Report for Project 1083 "Climate Informatics: New Machine Learning Methods for Climate Data and Climate Model Evaluation"

Project: **1083**
Project title: **Climate Informatics: New Machine Learning Methods for Climate Data and Climate Model Evaluation**
Project leader: **Prof. Dr. Jakob Runge**
Report period: **2023-07-01 to 2024-04-30**

Overview

In the last reporting period, we have developed bagged PCMCI+ algorithm to get a measure of confidence for the links based on bootstrapping. We have worked on developing new causal discovery algorithms for high-dimensional data sets and different regimes as well as methods to infer causally meaningful representations from data. We have applied existing causal inference algorithms to better understand the causal drivers for specific cloud regimes, atmospheric chemical-dynamical processes and Atlantic-Pacific interactions based on observations and climate model data. To detect extreme events, we used Gaussian mixture models and variational autoencoders.

Report for Task (i) Application of state-of-the-art causal discovery methods for observations and earth system model evaluation

We applied PCMCI, PCMCI+, and LPCMCI to identify causal relationships between variables involved in air-sea interactions within the Eddy Rich Earth System Models (EERIE) Horizon 2020 project and for modelling data of atmospheric chemical-dynamical processes for a project funded by the university of Bremen. We also started to apply Bagged-PCMCI+, the bootstrapped version of PCMCI+ (Debeire et al., 2024), to test the robustness of resulting causal networks (see Fig. 1).

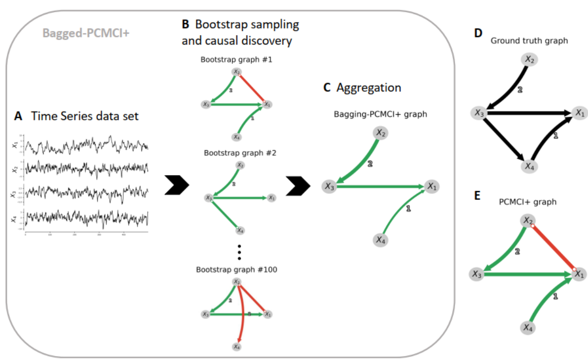

Figure 1: From (Debeire et al., 2024): Schematic example of Bagged-PCMCI+. (A) Time series for a model; (B) Ensemble of B = 100 estimated causal graphs obtained by applying PCMCI+ to each bootstrap dataset. Green links: true positives, red links: false positives, numbers next to the edges: non-zero time lags. (C) Graph obtained by aggregating the ensemble of graphs with confidence scores and visualized as the thickness of edges. (D) Ground truth causal graph of the model. (E) Estimated causal graph.

Report for Task (ii) Development and application of mixed-type causal discovery for observations and CMIP model evaluation

A mixed-type independence test was developed in (Popescu et al., 2023). For the analysis of causal drivers of cloud regimes, filter based regime oriented analysis (task (iii)) was used instead of mixed-type causal discovery.

Report for Task (iii) Development and application of methods for regime-dependent causal discovery

Günther et al. (2023) developed a PCMCI-based method for regime-dependent causal discovery. In Günther et al. (2023) applied methods from regime-based causal discovery to explore drivers of river discharge. Fons et al. (2023) investigated a regime-dependent problem of aerosol perturbations using causal inference.We applied filter based regime-dependent causal discovery to analyze major modes of climate variability and their teleconnections between

Atlantic and Pacific, evaluating CMIP model data (Karmouche et al., 2023b) and assessing the importance of external forcing (Karmouche et al., 2023a).

Report for Task (iv) Development and application of methods for causal inference with high-dimensional spatio-temporal data sets

Wahl et al. (2023) introduced a formal framework for causal discovery on vector-valued variables. Bing et al. (2024) discuss the problem of identifying latent causal representations from observational data. Debeire et al. (2024) introduces a technique for uncertainty quantification in causal discovery through bootstrapping. Kaltenborn et al. (2023) introduced a large-scale climate dataset as a test bed for causal inference and machine learning methods.

Report for Task (v) Application of extreme event detection machine learning techniques for CMIP model analysis

Building up on the analysis of heat waves based on Gaussian mixture models (GMMs) (Paçal et al., 2023), we started to use variational autoencoder (VAE) to detect extreme events in multivariate data. The standardized precipitation evapotranspiration index (SPEI) was computed to detect droughts and to analyze the dependency of changes of their frequency on climate sensitivity in climate model projections (Swaminathan et al., 2024).

References

S. Bing, U. Ninad, J. Wahl, and J. Runge. Identifying linearly-mixed causal representations from multi-node interventions. In *Proceedings of the Third Conference on Causal Learning and Reasoning*, volume 236 of *Proceedings of Machine Learning Research*, pages 843–867. PMLR, 01–03 Apr 2024.

K. Debeire, J. Runge, A. Gerhardus, and V. Eyring. Bootstrap aggregation and confidence measures to improve time series causal discovery, 2024. URL `https://arxiv.org/abs/2306.08946`. to be published in Proceedings of CLeaR 2024.

E. Fons, J. Runge, D. Neubauer, and U. Lohmann. Stratocumulus adjustments to aerosol perturbations disentangled with a causal approach. *npj Clim Atmos Sci*, 6(130), 2023.

W. Günther, U. Ninad, and J. Runge. Causal discovery for time series from multiple datasets with latent contexts. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 766–776. PMLR, 2023.

W. Günther, P. Miersch, U. Ninad, and J. Runge. Clustering of causal graphs to explore drivers of river discharge. *Environmental Data Science*, 2:e25, 2023. doi: 10.1017/eds.2023.17.

J. Kaltenborn, C. Lange, V. Ramesh, P. Brouillard, Y. Gurwicz, C. Nagda, J. Runge, P. Nowack, and D. Rolnick. Climateset: A large-scale climate model dataset for machine learning. In *Advances in Neural Information Processing Systems*, volume 36, pages 21757–21792. Curran Associates, Inc., 2023.

S. Karmouche, E. Galytska, G. A. Meehl, J. Runge, K. Weigel, and V. Eyring. Changing effects of external forcing on atlantic-pacific interactions. 2023a. URL `http://dx.doi.org/10.5194/egusphere-2023-1861`.

S. Karmouche, E. Galytska, J. Runge, G. A. Meehl, A. S. Phillips, K. Weigel, and V. Eyring. Regime-oriented causal model evaluation of atlantic–pacific teleconnections in cmip6. *Earth System Dynamics*, 14(2):309–344, 2023b. doi: 10.5194/esd-14-309-2023. URL `https://esd.copernicus.org/articles/14/309/2023/`.

A. Paçal, B. Hassler, K. Weigel, M. L. Kurnaz, M. F. Wehner, and V. Eyring. Detecting extreme temperature events using gaussian mixture models. *Journal of Geophysical Research: Atmospheres*, 128(18), Sept. 2023. ISSN 2169-8996. doi: 10.1029/2023jd038906. URL `http://dx.doi.org/10.1029/2023JD038906`.

O.-I. Popescu, A. Gerhardus, and J. Runge. Non-parametric conditional independence testing for mixed continuous-categorical variables: A novel method and numerical evaluation, 2023. URL `https://arxiv.org/abs/2310.11132`.

R. Swaminathan, J. Schewe, J. Walton, K. Zimmermann, C. Jones, R. A. Betts, C. Burton, C. D. Jones, M. Mengel, C. P. O. Reyer, A. G. Turner, and K. Weigel. Regional impacts poorly constrained by climate sensitivity, 2024. URL `https://arxiv.org/abs/2404.11939`. submitted to PNAS.

J. Wahl, U. Ninad, and J. Runge. Foundations of Causal Discovery on Groups of Variables, June 2023. arXiv:2306.07047 [math, stat].