8th International Workshop on Climate Informatics
September 19-21, 2018
CI 2018 Hosted by the National Center for Atmospheric Research in Boulder, CO

CAUSALITY ANALYSIS OF ECOLOGICAL TIME SERIES: A TIME-FREQUENCY APPROACH

Maha Shadaydeh[1], Yanira Guanche[1,2], Miguel Mahecha[2,3], Markus Reichstein[2,3] and Joachim Denzler[1,2]

***Abstract*—Attribution in ecosystems aims to identify the cause-effect relationships between the variables involved. Time series of ecological variables most often contain multiple periodical components, e.g. daily and seasonal cycles, induced by the meteorological forcing variables. Such components can significantly mask the underling endogenous causality structure of the biogeochemical cycle when using time domain analysis. This motivates the use of time-frequency analysis techniques such as short time Fourier transform or wavelet where causality inference can be investigated at different frequency bands or different time scales. In this study, we use the parametric spectral representation of the multivariate autoregressive Granger causality based on the generalized Partial Directed Coherence (gPDC) to investigate the cause-effect relationships between the meteorological observations of global radiation and air temperature, and the CO2 land fluxes of gross primary productivity and ecosystem respiration, at Hainich National Park-Germany. Preliminary results show that spectral domain causality analysis based on gPDC has promising potential in handling the presence of periodic components and in identifying the time variant cause-effect intensities between these variables at different time scales.**

I. INTRODUCTION

Local meteorological conditions have direct impact on CO2 fluxes and ecosystem respiration. Understanding the cause-effect relationships in such dynamical system is essential for the attribution of climate changes as well as for the development of intervention strategy to achieve desired prediction. The availability of high temporal resolution data along with the powerful computing platforms further enhance the capacity of data-driven methods in capturing the complex relationships between the variables of the underlying dynamical system. Time series of ecological variables most often contain multiple periodical components, e.g. daily and seasonal cycles, induced by the meteorological forcing variables. This can significantly mask the underling endogenous causality structure of the biogeochemical cycle when using time domain analysis. Filtering these periodic components as preprocessing step degrades causal inference [1]. This motivates the use of time-frequency processing techniques such as short time Fourier transform where the causality structure can be examined at different frequency bands or different time scales. In this study, we present a time-frequency approach for causality analysis applied to the meteorological observations and land flux eddy covariance data to investigate the causal-effect relationships between global radiation (Rg), air temperature (T), and the CO2 land fluxes: gross primary productivity (GPP) and ecosystem respiration (Reco). The coupling between the used variables is assumed to follow a multivariate autoregressive (MVAR) model. The cause-effect relationships are extracted using the MVAR Granger causality (MVAR-GC) [2], [3] based on the generalized partial directed coherence (gPDC) [4], [5]. We compare experimental results obtained using gPDC with those using time domain conditional MVAR-GC [3], [7] to highlight the advantages of using frequency analysis techniques. To account for the nonstationarity of the used variables, we also present the gPDC causality analysis using short time window approach and compare the time variant causal-effect intensities obtained over different seasons.

Corresponding author: M Shadaydeh, maha.shadaydeh@uni-jena.de [1]Department of Mathematics and Computer Science, Friedrich Schiller University Jena, Jena, Germany [2]Michael Stifel Center for Data-driven and Simulation Science Jena, Jena, Germany [3]Max Planck Institute for Biogeochemistry, Jena, Germany

II. METHODOLOGY

Various causality measures have been reported in literature. Among many other linear regression based models, Granger causality (GC) (Weiner 1956, Granger 1969)[2] is the most widely known method for causality analysis. GC assumes that causes both precede and help predict their effects.

Let $x_i, i = 1, \cdots, N$ denotes the time series of N Earth Observation variables. Each time series $x_i(n), n = 1, \cdots, m$ is a realization of length m real valued discrete stationary stochastic process $X_i, i = 1, \cdots, N$. These N time series can be represented by a pth order multivariate autoregressive model (MVAR(p)) of the form

$$\begin{bmatrix} x_1(n) \\ \vdots \\ x_N(n) \end{bmatrix} = \sum_{r=1}^{p} A_r \begin{bmatrix} x_1(n-r) \\ \vdots \\ x_N(n-r) \end{bmatrix} + \begin{bmatrix} \epsilon_1(n) \\ \vdots \\ \epsilon_N(n) \end{bmatrix}. \quad (1)$$

The residuals $\epsilon_i, i = 1, \cdots, N$ constitute a white noise stationary process with covariance matrix Σ. The model parameters at time lags $r = 1, \cdots, p$ is defined by

$$A_r = \begin{bmatrix} a_{11}(r) & \cdots & a_{1N}(r) \\ \vdots & \ddots & \vdots \\ a_{N1}(r) & \cdots & a_{NN}(r) \end{bmatrix}. \quad (2)$$

The model order can be estimated using Akaike or Bayesian Criterion. The model parameters $a_{ij}(r), i, j = 1, \cdots, N; r = 1, \cdots, p$, can then be estimated using the method of Least Square. It is worth noting that the use of the MVAR model (1) makes no assumption on the mechanism that produced the data (for example whether it is a linear or non-linear) except that the model itself exist and stable [6].

A. *Time Domain MVAR Granger Causality*

The conditional MVAR GC (3) of x_i on x_j quantifies the degree to which the past of x_i helps predict x_j, over and above the degree to which x_j is already predicted by its own past and the past of the variables other than x_i. Let Σ_j denote the covariance matrix of the residual ϵ_j associated to x_j using the model in (1), and let Σ_j^{i-} denotes the covariance matrix of the residual associated to x_j using the model (1) after eliminating x_i, i.e. eliminating the ith raw and column in (2). The time domain MVAR-GC of x_i on x_j conditioned on all other variables is defined by the likelihood ratio [3], [7],

$$\gamma_{i \to j} = \ln \frac{|\Sigma_j^{i-}|}{|\Sigma_j|}. \quad (3)$$

B. *Frequency Domain MVAR Granger Causality: Generalized Partial Directed Coherence*

The causal relation from x_i to x_j is described in the frequency domain via gPDC [4] by

$$g\pi_{i \to j}(f) = \frac{\frac{1}{\sigma_{jj}} \overline{A}_{ji}(f)}{\sqrt{\sum_{k=1}^{m} \frac{1}{\sigma_{kk}^2} \left|\overline{A}_{ki}(f)\right|^2}}, \quad (4)$$

where $\overline{A}_{ij}(f), i, j = 1 \cdots N$ are the elements of the matrix $\overline{A}(f) = I - A(f)$ where $A(f)$ is the Fourier transform of $A(r), r = 1, \ldots, p$:

$$A(f) = \sum_{r=1}^{p} \mathbf{A}_r z^{-r}|_{z=e^{i2\pi f}}, \quad (5)$$

and σ_{ii}^2 are the diagonal entries of the residual covariance matrix Σ. The value of $g\pi_{i \to j}(f)$ represents the causality strength of x_i on x_j at the normalized frequency f as compared to all of x_i's interactions to other variables. Nullity of $g\pi_{i \to j}(f)$ indicates absence of the Granger causality of x_i on x_j at the normalized frequency f.

III. Experimental results and discussion

Experiments are performed on the real half-hourly meteorological observations and land flux eddy covariance data measured at Hainich National Park spanning the seven years 2000-2006 using both time domain conditional MVAR GC and frequency domain gPDC. First the data of the seven years over all seasons were segmented into 90 days short segments with 50% overlap. The model order is estimated using Bayesian criterion and fixed for all segments. The model parameters are estimated for each segment using Least Square. Time domain causal intensities as defined in (3) are estimated using the MVAR GC toolbox [7] with statistical significance F-test. In case of the gPDC based frequency domain analysis defined in (4), we used the permutation test for statistical significance with confidence level 95%. The averages time and frequency domain causality strength between the four used variables of the real data are shown in Figures 1 and 2 respectively.

The time domain causality structure in Figure 1 show several spurious links, e.g. causal links of GPP as well as Reco on Rg and T. The frequency domain causality structure of GPP $\to$ Rg and GPP $\to$ T in Figure 2 show spectral peaks at frequency corresponding to the daily cycle (f= 0.0201 cycle/30min) which indicates that the spurious links in time domain causality are mainly due to the daily cycle induced by global radiation. Similarly for Reco, peaks occur on the time scale of the seasonal cycle, i.e. around $f = 0$ cycle/30min, which is the cause of the spurious links in the time domain causality analysis. Another advantage of frequency analysis is that it shows the time scale at which interaction between variables occurs. In Figure 2, the peak in the frequency plot of Rg $\to$ T indicates that although the causal effect of Rg on T occurs on time scales of half an hour up

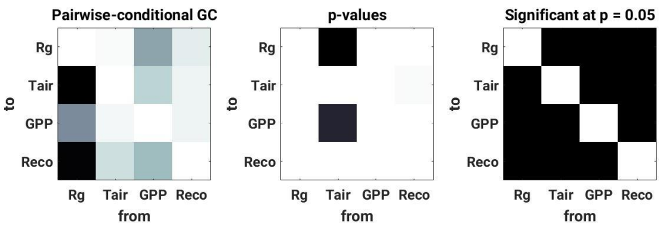

Fig. 1. Plots of the time domain MVAR Granger causality within the variables of the real data of Hainich National Park. The causal strength is visualized using gray levels with black for highest value.

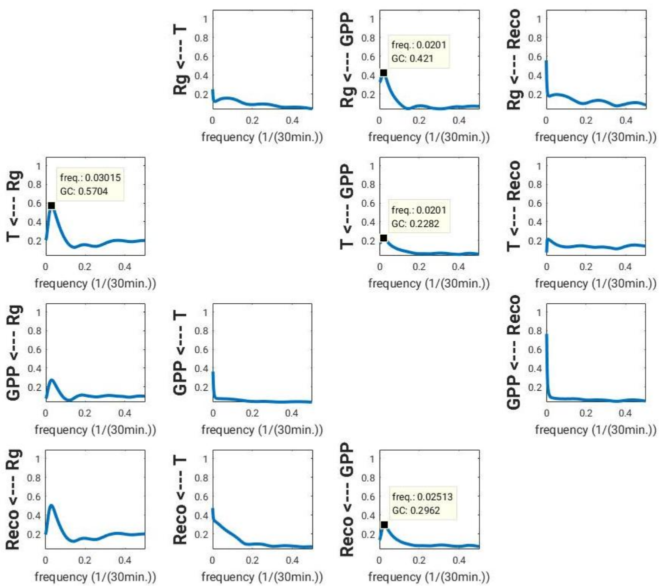

Fig. 2. Plots of the gPDC representing the spectral causal intensities between meteorological and land flux CO2 data of Hainich National Park (average of 40 time segments over all seasons of years 2000-2006).

till days, there is a clear peak around the time scale of 16 hours (f= 0.03015 cycle/30min) at the location of Hainich National Park. We can also notice a peak in the causal intensity of GPP on Reco on the time scale of 20 hours (f= 0.02513 cycle/30 min). The causal link of T on Reco exists over all the spectrum but with increased intensity on the time scale of two hours and more.

Similar experiments were repeated but on time segments of winter and summer seasons separately. Summer and winter gPDC spectral causality plots are shown in Figures 3 and 4 respectively. These figures show the time variant causal intensities between the four variables in different seasons. The causal intensity of T on Reco is higher in summer while the causal intensity of Rg on T and Reco is higher in winter.

To conclude, preliminary results show that the use of spectral GC based on gPDC is a promising method

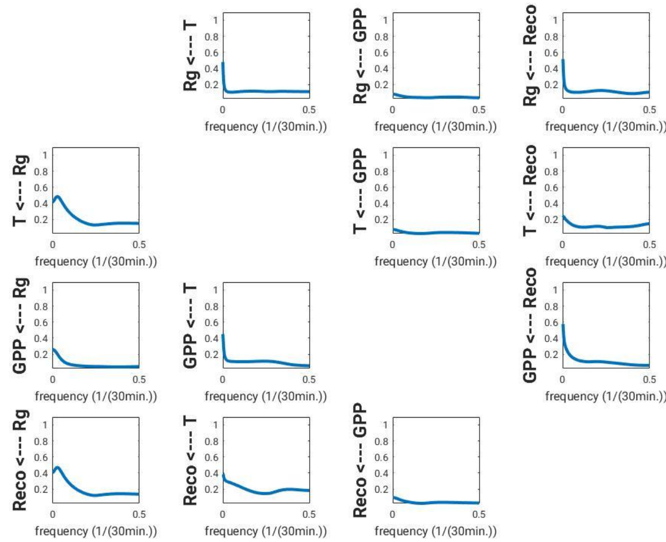

Fig. 3. Plots of the gPDC representing the spectral causal intensities between meteorological and land flux CO2 data of Hainich National Park during winter (average of 20 winter time segments from years 2000-2006)

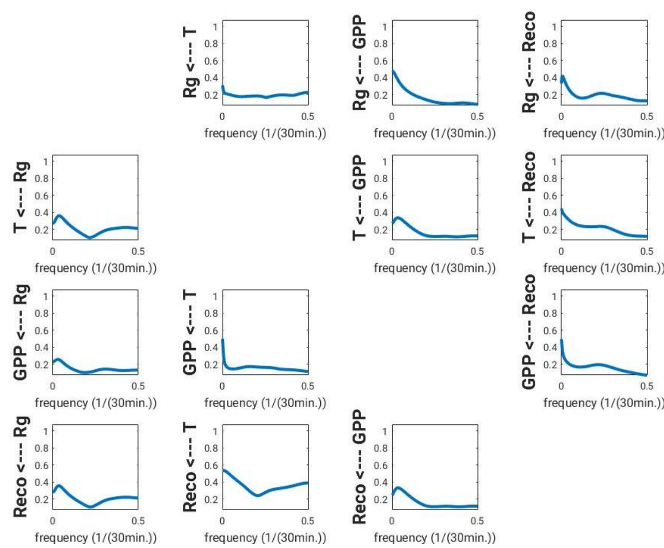

Fig. 4. Plots of the gPDC representing the causal strength within the variables of the real data of Hainich National Park during summer (average of 20 summer time segments from years 2000-2006).

for handling the presence of the periodic components necessary for accurate causality analysis at different time scales. Further ongoing research will be focused on the selection criteria of the model order as well as the selection of the sampling frequency.

ACKNOWLEDGMENTS

This study has been conducted within the framework of the project BACI: Towards a Biosphere Atmosphere

Change Index, funded by the European Union's Horizon 2020 research and innovation program under the grant agreement No 640176.

REFERENCES

[1] L. Barnett and A. K. Seth, "Behaviour of granger causality under filtering: Theoretical invariance and practical application," *Journal of Neuroscience Methods*, vol. 201, no. 2, pp. 404 – 419, 2011.

[2] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[3] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *Journal of the American statistical association*, vol. 77, no. 378, pp. 304–313, 1982.

[4] L. A. Baccalá, K. Sameshima, and D. Takahashi, "Generalized partial directed coherence," in *Digital Signal Processing, 2007 15th International Conference on*, pp. 163–166, IEEE, 2007.

[5] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, pp. 463–474, May 2001.

[6] T. Anderson, *The Statistical Analysis of Time Series*. Wiley Classics Library, Wiley, 1994.

[7] L. Barnett and A. K. Seth, "The mvgc multivariate granger causality toolbox: A new approach to granger-causal inference," *Journal of Neuroscience Methods*, vol. 223, pp. 50 – 68, 2014.