SAND2021-11449 R
LDRD PROJECT NUMBER: 224486
LDRD PROJECT TITLE: Causal Evaluations for Identifying Differences between Observations and Earth System Models
PROJECT TEAM MEMBERS: PI: Matt Peterson (1461), PM: Susan Altman (8140), Jake Nichol (1461), Kara Peterson (1442)

ABSTRACT:

We use a nascent data-driven causal discovery method to find and compare causal relationships in observed data and climate model output. We consider ten different features in the Arctic climate collected from public databases on observational and Energy Exascale Earth System Model (E3SM) data. In identifying and analyzing the resulting causal networks, we make meaningful comparisons between observed and climate model interdependencies. This work demonstrates our ability to apply the PCMCI causal discovery algorithm to Arctic climate data, that there are noticeable similarities between observed and simulated Arctic climate dynamics, and that further work is needed to identify specific areas for improvement to better align models with natural observations.

INTRODUCTION AND EXECUTIVE SUMMARY OF RESULTS:

The Arctic is changing rapidly and feedbacks between the ocean, atmosphere, and sea ice may be accelerating that change [12]. Accurate predictions of the future sea ice extent in the Arctic depend on understanding the impacts of greenhouse gas forcing and the superimposed internal variability of the complex Earth system. In particular, sea ice loss in the Arctic has been shown to have a linear relationship with global average surface temperature in both observational data and simulation data, with most predictions indicating that the Arctic will be seasonally ice free by mid-century [12,13]. The correlation is generally explained by a common dependency of temperature and sea ice concentration on greenhouse gas concentration, but causality has not typically been assessed. Other studies have found that internal variability in the climate system can accelerate or impede sea ice loss and there is currently no consensus on the dominant processes in the ocean and atmosphere that have the largest impact [14, 15, 16].

Earth system models (ESMs) are critical to our understanding of climate change, but the complex nature of the interactions between atmosphere, ocean, ice, and land can obscure causal relationships. Here, we investigate the causal relationships between Arctic climate features to better understand the complex feedbacks that result in rapid Arctic change and sea ice loss. This effort extends our feature analysis that identified features important for predicting yearly minimum sea ice concentration and compared feature importance between simulations and observations [1].

In [2], a recent review of causal discovery methods for complex systems, they argue that causal discovery is well-suited to improving climate models. In [3], authors provide an example analysis of a global climate model, though focus on a single feature in many separate regions of the globe. This work

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

builds on these publications by extending this nascent field to the U.S. Department of Energy's Energy Exascale Earth System Model (E3SM) [4] and including a multiple feature analysis within one common region. E3SM is a coupling of atmospheric, ocean, river, land, land ice, and sea ice numerical models. Its stated goal is to use exascale computing to output high-resolution simulations of natural and anthropogenic effects in the climate.

Commonly, causality is determined and quantified by interventionist experiments, usually in randomized trials. Because of the magnitude, complexity, and uniqueness of the Earth's climate, there are significant feasibility and ethical problems with controlling and intervening in the climate for experimentation. For this reason, climate science is largely studied with ESMs, which are coupled numerical models. Each model encapsulates subsystems and subprocesses coupled together to approximate the long-term climate.

The status-quo in ESM evaluation is based on descriptive statistics, like mean, variance, climatologies, and spectral properties of model output derived from correlation and regression methods [2]. These methods can be simple to implement and interpret but are often ambiguous or misleading; resulting associations can be spurious and the directions of effects is fundamentally unknown.

In recent decades, a rigorous mathematical framework has been developed for observational causal inference by Spirtes, Glymour, Scheines, Pearl, Rubin, and others [5, 6, 7, 8]. The framework for causal discovery is largely based on Reichenbach's [9] Common Cause Principle: that if two variables are statistically dependent, there must be a causal relationship between the two, or a third common driver of the two. Most importantly, causal discovery methods attempt to identify the direction of observed effects between variables and detect spurious correlations. Effectively understanding the causal drivers in the Arctic climate system is requisite for understanding the future of our climate and how we can mitigate or intervene in climate change.

In previous work, we used a random forest feature analysis to determine which summertime features in the Arctic are most predictive of yearly sea ice extent minimums in September [1]. We then compared results from observed data and simulation output data. This approach allowed us to discover and compare nonlinear relationships in the climate systems. Random forest feature importance values are correlations and direction can only be inferred from each feature to the single predictand. Therefore, inter-feature relationships in the model cannot be interpreted causally. This research expands on our previous work by identifying causal relationships in the data and comparing causal networks from historical simulations and observations.

Causal discovery of observational data is notoriously difficult because spurious correlations and incomplete data leads to spurious inferences. In this work we use conditional independence-based causal discovery, which relies on several assumptions for estimating causal links. One of which is causal sufficiency, that all confounding variables are observed. Because the complex dynamics of the Arctic system are actively researched, and there is no strong consensus on the dominant processes in the Arctic climate, we cannot validate causal sufficiency. We chose our variable set because of their strong correlation with sea ice extent and their success in predicting sea ice extent [17, 18, 1], and they serve as a good hypothesis for a sufficient set.

In our analysis, we were able to fit a network depicting conditional dependencies between features to each of six data sets, observed and five simulated. We then applied a similarity score to evaluate how well the simulated datasets agree with the observed data and each other. Finally, we discuss the next steps for this work and how to derive meaningful differences between the networks.

DETAILED DESCRIPTION OF RESEARCH AND DEVELOPMENT AND METHODOLOGY:

Data

We collected ten features of the Arctic climate. Each was a timeseries of monthly mean values, averaged spatially over the region above 60 degrees North latitude. The observed dataset consisted of natural observations and output from reanalysis products. Simulated data was from the five members, or runs, of the E3SM *historical* ensemble [4]. The *historical* ensemble is a set of runs simulating the Earth system from 1850 to 2014. These runs were initialized by a 500-year-long pre-industrial control simulations, named *piControl*. The selected features are a subset of the quantities E3SM models and were chosen to match observable natural quantities and have been shown in previous work to have strong correlations with sea ice extent [17, 18, 1]. Resulting are six separate datasets, one observational and five E3SM simulation datasets.

The specific quantities we used were mostly the same as outlined in our plan (as seen in Addendum A). We did choose to change a few details. Rather than limit each variable to the same temporal range, 1979-2014, we instead included all the data available for each. We used the entire 150-year span of the E3SM data. The observational timeseries' date range varied by each feature, though they all start in 1979 and continue at least through 2017. Additionally, the full 150-year surface zonal and meridional wind timeseries were not readily available, so we opted to use surface wind magnitude, SWind, in their place, which does not include a directional component. Lastly, we included monthly precipitation rate data from E3SM and from the National Centers for Environmental Prediction for the observational dataset. Full data details are in Addendum C.

Preprocessing

The method detailed below, PCMCI, assumes the data is statistically stationary, i.e., its summary statistics do not change in time. First, we tested each timeseries for stationarity. This consisted of using the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) and augmented Dickey-Fuller (ADF) hypothesis tests. KPSS tests the null hypothesis that a timeseries is stationary around a deterministic trend while ADF tests the null hypothesis that a timeseries is nonstationary around a deterministic trend. If KPSS fails to reject the null hypothesis, and the ADF test rejects, then we considered a timeseries stationary. We used an alpha value of 0.05 to determine significance and found that most features were nonstationary. To keep dependencies and inferences consistent, we applied a 12-month differencing transform to every timeseries. A 12-month difference transform is the process of subtracting a timeseries by itself lagged 12 months. Resulting is a timeseries of the original's change from one year to the next. Differencing removes trend and choosing 12-months will remove yearly seasonality in the data.

Causal network learning

Causal discovery is the process of reconstructing the causal structure from purely observational data [10]. Traditional causality research to determine the causal effect, inferences about the strength of effects between variables, is done when the causal structure is already known. Causal discovery is used when the causal structure is mostly unknown. The causal structure discovered is often represented as a directed acyclic graph in which the nodes represent observed variables, and the edges represent causal relationships.

Causal discovery generally makes four major assumptions: (1) the causal Markov assumption, that if two nodes, X and Y, are d-separated in a graph G, given a conditioning set Z, then X and Y are conditionally independent in their joint probability distribution, given Z; (2) the faithfulness assumption, that if two variables, X and Y, are conditionally independent, given a set of variables, Z, then their nodes in a graph, G, must be d-separated, given Z; (3) causal sufficiency, that there are not any unobserved confounding variables of any variables in the graph; and (4) acyclicity, that there are no cycles in the graph.

In this work, we applied the PCMCI algorithm [11]. PCMCI is an extension to the PC causal network learning algorithm [5], named for its authors Peter Spirtes and Clark Glymour. PC is known for a relatively high false positive rate and struggles with high dimensional, autocorrelated data [11]. In [11], Runge et al. adapted PC to use its skeleton discovery phase for condition selection and then utilize a momentary conditional independence (MCI) phase. PCMCI estimates the causal links between all variable pairs, including their temporal lags.

The first important determination in applying PCMCI is to choose a conditional independence test. The authors have implemented three, the partial correlation, a linear parametric test, gaussian process regression and distance correlation, a nonlinear parametric test, and conditional mutual information with a k-nearest-neighbors estimator, a nonlinear nonparametric test. Generally, the functional form of the dependencies in the feature set needs to be assumed and the appropriate test is chosen. In our case though, we knew it was likely that nonlinear dependencies existed in the data but could not assume if they remained after the data was transformed.

To estimate the dependencies' functional form, we plotted each feature with another one in a scatter plot. The resulting plot depicts how each feature varies with the other. With this, linearities and nonlinearities can be found by eye. Applying this process to the untransformed data, we indeed found several nonlinearities of various forms as well as linear dependencies. Applying it to the transformed data revealed no clear nonlinearities, and multiple clearly linear relationships. With this discovery, we selected the partial correlation parametric linear conditional independence test.

PCMCI has two primary hyperparameters for tuning. The first is the maximum lag, τ_{max}, the maximum lag to evaluate for each variable. τ_{max} is an estimate of the maximum time that every variable may have an effect on the others. The estimation of τ_{max} may come from prior knowledge or by analyzing the linear dependence of each variable with every other variable at a range of lags. The second parameter to estimate is the alpha significance threshold for edges in the graph. Every pairwise dependence is determined with conditional independence tests and has an associated p-value for its significance. Alpha is the threshold for whether the p-value of each link is small enough to be included in the final graph.

To estimate τ_{max}, we plotted the cross-dependencies between each variable at lags between 0 and 24 months and looked for dependence to reach zero for every graph. See Figure 1 for an example from the observed dataset. We repeated this process for each dataset and found that $\tau_{max} = 12$ months was adequate for each variable pair. To estimate alpha, we followed the procedure in [3], which selects from the list {*0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5*} by computing the Akaike information criterion (AIC) of the models fit by each value in the list. That list is slightly more extensive than in [3] because we found each graph was selecting 0.05 and wanted to be sure it was not just selecting the smallest available value.

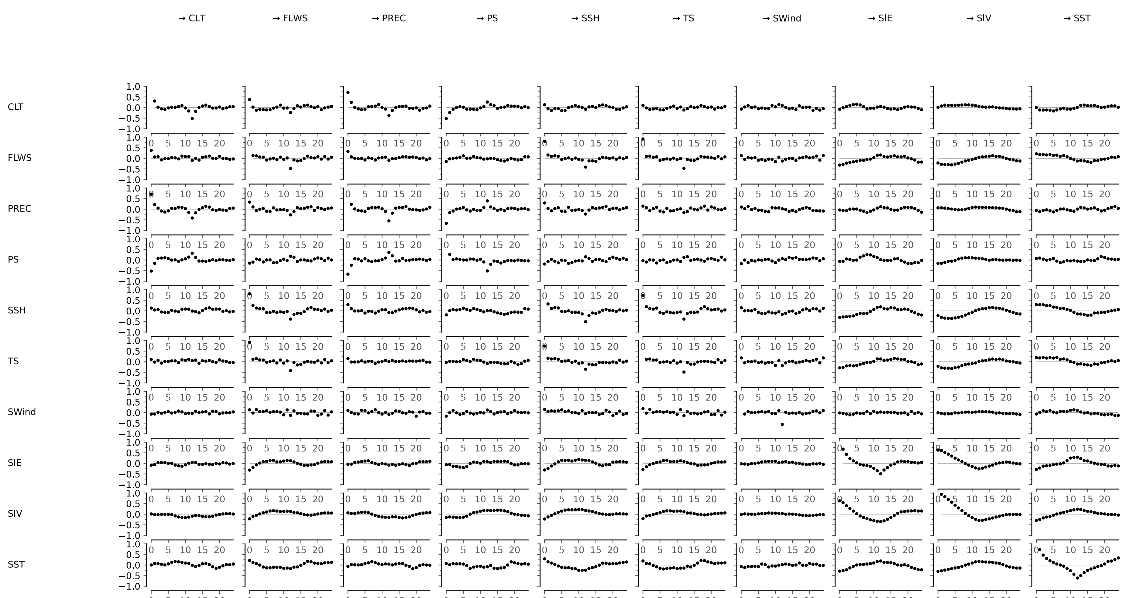


Figure 1: Plots of each feature as a function of each other feature's lags. The vertical axes denote linear dependence, and the horizontal axes denote the number of lags in months.

Causal network comparison

We utilized the F_1 score used in [3] to compare each pair of graphs. The F_1 score is a graph similarity metric with bounds [0,1], with 0 indicating no similarity and 1 indicating perfect similarity. The metric is computed from the precision, P, and recall, R, of a graph in comparison to a reference graph. Precisely, these values are computed as:

$$F_1 = \frac{2 \cdot P \cdot R}{P + R}$$

where

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN}$$

and TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. These terms often assume a ground truth, although because the observed graph is an estimated causal graph and not ground truth, it is important to consider this metric as a relative score and not absolute.

RESULTS AND DISCUSSION:

Before analyzing the results, we filtered links from each network with less than 0.001 significance. For each dataset, PCMCI independently selected pc-alpha value to be 0.05 via AIC. PCMCI evaluated lags between 0 and 12 months for each feature. The simplified graphs in Figure 2 and Figure 3 hide the nodes of each features' lags and only presents a single node per variable. The full timeseries graphs inferred by PCMCI include nodes for each feature's lags up to the maximum lag of 12 months. Because the date ranges on simulated and observed data are not the same, we present results from networks learned from the fully available date ranges, as well as from a homogenous range, 1979 to 2014. Although the algorithm has less data to learn from, this may be a fairer comparison to observed dynamics.

Simplified graphs label links with a list of the lags with significant dependency in order of magnitude. Node color depicts a feature's auto-dependency, how dependent a feature is on its lags. Edge color depicts cross dependency, how dependent a feature is on another feature. Negative, or blue, cross dependency indicates that as the parent's value increases or decreases, the child's value changes inversely. Positive dependence indicates parent and child values increase and decrease together. Because we used a linear conditional independence test, these relationships are linear. Since these colors span many lags, the color chosen for the simplified graphs is the maximum absolute link between two features or a feature and itself.

Figure 2 is a simplified causal network estimation, trained from the full range of observed data. Resulting is relatively sparse partially directed acyclic graphs, with only 5.3% of all possible links existing in the graph. Directed links represent discovered dependencies between features. Undirected links represent contemporaneous dependencies.

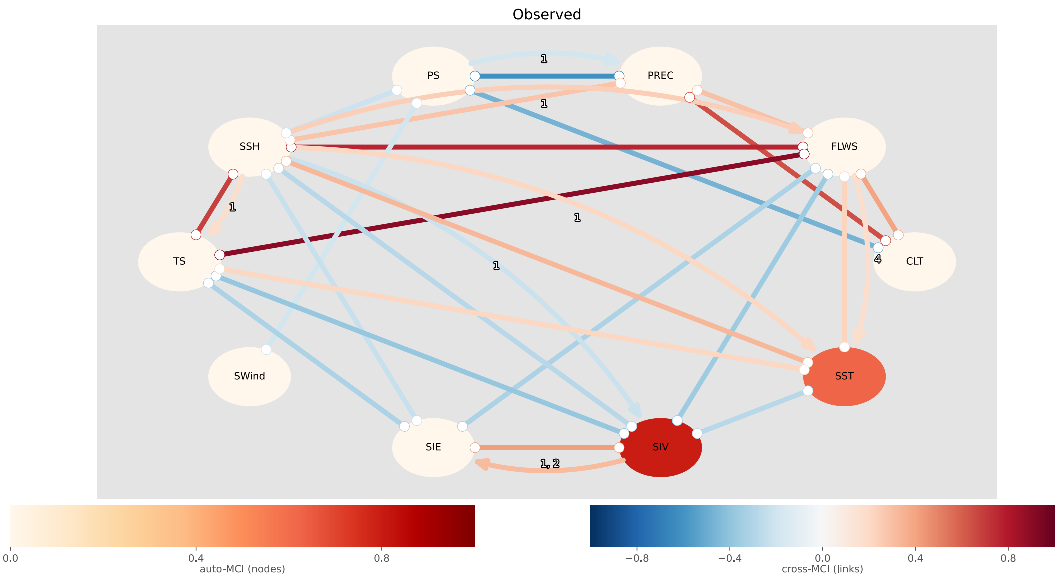


Figure 2: Simplified graph resulting from applying PCMCI with the partial correlation test on observational data in the fully available date range. The pc-alpha parameter was selected by AIC to be 0.05, the links are defined by a significance threshold of 0.001.

Figure 3 is the simplified graph fit by simulation 1 of the E3SM *historical* ensemble in the fully available date range. Although many similar links exist in this graph, it contains many more than the observed data graph. The remaining simulation graphs can be found in the Addendum C. They all differ but are more alike than the observed data graph and contain more links. An average of 8.6% of all possible links exist in the simulation graphs.

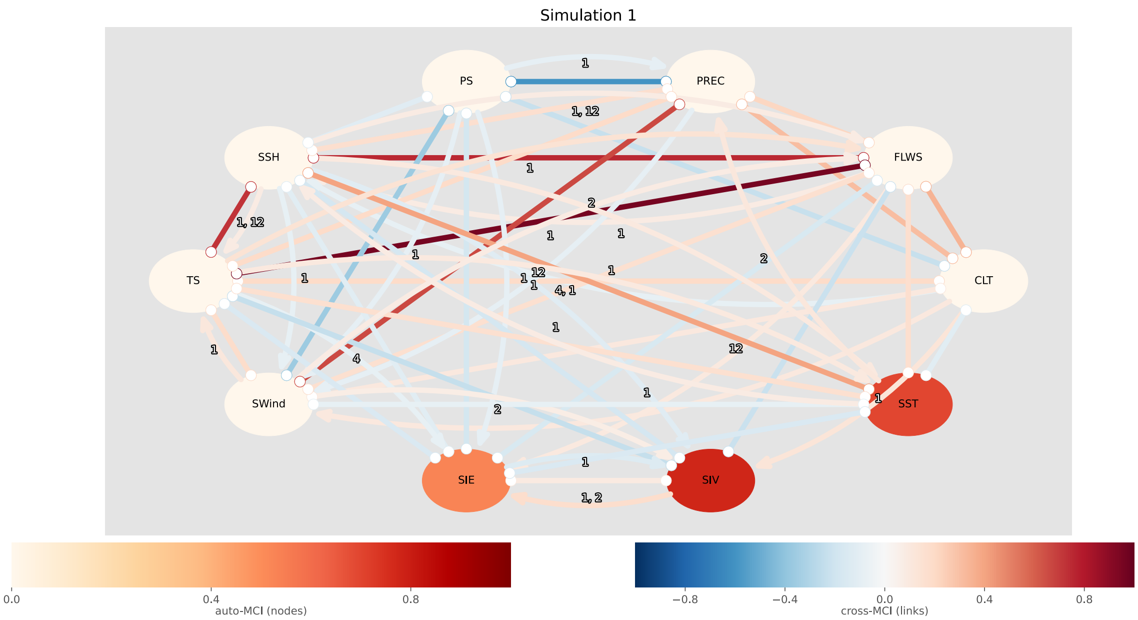
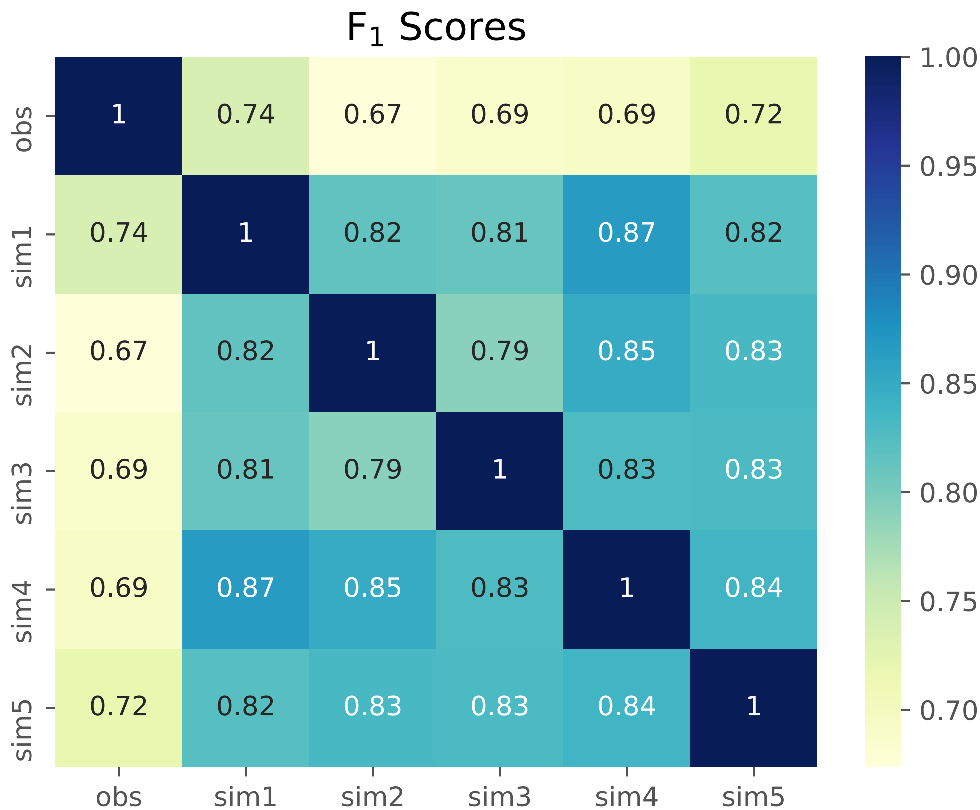


Figure 3: Simplified graph resulting from applying PCMCI with the partial correlation test on simulation 1's data in the fully available date range. The pc-alpha parameter was selected by AIC to be 0.05, the links are defined by a significance threshold of 0.001.

To better quantify the similarity between each graph, we computed the F_1 score of each pair of graphs. For this analysis, we included the fully detailed networks. These include a node for each lag of each feature. Figure 4 shows these results for the fully available date range graphs. The simulation networks are the most similar with each other, while the observed network is the most different from all other networks. The average simulation to simulation F_1 score is 0.83. The average simulation to observed F_1 score is 0.7.



	obs	sim1	sim2	sim3	sim4	sim5
obs	1	0.74	0.67	0.69	0.69	0.72
sim1	0.74	1	0.82	0.81	0.87	0.82
sim2	0.67	0.82	1	0.79	0.85	0.83
sim3	0.69	0.81	0.79	1	0.83	0.83
sim4	0.69	0.87	0.85	0.83	1	0.84
sim5	0.72	0.82	0.83	0.83	0.84	1

Figure 4: Matrix of F_1 similarity scores of each pair of graphs for the fully available date range graphs.

The homogenous date range changed the observed graph minimally but altered the simulated graphs noticeably. 5.5% of all possible links exist in the observed graph, while an average of only 4.8% exist in the simulation graphs for this date range. Figure 5 shows the F_1 similarity scores for the homogenous date range, 1979-2014. In this, the simulation networks lose some similarity, dropping to an average value of 0.71 simulation to simulation. The average similarity to the observed network increases slightly though, to 0.73. It is intuitive that the simulations would diverge in later years, after having been initialized equivalently, and eliminating the early years makes this apparent in their similarity scores.

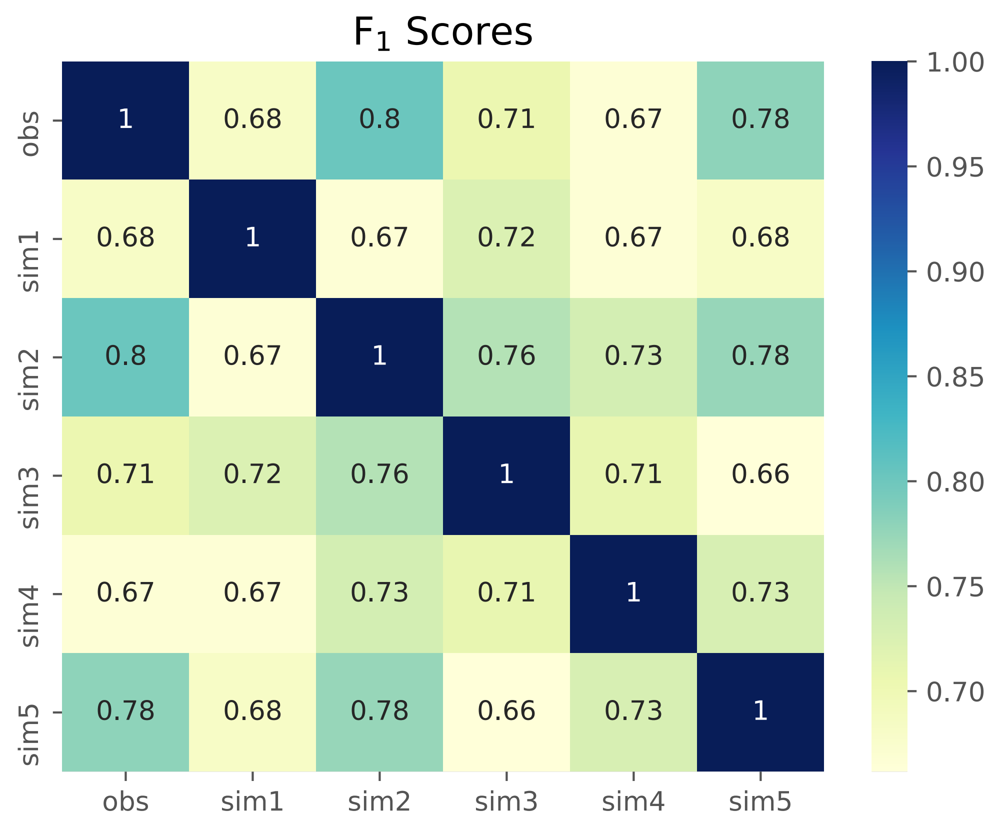


Figure 5: Matrix of F_1 similarity scores of each pair of graphs for the homogenous date range, 1979-2014.

In this work, we developed a strong foundation for applying conditional independence-based causal discovery algorithms. The differencing transforms we applied to the data were important for removing seasonality and trend, which removes the unobserved confounders driving them. We have found that we can apply causal discovery algorithms to Arctic climate data and find strong consistencies between observed and simulated timeseries. Although we cannot validate the causal sufficiency assumption with certainty, we can see that discovered conditional dependencies are similar in each dataset. In future work, we can develop and apply node-to-node similarity metrics to find which nodes are most responsibility for dissimilarity between graphs.

It is important to remember that each feature was transformed to create stationary timeseries. The 12-month differencing transform means that each timeseries is a series each month's deltas from that month's previous year. This means that a directed link from feature X to feature Y would be interpreted as the change in Y from year to year is dependent on the change in X from year to year.

The primary limitation of our findings is the inability to justify the causal sufficiency assumption. The remaining assumptions can be considered satisfied as they assume that an underlying causal structure exists in the data, and that cause and effect does not occur instantaneously. That is assured by the physical and temporal nature of these quantities. The challenge of causal sufficiency exists in any open complex system. We plan to apply causal discovery algorithms that do not rely on the causal sufficiency assumption, such as the Fast Causal Inference algorithm [5] or Latent PCMCI (LPCMCI) [19]. LPCMCI augments PCMCI to discover causal links in the presence of latent, or unobserved, features.

ANTICIPATED OUTCOMES AND IMPACTS:

During this project we presented our findings to the International Conference on Machine Learning (ICML) in the form of a workshop paper (as seen in Addendum A) and an online poster presentation. We also gave another presentation internally to the Validation and Verification of Machine Learning Models discussion group (as discussed in Addendum B). Later this fall there will be presentation at the Chesapeake Large-Scale Analytics Conference (CLSAC) about this work. These presentations allowed us to network with other groups around the labs and externally; organizations include 5493, 1463, 0515, and professors at the University of New Mexico (as discussed in Addendum B).

The major lesson learned in this project was that ground truth for artic climate dynamics is an ongoing research problem, which this work depends on for validating our results are causal. Currently we are relying heavily on climate experts to validate our causal models, but to fully develop metrics for comparing our models we need a concrete understanding of arctic climate dynamics as well as global dynamics. Once these climate dynamics are sufficiently validated, we can utilize these causal models to help us improve our simulated models.

This work will continue in the *CLimate impact: Determining Etiology thRough pAthways (CLDERA) Grand Challenge* project starting in FY22. We plan on improving and adding metrics for comparing similarities and differences between causal models. We are also looking into determining how well a given model fits the data used for training. Some other research areas we want to explore include incorporating spatial data features into our analysis. The work done in this project used averaged values over the entire arctic. We could have divided the data into subregions of the arctic, but with this being a Late-Start LDRD with limited time we decided it was best to simplify the problem space. This will be important for CLDERA because we will be working with data on a global scale and averaging values over the whole globe would not work as easily.

CONCLUSION: (400 word limit)

In this work, we found strong similarities between conditional dependencies discovered in observed and simulated climate dynamics. If the assumptions of causal discovery were to hold, we would find that E3SM climate simulation runs are causally similar to each other and, importantly, causally similar to observations. Although we cannot validate the causal sufficiency assumption, there is evidence that our feature set is a good hypothesis. The largest remaining sources of confounding may be from remaining seasonality and trend from external forcing such as periodic-natural and anthropogenic climate changes. A clear next step is to apply a causal discovery algorithm that does not require causal sufficiency and then compare results.

REFERENCES:

[1] J. Jake Nichol, Matthew G. Peterson, Kara J. Peterson, G. Matthew Fricke, and Melanie E. Moses. 2021. Machine learning feature analysis illuminates disparity between E3SM climate models and observed climate change. *J Comput Appl Math* 395, (October 2021), 113451. DOI:https://doi.org/10.1016/j.cam.2021.113451

[2] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Munoz-Mari, Egbert H. van Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Scholkopf, Peter Spirtes, George Sugihara, Jie Sun, Kun Zhang, and Jakob Zscheischler. 2019. Inferring causation from time series in Earth system sciences. *Nat Commun* 10, 1 (2019). DOI:https://doi.org/10.1038/s41467-019-10105-3

[3] Peer Nowack, Jakob Runge, Veronika Eyring, and Joanna D. Haigh. 2020. Causal networks for climate model evaluation and constrained projections. *Nat Commun* 11, 1 (2020), 1--11. DOI:https://doi.org/10.1038/s41467-020-15195-y

[4] Jean-christophe Golaz, Peter M Caldwell, Luke P Van Roekel, Mark R Petersen, Qi Tang, Jonathan D Wolfe, Guta Abeshu, Valentine Anantharaj, Xylar S Asay-davis, David C Bader, Sterling A Baldwin, Gautam Bisht, Peter A Bogenschutz, Marcia Branstetter, Michael A Brunke, Steven R Brus, Susannah M Burrows, Philip J Cameron-smith, Aaron S Donahue, Michael Deakin, Richard C Easter, Katherine J Evans, Yan Feng, Mark Flanner, James G Foucar, Jeremy G Fyke, Elizabeth C Hunke, Robert L Jacob, Douglas W Jacobsen, Nicole Jeffery, Philip W Jones, Noel D Keen, Stephen A Klein, Vincent E Larson, L Ruby Leung, Hong-yi Li, Wuyin Lin, William H Lipscomb, Po-lun Ma, Renata B Mccoy, Richard B Neale, Stephen F Price, Yun Qian, Philip J Rasch, J E Jack Reeves Eyre, William J Riley, Todd D Ringler, Andrew F Roberts, Erika L Roesler, Andrew G Salinger, Zeshawn Shaheen, Xiaoying Shi, Balwinder Singh, Milena Veneziani, Hui Wan, Hailong Wang, Shanlin Wang, and Dean N Williams. 2019. The DOE E3SM Coupled Model Version 1 : Overview and Evaluation at Standard Resolution. *J Adv Model Earth Sy* (March 2019). DOI:https://doi.org/10.1029/2018ms001603

[5] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. n.d. *Causation, prediction, and search*.

[6] Judea Pearl. 2009. Causal inference in statistics: An overview. *Statistics Surv* 3, September (2009), 96--146. DOI:https://doi.org/10.1214/09-ss057

[7] Peter Spirtes and Kun Zhang. 2016. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics* 3, 1 (2016), 3. DOI:https://doi.org/10.1186/s40535-016-0018-x

[8] Donald B. Rubin. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66, 5 (1974), 688–701. DOI:https://doi.org/10.1037/h0037350

[9] Hans Reichenbach. n.d. *The direction of time*. Univ of California Press.

[10] Clark Glymour, Kun Zhang, and Peter Spirtes. 2019. Review of Causal Discovery Methods Based on Graphical Models. *Frontiers Genetics* 10, (2019), 524. DOI:https://doi.org/10.3389/fgene.2019.00524

[11] Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. 2019. *Detecting and quantifying causal associations in large nonlinear time series datasets*. Retrieved from http://advances.sciencemag.org/

[12] Dirk Notz and Julienne Stroeve. The trajectory towards a seasonally ice-free Arctic ocean. *Current Climate Change Reports*, 4:407–416, 2018.

[13] Julienne Stroeve and Dirk Notz. Changing state of Arctic sea ice across all seasons. *Environ- mental Research Letters*, 13:103001, 2018.

[14] Dirk Olonscheck, Thorsten Mauritsen, and Dirk Notz. Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations. *Nature geoscience*, 12:430–434, 2019.

[15] Ana C. Ordonez, Cecilia M. Bitz, and Edward Blanchard-Wrigglesworth. Processes controlling Arctic and Antarctic sea ice predictability in the Community Earth System Model. *Journal of Climate*, 31:9771–9786, 2018.

[16] Qinghua Ding, Axel Schweiger, Michelle L'Heureux, Eric J. Steig, David S. Battisti, Nathaniel C. Johnson, Eduardo Blanchard-Wrigglesworth, Stephen Po-Chedley, Qin Zhang, Kirstin Harnos, Mitchell Bushuk, Bradley Markle, and Ian Baxter. Fingerprints of internal drivers of Arctic sea ice loss in observations and model simulations. *Nature Geoscience*, 12(1):28–33, 2019.

[17] Jiwon Kim, Kwangjin Kim, Jaeil Cho, Yong Q. Kang, Hong-Joo Yoon, and Yang-Won Lee. Satellite-based prediction of Arctic sea ice concentration using a deep neural network with multi-model ensemble. *Remote Sensing*, 11(19), 2019.

[18] Monica Ionita, Klaus Grosfeld, Patrick Scholz, Renate Treffeisen, and Gerrit Lohmann. September Arctic sea ice minimum prediction - a new skillful statistical approach. *Earth System Dynamics*, 10:189–203, 2019.

[19] Andreas Gerhardus and Jakob Runge. n.d. High-recall causal discovery for autocorrelated time series with latent confounders. In *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 12615–12625. DOI:https://doi.org/10.5194/egusphere-egu21-8259

ADDENDUM A: ICML WORKSHOP PAPER

Towards Knowing Why: Data-Driven Causal Evaluations of Climate Models

J. Jake Nichol [1 2] **Matthew G. Peterson** [2] **G. Matthew Fricke** [1]

Abstract

In this work, we plan to use nascent data-driven causal discovery methods to find and compare causal relationships in observed data and climate model output. We will look at ten different features in the Arctic climate collected from public databases and from the Energy Exascale Earth System Model (E3SM). In identifying and comparing the resulting causal networks, we hope to find important differences between observed causal relationships and simulated causal relationships. With these, climate modeling experts will be able to improve the coupling and parameterization of E3SM and other climate models.

1. Introduction

The Arctic climate has significant direct and indirect impacts on the ecology, geopolitics, and economics of not only the Arctic region, but the whole world (Hassol, 2004; Richter-Menge et al., 2019; Smith & Stephenson, 2013). Of particular importance, the volume and extent of Arctic sea ice are important indicators for the current state and projections of global climate change (Goosse et al., 2018; Sevellec et al., 2017; Runge et al., 2015; Cvijanovic et al., 2017). Because of this, effectively understanding the causal drivers in the Arctic climate system is requisite for understanding the future of our climate and how we can mitigate or intervene in climate change.

Commonly, causal effects are determined and quantified by interventionist experiments, usually controlling all but one variable at random, such as in randomized controlled trials. Because of the magnitude, complexity, and singularity of the Earth's climate, there are significant feasibility and ethical problems with controlling and intervening in the climate for experimentation. For this reason, climate science is largely studied with coupled numerical models. Each model ecapsulates subsystems and subprocesses that work together to determine the long-term climate.

The model we are interested in for this work is the United States Department of Energy (DOE) Energy Exascale Earth System Model (E3SM) (E3SM Project, 2018). This model is a coupling of atmospheric, ocean, river, land, land ice, and sea ice numerical models. Its goal is to use exscale computing to output high-resolution simulations of natural and anthroprogenic effects in the climate.

Climate models are in active development and the Coupled Model Intercomparison Project (CMIP) is a group that collects and curates modern climate models for world-wide collaboration. Researchers have found that models in phases 3 and 5 of CMIP underestimate the rate of Arctic sea ice loss on average (Rosenblum & Eisenman, 2017; Taylor et al., 2012; Stroeve et al., 2007). Given the importance of climate modeling for climate science, it is important to be able to evaluate models and determine why they fall short. Figure 1 shows the difference between observed sea ice extent and E3SM's modeled prediction.

The status-quo in Earth system model evaluation is based on simple descriptive statistics, like mean, variance, climatologies, and spectral properties of model output derived from correlation and regression methods (Runge et al., 2019a). These methods can be simple to implement and interpret but are often ambiguous or misleading; resulting associations can be spurious and direction of impact is fundamentally unknown. In previous work, we used random forest feature analysis to determine which summer-time features in the Arctic are most predictive of yearly sea ice extent minimums in September (Nichol et al., 2021). We then compared results from observed data and simulation output data. This approach allowed us to discover and compare nonlinear relationships in the climate systems Random forest feature importance values are correlational results and directionality can only be infered from each feature to the single predictand. Inter-feature relationships in the model cannot be interpretted causally.

In recent decades, causal inference has been established in a rigorous mathematical framework (Pearl, 2009). That framework can be leveraged to offer a deeper understanding by discovering causal relationships within existing data.

[1]Department of Computer Science, University of New Mexico, Albuquerque, New Mexico, USA [2]Sandia National Laboratories, Albuquerque, New Mexico, USA. Correspondence to: J. Jake Nichol <jefnich@sandia.gov>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

Most importantly, causal methods will identify the direction of causal effects between variables and determine which known correlations are spurious.

Finding different causal relationships between climate models and observed data will identify actionable problems with the models. Like in our previous work, we expect to find many similarities as well. Those similarities will validate the causal structure inherent in the expert-designed numerical climate models. We hope to find that our results from previous work are replicated in the causal analysis as well. It is possible that the causal discovery process will identify there are missing variables as well. Learning this will be a great help to determine that important drivers are missing from our selection or even E3SM itself. The results we find can be analyzed by subject matter experts in climatology and climate modeling to improve the coupling and and parameterization of E3SM and other climate models.

2. Data

We will use time series of ten features in the Arctic consisting of monthly mean values for each year of available data. Observed data will be collected from observational and reanalysis data products, and simulated data will be the output from five ensemble members of the E3SM *historical* dataset (E3SM Project, 2018; Golaz et al., 2019). The selected features are a subset of physical quantities simulated by E3SM in the Arctic and are the same ones used in our previous work with random forests, (Nichol et al., 2021). We originally chose these features because they match observable features in nature and we hypothesized they would be good predictors of sea ice loss. Each feature of the observed dataset is a time series beginning with the start of the satellite era in 1979 through 2018. The E3SM *historical* ensemble datasets span 1850 through 2014.

The observational data included monthly sea ice extent computed from gridded, daily, passive-microwave satellite observations of sea ice concentration provided by the National Snow & Ice Data Center (Peng et al., 2013). Sea ice concentration is a percentage value of ice in each grid cell, and sea ice extent (SIE) is computed as the total area of cells containing more than 15% ice. Sea ice volume (SIV) reanalysis data were provided by the Pan-Arctic Ice Ocean Modeling and Assimilation System (Schweiger et al., 2011). Atmospheric data, total cloud cover percentage (CLT), downward longwave flux at surface (FLWS), pressure at the surface (PS), near-surface specific humidity (SSH), temperature at the surface (TS), wind u component/zonal (uwind), and wind v component/meridional (vwind)) were from an atmosphere reanalysis provided by the National Centers for Environmental Prediction (NOAA et al., 2019a). Sea surface temperature (SST) was provided by the National Oceanic and Atmospheric Administration (NOAA et al., 2019b). For each of the atmospheric data variables, as well as SST, monthly Arctic area averages were computed from the global gridded fields.

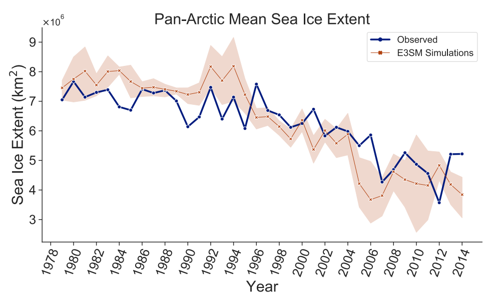


Figure 1. Comparison of observed, pan-Arctic mean September sea ice extent with predictions from E3SM's historical ensembles 1-5. The mean of E3SM simulations is shown with 95% confidence interval (shaded).

Climate simulation data in this work is from DOE's E3SM (E3SM Project, 2018; Golaz et al., 2019). E3SM version 1 was a fork of the community Earth system model (Kay et al., 2015), which was a part of the CMIP5 collection determined to underpredict the rate of sea ice loss (Rosenblum & Eisenman, 2017). E3SM is a global model comprised of submodels for land, atmosphere, land ice, sea ice, oceans, and rivers. Specifically, we will use data from E3SM's *historical* ensembles 1-5 at one-degree global resolution.

Figure 1 shows the difference between observed and E3SM's simulated sea ice extent in September each year between 1979 and 2014. September is when sea ice extent is at its minimum. The model generally predicts the same trend but fails to determine critical lows in yearly sea ice extent.

3. Approach

Causal inference is a mathematical framework for answering questions about why phenomena occur. Causal modeling is an effort to discover, describe, and analyze the relationships between cause and effect. The discovery and description of causal models is defined in two languages: a causal diagram, expressing what we know, and a symbolic language, expressing what we want to know (Pearl & Mackenzie, 2018). Causal inference methods attempt to identify causal relationships and interdependencies of an underlying system (Pearl, 2009; Spirtes & Zhang, 2016).

Many causal discovery algorithms produce a causal diagram or network. The diagram is a directed graph that represents the relationships between variables. Figure 2 is a diagram depicting correlation between variables in the observed dataset from our previous work, including only mean values from June in each year between 1979 and 2014. The PC algorithm (Spirtes et al., 2000) would take a diagram like this as input and iteratively remove spurious correlations and determine the causal direction between the remaining links.

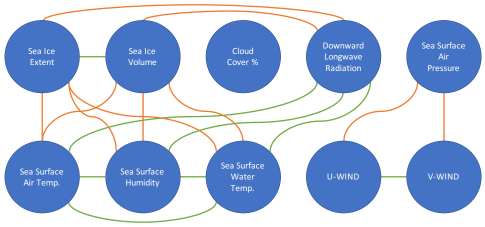


Figure 2. Diagram showing correlated relationships between variables in June from the observed dataset between 1979 to 2014. Green indicates a positive correlation and orange indicates a negative correlation. The correlation threshold is ± 0.6.

There are multiple methods for constructing causal networks that are candidates for investigation in this work. These include causal network learning algorithms, such as the Peter-Clark (PC) algorithm, structural causal model frameworks, such as LiNGAM, and the fast causal inference (FCI) algorithm. Each of these require sets of assumptions about the given data describing the system. We will need to determine which assumptions we can meet with the available data. Due to the nonlinear, stochastic, high-dimensional nature of the climate system, it is likely that causal network learning algorithm and structural causal models will be more effective.

3.1. First step: the PCMCI method

We plan to attempt our analysis with PCMCI (Runge et al., 2019b) first. PCMCI extends the PC-algorithm by adding momentary conditional independence (MCI) tests. These remove false-positives left by the PC algorithm and conditions on each variable's causal parent and its time-shifted parents as well. Thus, the algorithm is designed to remove spurious relationships and identify concurrent and time-lagged causal relationships. PCMCI was specifically designed for highly interdependent time series such as climate data.

In (Nowack et al., 2020), the authors used time series of sea level pressure data at 50 locations around the globe as the raw data. The authors then examined the relationship between precipitation and the causal network skill scores for sea level pressure to demonstrate that this method can help identify dynamic coupling mechanisms arising from underlying physical processes. The Nowack et al. study is one of the first causal network inference studies using large-scale spatiotemporal data and provides a proof-of-concept that such methods are viable in climate systems. They looked at a single variable in various regions, each region's data making up a time series that is input to the algorithm as a single variable. In our work, we plan to use PCMCI to do the opposite, analyze several different quantities in the same region.

3.2. Comparing and evaluating causal models

An obvious first approach for comparing causal diagrams is with standard graph comparison metrics such as global properties and summary statistics: edge density, global clustering coefficient, degree distribution, counts of subgraphs, hamming distance, etc. However, these are defined by correlation and do not address the causal nature of the networks. Other metrics grounded in information flow are more appropriate but possibly more difficult to interpret holistically. In (Runge, 2015), the authors present a framework for determining information flow from multivariate causal diagrams.

A different approach is to consider the resulting models' performance. This includes metrics such as true positive rate (TP), false positive rate (FP), accuracy, positive predictive value, false omission rate, and the G-measure and F1-score (metrics combining TP and FP), and the S-score. These require a baseline model, such as the causal diagram of the observed dataset, to measure the performance of a test model. These are easier to interpret than information flow but are relative measures and cannot be assessed independently.

4. Conclusions

The contributions of this work will bring climate modeling experts a step closer to understanding *why* E3SM does not model certain Arctic quantities well, such as sea ice extent. In our previous work, random forests were able to illucidate which features were more or less important for model predictability in observed and E3SM data. This work should support those results and help explain the causal drivers behind observed and E3SM results. Future research after this work could include: considering more features in the Arctic; other regions with known modeling biases, such as the Antarctic; and other climate modeling problems, such as determining the effects and sources of major climate events. Clear examples are volcanic erruptions and anthroprogenic climate change and intervention.

References

Cvijanovic, I., Santer, B. D., Bonfils, C., Lucas, D. D., Chiang, J. C. H., and Zimmerman, S. Future loss of Arctic sea-ice cover could drive a substantial decrease in california's rainfall. *Nature Communications*, 8(1947), 2017.

E3SM Project. Energy Exascale Earth System Model (E3SM). [Computer Software] https://dx.doi.org/10.11578/E3SM/dc.20180418.36, April 2018. URL https://dx.doi.org/10.11578/E3SM/dc.20180418.36.

Golaz, J.-c., Caldwell, P. M., Roekel, L. P. V., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-davis, X. S., Bader, D. C., Baldwin, S. A., Bisht, G., Bogenschutz, P. A., Branstetter, M., Brunke, M. A., Brus, S. R., Burrows, S. M., Cameron-smith, P. J., Donahue, A. S., Deakin, M., Easter, R. C., Evans, K. J., Feng, Y., Flanner, M., Foucar, J. G., Fyke, J. G., Hunke, E. C., Jacob, R. L., Jacobsen, D. W., Jeffery, N., Jones, P. W., Keen, N. D., Klein, S. A., Larson, V. E., Leung, L. R., Li, H.-y., Lin, W., Lipscomb, W. H., Ma, P.-l., Mccoy, R. B., Neale, R. B., Price, S. F., Qian, Y., Rasch, P. J., Eyre, J. E. J. R., Riley, W. J., Ringler, T. D., Roberts, A. F., Roesler, E. L., Salinger, A. G., Shaheen, Z., Shi, X., Singh, B., Veneziani, M., Wan, H., Wang, H., Wang, S., and Williams, D. N. The DOE E3SM Coupled Model Version 1 : Overview and Evaluation at Standard Resolution. 7 2019. doi: 10.1029/2018ms001603.

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., et al. Quantifying climate feedbacks in polar regions. *Nature Communications*, 9 (1919), 2018.

Hassol, S. *Impacts of a warming Arctic-Arctic climate impact assessment*. Cambridge University Press, 2004.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349, 2015. doi: 10.1175/BAMS-D-13-00255.1. URL https://doi.org/10.1175/BAMS-D-13-00255.1.

Nichol, J. J., Peterson, M. G., Peterson, K. J., Fricke, G. M., and Moses, M. E. Machine learning feature analysis illuminates disparity between E3SM climate models and observed climate change. *Journal of Computational and Applied Mathematics*, 395:113451, 10 2021. ISSN 0377-0427. doi: 10.1016/j.cam.2021.113451.

NOAA, OAR, and ESRL-PSD. Ncep-doe reanalysis 2, 2019a. URL https://www.esrl.noaa.gov/psd/. NCEP_Reanalysis 2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA.

NOAA, OAR, and ESRL-PSD. Noaa extended reconstructed sea surface temperature, 2019b. URL https://www.esrl.noaa.gov/psd/. NOAA_ERSST_V4 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA.

Nowack, P., Runge, J., Eyring, V., and Haigh, J. D. Causal networks for climate model evaluation and constrained projections. *Nature Communications 2020 11:1*, 11 (1):1—11, 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15195-y. URL http://www.nature.com/articles/s41467-020-15195-y.

Pearl, J. Causal inference in statistics: An overview. *Statistics Surveys*, 3(September):96—146, 2009. ISSN 19357516. doi: 10.1214/09-ss057.

Pearl, J. and Mackenzie, D. *The Book of Why*. Basic Books, New York, 2018. ISBN 978-0-465-09760-9.

Peng, G., Meier, W. N., Scott, D. J., Savoie, M. H., and Snow, N. A long-term and reproducible passive microwave sea ice concentration data record for climate studies and monitoring. *Earth System Science Data*, pp. 311—318, 2013. doi: 10.5194/essd-5-311-2013.

Richter-Menge, J., Druckenmiller, M. L., and M. Jeffries, E. Arctic Report Card 2019. Technical report, National Oceanic and Atmospheric Administration, 2019. URL https://www.arctic.noaa.gov/Report-Card.

Rosenblum, E. and Eisenman, I. Sea ice trends in climate models only accurate in runs with biased global warming. *Journal of Climate*, 30(16):6265—6278, 2017. ISSN 08948755. doi: 10.1175/jcli-d-16-0455.1.

Runge, J. Quantifying information transfer and mediation along causal pathways in complex systems. *Physical Review E*, 92(6):062829, 2015. ISSN 1539-3755. doi: 10.1103/physreve.92.062829.

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M., and Kurths, J. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6(1):8502, 2015. doi: 10.1038/ncomms9502.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Munoz-Mari, J., Nes, E. H. v., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Scholkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J. Inferring causation from time series in Earth system sciences. *Nature Communications*, 10(1), 2019a. ISSN 20411723. doi: 10.1038/s41467-019-10105-3.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets.

Technical report, 2019b. URL http://advances.sciencemag.org/.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R. Uncertainty in modeled Arctic sea ice volume. *Journal of Geophysical Research: Oceans*, 116(9):1—21, 2011. ISSN 21699291. doi: 10.1029/2011jc007084.

Sevellec, F., Fedorov, A. V., and Liu, W. Arctic sea-ice decline weakens the atlantic meridional overturning circulation. *Nature Climate Change*, 7:604–610, 2017.

Smith, L. C. and Stephenson, S. R. New trans-Arctic shipping routes navigable by midcentury. *PNAS*, 110(13): 4871–4872, 2013.

Spirtes, P. and Zhang, K. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics*, 3(1):3, 2016. doi: 10.1186/s40535-016-0018-x.

Spirtes, P., Glymour, C. N., Scheines, R., and Heckerman, D. *Causation, prediction, and search*. MIT press, 2000.

Stroeve, J., Holland, M. M., Meier, W., Scambos, T., and Serreze, M. Arctic sea ice decline: Faster than forecast. *Geophysical Research Letters*, 34(9), 2007. ISSN 00948276. doi: 10.1029/2007gl029703.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. An Overview of CMIP5 and the Experiment Design. *American Meteorological Society*, 3(april):485—498, 2012. doi: 10.1175/bams-d-11-00094.1.

ADDENDUM B: REPORT PRESENTATION

OFFICIAL USE ONLY

Energy & Homeland Security

Causal Evaluations for Identifying Differences between Observations and Earth System Models, 224486

Presented by

PI: Matt Peterson (1461), PM: Susan Altman (8140)

Team: Jake Nichol (1461), Kara Peterson (1442)

Budget: $99k

OFFICIAL USE ONLY
May be exempt from public release under the Freedom of Information
Act (5 U.S.C. 552), exemption number and category:

Department of Energy review required before public release.

Name/Org: Date:
Guidance (if applicable) ______________

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525

OFFICIAL USE ONLY

SAND2020-11795 PE

OFFICIAL USE ONLY

2 PROJECT OVERVIEW

Project Overview

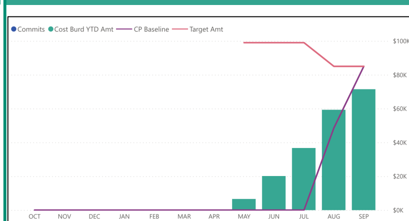
PROJECT SUMMARY:

- *Apply data-driven causal inference methods to quantify the critical differences between climate simulations of Arctic sea ice extent and observations.*
- *Accurate prediction of Arctic sea ice loss is a critical challenge and addressing it aligns with Sandia's Arctic Science and Security Initiative*

IMPACT:

- *Causal relationships will enhance findings in studies of correlational relationships between Arctic variables such as those in our recently completed Arctic LDRD (SAND2020-9932, J. Nichol, et al.)*
- *The proof-of-concept methods from this late start can be applied in the proposed CLDERA grand challenge project*

Financial Performance at Year-End



Data as of: 9/3/21	Budget: $85K
YTD Cost: $71.5K	Commits: $0K
% C+C: 84%	Remaining: $13.5K

Major Cost Elements	$ Est.
Budget Decrease	$-14K

Accomplishments & Plan

ACCOMPLISHMENTS:

- *Determined which causal methodology performs best with Arctic climate data*
- *Applied Chosen Methodology on our data sets*
- *Compared causal network graphs generated by observed data and the 5 E3SM historical ensemble members*
- *Lessons Learned – Ground Truth for Arctic Climate Dynamic is Rare:*
  - *Validating causal discovery relies heavily on subject matter expertise and Arctic climate dynamics are not well defined*
  - *Graph metrics utilize ground truth*

UPCOMING ACTIVITIES:

Work will continue on with the CLDERA Grand Challenge

- *Improve comparison and evaluation metrics that were applied in this project*

Risks, Issues, and Opportunities

RISKS:

- IF our causal assumptions are not validated, THEN our estimated links between variables will have little impact on improving our climate simulation models

ISSUES:

- Validating the ground truth climate dynamics is an ongoing research topic, until that is well understood we cannot fully be certain that our causal assumptions are correct

OPPORTUNITIES:

- Any effort that is designed to improve simulations with observational data
  - Stockpile flight test and lab test data
  - Particle physics simulations
  - Global Climate problems

OFFICIAL USE ONLY

LABORATORY DIRECTED RESEARCH & DEVELOPMENT

WHERE INNOVATION BEGINS

MILESTONES

Status of FY21 Project Milestones	Completion Date	Changes
Determined which causal methodology performs best with Arctic climate data - Literature Review - Ask SMEs about data	May/2021	
Applied Chosen Methodology on our data sets - Preprocessing data - Parameter tuning - Interpret Output	July/2021	
Compare causal network graphs generated by observed data and the 5 E3SM historical ensemble members	August/2021	Focused on Metrics for determining: 1. Model fit to Data 2. Model to Model distance/difference

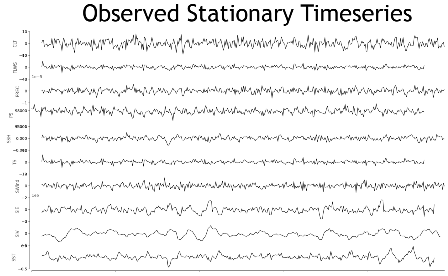
OFFICIAL USE ONLY

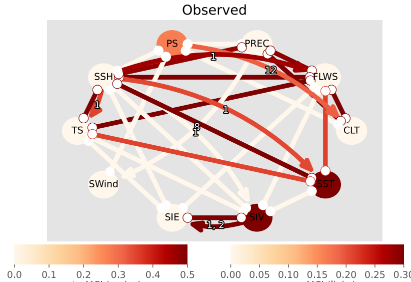
TECHNICAL RESULTS

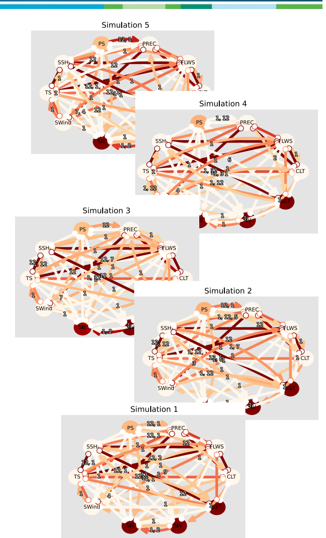
Steps

- Preprocessing
  - Create a time series of each variable
  - Timeseries stationarity is needed because the algorithm must assume that deviations from the mean/variance are due to internal influences rather than some external seasonality or long-term trend
  - Transform time series to make them all stationary
- Parameterization Tuning
  - Choose a maximum lag to include
  - Choose the alpha significance threshold for independence tests
- Causal Network Learning
  - Fit the PCMCI [1] causal discovery algorithm to each dataset
  - Analyze resultant networks

Observed Stationary Timeseries







[1]Peer Nowack, Jakob Runge, Veronika Eyring, and Joanna D. Haigh. 2020. Causal networks for climate model evaluation and constrained projections. *Nat Commun* 11, 1 (2020), 1--11. DOI:https://doi.org/10.1038/s41467-020-15195-y

LABORATORY DIRECTED RESEARCH & DEVELOPMENT

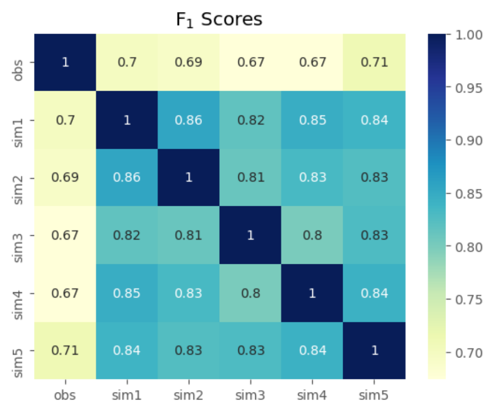
WHERE INNOVATION BEGINS

5 TECHNICAL RESULTS

The F_1 Score is a similarity metric computed from existence of edges in a pair of networks

Future work includes more metrics:

- An average goodness of fit score for each network
  - Each edge has a goodness of fit and a significance value to determine if it should exist in the network
  - Combining these could be a good metric for overall fit
- Some node-node similarity metrics
  - Node-node F_1 Score
  - Others
- Node level metrics will identify where the differences occur and more meaningful inferences may be possible
- Computing Causal Effects with the discovered graph
  - Strength of pathway from node A to node B



6 PROJECT LEGACY

- Publications
  - Nichol, J. Jake, Matthew G. Peterson, G. Matthew Fricke, and Kara J. Peterson. "Learning Why: Data-Driven Causal Evaluations of Climate Models." ICML 2021 Workshop: tackling Climate Change with Machine Learning. SAND2021-8028 C
    - https://www.climatechange.ai/papers/icml2021/80
- Presentations: Workshops, conferences, Industry Days
  - ICML 2021 Workshop: Tackling Climate Change with Machine Learning
    - "Learning Why: Data-Driven Causal Evaluations of Climate Models" presented by Jake Nichol, SAND2021-8130 C
  - Presentation on causal analysis and causal discovery to Sandia's Verification and Validation of Machine-Learned Models Used in Science and Engineering discussion group (5954).

LABORATORY DIRECTED RESEARCH & DEVELOPMENT

WHERE INNOVATION BEGINS

OFFICIAL USE ONLY

7 CAREER DEVELOPMENT & CAPABILITIES STEWARDSHIP

- Establishment of Capabilities expected to impact future work
  - New knowledge base of how to apply Causal Modeling for climate data
    - Preprocessing data
    - Parameter Tuning of Causal Models
    - Generating Comparison and Evaluation Metrics
- Career Development & Capabilities Stewardship
  - Matt Peterson (PI), 1461, is an early career staff and first PI position on an LDRD
  - Jake Nichol ,1461, is a Year-round PhD student intern being mentored by Matt Peterson and this work is directly related to his dissertation topic

OFFICIAL USE ONLY

OFFICIAL USE ONLY

8 PROGRAMMATIC RESULTS

- Engagement with potential customers
  - Will be working on the CLDERA Grand Challenge to continue this work; Diana Bull (0515)
- New research teams or collaborations formed
  - Collaborated with other Causal Network folks; Mark Smith (5493) and Laura Swiler (1463)
  - Presented at Validation and Verification Discussion group, hosted by Erin Acquesta (5954)
  - This work ties into Jake's dissertation at UNM with his advisor Melanie Moses

OFFICIAL USE ONLY

ADDENDUM C: DATA AND RESULTS

Observed Data

Feature	*Date Range*	*Source*
Cloud cover percentage (CLT)	1979-2019	https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html
Downward longwave radiation flux at the surface (FLWS)	1979-2017	https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html
Precipitation rate (PREC)	1979-2019	https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html
Air pressure at the surface (PS)	1979-2017	https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html
Sea ice extent (SIE)*	1979-2018	https://nsidc.org/data/seaice_index/archives
Sea ice volume (SIV)	1979-2019	http://psc.apl.uw.edu/research/projects/arctic-sea-ice-volume-anomaly/data/
Sea surface humidity (SSH)	1979-2019	https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html
Sea surface temperature (SST)	1979-2018	https://psl.noaa.gov/data/gridded/data.noaa.ersst.v4.html
Surface wind speed (SWind)**	1979-2019	https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html
Air temperature at the surface (TS)	1979-2017	https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html

* Computed from directional zonal and meridional wind velocities.

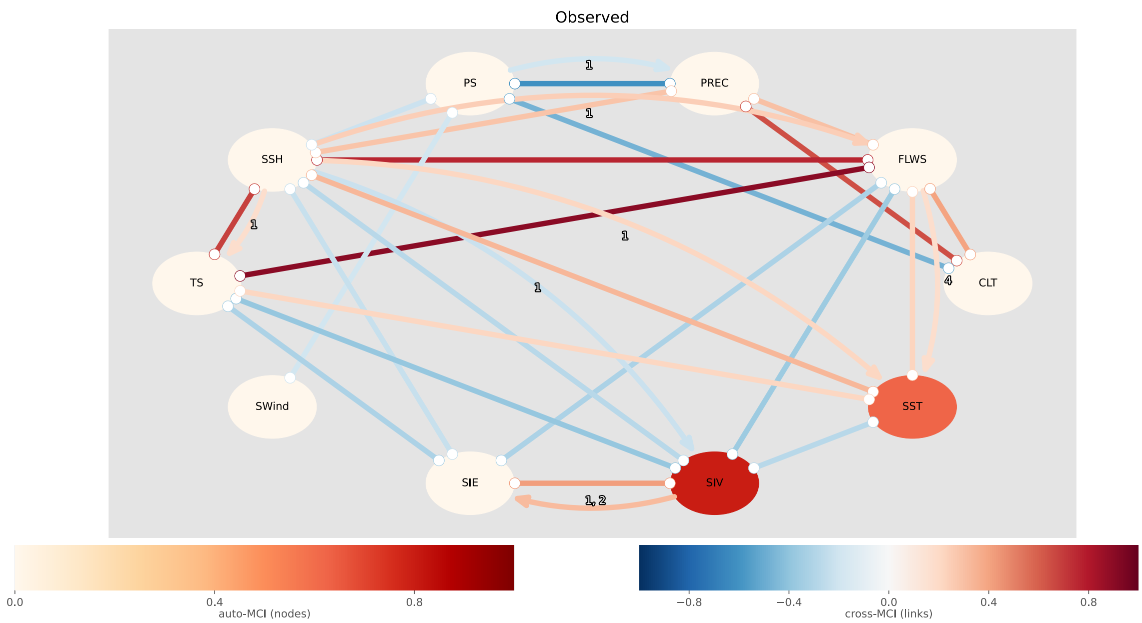
** Sea ice extent data was computed from sea ice concentration values. Sea ice concentration is a percentage value of ice in each grid cell, and extent was computed as the total area of cells containing more than 15% ice.

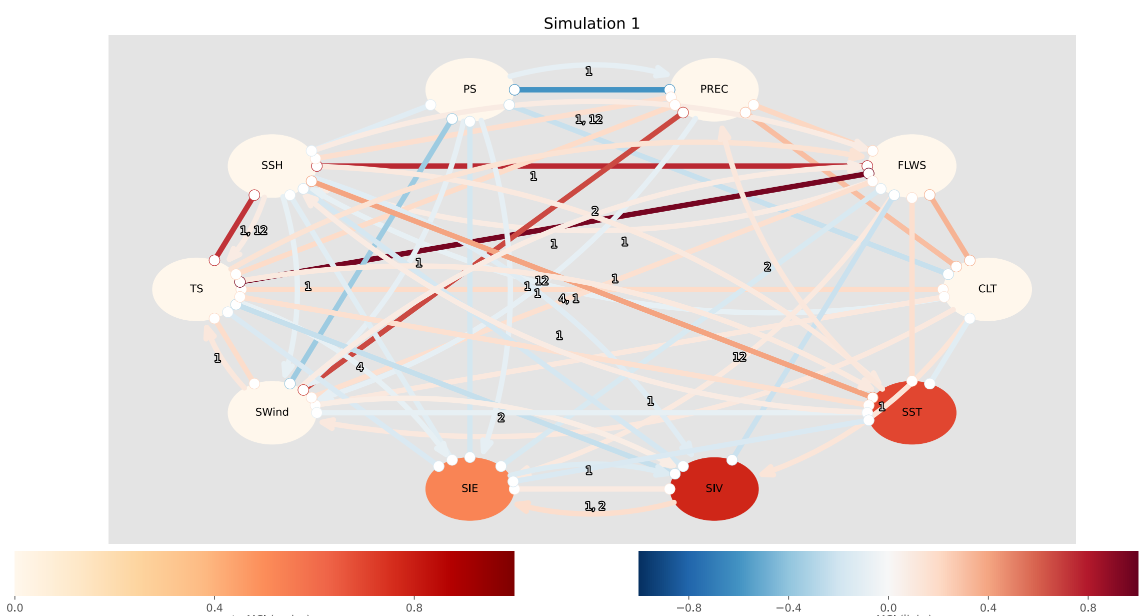
E3SM Simulated Data

E3SM historical ensemble data was collected from https://esgf-node.llnl.gov/search/e3sm/

Feature	*Date Range*
Cloud cover percentage (CLT)	1850-2014
Downward longwave radiation flux at the surface (FLWS)	1850-2014
Precipitation rate (PREC)	1850-2014
Air pressure at the surface (PS)	1850-2014
Sea ice extent (SIE)	1850-2014
Sea ice volume (SIV)	1850-2014
Sea surface humidity (SSH)	1850-2014
Sea surface temperature (SST)	1850-2014
Surface wind speed (SWind)	1850-2014
Air temperature at the surface (TS)	1850-2014

Results for the full 1850-2019, nonoverlapping date range

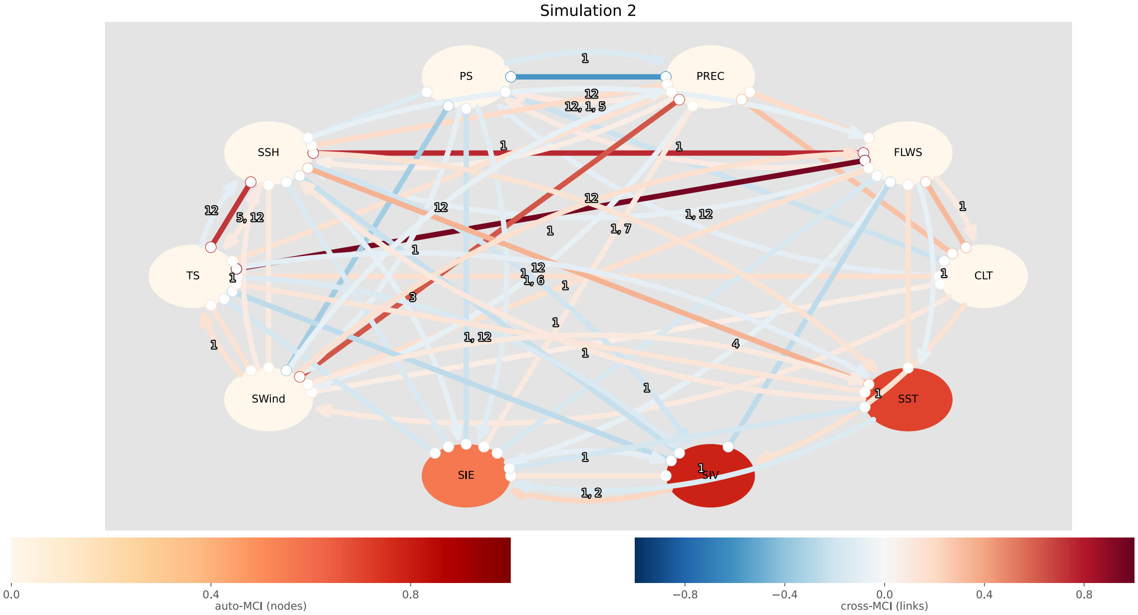

Observed
PS
PREC
SSH
FLWS
TS
CLT
SWind
SST
SIE
SIV
1
1
1
1
1
4
1, 2
0.0
0.4
0.8
auto-MCI (nodes)
−0.8
−0.4
0.0
0.4
0.8
cross-MCI (links)


Simulation 1
PS
PREC
SSH
FLWS
TS
CLT
SWind
SST
SIE
SIV
1
1, 12
1, 12
1, 2
4, 1
12
2
4
0.0
0.4
0.8
auto-MCI (nodes)
−0.8
−0.4
0.0
0.4
0.8
cross-MCI (links)

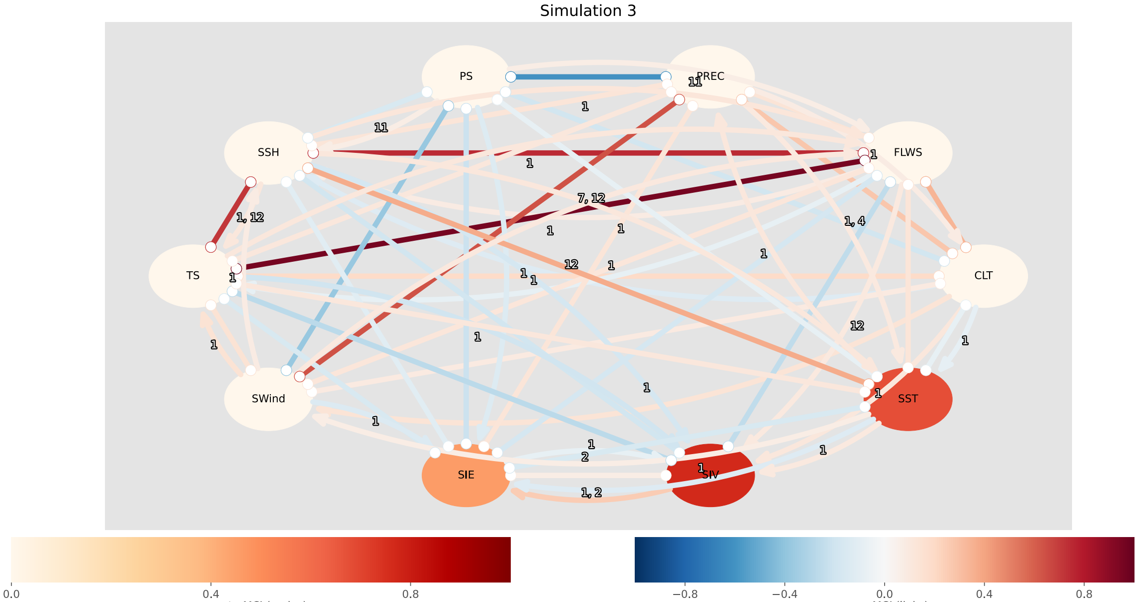
LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT

WHERE INNOVATION BEGINS

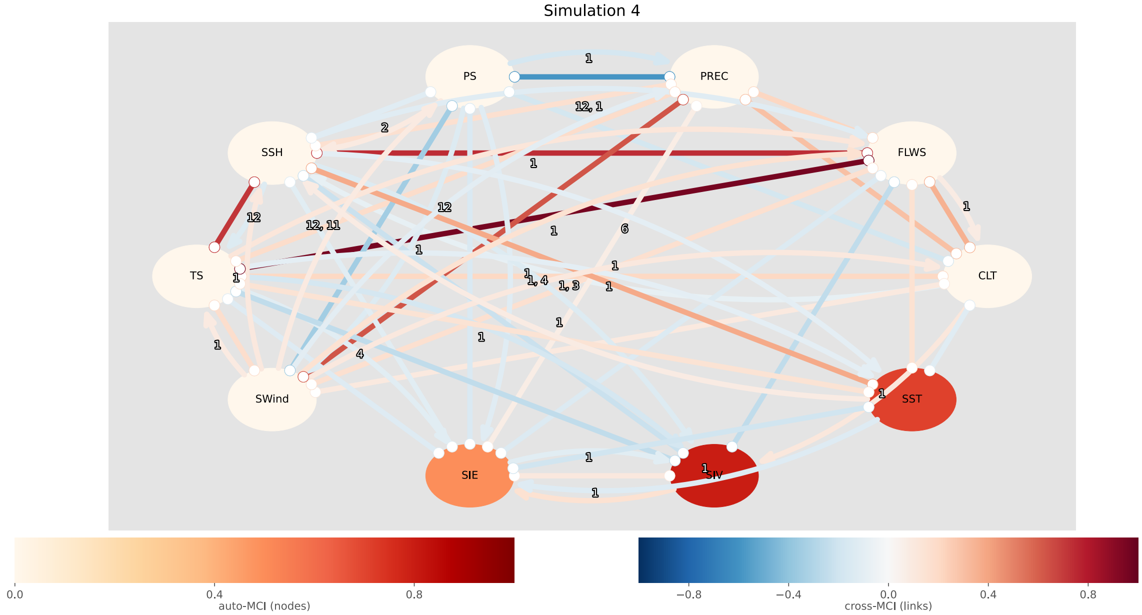
Simulation 2



Simulation 3



Simulation 4

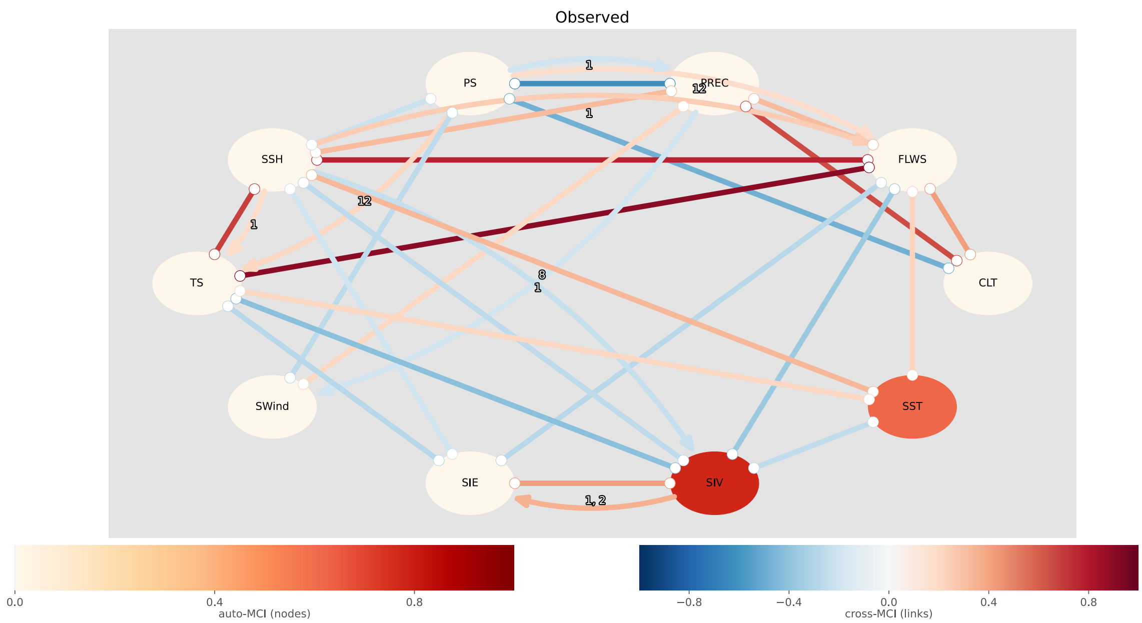


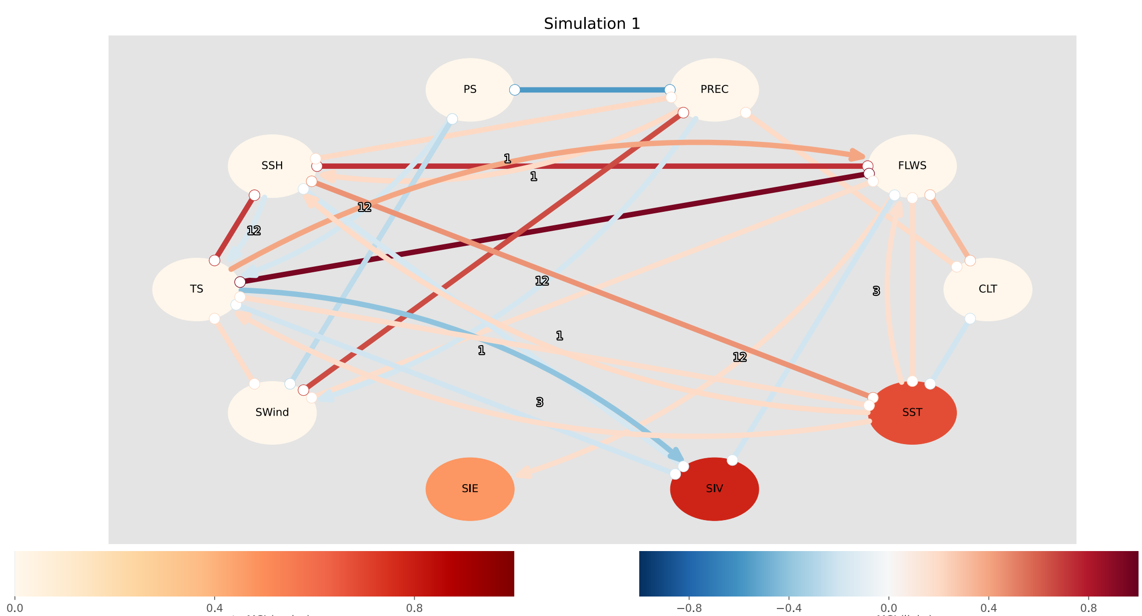
LABORATORY DIRECTED
RESEARCH & DEVELOPMENT
WHERE INNOVATION BEGINS

Results for the full 1850-2019, nonoverlapping date range

LABORATORY DIRECTED RESEARCH & DEVELOPMENT

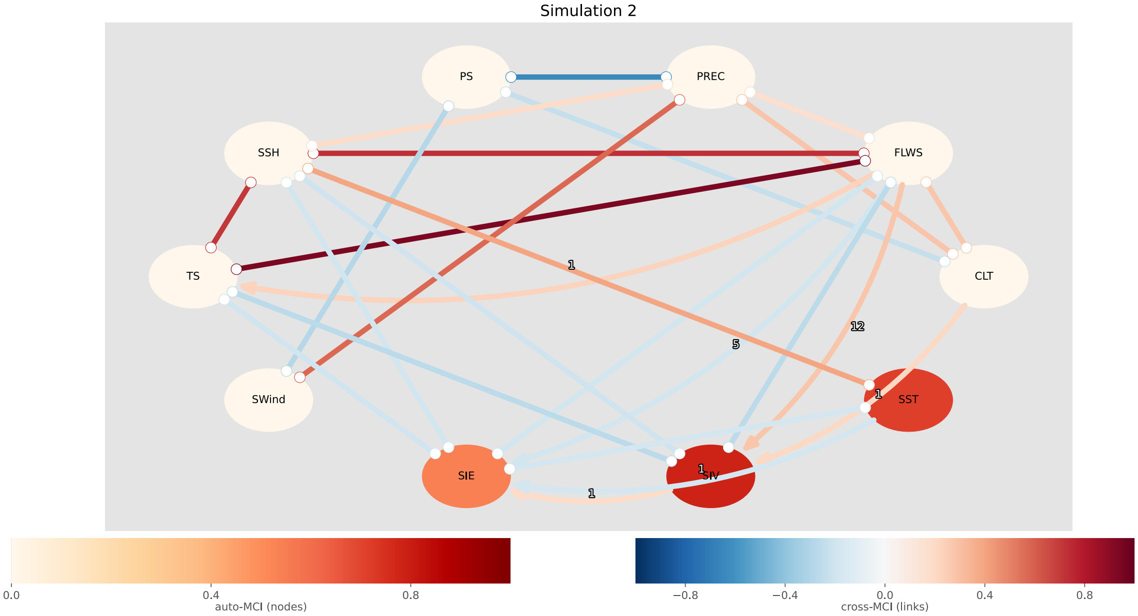
WHERE INNOVATION BEGINS



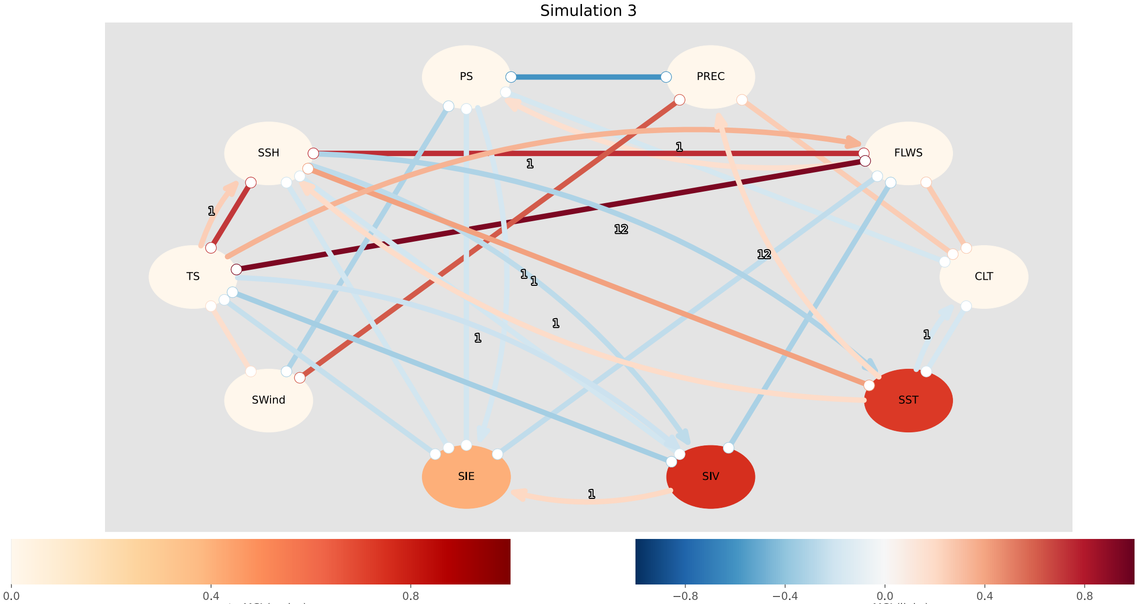


LABORATORY DIRECTED
RESEARCH & DEVELOPMENT
WHERE INNOVATION BEGINS

Simulation 2



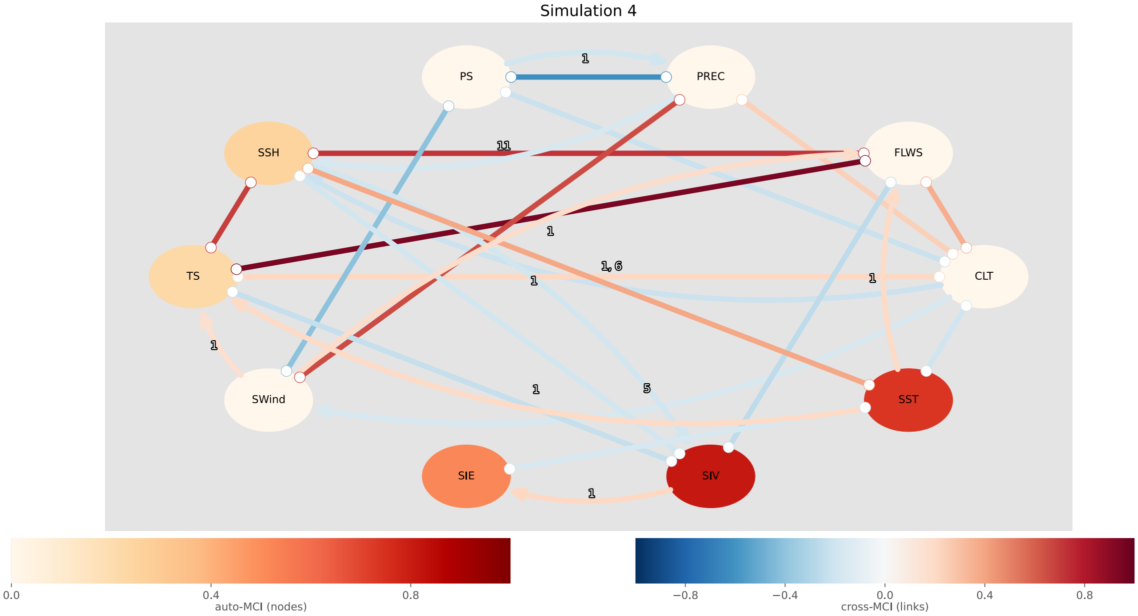
Simulation 3



LABORATORY DIRECTED RESEARCH & DEVELOPMENT

WHERE INNOVATION BEGINS

Simulation 4



Simulation 5