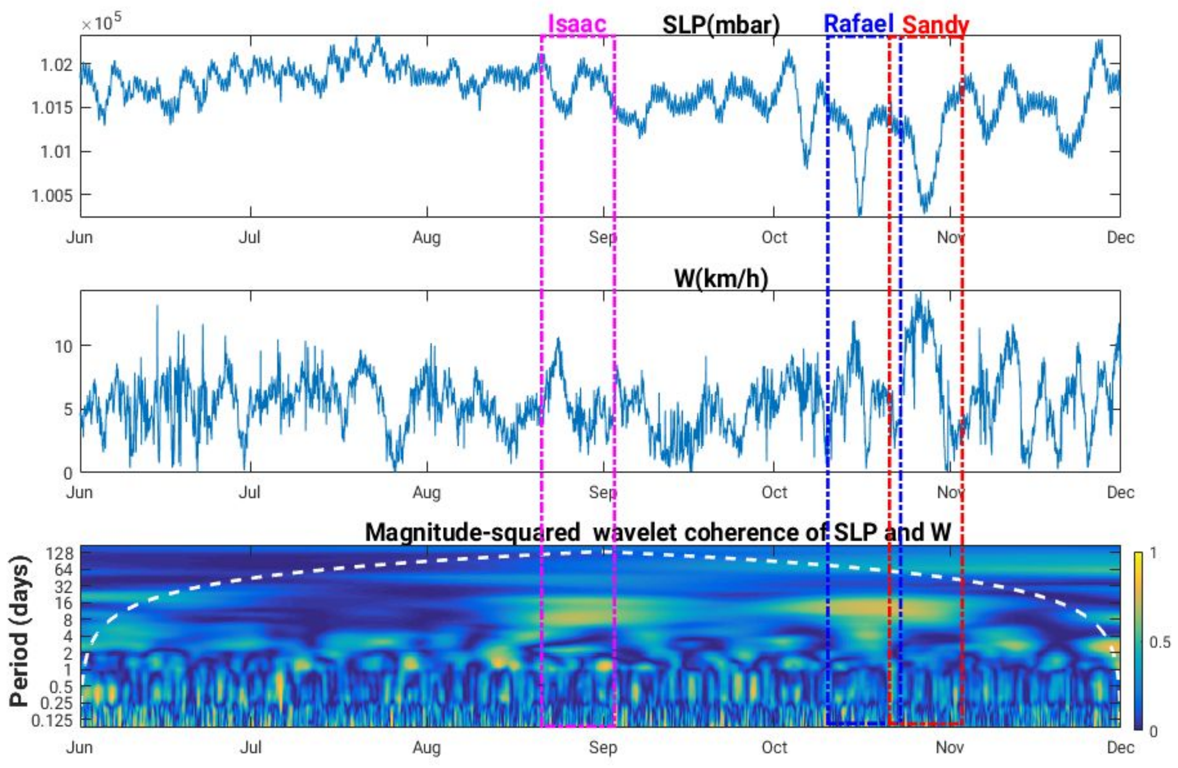
Understanding Changes in Environmental Time Series with Time-Frequency Causality Analysis

Maha Shadaydeh[1], Yanira Guanche[2,4], Miguel Mahecha[3,4], Joachim Denzler[1,4]

[1] Friedrich Schiller University, Jena [2] Institute of Data Science, German Aerospace Center, Jena [3] Max Planck Institute for Biogeochemistry, Jena [4] Michael Stifel Center for Data driven and Simulation Science, Jena, Germany

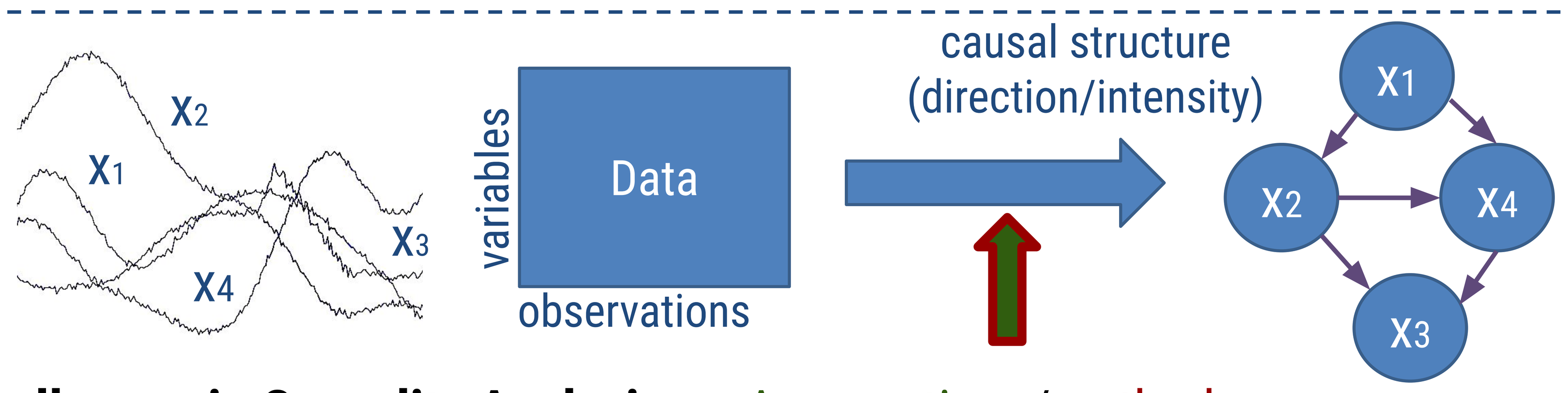
Research Objective

Can we detect and attribute anomalous events based on the analysis of changes in the causal effect relationships?



The high magnitude wavelet coherence of sea level pressure (SLP) and wind speed (W) correlates well with the date and time-scale (8-16 days) of the historic hurricanes in year 2012.

Method



Challenges in Causality Analysis → Assumptions/methods

- Nonstationary → sliding time window/adaptive system
- Presence of periodic components → frequency domain analysis
- Multi-scale cause effect relations → time-frequency analysis
- Nonlinearity → separability of cause and effect, no prior assumption on the linearity of the cause-effect relation.
- Presence of hidden cofounders → assume no hidden cofounders

Time Domain Vector Autoregressive (VAR) Granger Causality

The time series of the *N* environmental variables are represented by a p^{th} order Vector Autoregressive (VAR) model:

$$\begin{bmatrix} x_1(n) \\ \vdots \\ x_N(n) \end{bmatrix} = \sum_{r=1}^{p} A_r \begin{bmatrix} x_1(n-r) \\ \vdots \\ x_N(n-r) \end{bmatrix} + \begin{bmatrix} \epsilon_1(n) \\ \vdots \\ \epsilon_N(n) \end{bmatrix} \qquad A_r = \begin{bmatrix} a_{11}(r) & \dots & a_{1N}(r) \\ \vdots & \ddots & \vdots \\ a_{N1}(r) & \dots & a_{NN}(r) \end{bmatrix}$$

The causal relation from x_i to x_j conditioned on other variables is [1][2]

$$\gamma_{i \to j} = \ln \frac{|\Sigma_j^{i-}|}{|\Sigma_j|}$$

Σ_j^{i-}: covariance matrix of the residual j associated to x_j using the reduced model (after eliminating x_i)

Σ_j: covariance matrix of the residual j associated to x_j using the full model

Frequency Domain VAR-GC: The generalized Partial Directed Coherence (gPDC)

The causal relation from x_i to x_j is described in the freq. domain via the gPDC [3]:

$$g\pi_{i \to j}(f) = \frac{\frac{1}{\sigma_{jj}} \overline{A}_{ji}(f)}{\sqrt{\sum_{k=1}^{m} \frac{1}{\sigma_{kk}^2} \left|\overline{A}_{ki}(f)\right|^2}} \qquad \overline{A}(f) = I - A(f) \qquad A(f) = \sum_{r=1}^{p} \mathbf{A}_r z^{-r}|_{z=e^{i2\pi f}}$$

Nullity of $g\pi_{i \to j}(f)$ indicates an absence of Granger causality from x_i to x_j at normalized freq. f.

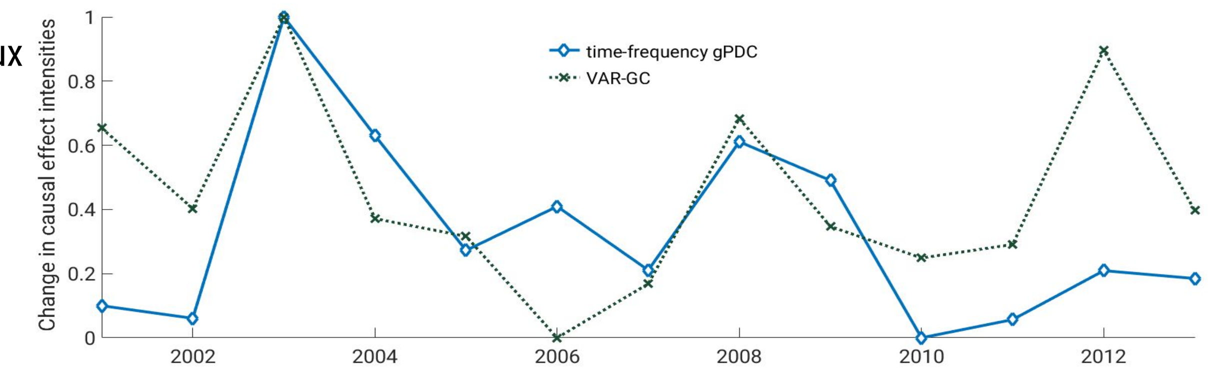
Experimental Results

Event Detection in Ecological Time Series [4]

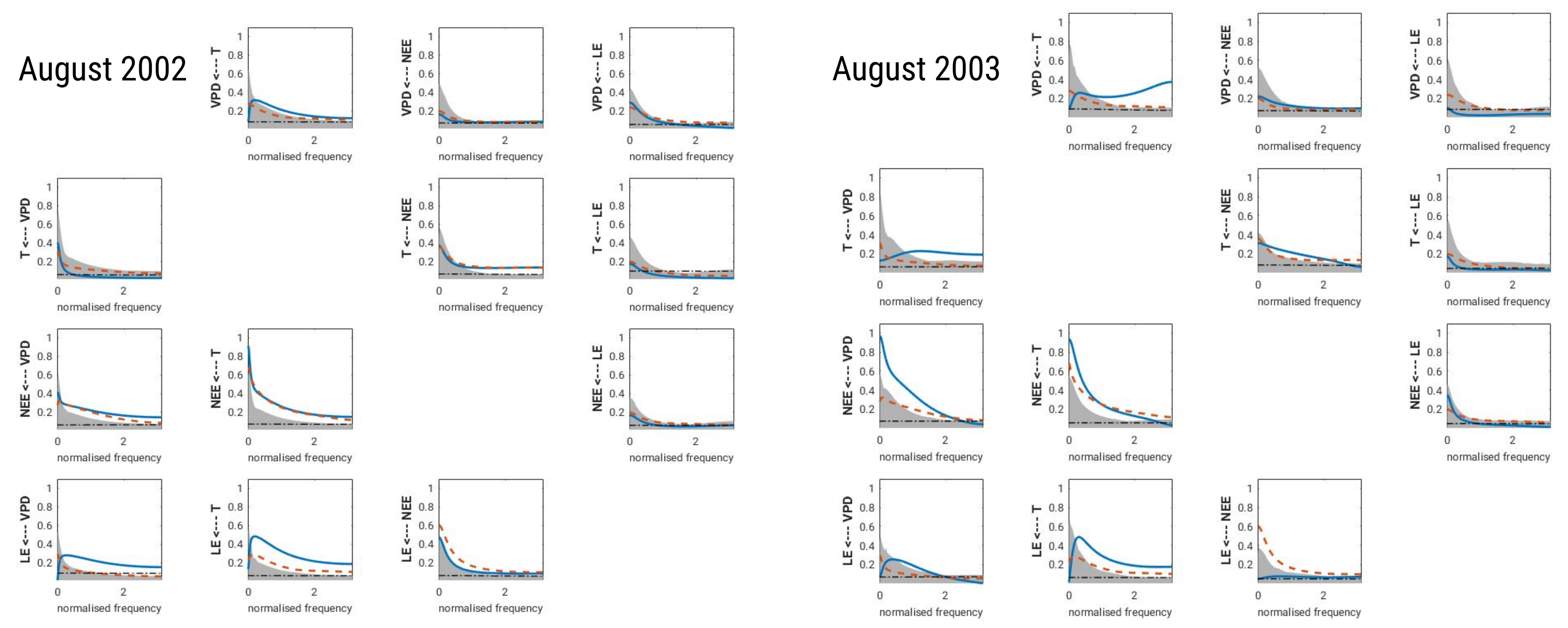
How ecosystem functioning is affected during unusual hydro-meteorological conditions?

Half hourly data measured at the flux tower of Puechabon-France

Variables:
Air Temperature (T)
Vapor Pressure Deficiency (VPD)
Latent Energy (LE)
Net Ecosystem Exchange (NEE)



The deviation of the causal effect intensities from the average ones during the months of August in years 2001-2013. The peak in year 2003 is due to the 2003 summer heatwave in France.

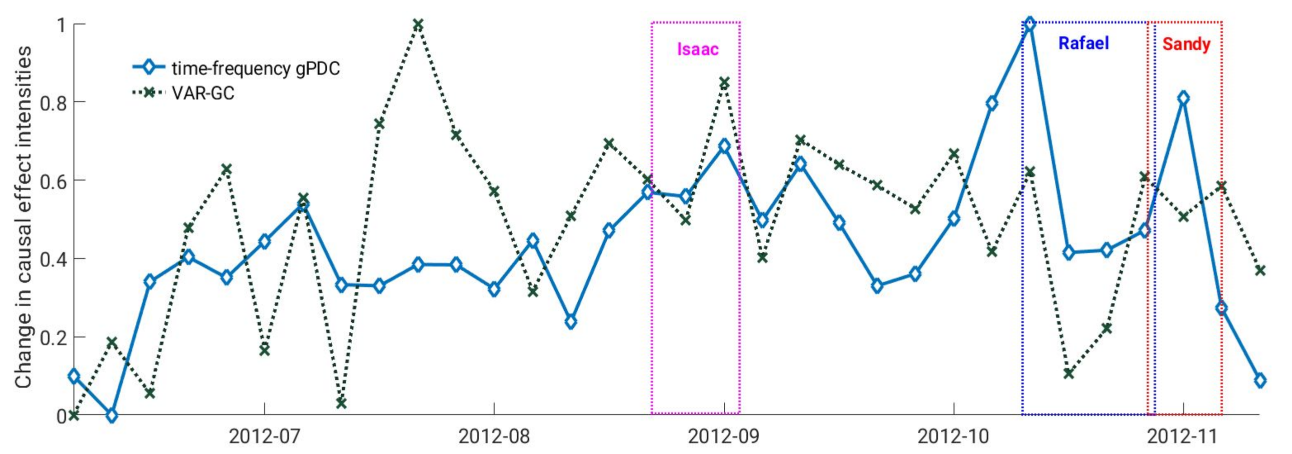


During the heatwave of August 2003 in France, we can notice deviation of the causal effect intensities (blue solid line) from to the average ones of similar summer period within years 2001-2013 (red dashed line): clear increase in the causal intensity of VPD on NEE at the low frequency range (long term change) pointing towards an increase in water stress on ecosystem functioning; Note also the increased effect of T on VPD as short term change (high freqs.). The threshold for statistical significance estimated using permutation test and the FT surrogate test is shown in the dashed-dotted line and gray area respectively.

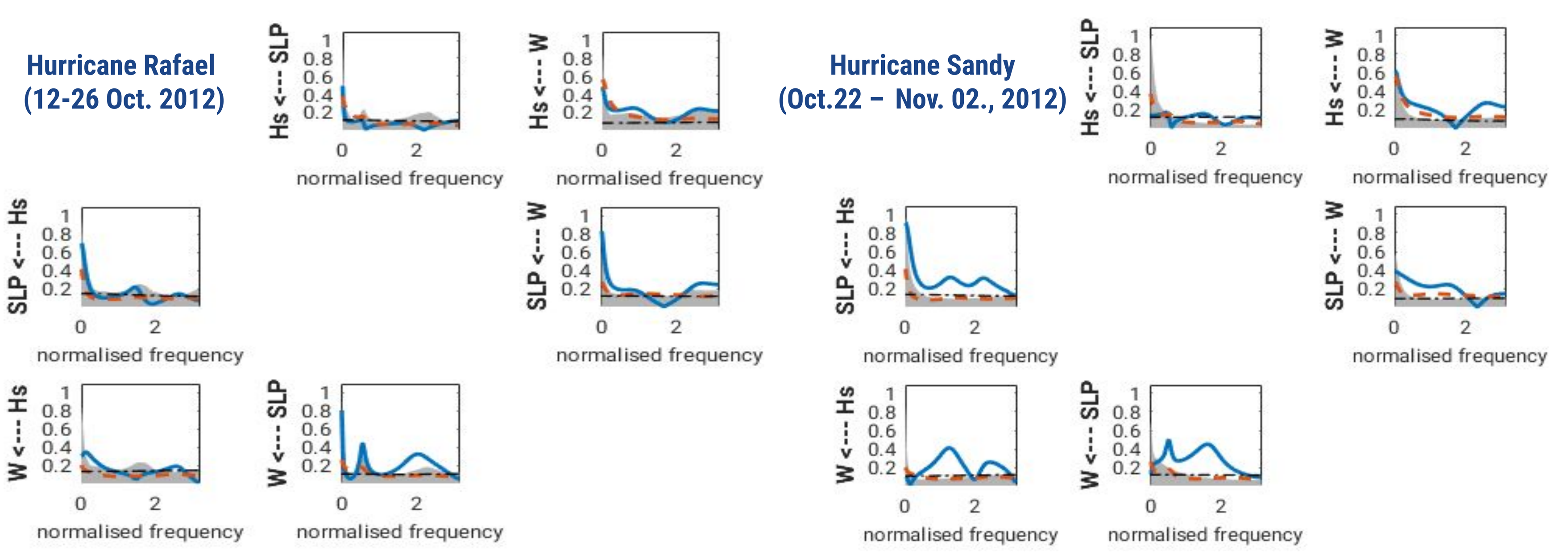
Event Detection in Marine Climate Time Series [4]

Hourly data from a buoy located near the Bahamas in the Atlantic Sea (23.838° N, 68.333° W).

Variables:
Sea level pressure (SLP)
Wind speed (W)
Wave height (Hs)



Changes in the causal effect intensities. For the time-frequency analysis using the gPDC, the high change values correlate better with the three hurricanes when compared to time domain VAR-GC.



The causal effect intensities calculated using the gPDC over sliding time window (blue line) when compared to the average causal intensities (red dashed line). The threshold for statistical significance estimated using permutation test and the FT surrogate test is shown in the dashed-dotted line and gray area respectively. Note the increased effect of SLP on W, the sharp peak corresponds to the semi-diurnal cycle of the atmospheric pressure.

Acknowledgment & References

This study is supported by the Carl Zeiss Foundation within the scope of the program line "Breakthroughs: Exploring Intelligent Systems" for "Digitization — explore the basics, use applications".

References

[1] Granger, C.W. 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, pp. 424-438.
[2] Geweke, J. 1982. Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association, vol. 77, pp. 304-313.
[3] Baccalá, L.A., Sameshima, K., and Takahashi, D. 2007. Generalized partial directed coherence. IEEE 15th International Conference in Digital Signal Processing, pp.163-166.
[4] Shadaydeh M., Denzler J., García Y.G., Mahecha M. (2019) Time-Frequency Causal Inference Uncovers Anomalous Events in Environmental Systems. Lecture Notes in Computer Science, vol 11824. Springer, Cham

Conclusions & Future Work

- An anomalous event can be detected as the one where the causal intensities between the variables differ from the average causal intensities.
- Time frequency analysis allows for causality analysis at different frequency components and different time scales. Further details can be found in [4].
- The analysis of the spectral causal effect patterns allows for understanding these events and define the time scale on which changes occur.
- Further work: 1. Extension to state space model with latent variables, 2. Extension to spatiotemporal data, and 3. Adaptive window size selection based on stationarity test.

Deutsches Zentrum für Luft- und Raumfahrt DLR | Carl Zeiss Stiftung | Max Planck Institute for Biogeochemistry | MichaelStifelCenterJena for Data-Driven and Simulation Science | Computer Vision Group | Friedrich-Schiller-Universität Jena