Exploring the Causal Relationship between Environment, Clouds, Aerosol, and Precipitation Properties using Machine Learning

Piyush Garg, Virendra P. Ghate, Maria Cadeddu, Bethany Lusch
Argonne National Laboratory, Lemont, IL
Contact: pgarg@anl.gov

INTRODUCTION AND MOTIVATION

Current weather and climate models do not accurately represent low-level cloudiness and the associated precipitation. Detailed observations made at the Atmospheric Radiation Measurement (ARM) observatories provide an opportunity to identify causal relationships between large-scale environments, clouds, aerosol, and drizzle properties, which would help us constrain warm rain initiation in the microphysics schemes.

Bayesian networks or causal probabilistic networks, are graphical models used to represent and analyze probabilistic relationships between variables. A Bayesian network consists of nodes (for each variable) and directed edges (connecting the nodes), which represent the probabilistic dependencies between variables. Structural learning is the process of inferring the causal structure of a Bayesian network from the data. We applied structural learning with Bayesian networks by implementing directed acyclic graphs using neural networks.

Objective: Apply Deep Learning to create Bayesian Networks to learn the relationship between environment, cloud, and aerosol properties with drizzle characteristics.

ARM DATASET

ARM East North Atlantic (ENA; **28°W, 39.5°N**, 25 m ASL) site has been operational since 2013 and has multiple instruments to make detailed observations of surface, dynamic, thermodynamic, radiative, aerosol, and cloud fields. We used the ARM dataset from 2015 – 2021 for all weather regimes to obtain a robust causal relationship.

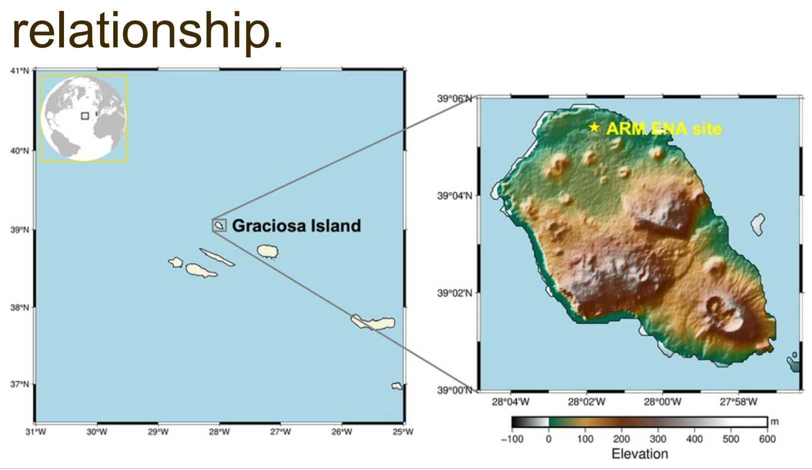


Figure 1: ARM-ENA Site Courtesy: Tozer et al., 2019)

CAUSAL ML

We applied Directed Acyclic Graphs (DAG) with Non-combinatorial Optimization via Trace Exponential and Augmented lagRangian for Structure learning (NOTEARS; Zheng et al., 2018) using MLP neural networks using *PyTorch* library. In this method, continuous optimization is performed by applying global updates in each iteration, thus avoiding assumptions related to the local structure of the graph. We also compared our causal ML results with *Random Forest (RF)* algorithm, which is a traditional ML algorithm.

Predictors	700 hPa Susidence (ω_{700})	Cloud Liquid Water Path (LWP_c)	CCN at 0.2% Supersaturation ($SS_{2\%}$)	Low Tropospheric Stability (LTS)	Sensible Heat flux (SHF)	Temperature Advection (T_{adv})	Total aerosol concentration (#C)	Moisture Advection (Q_{adv})	1 μm Scattering Coefficient ($SC_{1\mu}$)	Latent Heat Flux (LHF)	10 μm Scattering Coefficient ($SC_{10\mu}$)	Precipitable Water Vapor (PWV)	Net Radiative Cooling (NET)
Predictands	Cloud Liquid Water Path (LWP_C)	Drizzle Diameter (D_{Diam})	Drizzle Rain Rate (D_{RR})	Cloud Optical Depth (COD)									

CAUSAL DAG FOR LWP_C

We have performed our analysis on the lower (25th) and upper (75th) quantile and have categorized our predictand variables in two separate classes to identify causal relationships for the tail ends of the distribution. In this way, we avoided any overlap between the two classes due to instrument-related uncertainty in the dataset.

Figure 2 depicts the causal DAG for LWP_c, where the edges are colored according to their normalized edge weights. Conditional Average Treatment Effect (CATE) scores further provide an idea of how strong the "effect" is between the two nodes.

Note that LWP_c is strongly influenced by environment (LTS, T_{adv}, PWV), followed by aerosol (SC, #C) and cloud (NET) properties.

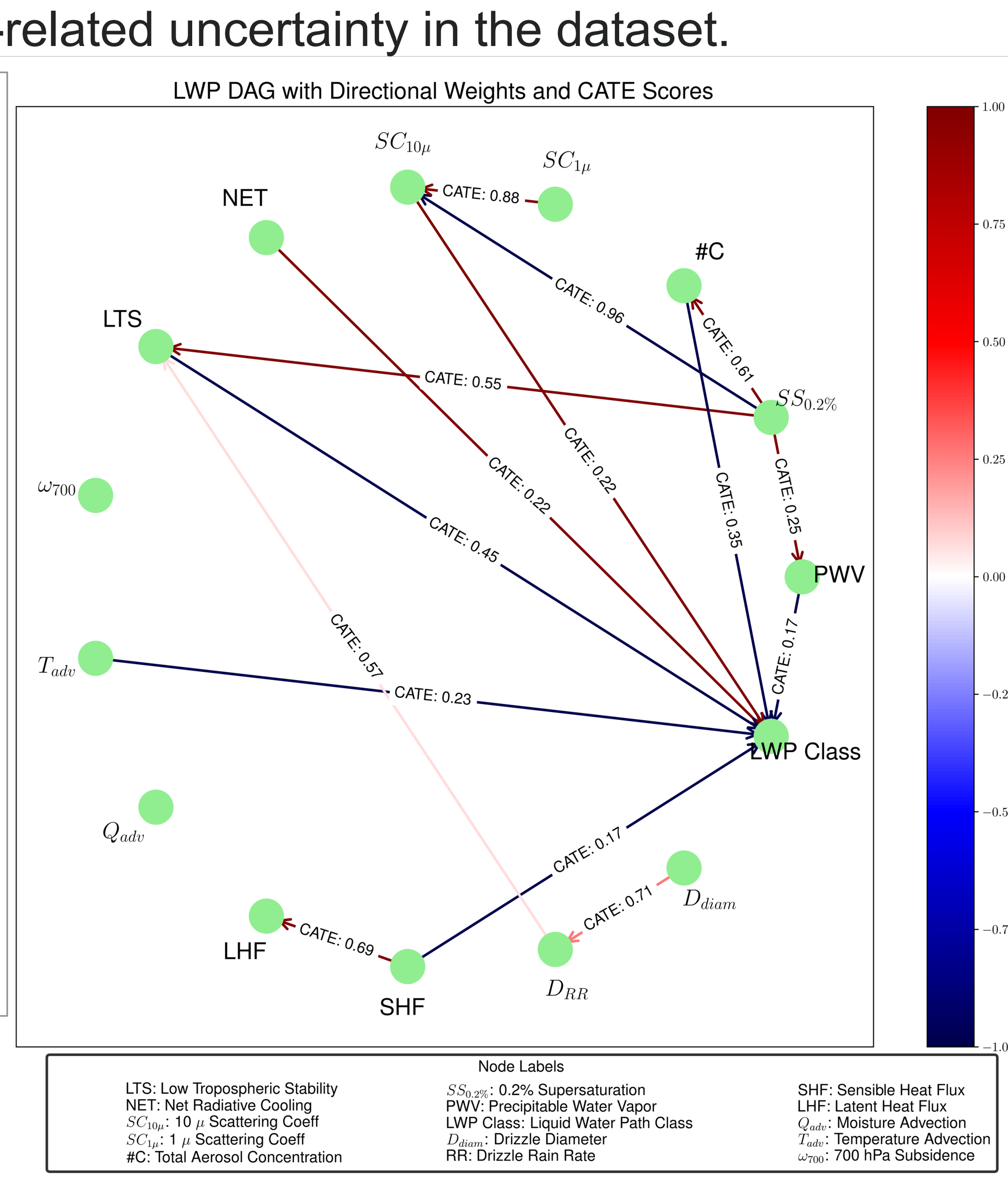


Figure 2: Causal DAG with normalized edge weights (colors) and CATE scores for LWP_c

DAG-NOTEARS v/s RANDOM FOREST

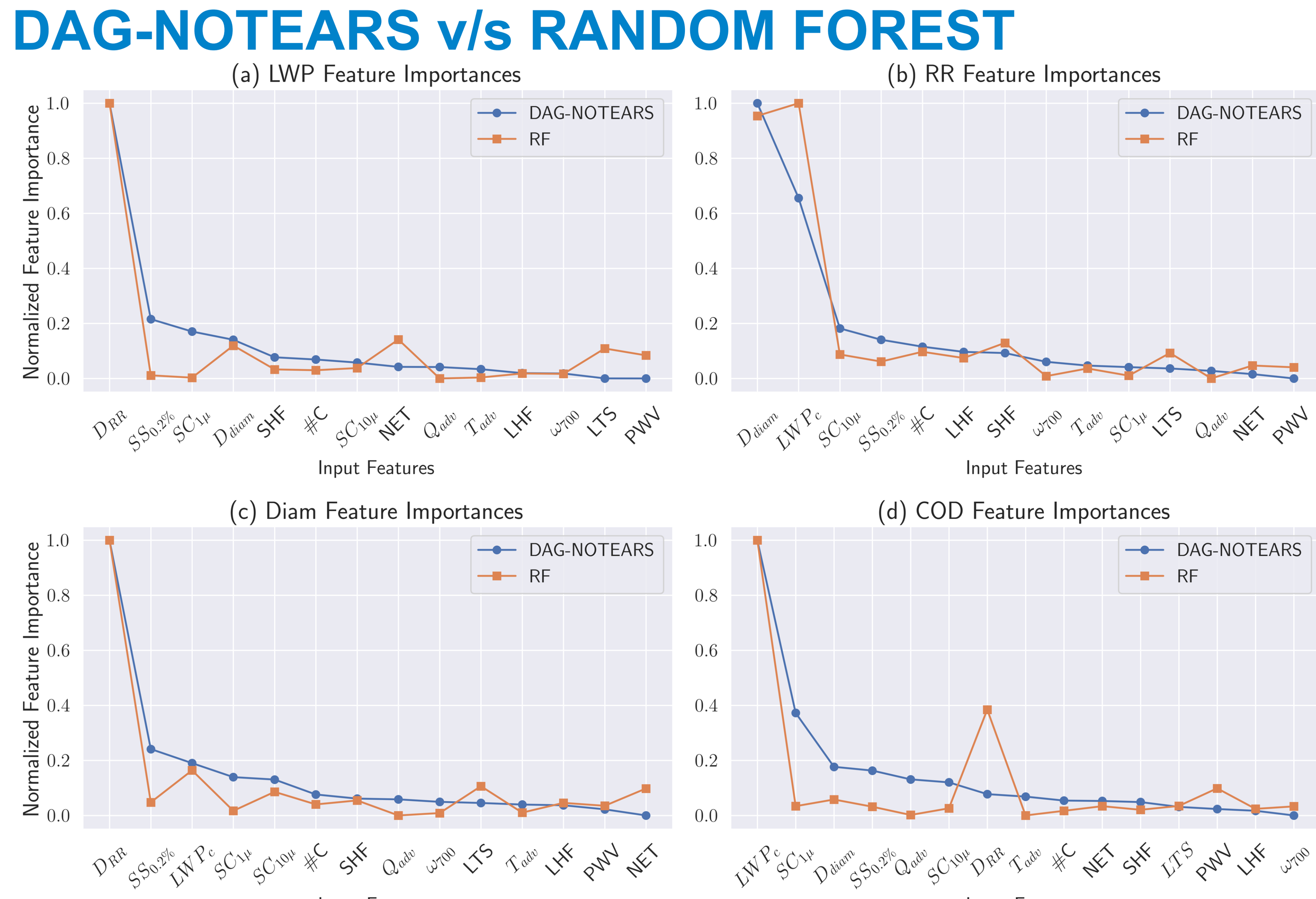
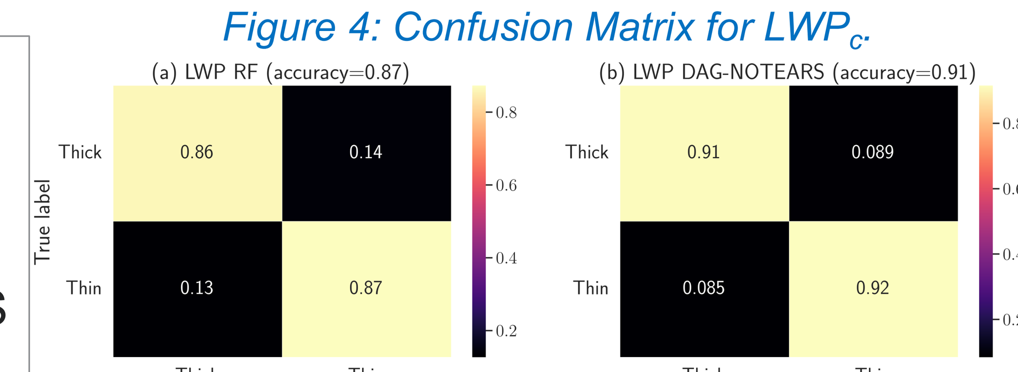


Figure 3: DAG-NOTEARS and RF Feature Importance for LWP_c, D_{RR}, D_{Diam}, COD.

- Note that DAG-NOTEARS weighs the input variables differently than RF.
- Improved accuracy is achieved in DAG-NOTEARS as seen in the confusion matrix for LWP_c

Figure 4: Confusion Matrix for LWP_c.



CONCLUSIONS

- DAG-NOTEARS Framework was successfully implemented to ARM observations to identify causal relationships between environment, cloud, aerosol, and precipitation properties.
- Cloud liquid water path (LWP_c) is primarily impacted by environmental (temperature advection, LTS, PWV), cloud (NET), and aerosol (number conc, scattering) properties, thus suggesting a concurrent role of local and large-scale environment.
- Rain rate prediction has the highest accuracy trailed by cloud optical depth, cloud liquid water path and diameter.

CONCLUSIONS (CONTD)

- Rain rate shows a high conditional treatment effect by Cloud liquid water path, followed by environmental (LTS, T_{adv}), aerosol number concentration, and cloud top net radiative cooling.
- Drop diameter has the most number of direct edges with a dominant effect from cloud liquid water path, followed by rain rate and LTS.
- Cloud optical depth also depicts high CATE and weights from cloud liquid water path, rain rate, precipitable water vapor and aerosol number concentration.

REFERENCES

- Zheng, Xun, Bryon Aragam, Pradeep K. Ravikumar, and Eric P. Xing. "DAGs with NOTEARS: Continuous optimization for structure learning." Advances in neural information processing systems 31 (2018).
- Tozer, B, Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., & Wessel, P. (2019). Global bathymetry and topography at 15 arc sec: SRTM15+. *Earth and Space Science*, 6, 1847–1864. https://doi.org/10.1029/2019EA000658