Causal Structure of Earth: Learning System Dynamics with Hypergraphs

Abstract

Abrupt shifts in the Earth system such as Amazon forest dieback, Arctic ice loss, or monsoon collapse arise from complex teleconnections [Beamish, 1995], where variability in one region drives ecological responses elsewhere. Standard causal discovery methods, including Granger causality and PCMCI+, represent dependencies with pairwise graphs . While effective for low-dimensional signals, they cannot capture the higher-order interactions that often trigger regime shifts, for example when both tropical Atlantic warming and Pacific circulation anomalies jointly cause Amazon drought [Marengo et al., 2008].

We introduce a framework that models causal structure as a hypergraph [Ma et al., 2022] $H = (V, \mathcal{E})$, where $\mathcal{E} \subseteq 2^V \setminus \{\varnothing\}$ encodes group-level edges linking subsets of variables to outcomes. For any subset $S \subseteq V$, we define the set-wise causal effect

$$\tau(S \to Y; x_S, x'_S) = \mathbb{E}[Y \mid \mathrm{do}(X_S = x_S)] - \mathbb{E}[Y \mid \mathrm{do}(X_S = x'_S)],$$

which generalizes pairwise interventions to multi-variable causes. This representation allows us to test whether sets of factors rather than individual variables drive critical transitions.

Our method extends constraint-based and score-based discovery by lifting conditional independence tests and sparsity penalties from pairwise to set-wise relations. Candidate hyper-edges are screened via conditional mutual information $I(Y; X_S \mid Z)$ over spatio-temporal contexts, and group-sparse penalties control combinatorial growth. We implement this directly on multi-scale latent features from a recent Earth-system foundation models, enabling discovery under non-stationarity and across spatial scales.

To address data-poor regions such as the Arctic, polar oceans, or tropical rainforests, we integrate hypergraph discovery with causal transportability [Pearl, 2010]. By constructing selection diagrams that encode cross-domain differences, we identify conditions under which causal relations transfer from data-rich basins (e.g., the Pacific Ocean) to under-observed ecosystems. This extends counterfactual reasoning [Pearl, 2010] to regions where direct observations are limited, enabling forecasts of coral reef collapse, Arctic ice decline, or rainforest tipping points.

Preliminary results show that hypergraph discovery (i) recovers known teleconnections such as ENSO–East Africa rainfall and Arctic –midlatitude winter links , (ii) reveals higher-order interactions (triplets and quartets) that outperform pairwise models in explaining regime shifts, and (iii) enables counterfactual experiments that assess scenarios such as reduced Arctic warming trajectories.

By moving beyond pairwise causality, this work demonstrates that hypergraph-based causal discovery in foundation models yields more interpretable and actionable forecasts of cascading climate risks, offering a new pathway to earlier warnings and more targeted adaptation strategies.

References

Richard James Beamish. *Climate change and northern fish populations.* Number 121. NRC Research Press, 1995.

Jing Ma, Mengting Wan, Longqi Yang, Jundong Li, Brent Hecht, and Jaime Teevan. Learning causal effects on hypergraphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1202–1212, 2022.

JA Marengo, CA Nobre, John Tomasella, MF Cardoso, and MD Oyama. Hydro-climatic and ecological behaviour of the drought of amazonia in 2005. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1498):1773–1778, 2008.

Judea Pearl. Causal inference. *Causality: objectives and assessment*, pages 39–58, 2010.