Paper ID: 88262
oral
student

IAF EARTH OBSERVATION SYMPOSIUM (B1)
Interactive Presentations - IAF EARTH OBSERVATION SYMPOSIUM (IP)

Author: Ms. Allegra Farrar
Massachusetts Institute of Technology (MIT), United States

# CAUSAL MODELS OF CYCLOGENESIS: AN INFORMED STRUCTURAL LEARNING APPROACH FOR EFFICIENT DATA-DRIVEN INFERENCE

**Abstract**

The evolving behavior of Tropical Cyclones (TCs) and their formation (cyclogenesis) makes them the most destructive and uncertain weather phenomenon. The governing physics of cyclogenesis is high-dimensional, nonlinear, and influenced by intricate atmosphere-ocean coupling. Furthermore, the evolving climate is changing TC frequency and intensity making the scale to which environmental factors are causing cyclogenesis a persistent area of uncertainty. This complexity is not well captured by first principal physics models; however, we can study data to improve understanding of the governing physics.

Rising popularity of data-driven methods has demonstrated success at managing and performing inference on such high-dimensional, nonlinear dynamical systems. However, lack of interpretability and concerns of accuracy limit their integration in Earth system science. To improve data-driven modeling of environmental factors contributing to cyclogenesis, this work generates a causal model that estimates the probabilistic relationships between variables in the input dataset. The causal model will generate a directed acyclic graph (Bayesian network) that encodes the dependence relationships among observed variables adding a layer of interpretability to the learning.

To generate the causal model, we implement Monte Carlo Tree Search (MCTS) to traverse the large space of candidate structures to build out the graph. MCTS is a sample-based search method known for its efficiency by prioritizing the search in likely regions of optimal solutions. We then advance the MCTS by developing a novel heuristic that guides the search toward candidate structures that improve the posterior probability of the incumbent structure's likelihood. Rather than exploring the entire space, the heuristic focuses the search effort on promising candidates by limiting the expansion of outlier cases. Of the work on causal models in this domain, few studies implement informed search techniques that improve efficiency when reasoning high-dimensional data.

This approach learns from emerging open-source time-series data from active Earth-observing satellites (e.g., NASA's Aqua/Terra MODIS instrument). To account for gaps in this data, we supplement with reanalysis data (MERRA-2) as needed. This approach reduces the trade-off between spatial and temporal resolution, essential for proper model inferencing. We then implement a binary variable indicating TC occurrence in the time series. To evaluate the consistency of the learned structure, the model tests different time series datasets that encompass TC events. Consistent structures over different time series indicate causality over correlation. In the end, we demonstrate a reliable model architecture capable of quantifying the probability of causal pathways useful in generating metrics for adaptation policies.