5 - 3.5Z[k] - 0.12Y[k]]$$

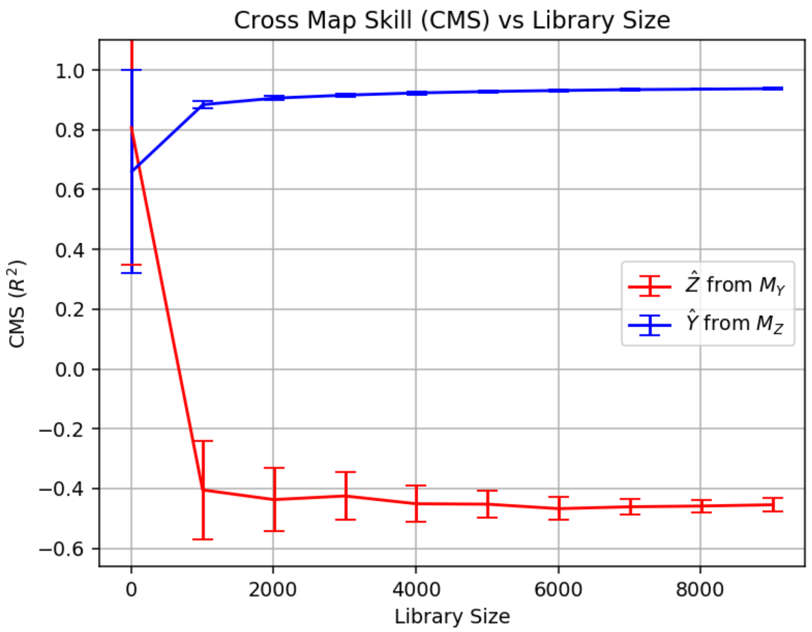
$$X[0] = 0.8, Y[0] = 0.2, Z[0] = 0.4$$

True network









CCM Detects only Direct links: Logistic Map

- CCM identifies only the presence of **total causal connections**

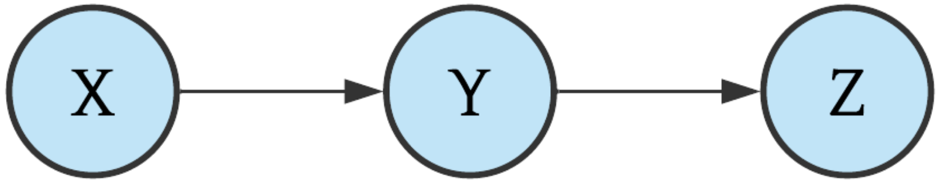
$$X[k+1] = X[k][3.9 - 3.9X[k]]$$

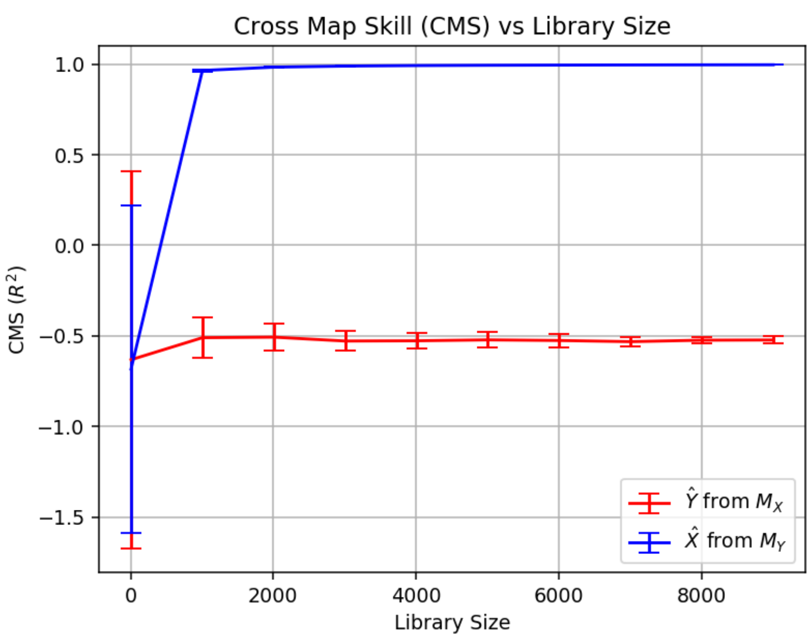
$$Y[k+1] = Y[k][3.3 - 3.3Y[k] - 0.15X[k]]$$

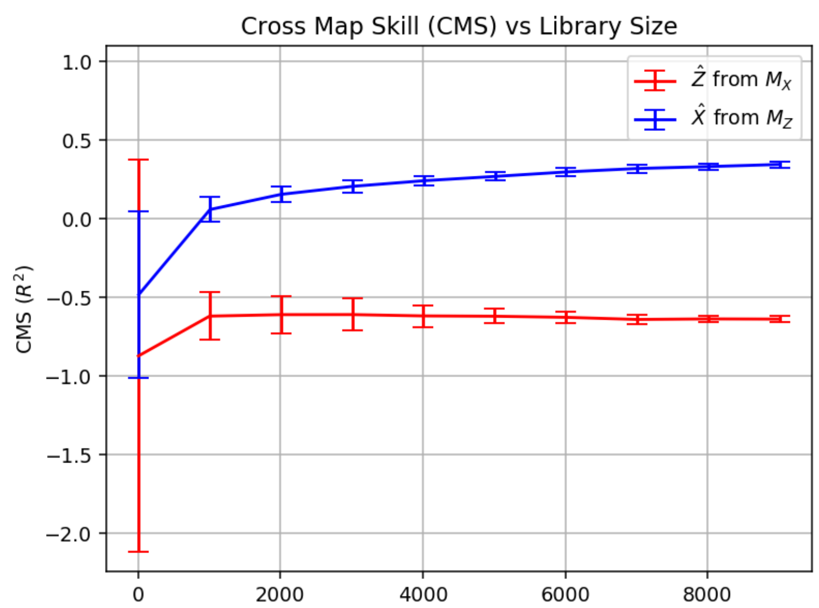
$$Z[k+1] = Z[k][3.5 - 3.5Z[k] - 0.12Y[k]]$$

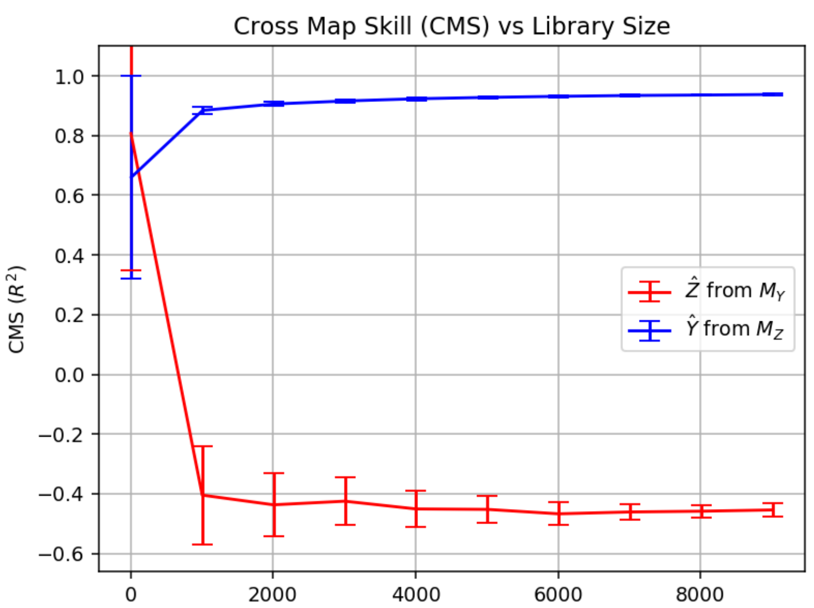
$$X[0] = 0.8, Y[0] = 0.2, Z[0] = 0.4$$

True network

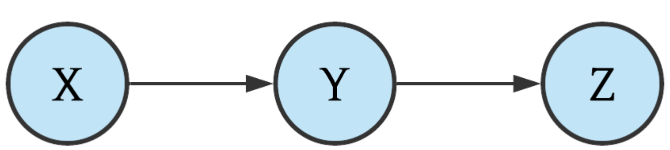


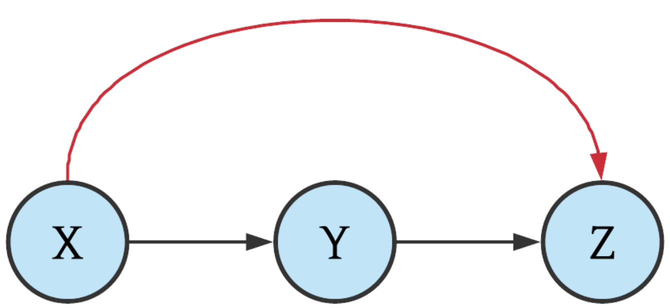






Possible causal graphs





Multivariable State Space Reconstruction[1]

A typical shadow manifold is given by:

$$\mathcal{M}' = (s_1[k-\tau_{11}], s_1[k-\tau_{12}], ...s_p[k-\tau_{pm_p}])$$

Future vector & Candidate set:

$$v_F = (s_1[k+1], s_1[k+2], ..., s_p[k+T_p])$$

$$C = \{s_1[k], s_1[k-1], ..., s_p[k-L_p]\}$$

j^{th} embedding cycle:

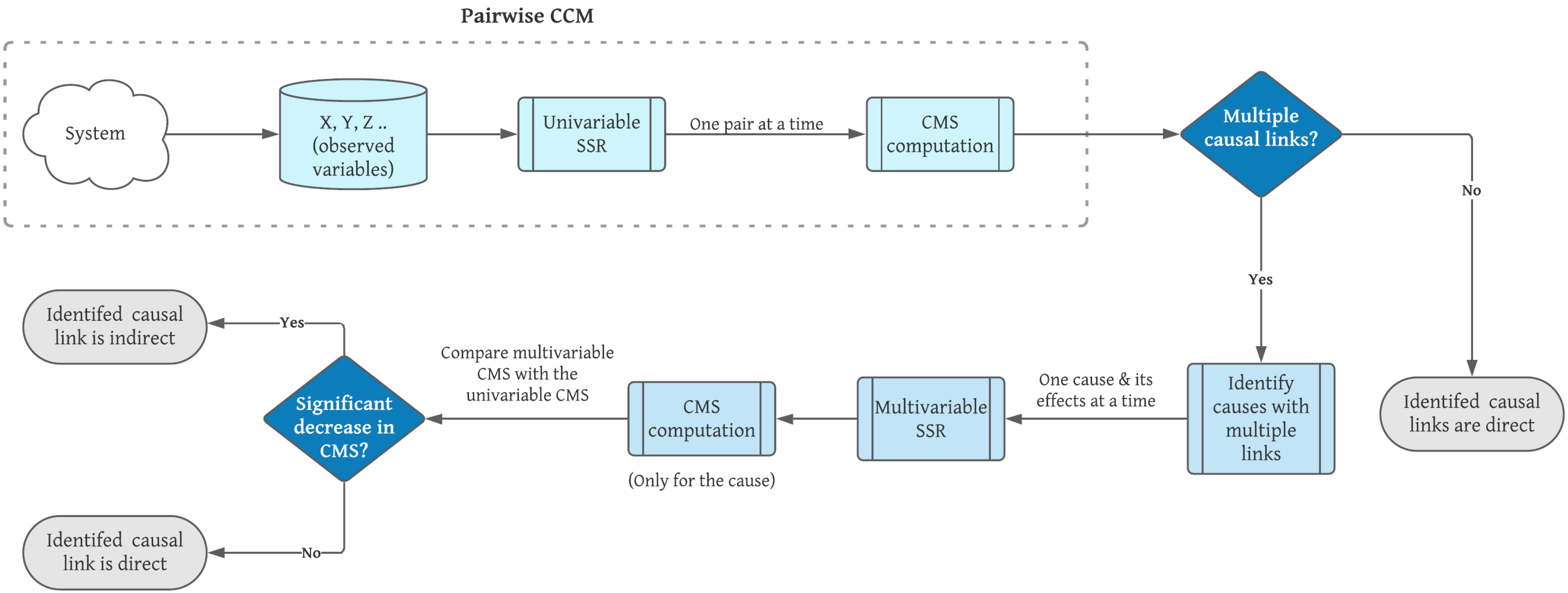
- $c_j \leftarrow \max_{c_i \in C} I(v_F, c_i | \mathcal{M}_{j-1})$
- Stop iteration when $I(v_F, \mathcal{M}_{j-1})/I(v_F, \mathcal{M}_j) > A$

M' must be such that

1. its components are least dependent on each other
2. it best explains the dynamics of the system, i.e. has knowledge of one or several time steps into the future

Vlachos, Ioannis, and Dimitris Kugiumtzis. Nonuniform State-Space Reconstruction and Coupling Detection. Physical Review E, vol. 82, no. 1, July 2010, p. 016207. DOI.org (Crossref), doi:10.1103/PhysRevE.82.016207.

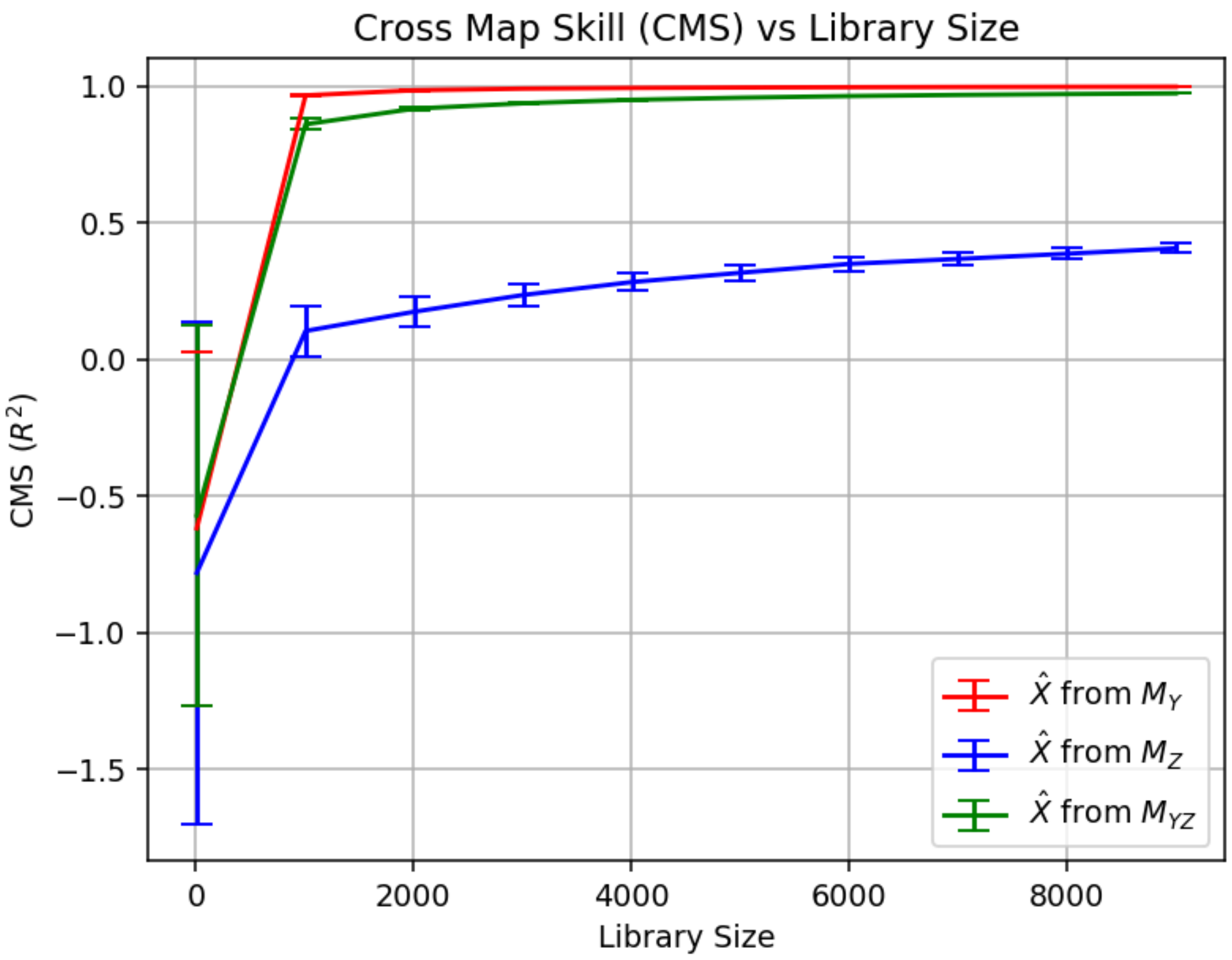
Multivariable CCM

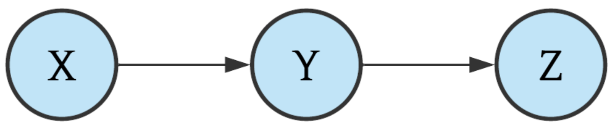


- The multivariable state-space reconstruction (SSR) method by Vlachos *et al* (2010) is used to reconstruct the causal variable of interest from multiple effects
- Improvement in the reconstructions from individual effects vs. combination of effects is measured

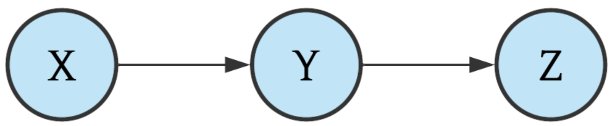
S. Nithya and A.K. Tangirala (2021). Multivariable Causal Analysis of Nonlinear Dynamical Systems using Convergent Cross Mapping, 7th Indian Control Conference, ICC 2021 Proceedings, 436–441

Multivariate Causal Discovery in NL DD Systems





Original causal structure



Reconstructed causal structure

- A significant improvement is seen in the reconstructed cause $\hat{X} \leftarrow M_Z$ as compared to $\hat{X} \leftarrow M_{YZ}$
- However, no significant improvement is seen in $\hat{X} \leftarrow M_Y$ vs $\hat{X} \leftarrow M_{YZ}$
- Thus, X directly affects Y and indirectly affects Z

Objective

- To investigate the spatiotemporal evolution of synchronous/concurrent rainfall extremes during the Indian summer monsoon over the past century, employing complex networks (CN) and event synchronization techniques to identify regions of high connectivity.

Data

Primary data: 0.25° x 0.25° daily rainfall gridded dataset for a time period of 119 years (1901-2019) has been provided by India Meteorological Department (IMD)

For validation: APHRODITE [1951-2015] **(grid-matched at same resolution)**

Spatiotemporal Patterns of Synchronous Extremes

Objective

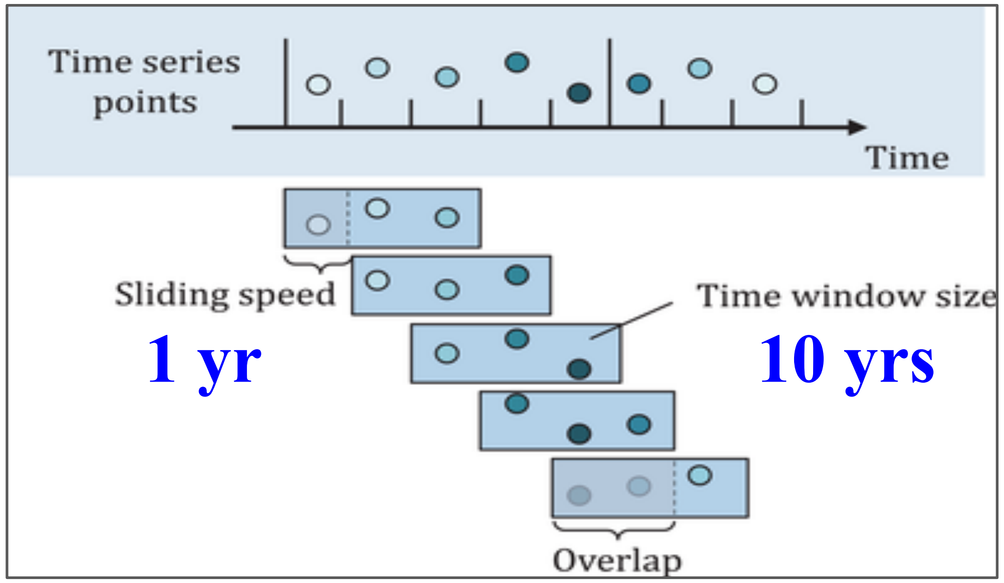
- To investigate the spatiotemporal evolution of synchronous/concurrent rainfall extremes during the Indian summer monsoon over the past century, employing complex networks (CN) and event synchronization techniques to identify regions of high connectivity.

Data

Primary data: 0.25° x 0.25° daily rainfall gridded dataset for a time period of 119 years (1901-2019) has been provided by India Meteorological Department (IMD)

For validation: APHRODITE [1951-2015] **(grid-matched at same resolution)**

Moving Window Approach



Schematic representation of moving window

DETERMINING AND MINING THE SYNC CLIMNET

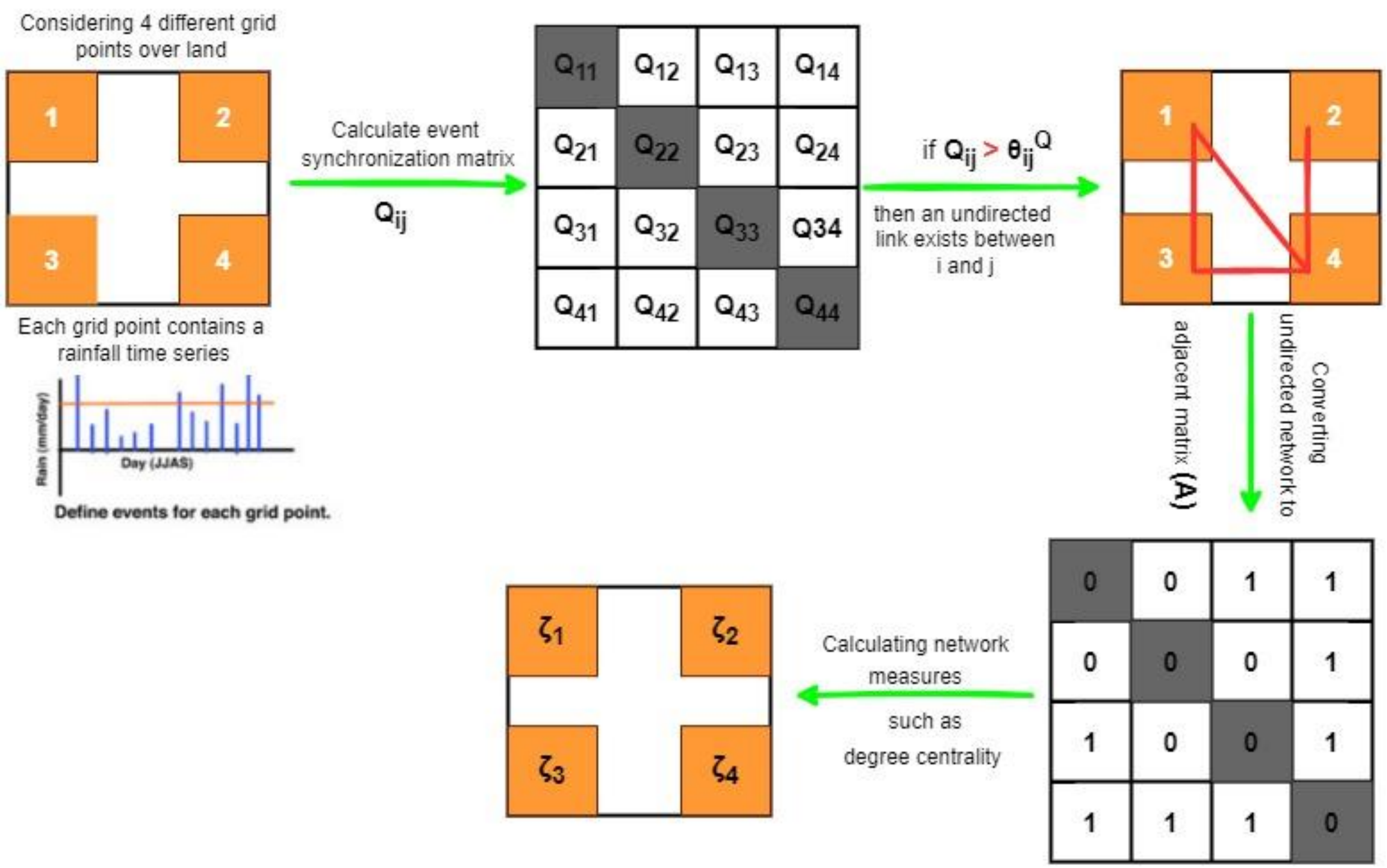
- Extreme Event Series:

  Series of events that are above a given threshold.

- Event Synchronization:

  A nonlinear measure to quantify the synchronicity of rainfall events at different grid points and their delay behaviour.

- Network Construction:
  - Adjacency matrix
  - Network measures



Schematic flow diagram for determining and mining the climate network

Methodology: Event Synchronization

$$C(i/j) = \sum_{l=1}^{s_i} \sum_{m=1}^{s_j} J_{ij}$$

$$J_{ij} = \begin{cases} 1 & \text{if } 0 < t_l^i - t_m^j < \tau_{lm}^{ij} \\ 1/2 & \text{if } t_l^i = t_m^j \\ 0 & \text{else,} \end{cases}$$

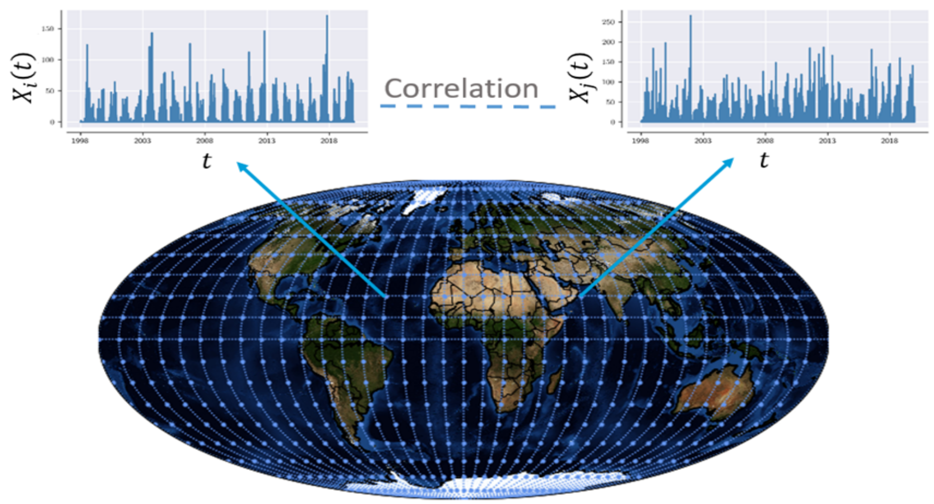
Time lag

$$\tau_{lm}^{ij} = min(t_{l+1}^i - t_l^i, t_l^i - t_{l-1}^i, t_{m+1}^j - t_m^j, t_m^j - t_{m-1}^j)/2$$

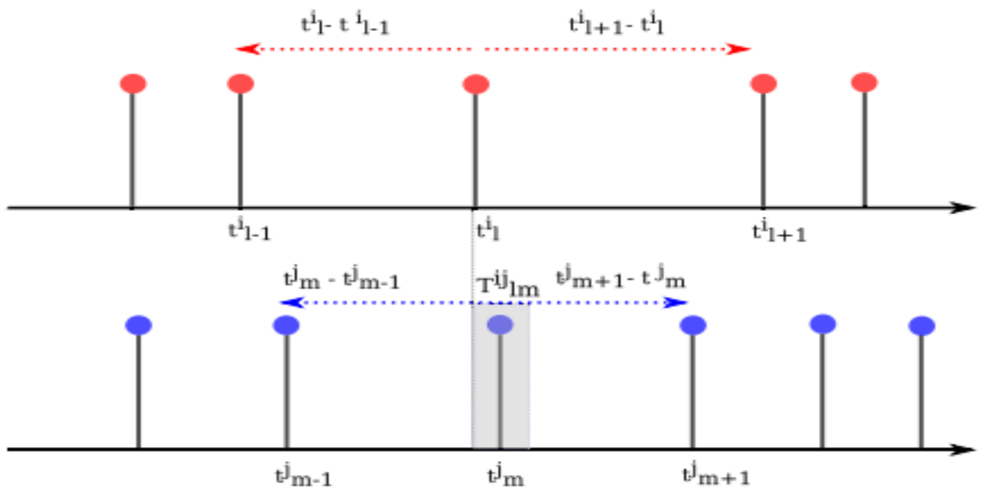
Strength of synchronization

$$Q_{ij} = \frac{c(i/j) + c(j/i)}{\sqrt{s_i s_j}}$$

- c(i/j) = No.of times the event occurs at i, once it has already appeared at j
- Q =1 (complete synchronization) and Q=0 (no synchronization)
- s_i and s_j total number of events at the i^{th} and j^{th} grid sites



Schematic representation of two time series.

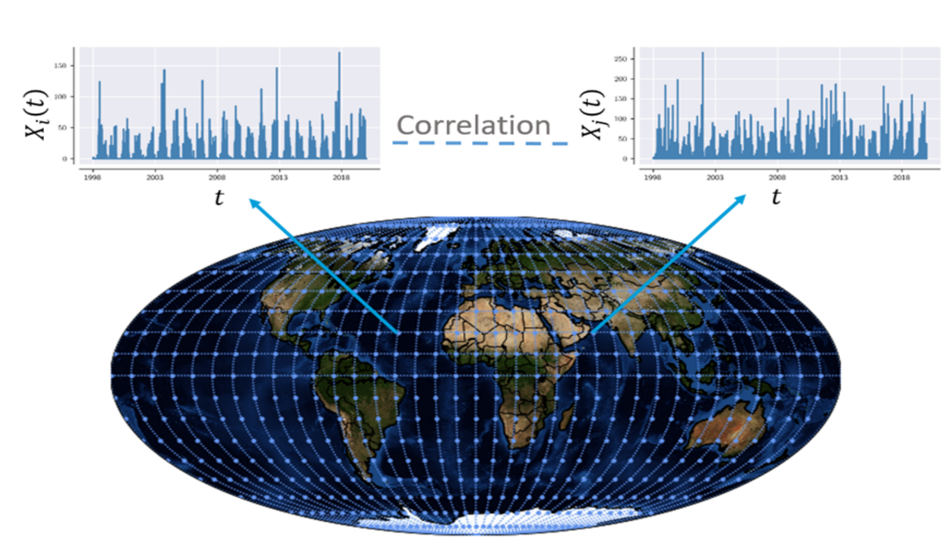


Schematic representation of synchronization between two event series

Malik. et al., 2012, Climate dynamics

METHODOLOGY: EVENT SYNCHRONIZATION

$$C(i/j) = \sum_{l=1}^{s_i} \sum_{m=1}^{s_j} J_{ij} \qquad J_{ij} = \begin{cases} 1 & \text{if } 0 < t_l^i - t_m^j < \tau_{lm}^{ij} \\ 1/2 & \text{if } t_l^i = t_m^j \\ 0 & \text{else,} \end{cases}$$



Schematic representation of two time series.

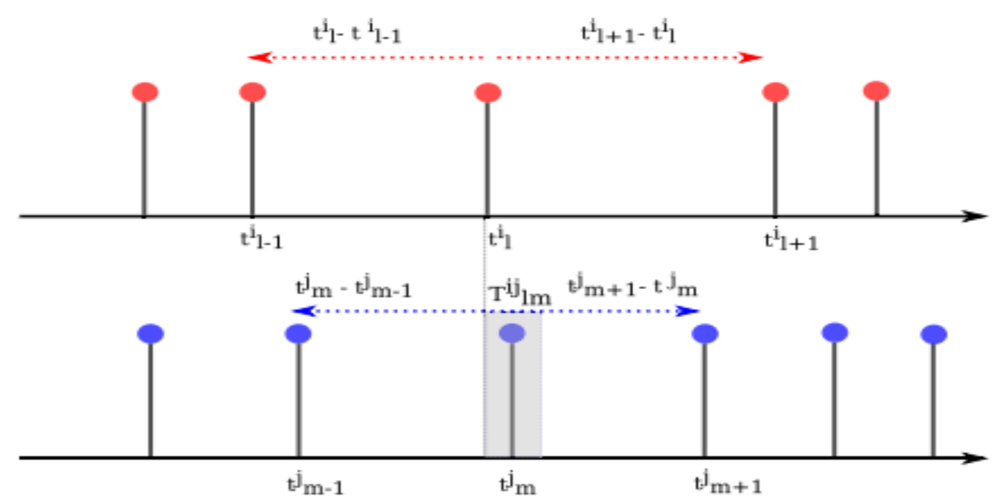
Time lag

$$\tau_{lm}^{ij} = min(t_{l+1}^i - t_l^i, t_l^i - t_{l-1}^i, t_{m+1}^j - t_m^j, t_m^j - t_{m-1}^j)/2$$

Strength of synchronization

$$Q_{ij} = \frac{c(i/j)+c(j/i)}{\sqrt{s_i s_j}}$$

- c(i/j) = No.of times the event occurs at i, once it has already appeared at j
- Q =1 (complete synchronization) and Q=0 (no synchronization)
- s_i and s_j total number of events at the i^{th} and j^{th} grid sites



Schematic representation of synchronization between two event series

Adjacency Matrix

$$A_{ij} = \begin{cases} 1 & \text{if } Q_{ij} > \theta_{ij}^Q \\ 0 & \text{else,} \end{cases}$$

- Only 5% strongest connections were used in order to deal with only highly synchronous event.

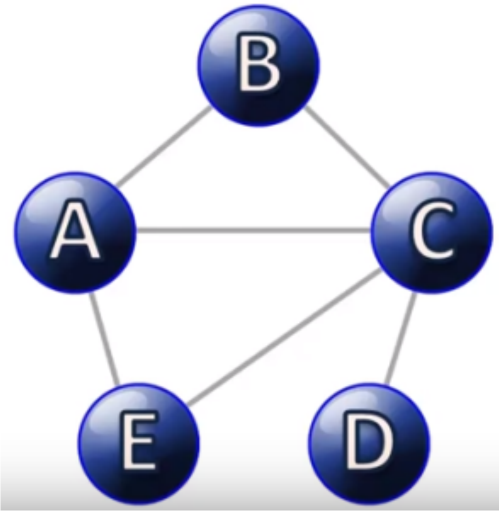
Malik. et al., 2012, *Climate dynamics*

Network Topological Measures

1. Degree Centrality of a node 'j' is:

$$DC_j = \frac{\sum_{i=1}^{N} A_{ij}}{N-1}$$

- Number of grids displaying **synchrony** in the occurrence of extreme rainfall events with respect to a node.
- Higher the degree (or DC), higher the connections of a node.



to

from

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

2. Local clustering coeff. of a given node 'i':

$$LCC_i = \frac{2L_i}{k_i(k_i-1)}$$

where L_i is represents the number of links between k_i neighbours of node i.

- LCC measures a regional connectivity among neighboring grid points.
- **Mean or global clustering coefficient** is average of LCC_i

3. Geographical link length between two connected nodes (i,j):

$$L_{ij} = R\sqrt{(\delta\phi_{ij})^2 + (cos(\phi_m)\delta\lambda_{ij})^2}$$

- *where* $\boldsymbol{\delta\phi_{ij}}$ *and* $\boldsymbol{\delta\lambda_{ij}}$ *are the differences in latitude and longitude in radians between grid points i and j.*
- ϕ_m *is the mean of latitudes of i and j. R is the radius of earth.*
- The average link length of the a given node i is average of L_{ij} of i^{th} nodes links

SYNCHRONIZATION OF RAINFALL EXTREMES: DEGREE CENTRALITY

- Rainfall extremes (REs) are highly synchronized and spatially localized over the **central India** (CI)
- Spatiotemporal clustering of REs over CI are associated with the synoptic events such as LPS that get clustered over it (***Krishnamurthy, V., et al., 2010 ,Thomas et al.,2021***)
- Over CI, around 60% (to 80% of rainfall extremes are associated with LPS (***Krishnamurthy, V., et al., 2010***)
- Northeast India (NEI) and Western Ghats (WG) shows low degree centrality (*although high frequency of rainfall extremes*) due to;
  - Lag in the monsoonal rainfall between NEI and the rest of the country (***Rajeevan, M.,et al.,2010***)
  - Barrier of high mountains.
- The degree distributions are unimodal with peak at 150 (of degree) for both 90th%ile and 95th%ile rainfall extremes.

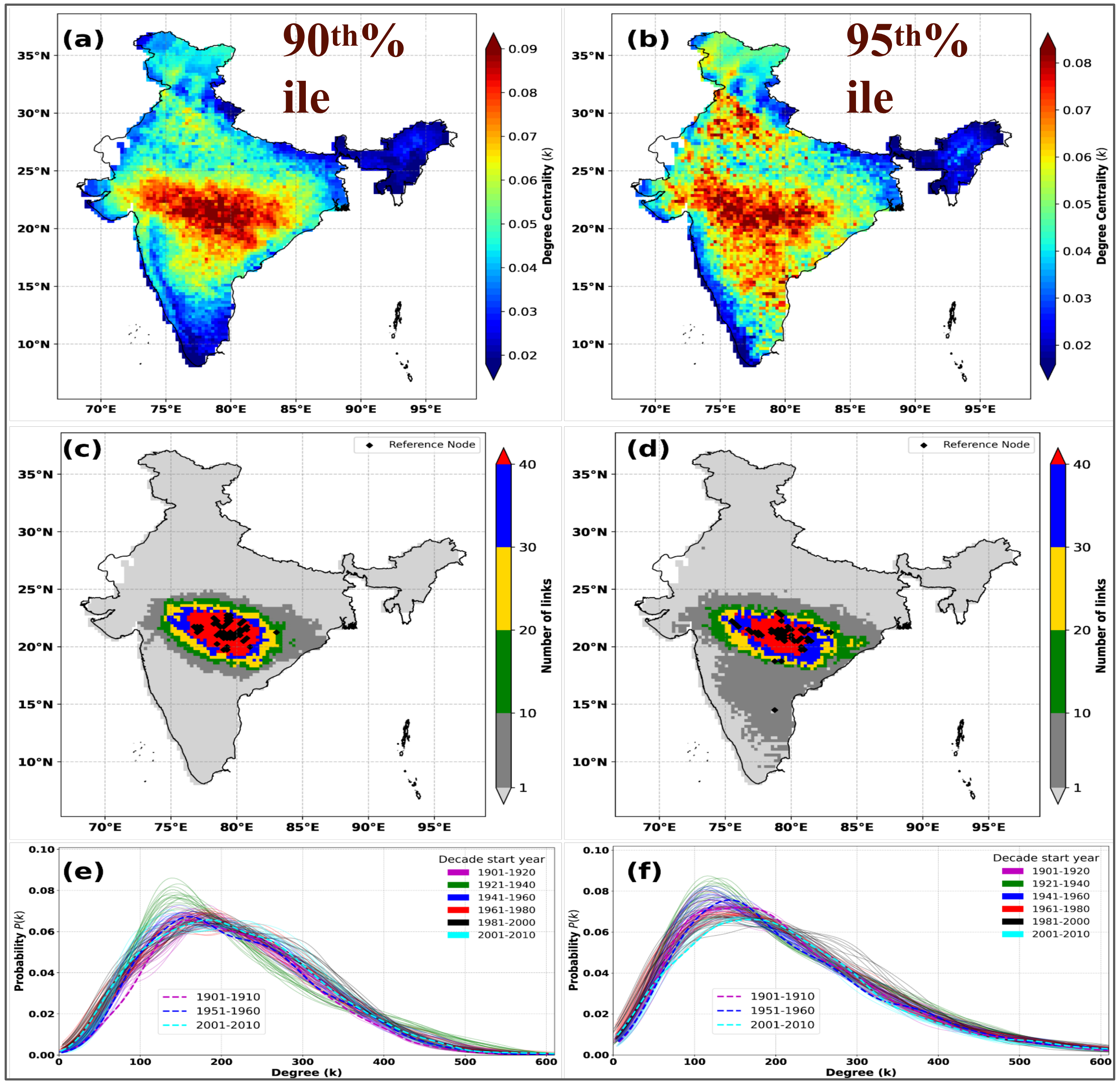


Fig 11: Degree centrality and its distribution

Statistical Testing of Degree Distributions

- The spatial average degree centrality and its std.dev. does not show any significant increasing or decreasing trend (p>0.01)
- In order to check the divergence of degree distribution over different decades:

- ***Kolmogorov-Smirnov (KS)*** *test to check whether degree distributions are identical or not.*
- ***Jeffreys divergence*** *test to measure the extent of divergence between the degree distribution.*

- **Although the distributions are not identical, the divergence of one distribution from others is very low.**

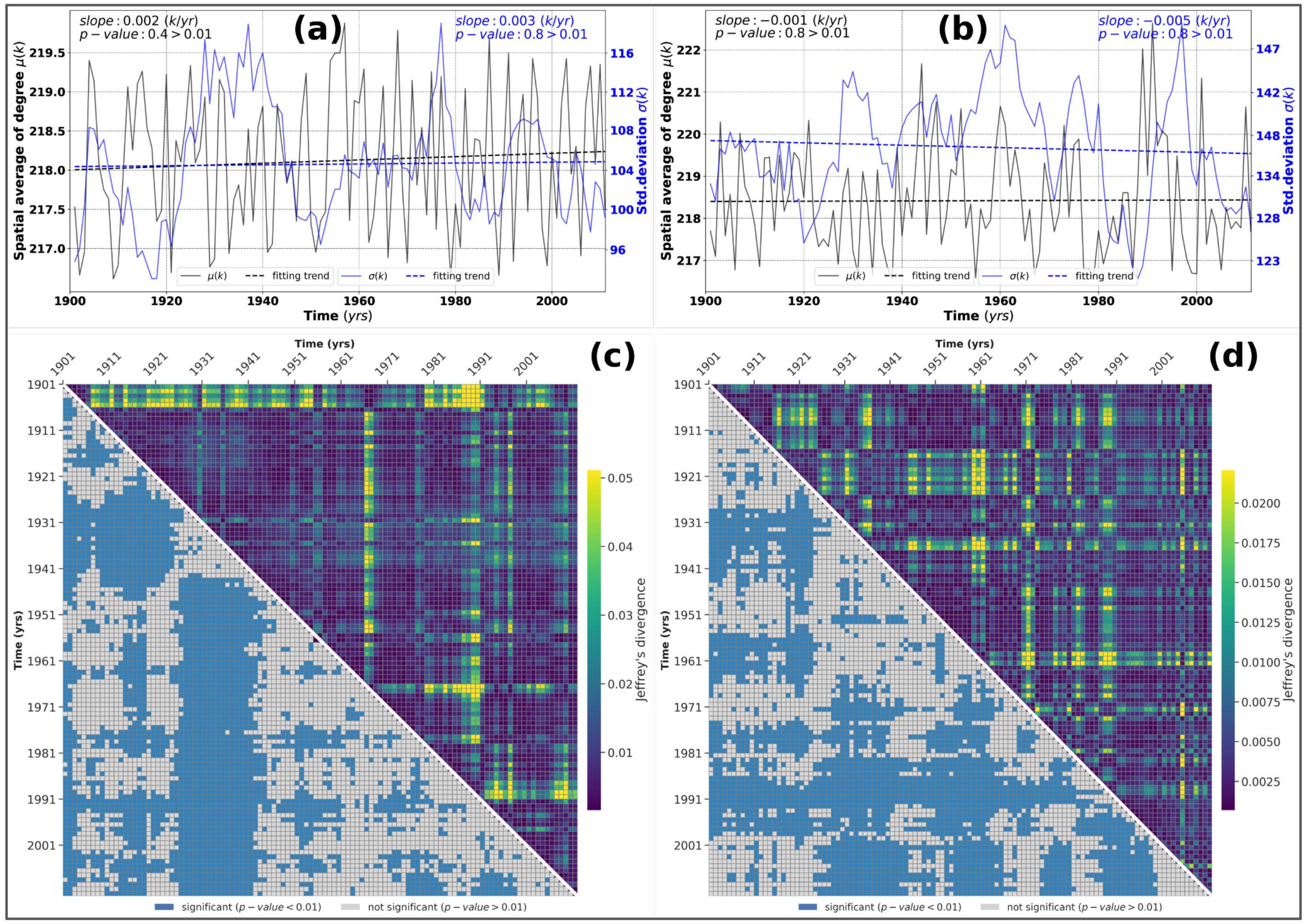
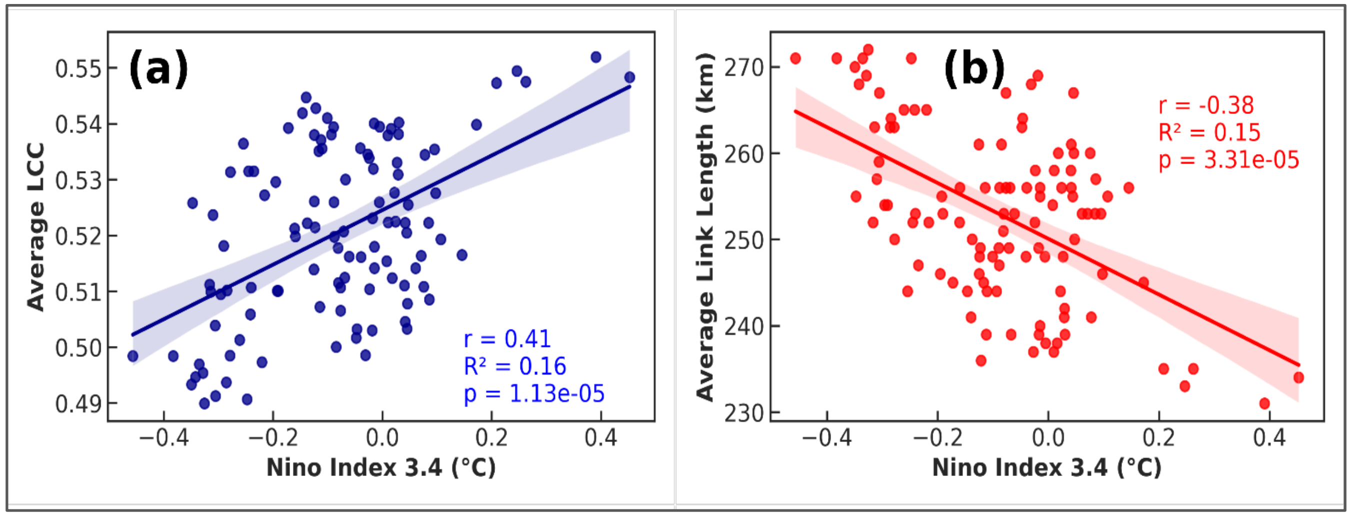
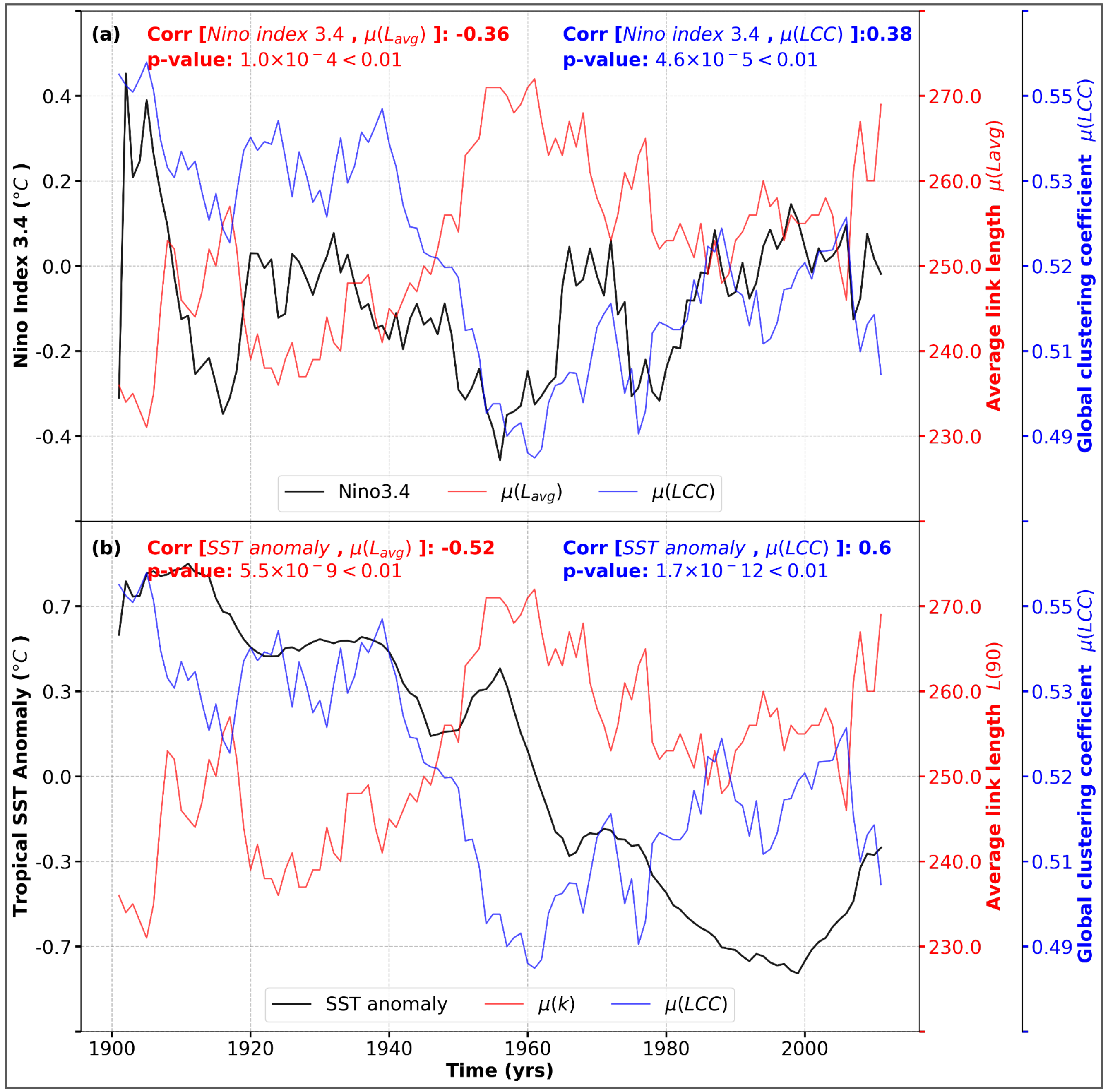


Fig 14: Spatial average of degree and divergence of degree distributions

Relationship Between N/W Measures & SST Anomalies



Regression between ENSO Index and topological measures



Relationship between tropical SST and topological measures

ENSO (Nino-Index 3.4) shows a significant ($p<0.01$) correlation of 0.38 and -0.36 with GCC and LL respectively.

- **La-Nina-phase:** GCC decreases and LL increases, means ***synchronous rainfall extremes occur at larger scale.***
- **El-Nino-phase**: GCC increases and LL decreases , which implies ***localization of synchronous rainfall extremes.***

Acknowledgements

- PhD, MS, M. Tech. and UG *students at IITM and other institutions*
- Collaborators across institutes in India
  - Prof. Udit Bhatia (IITGN)
- Dept. of Chemical Engg., Dept. of DS & AI, RBCDSAI (IITM) & IITM
- IIT Tirupati
- Funding agencies

Questions?