Learning Why: Data-Driven Causal Evaluations of Climate Models

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

Abstract

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 those in climate models. With these, climate modeling experts will be able to improve the coupling and parameterization of E3SM and other climate models.

1 Introduction

Climate models are critical to our understanding of climate change. We believe there is an opportunity to apply causal inference methods to these models to improve predictions. We can understand the quality of a model by comparing it with observations of the natural phenomena being simulated. From there, we can make the necessary improvements to the model, but where to start? Currently models are developed using a trial and error approach, in which a model is designed and parameterized and the resulting accuracy is observed. For computationally expensive models this approach quickly becomes inefficient. We propose to investigate the causal relationships between features and their weights to better target reparameterization and feature selection efforts. We propose to focus on the pan-Arctic region because we previously studied Earth system model (ESM) prediction discrepancies there (Nichol et al., 2021). The Arctic climate, though important in itself, also has global climate implications.

In Runge et al. (2019a), a recent review of causal methods, they argue that causal discovery is well-suited to improving climate models. Nowack et al. (2020) provide an example analysis of a global climate model. This work proposes to build these publications, by extending this nascent field to Energy Exascale Earth System Model (E3SM) (E3SM Project, 2018) and a including multiple feature analysis.

In contrast to methods based in statistical correlations, causal inference tells us *why* systems behave the way they do. Discovering the underlying causal structure in data and then comparing those structures from observed and simulated datasets will give us a richer understanding of the differences between the data sources.

Commonly, causal effects are 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 coupled numerical models. Each model encapsulates subsystems and subprocesses that work together to determine the long-term climate.

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 the directions of effects is fundamentally unknown.

In recent decades, a rigorous mathematical framework has been developed for observational causal inference by Pearl, Spirtes, Glymour, Scheines, and others (Spirtes et al., 2000; Pearl, 2009; Spirtes & Zhang, 2016). The framework is largely based on Reichenbach's (Reichenbach, 1991) Common Cause Principle: that if two variables are dependent, there must be a causal relationship between the two or a third common driver of the two. Most importantly, causal methods identify the direction of observed effects between variables and detect spurious correlations.

The model we are interested in for this work is the United States Department of Energy 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 exascale computing to output high-resolution simulations of natural and anthropogenic effects in the climate.

The Arctic climate has significant direct and indirect impacts on global climate, ecology, geopolitics, and economics (Hassol, 2004; Richter-Menge et al., 2019; Smith & Stephen-

[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>.

Tackling Climate Change with Machine Learning Workshop at ICML 2021.

son, 2013). In particular, 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.

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). Figure 1 shows the difference between observed sea ice extent and E3SM's modeled prediction.

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 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. Finding differences between in causal relationships between climate models and observed data will identify clear, actionable problems with the models.

2 Data

We selected time series data for ten features in the Arctic consisting of monthly mean values for each year of available data. Empirical data was collected from observational and reanalysis data products, and simulated data were taken 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. Through feature analysis, we discovered that some inputs were far more predictive than others, but we did not have a causal inference framework to explain why. Each feature of the observed dataset is a time series beginning with the start of the satellite era in 1979 to 2018. The E3SM *historical ensembles* span 1850 to 2014.

The observational data includes 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. Simulated data features were selected to match the observation dataset.

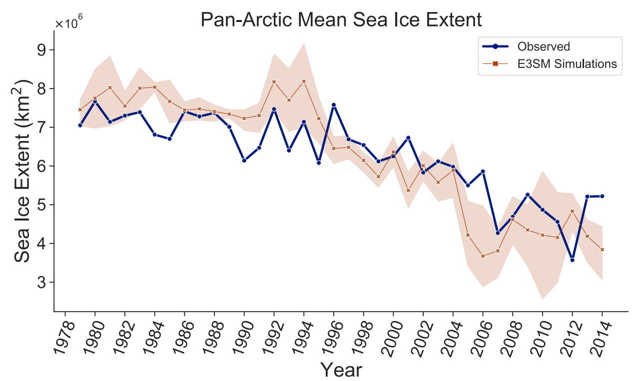

Figure 1. Comparison of observed, pan-Arctic mean September sea ice extent with predictions from E3SM's historical ensembles 1-5. The mean of E3SM simulations is shown with 95% confidence interval (shaded).

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. While the simulations generally predict sea ice extent well, there are significant departures (fall outside the 95% CI) in particular years. For example, in 2012 there was a reversal between simulation, which predicted a year-over-year increase in sea ice, but instead a record low was observed. Since sea ice extent has a non-linear effect on the global climate, providing a causal explanation for these departures is critical.

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 (Pearl, 2009; Spirtes & Zhang, 2016). The calculus of causation 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). The methods we propose derive a causal diagram from the given data.

A causal diagram is a directed graph where arcs represent the causal relationships between variables. Figure 2 is a diagram depicting correlations between variables in the observed dataset from our previous work. Only mean values from June in each year between 1979 and 2014 were included. For example, the PC algorithm (Spirtes et al., 2000) could take a diagram such as the one in Figure 2 as input and iteratively remove spurious correlations and determine the causal direction between the remaining links.

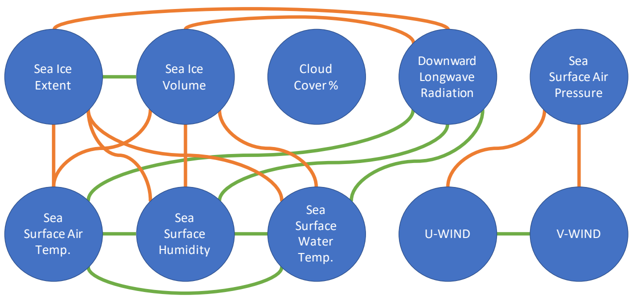

Figure 2. Diagram showing correlated relationships between variables in June from the observed dataset between 1979 to 2014. Green indicates a positive correlation and orange indicates a negative correlation. The correlation threshold is ± 0.6.

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 (Spirtes & Glymour, 1991), structural causal model frameworks, such as LiNGAM (Shimizu et al., 2006), 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 algorithms and structural causal models will be more effective.

3.1 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 data for sea level pressure data collected at 50 locations around the globe. 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 for analyzing climate systems. They looked at a single variable in various regions. In contrast, we plan to use PCMCI to 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 theory, such as information flow, are more appropriate for causal networks 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, the S-score, and the G-measure and F1-score (metrics combining TP and FP). 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 Anticipated Contributions

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 elucidate 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 eruptions and anthropogenic climate change and intervention. Developing more informative analytics for climate models will hasten their improvement and better inform policy decisions to mitigate and combat global climate change.

Acknowledgements

This work is supported by Sandia Earth Science Investment Area Laboratory Directed Research and Development funding. 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.

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

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.

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.

Reichenbach, H. *The direction of time*, volume 65. Univ of California Press, 1991.

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.

Shimizu, S., Hoyer, P. O., Hyvarinen, A., and Kerminen, A. A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research*, 7(72):2003–2030, 2006. URL https://www.jmlr.org/papers/volume7/shimizu06a/shimizu06a.pdf.

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

Spirtes, P. and Glymour, C. An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, 9(1):62–72, 1991. doi: 10.1177/089443939100900106. URL https://doi.org/10.1177/089443939100900106.

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.