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: **2025-05-01 to 2026-04-30**

Overview

In the last reporting period, we worked on multiple frameworks for handling high spatiotemporal dimensionality (task iv) and endogenous regimes (task iii) for causal discovery and effect estimation. We additionally applied existing causal inference algorithms to better understand the causal drivers for specific cloud regimes (task iii), atmospheric chemical-dynamical processes and air-sea interactions (task i) based on observations and climate model data. Finally, we improved parameterizations for earth system models with machine learning (task ii) and used variational autoencoders to characterize and analyze heatwaves and to detect and analyze droughts in climate model projections (task v).

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

We used PCMCI+ to analyze causal networks for variables related to air-sea interactions in the Southern Ocean from observation based data and high-resolution Earth System Model daza from the Eddy Rich Earth System Models (EERIE) Horizon 2020 project. A causal inference framework that combines causal discovery with causal effect estimation was used to investigate the coupling between chemical and dynamical processes driving tropical middle stratospheric ozone variability (Galytska et al., 2025). Debeire et al. (2025) analyzed the impact of perturbations on both large- and small-scale variables to enhance the understanding of complex systems.

Task (ii) Development and application of improved parameterizations for earth system models with causal informed neural networks and equation discovery

We concluded the experiments with machine learning stochastic and deterministic multi-member convection parameterizations for the EERIE project. Additionally, we completed the sea-ice albedo parameterization based on data-driven equation discovery using symbolic regression on satellite and reanalysis data and analyzed the results (Atmojo et al., 2025). In (Bonnet et al., 2025), an automated, ML-based tuning was applied to adjust uncertain parameters in the model.

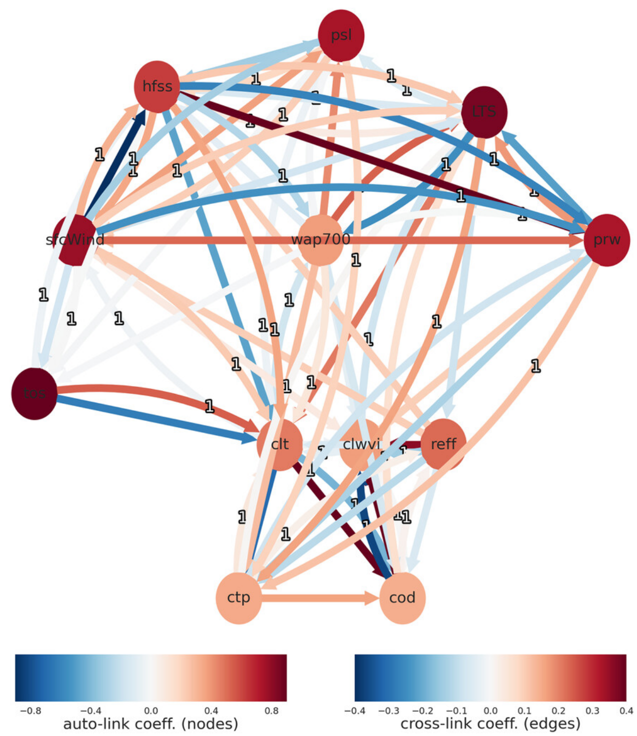

Figure 1: From (Bock et al., 2026): Full causal graph including estimated link strengths showing all links found between all CCFs and cloud properties: Total cloud fraction (clt), total cloud water path (clwvi), cloud optical depth (cod), cloud effective radius (reff), cloud-top pressure (ctp), sea surface temperature (tos), water vapor path (prw), vertical velocity at 700 hPa (wap700), lower tropospheric stability (LTS), sea level pressure (psl), sensible heat flux at the surface (hfss), 10-m horizontal wind speed (sfcWind).

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

We extended the framework for causal discovery with endogenous regimes established in Günther et al. (2024) to the time series setting (Popescu et al., 2026), and developed a modular framework for handling non-stationarity, regimes, and spatiotemporal patterns simultaneously for causal discovery (Rabel and Runge, 2026). As part of the European Space Agency (ESA) CMUG (Climate Modelling User Group) project, we used causal inference to better understand the causal drivers for cloud properties (Bock et al., 2026). We applied the causal discovery algorithm LPCMCI on the ESA Climate Change Initiative (CCI) and ERA5 data to estimate causal links from cloud-controlling factors (CCFs) to cloud properties (see Fig. 1) and quantified them using causal effect estimation focused on the marine stratocumulus region in western South America.

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

We began work developing and testing a causal discovery algorithm for dynamical systems that reduces dimensionality over time, and we worked on several approaches for handling high dimensionality over space: We developed structural causal bottleneck models that provide a flexible framework for task-specific dimension reduction (Bing et al., 2026). We developed a conceptual and computational framework that employs clustering for spatiotemporal causal effect estimation in complex dynamical systems (Herman et al., 2026), and explored the use of clustering or aggregation in the context of causal discovery (Ninad et al., 2025).

Task (v) Application of extreme event and tipping points detection machine learning techniques.

Lanson and Runge (2025) use causal inference to distinguish between autocorrelation due to autocorrelation of causal drivers and that due to widening of the potential well to detect tipping points more reliably. Changes in drought characteristics and their occurrence in harvest areas were analyzed based on projections from Earth system models (Lindenlaub et al., 2026, now published). Patterns related to European heatwaves were detected in ERA5 reanalysis data using a variational autoencoder (VAE) and analyzed based on their latent space (Paçal et al., 2025).

References

D. W. Atmojo, K. Weigel, A. Grundner, M. M. Holland, D. Sidorenko, and V. Eyring. Data-driven equation discovery of a sea ice albedo parametrisation. *EGUsphere*, 2025:1–34, 2025. doi: 10.5194/egusphere-2025-3556. URL `https://doi.org/10.5194/egusphere-2025-3556`.

S. Bing, J. Wahl, and J. Runge. Structural causal bottleneck models, 2026. URL `https://arxiv.org/abs/2603.08682`.

L. Bock, A. Lauer, and J. Runge. Quantifying the causal effect of cloud-controlling factors on marine stratocumulus clouds. *Journal of the Atmospheric Sciences*, 83 (2):255–264, Feb. 2026. ISSN 1520-0469. doi: 10.1175/jas-d-25-0153.1. URL `http://dx.doi.org/10.1175/JAS-D-25-0153.1`.

P. Bonnet, L. Pastori, M. Schwabe, M. Giorgetta, F. Iglesias-Suarez, and V. Eyring. Tuning the icon-a 2.6.4 climate model with machine-learning-based emulators and history matching. *Geoscientific Model Development*, 18(12):3681–3706, 2025. doi: 10.5194/gmd-18-3681-2025. URL `https://gmd.copernicus.org/articles/18/3681/2025/`.

K. Debeire, A. Gerhardus, R. Bichler, J. Runge, and V. Eyring. Uncertainty bounds for long-term causal effects of perturbations in spatiotemporal systems. *Environmental Data Science*, 4:e33, 2025. doi: 10.1017/eds.2025.10007.

E. Galytska, B. Hassler, C. Arosio, M. P. Chipperfield, S. S. Dhomse, K. Dubé, W. Feng, F. Iglesias-Suarez, and J. Runge. Causal inference for stratospheric chemistry: insights into tropical middle stratospheric ozone variability. *EGUsphere*, 2025. doi: 10.21203/rs.3.rs-6426983/v2. URL `http://dx.doi.org/10.21203/rs.3.rs-6426983/v2`.

W. Günther, O.-I. Popescu, M. Rabel, U. Ninad, A. Gerhardus, and J. Runge. Causal discovery with endogenous context variables. *Advances in Neural Information Processing Systems*, 37:36243–36284, 2024.

R. Herman, U. Ninad, and J. Runge. Spatiotemporal causal effect estimation for complex dynamical systems. in preparation, 2026.

A. Lanson and J. Runge. Causal effect estimation for robust detection of critical slowing down, 2025. URL `https://doi.org/10.5194/egusphere-egu25-10947`.

L. Lindenlaub, K. Weigel, B. Hassler, C. Jones, and V. Eyring. Characteristics of agricultural droughts in cmip6 historical simulations and future projections. *Earth System Dynamics*, 17(1):81–105, Jan. 2026. ISSN 2190-4987. doi: 10.5194/esd-17-81-2026. URL `http://dx.doi.org/10.5194/esd-17-81-2026`.

U. Ninad, J. Wahl, A. Gerhardus, and J. Runge. Causal discovery on vector-valued variables and consistency-guided aggregation, 2025. URL `https://arxiv.org/abs/2505.10476`.

A. Paçal, B. Hassler, K. Weigel, M.-A. Fernández-Torres, G. Camps-Valls, and V. Eyring. Understanding european heatwaves with variational autoencoders. *EGUsphere*, 2025. doi: 10.5194/egusphere-2025-2460. URL `http://dx.doi.org/10.5194/egusphere-2025-2460`.

O.-I. Popescu, W. Günther, M. Rabel, and J. Runge. Causal discovery with endogenous context variables for time-series. Accepted at ICML. Equal contribution from Popescu and Günther, 2026.

M. Rabel and J. Runge. Context-specific causal graph discovery with unobserved contexts: Non-stationarity, regimes and spatio-temporal patterns, 2026. URL `https://arxiv.org/abs/2511.21537`.