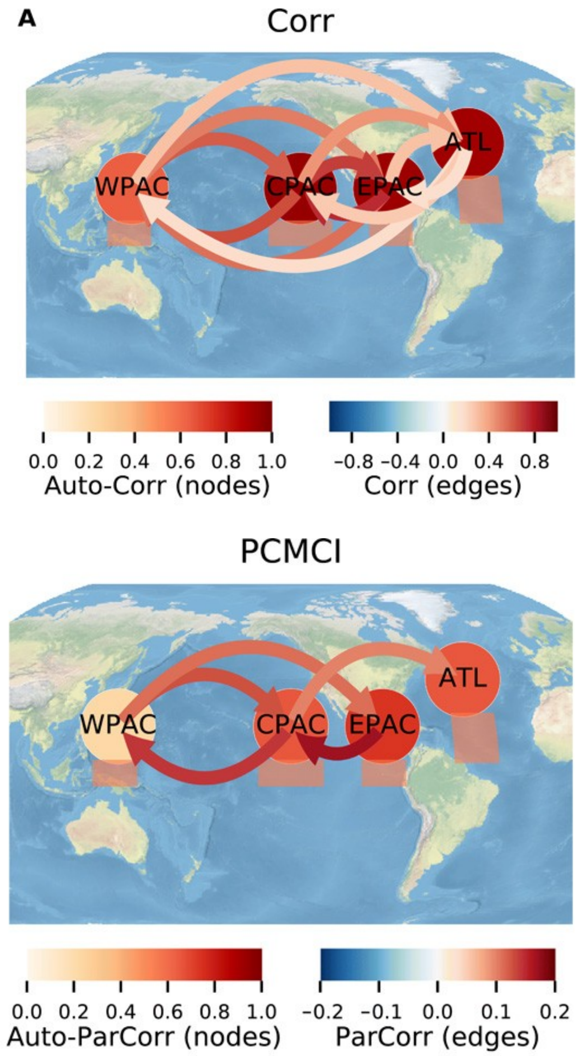
Spatiotemporal model for benchmarking causal discovery algorithms

Xavier-Andoni Tibau[1], Christian Reimers[1,2], Veronika Eyring[3,4], Joachim Denzler[2,5], Markus Reichstein[6], and Jakob Runge[1]

[1] German Aerospace Center (DLR), Institute of Data Science, Jena, Germany
[2] Computer Vision Group, Friedrich-Schiller-Universität Jena, Germany
[3] Institute for Atmospheric Physics, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[4] Institute of Environmental Physics, University of Bremen, Bremen, Germany
[5] Michael Stifel Center Jena for data-driven and simulation science, Jena
[6] Max-Planck-Institute for Biogeochemistry, Jena, Germany



Runge et al. 2019 *Science Advances*

Causality on climate and weather

Motivation:

To understand weather and climate forecasting the causal understanding of climate interactions is vital

Observational causal inference is a major current topic in machine learning as well as other domains

Problem:

There is no ground truth in climate and weather data for emergent properties as modes of variability and teleconnections

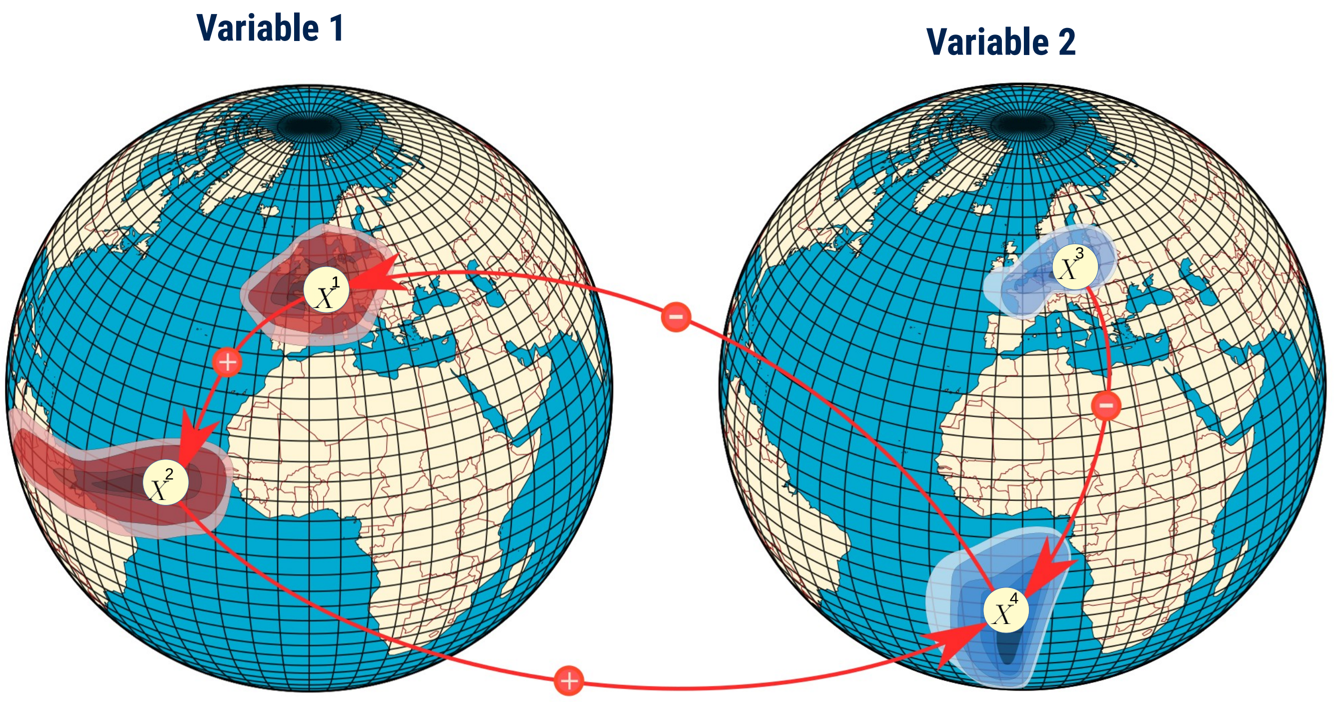
Our solution

Spatially Averaged Vector AutoRegressive model (SAVAR)

A spatiotemporal model system that encodes causal relationships among well-defined modes of variability

The model can be thought of as an extension of Vector AutoRegressive models well-known in time series analysis

What is a SAVAR model



- **Spatio-temporal model**
  With dynamics in space and time
- **High dimensional**
  Several grid-points + climate variables
- **Aggregate variables**
  Lower dimensional representation of the variables
- **Causal dynamics**
  A causal model for latent variables

What is a SAVAR model

Modes of variability or latent variables

$$x_t^i = \sum_{\ell}^{L} w^{i\ell} y_t^\ell$$

Underlying causal model

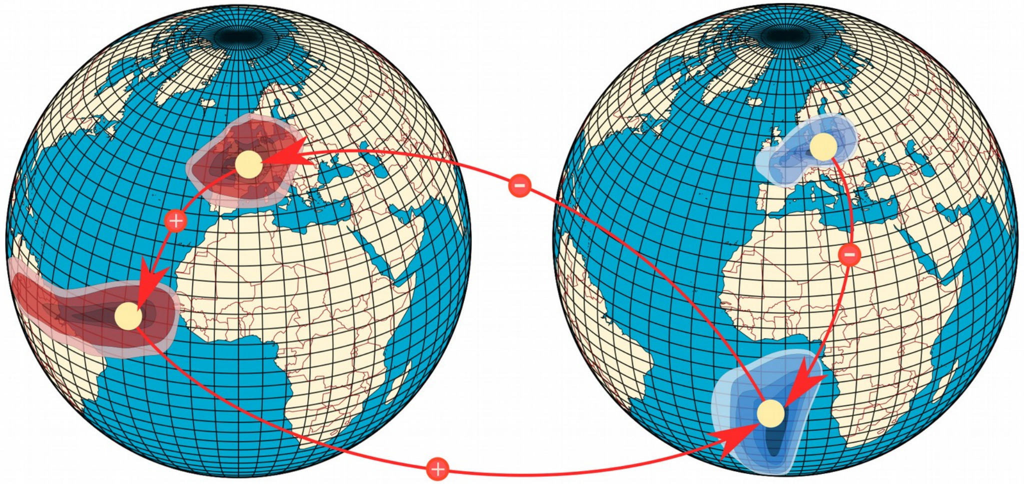
$$x_t^j = \sum_{i=1}^{N} \sum_{\tau=1}^{\tau_{max}} \phi^{ji}(\tau) x_{t-\tau}^i$$

A VAR(p)

SAVAR

$$y_t^\ell := \sum_{j=1}^{N} u^{\ell j} \sum_{i=1}^{N} \sum_{\tau=1}^{\tau_{max}} \phi^{ji}(\tau) \sum_{\ell=1}^{L} w^{i\ell} y_{t-\tau}^\ell + \epsilon_t^\ell$$

$$\mathbf{y}_t = W^+ \sum_{\tau=1}^{\tau_{max}} \Phi(\tau) W \mathbf{y}_{t-\tau} + \epsilon_t$$



$$WW^+W = W$$

$\mathrm{y}_t^T = (y_t^1, \ldots, y_t^L)$, $u^{\ell j} \in W^+$, $\phi^{ij}(\tau) \in \Phi(\tau)$, $w^{i\ell} \in W$ and $\epsilon_t^\ell \in \epsilon_t$

What is a SAVAR model

Y stands for the variables at grid level
X stands for aggregated variables
W stands for mode weights
Epsilon stands for the noise term

$$\mathbf{y}_t = W^+ \sum_{\tau=1}^{\tau_{max}} \Phi(\tau) W \mathbf{y}_{t-\tau} + \epsilon_t$$

The noise term follows a multivariate Gaussian
Distribution with the same spatial distribution of the modes, this is:

$$\epsilon_t \sim N(0, \Sigma) \qquad \Sigma = W^+ (W^+)^\intercal + \mathbf{I}_\sigma$$

Mathematical properties

Proposition 1
If a VAR(p) with a coefficient matrix Φ is stationary then a SAVAR model with the same coefficient matrix Φ is also stationary, independently of W

Stationarity

$$\forall\, t: \; \mathbb{E}(\mathbf{y}_t) = \overline{\mathbf{y}}$$

$$\forall\, t, \tau: \; \mathbb{E}[(\mathbf{y}_t - \overline{\mathbf{y}})(\mathbf{y}_{t-\tau} - \overline{\mathbf{y}})^T] = \Omega(\tau)$$

Proposition 2
Given a stationary SAVAR process, if $\epsilon_t \sim \mathcal{N}(0, \Sigma)$ then $\mathbb{E}(\mathbf{y}_t) = 0$

$$\mathbf{y}_t = W^{+} \sum_{\tau=1}^{\tau_{max}} \Phi(\tau) W \mathbf{y}_{t-\tau} + \epsilon_t$$

Mathematical properties

Reduced from of SAVAR
SAVAR can be expressed as a VAR(1) model:

$$\mathbf{y}_{t:\tau_{max}} = A_{\tilde{\Phi}} \mathbf{y}_{t-1:\tau_{max}} + e_t$$

Autocovariance function of SAVAR

$$\Omega(\tau) = A_{\tilde{\Phi}}^{\tau} \Omega(0)$$

Preposition 3
Given a reduced form of a SAVAR process with coefficient matrix $A_{\tilde{\Phi}}$ it is possible to identify A_{Φ} from a VAR(p) process up to **similarity**

$$B = P^{-1} A P$$

Similarity: A and B are similar if P exists. Then, they share the characteristic polynomial.

$$A_{\tilde{\Phi}} = \begin{pmatrix} \tilde{\Phi}(1) & \tilde{\Phi}(2) & \cdots & \tilde{\Phi}(\tau_{max}-1) & \tilde{\Phi}(\tau_{max}) \\ 1 & 0 & \ddots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

$$\mathbf{y}_t = W^{+} \sum_{\tau=1}^{\tau_{max}} \Phi(\tau) W \mathbf{y}_{t-\tau} + \epsilon_t$$

$$\tilde{\Phi}(\tau) = W^{+} \Phi(\tau) W$$

$$\mathbf{y}_{t:\tau_{max}}^{T} = (y_t^1, y_t^2, \ldots, y_t^L, y_{t-1}^1, \ldots, y_{t-\tau_{max}}^L)$$

Goal of experiments

Show the performance of different Causal Discovery algorithms that involve dimensionality reduction steps

Methods

	Dimensionality Reduction	Link estmation	Link coefficient estimation
Method 1	Varimax	Unconditional Correlation	Univariate linear regression
Method 2	Varimax	PCMCI	Multivariate Linear regression
Method 3	PCA	Unconditional Correlation	Univariate linear regression
Method 4	PCA	PCMCI	Multivariate Linear regression

Metrics

Mean absolute correlation coefficient of weights $= \frac{1}{I}\sum_{i}^{I}|\rho(W_i, \tilde{W}_i)|$

Mean absolute correlation coefficient of Aggregated variables $= \frac{1}{I}\sum_{i}^{I}|\rho(X_i, \tilde{X}_i)|$

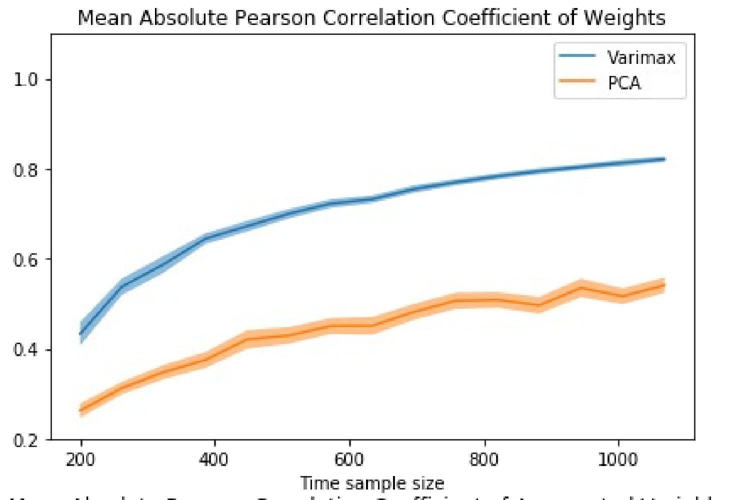
Mean Square Error $= \frac{1}{N}\sum_{i,j,\tau}(\phi^{ji}(\tau) - \tilde{\phi}^{ji}(\tau))^2, \forall\phi^{ji}(\tau) \neq 0$

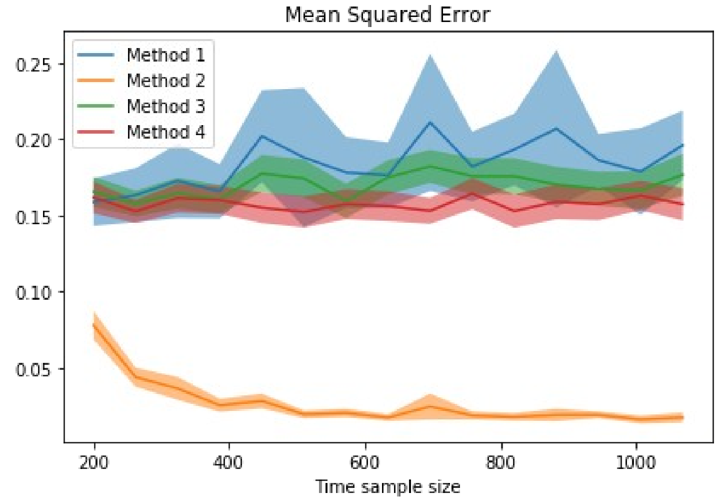
Relative Mean Absolute Error $= \frac{1}{N}\sum_{i,j,\tau}\frac{|\phi^{ji}(\tau) - \tilde{\phi}^{ji}(\tau)|}{|\phi^{ji}(\tau)|}, \forall\phi^{ji}(\tau) \neq 0$

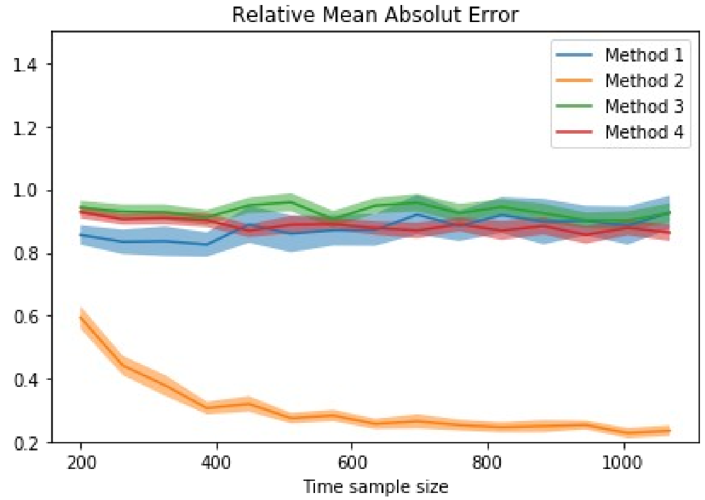
Precision of causal graph $= \frac{TP}{TP + FP}$

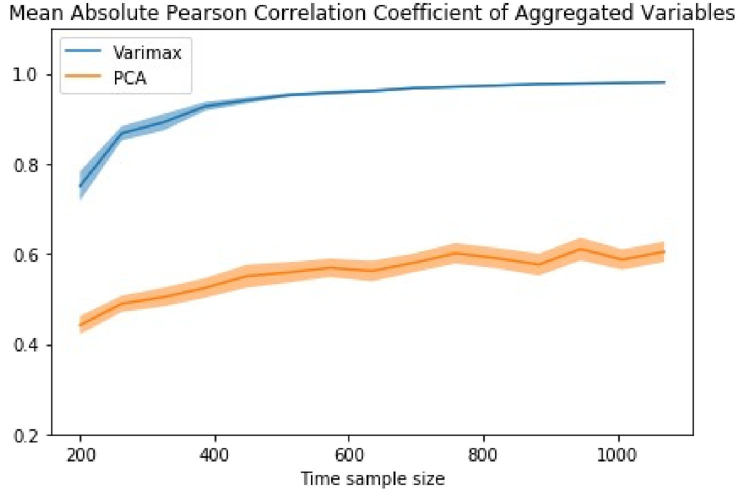
Recall of causal graph $= \frac{TP}{TP + FN}$

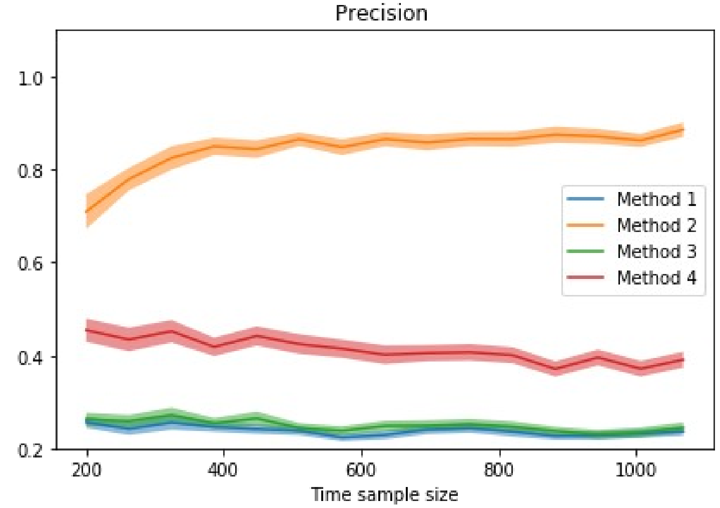
Time sample size X axis is the number of time sample sizes

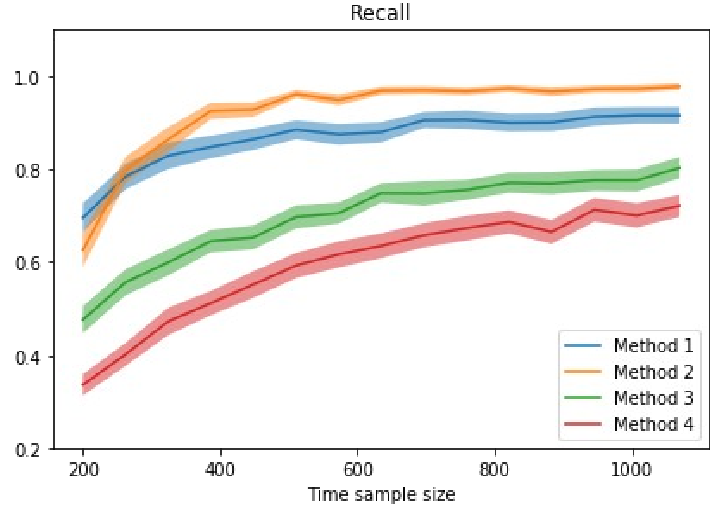






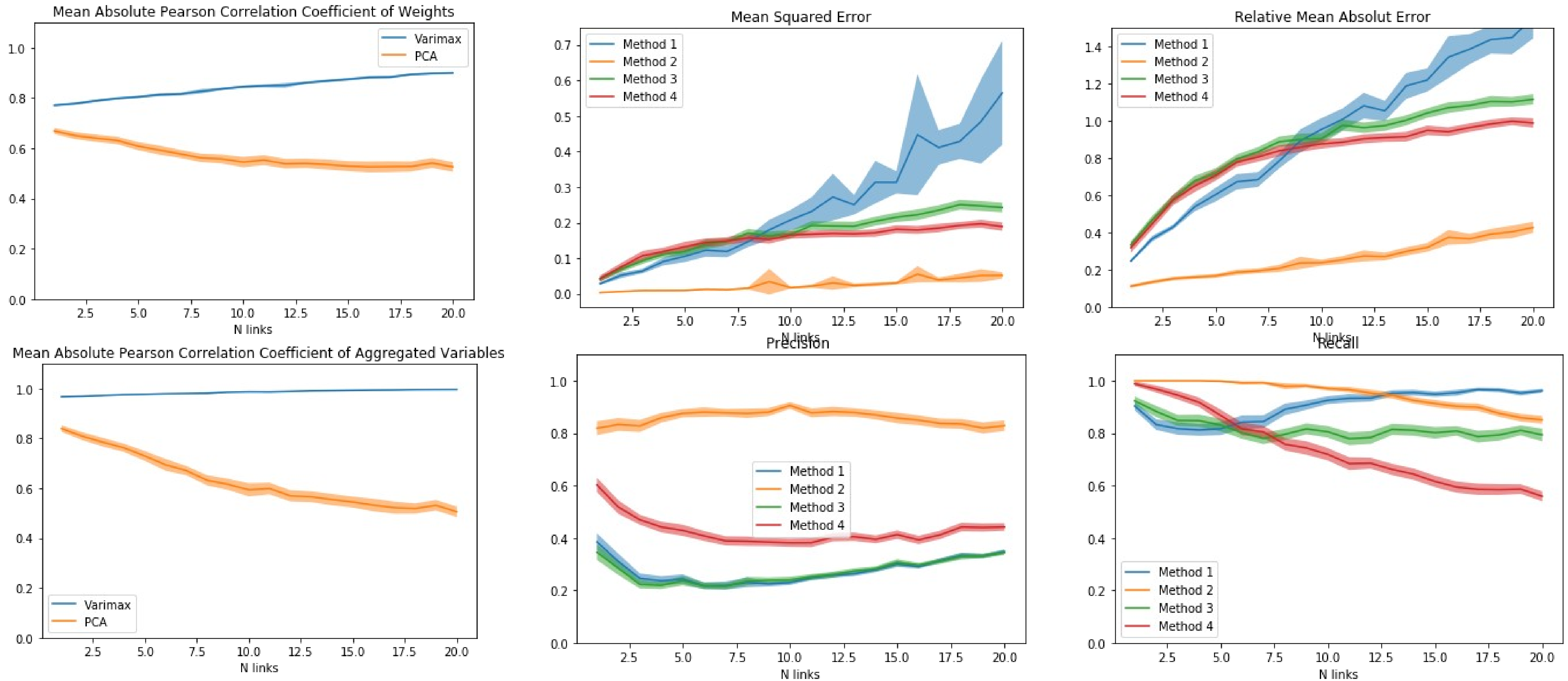






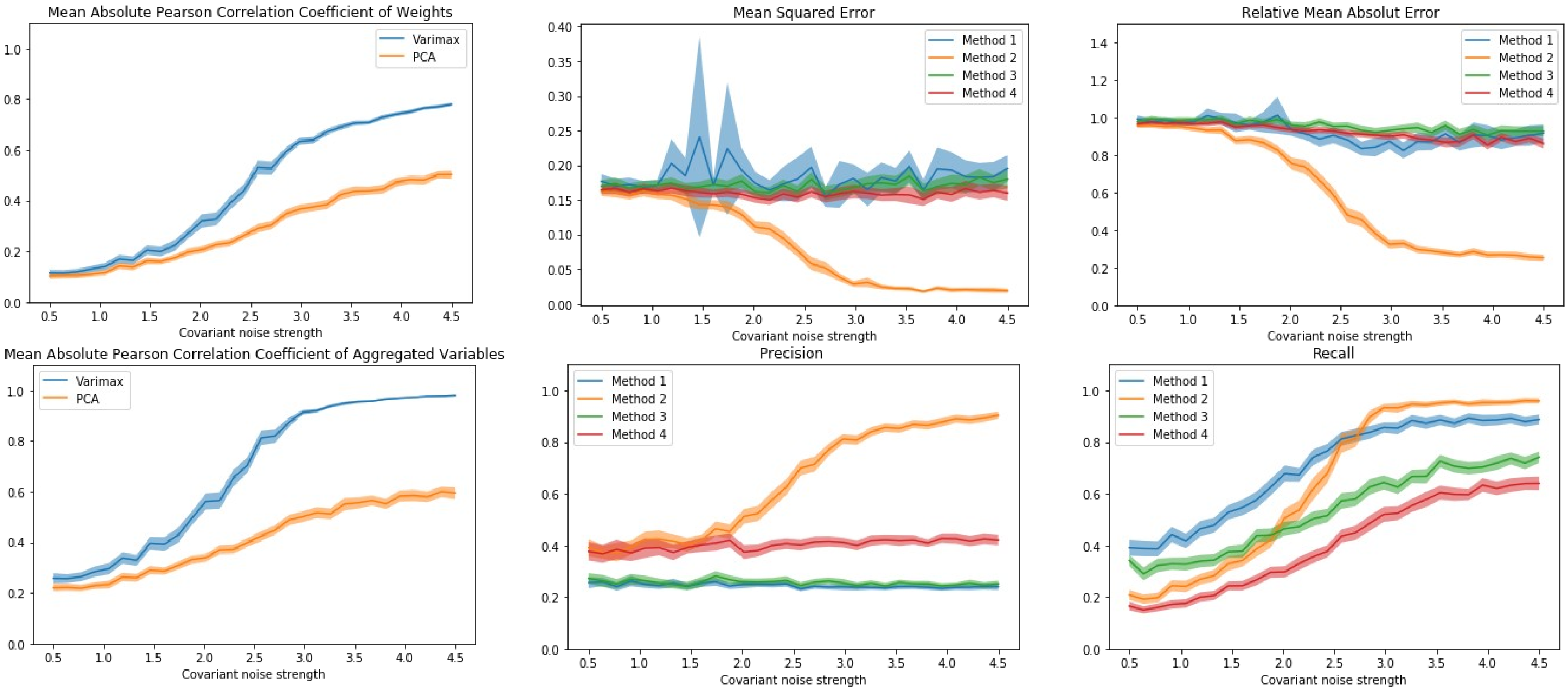
Each experiment has been done 100 times. Shadow areas show 95% confidence level

Number of links X axis is the number of links of the model



Each experiment has been done 100 times. Shadow areas show 95% confidence level

Noise strength X axis is the strength of the spatial covariance of noise term



Each experiment has been done 100 times. Shadow areas show 95% confidence level

Remarks

Causal inference is relevant for understanding climate and weather systems

Climate models and observations have no ground truth for emergent properties such as modes of variability

SAVAR model is a good representation of climate modes of variability

SAVAR has similar properties as VAR(p)

SAVAR can be used to create benchmark data for causal discovery algorithms