Understanding drivers of climate extremes using regime-specific causal graphs

Climate Change AI Workshop at ICLR 2024

Oana Popescu[1], Wiebke Günther[1], Raed Hamed[3], Dominik Schumacher[4], Martin Rabel[1], Dim Coumou[3], Jakob Runge[1,2]

[1]German Aerospace Center (DLR), Institute of Data Science; [2]TU Berlin; [3]Institute for Environmental Studies, VU Amsterdam; [4]Institute for Atmospheric and Climate Science, ETH Zürich

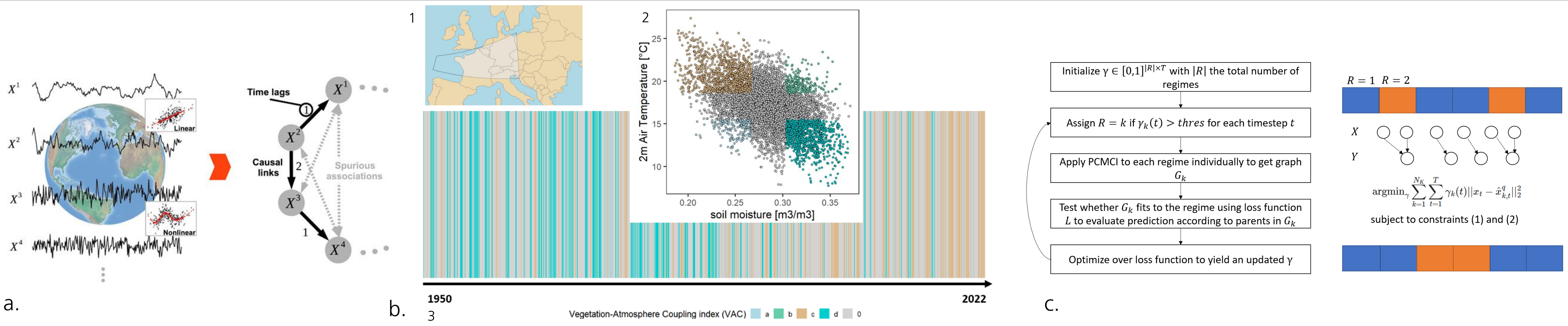

Figure 1. a. Causal discovery can uncover causal graphs from observed time-series data. Taken from [7]. **b.1**. The dataset used in this tutorial contains daily values for the highlighted region in Western Europe across the period 1950-2022. **b.2.** Soil moisture and air temperature are used to generate the VAC index [4] using thresholding. **b.3.** The daily evolution of the VAC index (yearly values from March to September). **c.** The steps of the Regime-PCMCI [1] algorithm.

Motivation

- The rise in global temperature has been accompanied by a rise in extreme weather events such as temperature anomalies.
- In western and central Europe, soils typically contain sufficient water to sustain high evaporation and limit surface sensible heating. In recent years, a decrease in soil moisture has been observed (Fig. 1.b.3.). This leads to soil-moisture temperature feedbacks, which can amplify hot extremes beyond greenhouse-gas induced global warming.
- Causal discovery methods can identify the underlying causal relationships between various climate variables to discern the factors contributing to unusual weather patterns.
- We frame the problem of uncovering soil moisture drivers in Western Europe using regime-specific causal discovery and apply Regime-PCMCI [1], implemented in the Tigramite Python package (https://github.com/jakobrunge/tigramite).

Causal discovery for climate anomalies

- Data with distribution shifts such as anomalies can have periods during which the data exhibits different statistical patterns.

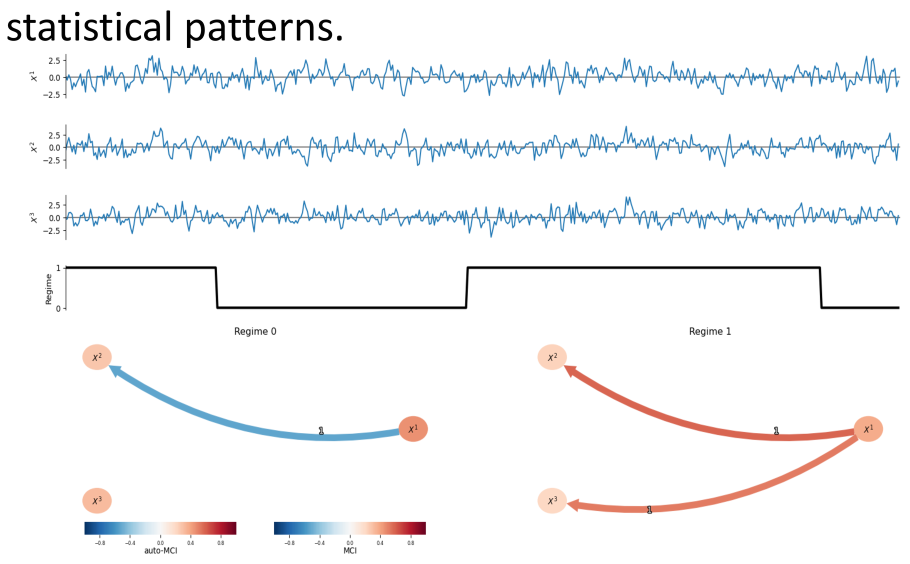

Figure 2. Example of two regime-specific graphs and time-series data generated using these causal graphs. Taken from the Tigramite repository.

- Each regime is characterized by its own set of statistical properties: the "normal" causal structure or causal mechanisms are altered, resulting in different regimes governing the "normal" and "anomalous" state, as exemplified in Fig. 2.
- We model each regime using a structural causal model (SCM), under the assumption that an exogeneous variable R describes the N_K regimes.
- The regime-specific SCM for regime k with $k = 1, \dots, N_K$ at time t is written as

$$\boldsymbol{X}_{k,t} = f^k\left(Pa^k(\boldsymbol{X}_t), \boldsymbol{U}_{k,t}\right)$$

  - $Pa^k(\boldsymbol{X}_t)$ – the vector of variables which are direct causes of $X_{i,t} \in \boldsymbol{X}_t$
  - $\boldsymbol{U}_{k,t}$ – exogeneous noise.

$$X_i \xrightarrow{X_j = f_j(X_i, U_j)} X_j$$

Figure 3. Example of a two-variable causal graph, where X_i causes X_j**.** X_j can be written as function f_j of the parent X_i and its exogeneous noise U_j.

Regime-PCMCI

- Regime-PCMCI [1] builds upon the PCMCI [2] algorithm. PCMCI uses an independence-testing approach adapted to the time-series case for constaint-based causal discovery, and additionally orients edges from past to future.
- Regime-PCMCI finds a regime assignment for each sample by alternating between two steps, as also depicted in Fig.1.c.:
  1. **Discovering regime-specific causal graphs** using PCMCI. At first, a random assignment of regimes is used.
  2. **Detecting regimes:**
     - At iteration q, predict $\hat{x}_{k,t}^q = \hat{G}_t(Pa_k(\boldsymbol{X}_{k,t}, \Theta_t)$ at time step t using parents $Pa^k(\boldsymbol{X}_t)$ for regime k using the functional with parameters Θ_t.
     - Predict the regime labels γ_k using the distance between the real and predicted data for a new iteration of the optimization, subject to constraints (1) and (2), where N_C is the number of transitions between regimes (user-defined):

$$\arg min_\gamma \sum_{k=1}^{N_K} \sum_{t=1}^{T} \gamma_k(t) |x_t - \hat{x}_{k,t}^q|_2^2$$

$$(1) \sum_{k=1}^{N_K} \gamma_k(t) = 1 \quad \forall t \; with \; \gamma_k(t) \in [0,1]$$

$$(2) \sum_t^{T-1} |\gamma_k(t+1) - \gamma_k(t)| \leq N_C.$$

Dataset description

- The dataset contains daily values for the time period 1950-2022 in Western Europe (see Fig.1.b.1.) from ERA5 [3] for the following:
  - Mean air temperature 2m above ground (T2m),
  - Root-zone (1m) soil moisture (SM),
  - Surface sensible heat flux (SH),
  - Surface sensible heat flux (SH),
  - Surface latent heat flux (LH),
  - Large-scale circulation function at 250 hPa (Stream),
  - Shortwave downward radiation (SW),
  - The vegetation-atmosphere coupling (VAC) index [4] computed using thresholding of SM and T2m anomalies as in Fig. 1.b.1.

Experimental setup

- We search for the causal links between LH, SH, and SM in the moist and dry regimes. We hypothesize the unionized causal graph in Fig. 5:
  - In the dry regime, drier soils lead to a reduction in LH flux (the energy used for evaporation), which leads to an increase in SH flux. These conditions can further exacerbate hot and dry conditions [5].
  - In moist regimes, LH is mostly insensitive to SM variations. Evaporation is instead controlled by other factors, such as cloud cover or sunshine hours [6].
- We aggregate data points using three-day averages in the time period 1993 -2022.
- We search for three regimes: dry, moist, and one regime for all samples that do not fit into the others.

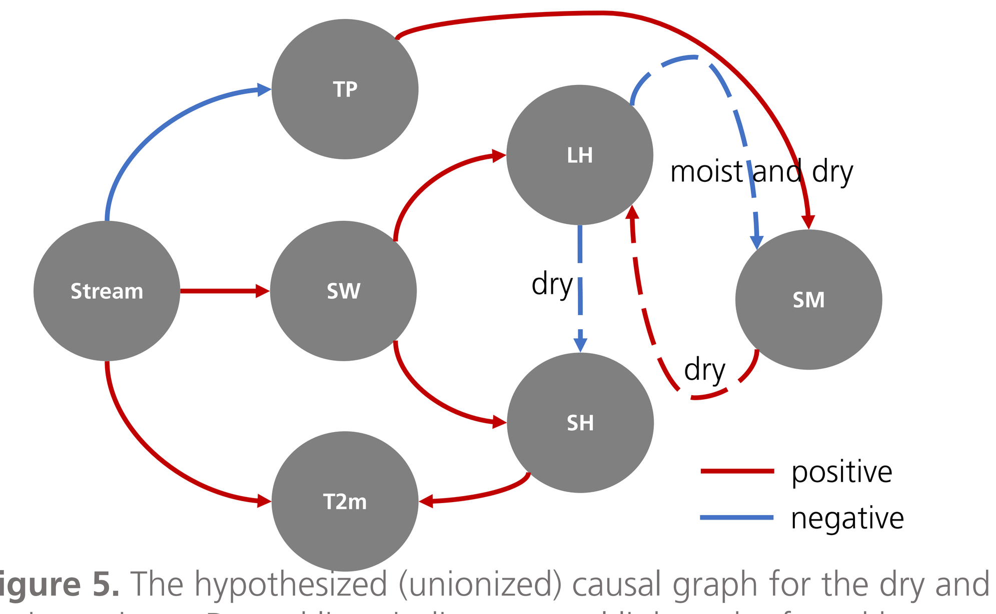

Figure 5. The hypothesized (unionized) causal graph for the dry and moist regimes. Dotted lines indicate causal links to be found by Regime-PCMCI. For each of the dotted line, we indicate which regime it corresponds to.

Results

- We present results for the moist regime (regime 1) and the dry regime (regime 2), as for regime 0 no further links were found.

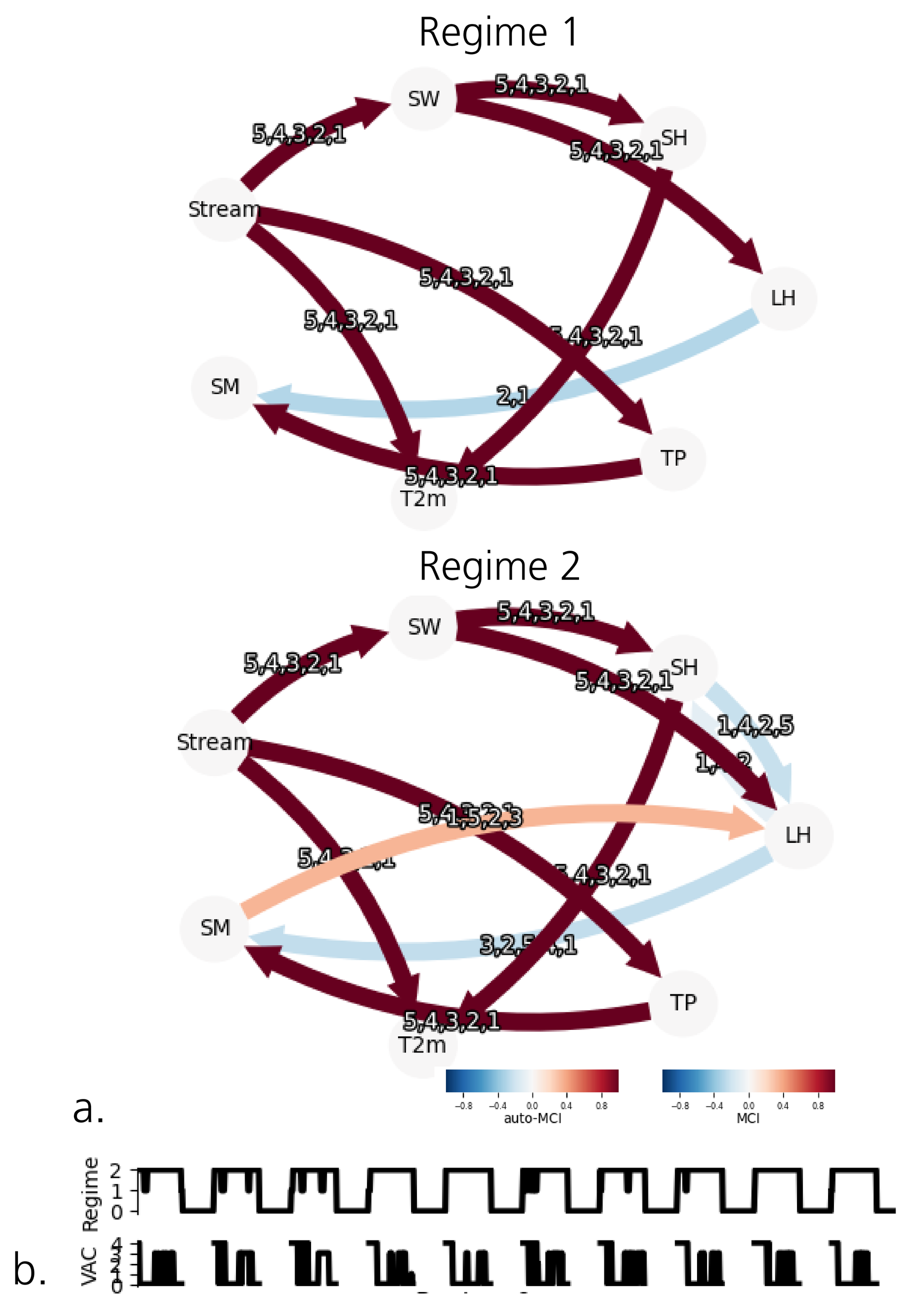

Figure 6. a. The two regimes of interest discovered with Regime-PCMCI. **b.** Comparison of the regimes assigned by Regime-PCMCI (above) with the VAC index (below).

Discussion

- We discover the key causal links which characterize the moist and dry regimes in a simplified setting
- Regime-based causal discovery offers a promising approach to understanding anomalies, but can be particularly challenging when data stems from a high-dimensional, strongly coupled system.
- Challenges also arise due to limitations of the Regime-PCMCI algorithm, such as:
  - Strong assumptions for causal discovery, such as causal sufficiency, no unmeasured confounders, no acyclicity,
  - Assumptions of the conditional independence tests,
  - Sensitivity to selected time-scale,
  - Computational complexity of the algorithm, as well as possible model and algorithm bias.

References

[1] Saggioro, E., de Wiljes, J., Kretschmer, M., & Runge, J. (2020). Reconstructing regime-dependent causal relationships from observational time series. Chaos, 30 11, 113115.

[2] Runge, J., Nowack, P.J., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2017). Detecting and quantifying causal associations in large nonlinear time series datasets. Science Advances, 5.

[3] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G.D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R.G., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hogan, R.J., Holm, E.V., Janisková, M., Keeley, S.P., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P.D., Rozum, I., Vamborg, F., Villaume, S., & Thepaut, J. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146, 1999 - 2049.

[4] Zscheischler, J., Orth, R., & Seneviratne, S.I. (2015). A submonthly database for detecting changes in vegetation-atmosphere coupling. *Geophysical Research Letters, 42*, 9816 - 9824.

[5] Seneviratne, S.I., T. Corti, E.L. Davin, M. Hirschi, E.B. Jaeger, I. Lehner, B. Orlowsky, and A.J. Teuling, 2010: Investigating soil moisture-climate interactions in a changing climate: A review. Earth-Science Reviews, 99, 3-4, 125-161, doi:10.1016/j.earscirev.2010.02.004.

[6] Benson, D. & Dirmeyer, P. Characterizing the Relationship between Temperature and Soil Moisture Extremes and Their Role in the Exacerbation of Heat Waves over the Contiguous United States. J. Clim. 34, 2175–2187, doi=10.1175/JCLI-D-20-0440.1, 2021.

[7] Runge, J., Nowack, P.J., Kretschmer, M., Flaxman, S., & Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances, 5*. 2017.