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: **2022-05-01 to 2023-04-30**

Overview

In the last reporting period, we have completed the task on applying causal discovery to systematically find emergent constraints. We have worked on developing new causal inference algorithms for high-dimensional data sets, as well as novel conditions independence tests that these algorithms employ. We have applied existing causal inference algorithms to analyse and evaluate climate model data from the Coupled Model Intercomparison Project (CMIP).

Report for Task (i) Development and application of latent causal discovery methodology for observations CMIP model evaluation

The ESMValTool (Eyring et al., 2020) preprocessor as well as existing and newly created diagnostics are used to preprocess ECMWF ERA-5 reanalysis and CMIP simulations. These data are used to study causal networks using PCMCI. Several topics are addressed including the evaluation of CMIP6 models based on causal networks related to Arctic-midlatitude teleconnections (Galytska et al., 2022). For Arctic-midlatitude teleconnections as well as major modes of climate variability and their teleconnections between Atlantic and Pacific, latent PCMCI (LPCMCI), that allows for latent confounders, was tested, but the final analysis are based on PCMCI+, because LPCMCI was not well applicable in these cases due to too small sample sizes (Karmouche et al., 2023).

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

The analysis of modes of climate variability was taken further to include non-continuous data due to distinct regimes with different causal networks. Karmouche et al. (2023) used filters based on smoothed indices for Atlantic (AMV: Atlantic Multidecadal Variability) and Pacific (PDV: Pacific Decadal Varibility) variability to do Regime-oriented causal discovery using PCMCI+ on Reanalysis and CMIP6 Large ensembles (LEs), see Fig. 1. The networks and network similarity between LE simulations and reanalysis data differed for the selected regimes, with the best similarity when PDV and AMV are out of phase. This work shows, that causal discovery complements the available diagnostics and statistics metrics of climate variability providing a powerful tool for climate model evaluation.

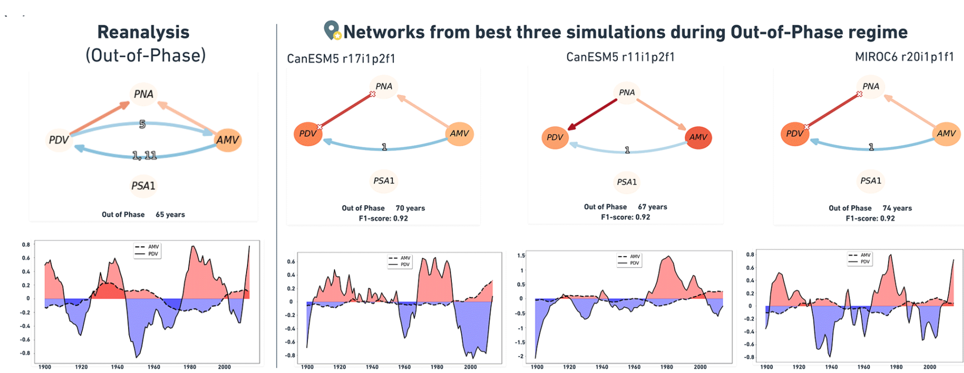

Figure 1: From (Karmouche et al., 2023): Reference causal network estimated from reanalysis during the Out-of-Phase regime (left, with low-pass AMV and PDV time series below) compared to networks and time series from three CMIP6 simulations (right, with simulated low-pass AMV and PDV time series below each network) with the best network similarity i.e. highest F1-score.

Report for Task (iii) Application of Causal Discovery to systematically find emergent constraints

We completed the work on a simplified stochastic climate model that outputs gridded data and represents climate modes and their teleconnections through a Spatially Aggregated Vector-Autoregressive (SAVAR) model (Tibau et al., 2022). This model is used to construct benchmarks and evaluate causal discovery methods on spatiotemporal data, see e.g. (Debeire et al., 2023 in preparation).

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

In Wahl* et al. (2022 accepted at AAAI), a novel density estimation based technique was developed to distinguish cause from effect, when the variables are high dimensional. We tested the method on the effect of the surface temperature of the ENSO 3.4 region over British Columbia. We also developed a novel conditional independence test (Günther et al., 2022) when the data is heteroscedastic.

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

Extreme temperature events were detected using Gaussian mixture models (GMMs) to analyze the change in their frequency using CMIP6 data sets under different global warming level scenarios (Pacal et al., 2023 under review for JGR - Atmosphere.). Additionally, GMMs are used to show how the temperature distribution changes for future scenarios. Classical drought indices (standardized precipitation index SPI and standardized precipitation evapotranspiration index SPEI) were computed with ESMValTool (Weigel et al., 2021) and compared between CMIP6 and reanalysis data (Weigel et al., 2023 in preparation).

References

K. Debeire et al. Estimating long-term effects of a perturbation in an aggregated auto-regressive model with uncertainty., 2023 in preparation.

V. Eyring et al. Earth system model evaluation tool (ESMValTool) v2.0 – an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of earth system models in CMIP. *Geoscientific Model Development*, 13(7):3383–3438, July 2020. doi: 10.5194/gmd-13-3383-2020. URL `https://doi.org/10.5194/gmd-13-3383-2020`.

E. Galytska, K. Weigel, D. Handorf, R. Jaiser, R. H. Köhler, J. Runge, and V. Eyring. Causal model evaluation of Arctic-midlatitude teleconnections in CMIP6. *Earth and Space Science Open Archive*, page 33, 2022. doi: 10.1002/essoar.10512569.1. URL `https://doi.org/10.1002/essoar.10512569.1`.

W. Günther, U. Ninad*, J. Wahl*, and J. Runge. Conditional independence testing with heteroskedastic data and applications to causal discovery. In *Advances in Neural Information Processing Systems*, volume 35, pages 16191–16202. Curran Associates, Inc., 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/file/6739d8df16b5bce3587ca5f18662a6aa-Paper-Conference.pdf`.

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, 2023. doi: 10.5194/esd-14-309-2023. URL `https://esd.copernicus.org/articles/14/309/2023/`.

A. Pacal, B. Hassler, K. Weigel, M. Wehner, M. L. Kurnaz, and V. Eyring. Detecting extreme events using gaussian mixture models, 2023 under review for JGR - Atmosphere.

X.-A. Tibau, C. Reimers, A. Gerhardus, J. Denzler, V. Eyring, and J. Runge. A spatiotemporal stochastic climate model for benchmarking causal discovery methods for teleconnections. *Environmental Data Science*, 1:e12, 2022. doi: 10.1017/eds.2022.11.

J. Wahl*, U. Ninad*, and J. Runge. Vector causal inference between two groups of variables. *arXiv:2209.14283*, 2022 accepted at AAAI.

K. Weigel et al. Earth system model evaluation tool (ESMValTool) v2.0 – diagnostics for extreme events, regional and impact evaluation, and analysis of earth system models in CMIP. *Geoscientific Model Development*, 14(6): 3159–3184, June 2021. doi: 10.5194/gmd-14-3159-2021. URL `https://doi.org/10.5194/gmd-14-3159-2021`.

K. Weigel et al. Evaluation of extreme droughts in cmip6 model simulations, 2023 in preparation.