EGU2020-9604
EGU General Assembly 2020

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 (xavier.tibau@dlr.de)
[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

We propose a spatiotemporal model system to evaluate methods of causal discovery. The use of causal discovery to improve our understanding of the spatiotemporal complex system Earth has become widespread in recent years (Runge et al., Nature Comm. 2019). A widespread application example are the complex teleconnections among major climate modes of variability.

The challenges in estimating such causal teleconnection networks are given by (1) the requirement to reconstruct the climate modes from gridded climate fields (dimensionality reduction) and (2) by general challenges for causal discovery, for instance, high dimensionality and nonlinearity. Both challenges are currently being tackled independently. Both dimensionality reduction methods and causal discovery have made strong progress in recent years, but the interaction between the two has not yet been much tackled so far. Thanks to projects like CMIP a vast amount of climate data is available. In climate models climate modes of variability emerge as macroscale features and it is challenging to objectively benchmark both dimension reduction and causal discovery methods since there is no ground truth for such emergent properties.

We propose 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. This model provides a framework for experimenting with causal discovery in large spatiotemporal models. For example, researchers can analyze how the performance of an algorithm is affected under different methods of dimensionality reduction and algorithms for causal discovery. Also challenging features such as non-stationarity and regime-dependence can be modelled and evaluated. Such a model will help the scientific community to improve methods of causal discovery for climate science.

Runge, J., S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, E. H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, and J. Zscheischler (2019). Inferring causation from time series in earth system sciences. Nature Communications 10 (1), 2553.