Causal networks of biosphere–atmosphere interactions

Christopher Krich[1,2], Jakob Runge[3], Diego G. Miralles[2], Mirco Migliavacca[1], Oscar Perez-Priego[1], Tarek El-Madany[1], Arnaud Carrara[4], and Miguel D. Mahecha[1,5]

[1]Max Planck Institute for Biogeochemistry, 07745 Jena, Germany
[2]Laboratory of Hydrology and Water Management, Ghent University, Ghent 9000, Belgium
[3]German Aerospace Center, Institute of Data Science, 07745, Jena, Germany
[4]Fundación Centro de Estudios Ambientales del Mediterráneo (CEAM), 46980 Paterna, Spain
[5]German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103 Leipzig, Germany

Correspondence: Christopher Krich (ckrich@bgc-jena.mpg.de)

Abstract. Local meteorological conditions and biospheric activity are tightly coupled. Understanding these links is an essential prerequisite for predicting the Earth system under climate change conditions. However, many empirical studies on the interaction between the biosphere and the atmosphere are based on correlative approaches that are not able to deduce causal paths, and only very few studies apply causal discovery methods. Here, we use a recently proposed causal graph discovery algorithm, which aims to reconstruct the causal dependency structure underlying a set of time series. We explore the potential of this method to infer temporal dependencies in biosphere-atmosphere interactions. Specifically we address the following questions: How do periodicity and heteroscedasticity influence causal detection rates, i.e. the detection of existing and non-existing links? How consistent are results for noise-contaminated data? Do results exhibit an increased information content that justifies the use of this causal-inference method? We explore the first question using artificial time series with well known dependencies that mimic real-world biosphere-atmosphere interactions. The two remaining questions are addressed jointly in two case studies utilizing observational data. Firstly, we analyse three replicated eddy covariance datasets from a Mediterranean ecosystem at half hourly time resolution allowing us to understand the impact of measurement uncertainties. Secondly, we analyse global NDVI time series (GIMMS 3g) along with gridded climate data to study large-scale climatic drivers of vegetation greenness. Overall, the results confirm the capacity of the causal discovery method to extract time-lagged linear dependencies under realistic settings. The violation of the method's assumptions increases the likelihood to detect false links. Nevertheless, we consistently identify interaction patterns in observational data. Our findings suggest that estimating a directed biosphere-atmosphere network at the ecosystem level can offer novel possibilities to unravel complex multi-directional interactions. Other than classical correlative approaches, our findings are constrained to a few meaningful set of relations which can be powerful insights for the evaluation of terrestrial ecosystem models.

1 Introduction

Understanding biosphere-atmosphere interactions is a prerequisite to accurately quantify feedbacks in the Earth system (Bonan, 2015). On the one hand, the biosphere responds to the atmospheric drivers such as radiation intensity, temperature, vapour pressure deficit, and composition of trace gases. On the other hand, the biosphere influences the atmosphere via partitioning

5   the incoming net radiation into sensible, latent, and ground heat fluxes as well as via controlling the exchange of trace gases and volatile organic compounds. Over the past decades, many of these processes have been identified and their physical, chemical and biological effects have been investigated (see e.g. Monson and Baldocchi, 2014; McPherson, 2007, for overviews). However, the synergistic effects of these processes and the specific cause-effect interactions underlying these effects are still not fully understood (Baldocchi et al., 2016; Miralles et al., 2018).

10   Today, there is a manifold of monitoring systems operating at various spatial and temporal scales. Networks of eddy covariance towers offer high temporal resolution to study the dependence between carbon, water, and energy fluxes and the atmosphere under a variety of normal and stress conditions or in response to disturbances (Baldocchi, 2014). Satellite remote sensing data complement this picture. They typically only monitor vegetation states at multi-day resolutions and some products offer nearly complete global coverages (Justice et al., 2002; Woodcock et al., 2008). The actual and future satellite missions

15   are leading to rapid development in the field with ever higher spatial, temporal, and spectral measurements (Malenovský et al., 2012; Guanter et al., 2015; Qi and Dubayah, 2016).

The study of biosphere-atmosphere interactions using observations typically relies on correlative approaches, or is based on model-data i.e. requires *a-priori* knowledge. Only few attempts have been made to learn directional dependencies in a data-driven manner in biogeosciences. For instance, Ruddell and Kumar (2009) used transfer entropy, a bivariate information

20   theoretic measure, to estimate networks of information flow. These networks constructed for a corn-soybean ecosystem under drought and normal conditions showed substantial differences in connectivity. The decoupling during drought between variables describing both land and atmospheric conditions was attributed to changes in the feedback patterns for the two conditions. In recent years, a new branch in statistics aiming for causal inference from empirical data has experienced substantial progress. The idea of causal inference emerged already in the early 20th century (Wright, 1921). Later, Granger suggested one of the

25   first applicable formalisms (Granger, 1969); since then, several efforts in ecology and climate science have concentrated on the bivariate form of Granger causality (Elsner, 2006, 2007; Kodra et al., 2011; Attanasio, 2012; Attanasio et al., 2012).

Aiming to mitigate some of the limitations of the traditional Granger causality, Detto et al. (2012) used a conditional spectral Granger causality framework that allows to disentangle system inherent periodic couplings from external forcing. The disentanglement is enabled via decomposition into the frequency domai using wavelet theory. This method enabled the finding

30   that soil respiration in a pine and hardwood forested ecosystem in winter is not influenced by canopy assimilation but only by temperature, a result that would not be detectable via lagged correlation or bivariate Granger causality. Green et al. (2017) used a similar approach to investigate biosphere-atmosphere feedback loops. It is found that they can explain up to 30% of variance in radiation and precipitation in certain regions. Recently, Papagiannopoulou et al. (2017a) applied a non-linear multivariate conditional Granger causality framework to study climatic drivers of vegetation at the global scale. This approach revealed

that water limitations on plant productivity play a considerably larger role than previously expected (Papagiannopoulou et al., 2017b). These examples show that unexpected interactions patterns can be found from data only and may challenge our theoretical assumptions. In fact, in the last years the science of causal inference has developed a strong theoretical foundation and several algorithms have been proposed (Spirtes et al., 2001; Pearl, 2009; Peters et al., 2017; Runge et al., 2019). However, we

5   only find few studies that test the latest generation of such methods in ecosystem ecology (see e.g. Shadaydeh et al., 2018; Christiansen and Peters, 2018).

Ecological and climate data is usually time ordered. Using the concept of time series graphs (Ebert-Uphoff and Deng, 2012), the time order can be exploited to create efficient causal graph discovery algorithms (Runge, 2018a). Recently, Runge et al. (2018) introduced such an algorithm, called PCMCI, a combination of the PC algorithm (named after its inventors Peter and

10 Clark, Spirtes and Glymour, 1991) and the Momentary Conditional Independence (MCI) test. PCMCI has been applied to artificial tests (Runge et al., 2018) and climatological case studies (Runge et al., 2014; Kretschmer et al., 2016). Hence, this method could be potentially of very high relevance for learning the casual dependency structure in the complex biosphere-atmosphere system.

In this study, we explore the potential of PCMCI for disentangling and quantifying interactions and feedbacks between

15 terrestrial biosphere state and fluxes and meteorological variables. The study is structured as follows: Firstly, we describe the performance of the method on artificial time series data with well-known dependencies that mimic some basic properties of observed land surface fluxes such as heteroscedasticity (Sect.3.1). Secondly, we explore three replicated eddy covariance measurement towers in a Mediterranean ecosystem and explore how the identified interdependencies of carbon and energy fluxes and micro-meteorological observations vary over time (Sect.3.2). Thirdly, we use global satellite data of vegetation greenness

20 to understand the lagged dependency of ecosystems with respect to climatic drivers (Sect.3.3). Based on these results, we discuss the potentials and limitations of PCMCI for other applications in land-atmosphere studies and give recommendations for further methodological developments (Sect.4).

2 Method and Data

2.1 Causal Discovery via PCMCI

PCMCI assesses the causal structure of a multivariate dataset or process $\mathbf{X}$ by estimating its time series graph. In short, the nodes of a time series graph $\mathcal{G}$ represent single variables $X^1, X^2, ... \in \mathbf{X}$ at specific (lagged) time steps $t, t-1, .., t-\tau_{max}$. A

5 lag is the number of time steps τ between two nodes $X^i_{t_1}$ and $X^j_{t_2}$, i.e. $\tau = t_1 - t_2$ for $t_1 > t_2$. To generate a causal graph from observational data through statistical dependencies, a series of assumptions must be adopted. Here, we assume: time order, the causal Markov condition, faithfulness, causal sufficiency, and no contemporaneous causal effects (Runge et al., 2018). PCMCI is applied in combination with the ParCorr linear independence test based on partial correlations (cf. Sect. 2.1.1). This application additionally requires stationarity in mean and variance and linear dependencies. In the following, we briefly discuss

10 these assumptions (further details in Runge, 2018a) and refer to a variable X^i that is causally affecting a variable X^j as 'parent' or 'driver' and the latter as 'receiver' or 'target'.

The time order within the time series allows to orient directed links which are only pointing forward in time. This accounts for causal information propagating forward in time only, i.e. the cause shall precede the effect. Therefore, a directed causal link $X^i_{t_1} \rightarrow X^j_{t_2}$ can only exist between two nodes $X^i_{t_1}, X^j_{t_2}$ if $t_1 < t_2$. When a contemporaneous link is found, i.e. $t_1 = t_2$, it is

15 considered to be undirected. The connection between the graph $\mathcal{G}$ and its process $\mathbf{X}$ is given by the causal Markov condition together with the faithfulness. The Markov condition says that if there is no link in the graph, there is conditional independence in the distribution, and if there is no conditional independence, there is connectedness in the graph; the reverse relations, if there is conditional independence, there is also separation in the graph and if there is connectedness in the graph, there is no conditional independence, are given by the faithfulness assumption. With these assumptions the graph is a visualisation of the

20 conditional dependence and independence relationships among the variables including their lags. Finally, the causal sufficiency assumption implies that, every common cause of two or more variables $X^i \in \mathbf{X}$ is included in $\mathbf{X}$. If this is not the case, detected links may be indirect or due to an unobserved common driver. However, the absence of a link in the detected graph still implies that no direct link is present (as this only requires the assumption of faithfulness).

Several causality methods suffer from high dimensionality, i.e. multivariate transfer entropy (Runge et al., 2012). They

25 use the whole past $\mathbf{X}^-_t \backslash X^i_{t-\tau}$ of X^j_t with $\mathbf{X}^-_t = (\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, ...)$ as a conditioning set $\mathcal{S}$ in the estimation of the test statistic $I(X^i_{t-\tau}, X^j_t | \mathcal{S}) : \mathcal{S} = \mathbf{X}^-_t \backslash X^i_{t-\tau}$. PCMCI addresses this issue by reducing the set of conditions $\mathcal{S}$. PCMCI first estimates $\mathcal{S}$ for each X^j_t. The estimation removes all variables that are either independent or dependent due to indirect paths or common drivers only and leaves only relevant conditions converging to the true causal parents in the limit of infinite sample size. This is done using a variant of the PC algorithm (Spirtes et al., 2001). The conditioning set is subsequently used in the MCI step to test

30 every possible link; MCI also attributes a link strength. These two steps as well as the concept of independence tests, which are at the core of both PC and MCI algorithms, are explained in more detail in the following.

2.1.1 Independence Test

At the core of PCMCI, there are conditional independence tests $CI(X_{t-\tau}^i, X_t^j, \mathcal{S})$ testing for $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \mathcal{S}$. Within the PCMCI software package Tigramite (Runge, 2018b), several independence tests are implemented. Here, we focus on the linear independence test called ParCorr. The ParCorr conditional independence test is based on partial correlations and a t-test. This

5 assumes the model

$$X = \mathcal{S}\boldsymbol{\beta_x} + \epsilon_X, \quad Y = \mathcal{S}\boldsymbol{\beta_y} + \epsilon_Y, \tag{1}$$

with coefficients $\boldsymbol{\beta}$ and Gaussian noise ϵ. This leads to the residuals

$$r^x = X - \mathcal{S}\hat{\boldsymbol{\beta_x}}, \quad r^y = Y - \mathcal{S}\hat{\boldsymbol{\beta_y}} \tag{2}$$

with estimated $\hat{\boldsymbol{\beta}}$. ParCorr removes the influence of $\mathcal{S}$ on X and Y via ordinary least squares regression and tests for independence of the residuals using the Pearson correlation with a t-test. The independence test returns a p-value and test statistic value

10 I, i.e. the correlation coefficient in case of ParCorr.

2.1.2 PC algorithm

The PC algorithm estimates a set of parents for each variable of the process $X_t^j \in \mathbf{X}_t$ which are used as low-dimensional conditions in the MCI algorithm. A comprehensive pseudo code of this procedure is given in Runge et al. (2018). The algorithm

15 starts with a fully connected graph and iteratively removes links if conditional independence is found. At first, a preliminary set of parents $\widetilde{\mathcal{P}}$ for each X_t^j equal to the whole past of the process $\mathbf{X}_t$ is initialised: $\widetilde{\mathcal{P}}(X_t^j) = \mathbf{X}_t^- = \{X_{t-\tau}^i : i = 1, ..., N, \tau = 1, ..., \tau_{max}\}$. Next, conditional independence $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \mathcal{S}$ is tested for all $X_{t-\tau}^i \in \widetilde{\mathcal{P}}$. $\mathcal{S}$ is a subset of $\widetilde{\mathcal{P}}(X_t^j) \backslash \{X_{t-\tau}^i\}$. The conditional independence test is embedded into two loops. The outer loop iteratively increases the cardinality p of $\mathcal{S}$ starting from 0, the empty set. Thus it is first tested, whether X^i and X^j are independent. Further, only the p strongest conditions are selected. The inner

20 loop iterates through all $X_{t-\tau}^i \in \widetilde{\mathcal{P}}(X_t^j)$. The cardinality is increased until $|\mathcal{S}| = |\widetilde{\mathcal{P}}(X_t^j) \backslash \{X_{t-\tau}^i\}| - 1$. At the end of each iteration in the outer loop the non-significant parents are removed from $\widetilde{\mathcal{P}}(X_t^j)$. Every conditional independence test is evaluated at a significance threshold called α_{pc}, which is usually set to a liberal value between 0.1 and 0.4. Alternatively, one can let α_{pc} unspecified. PCMCI then evaluates the best choice of $\alpha_{pc} \in \{0.1, 0.2, 0.3, 0.4\}$ based on the Akaike information criterion which is further explained in (Runge et al., 2018).

2.1.3 MCI tests

MCI is the actual causal discovery step that ascribes a p-value and strength to each possible link. MCI iterates through all pairs $(X_{t-\tau}^i, X_t^j) : i = 1, ..., N, \ \tau = 0, ..., \tau_{max}$ and calculates $CI(X_{t-\tau}^i, X_t^j, \mathcal{S})$. $\mathcal{S}$ consists of the two potentially multivariate sets of parents $\widetilde{\mathcal{P}}(X_t^j)$ and $\widetilde{\mathcal{P}}(X_{t-\tau}^i)$ obtained in the PC step. $\widetilde{\mathcal{P}}(X_{t-\tau}^i)$ is constructed by shifting the time series of $\widetilde{\mathcal{P}}(X_t^i)$ by τ. In case $X_{t-\tau}^i \in \widetilde{\mathcal{P}}(X_t^j)$, $X_{t-\tau}^i$ has to be removed from $\widetilde{\mathcal{P}}(X_t^j)$. If $\tau = 0$, conditional dependence is estimated for contemporaneous nodes

30 X_t^j and X_t^i. Due to missing time order, a dependence would be left undirected. Further, as the parents $\widetilde{\mathcal{P}}(X_t^i)$ and $\widetilde{\mathcal{P}}(X_t^j)$ used in

each conditional dependence test are defined to lie in the past of X_t^i and X_t^j, links, both contemporaneous and lagged, can be spurious due to contemporaneous common drivers or contemporaneous indirect paths. The absence of a link, though, means that a physical (contemporaneous) link is unlikely (assuming faithfulness, cf. Runge et al. (2018)).

The link strength is given by the effect size of the conditional independence test statistic measure CI used in combination

5 with MCI. In case of ParCorr, the effect size is given by the partial correlation value, which is between -1 and 1. This value is shown in Runge et al. (2014) to directly depend on the receiver's and driver's variance as well as the coupling coefficient. Even non-linear links can be detected with ParCorr as they can often be linearly approximated. In case the linear part is even stronger than the non-linear part, ParCorr can have better detection rate than a non-linear independence test (Runge et al., 2018).

2.2 Data

2.2.1 Artificial Dataset - Test Model

We tested the algorithm on artificial datasets prior to its application to real world data. Using time series of measured global radiation ($\mathcal{R}g$) we created three artificial time series that conceptually represent temperature ($\mathcal{T}$), gross primary production

5 ($\mathcal{GPP}$) and ecosystem respiration ($\mathcal{R}eco$). Note that this test model is not intended to accurately represent observed land-atmosphere fluxes, but only serves to test the procedure. The model incorporates one linear auto dependence $\mathcal{T}_{t-\tau_1} \rightarrow \mathcal{T}_t$, one linear additive cross-dependence $\mathcal{R}g_{t-\tau_2} \rightarrow \mathcal{T}_t$ and two non-linear dependencies, multiplicative $\mathcal{R}g_{t-\tau_3} \bullet \mathcal{T}_{t-\tau_4} \rightarrow \mathcal{GPP}_t$, and multiplicative exponential $\mathcal{GPP}_{t-\tau_5} \bullet c^{\mathcal{T}_{t-\tau_6}} \rightarrow \mathcal{R}eco_t$ (cf. Fig. 1) according to the equations:

$$\mathcal{R}g_{mo} = \mathcal{R}g_{obs} \tag{3}$$

10 $$\mathcal{T}_{mo}(t) = c_1\, \mathcal{T}_{mo}(t-\tau_1)\, + c_2\, \mathcal{R}g_{mo}(t-\tau_2)\, +\, \xi_T \tag{4}$$

$$\mathcal{GPP}_{mo}(t) = c_3\, \mathcal{R}g_{mo}(t-\tau_3)\, *\, \mathcal{T}_{mo}(t-\tau_4)\, +\, \xi_{GPP} \tag{5}$$

$$\mathcal{R}eco_{mo}(t) = c_4\, \mathcal{GPP}_{mo}(t-\tau_5)\, *\, c_5^{\frac{\mathcal{T}_{mo}(t-\tau_6)-T_{ref}}{10}}\, +\, \xi_{\mathcal{R}eco} \tag{6}$$

The parameters $c_1, c_2, ..., c_5$ are referred to as coupling coefficients, and the time lags are noted as $\tau_1, \tau_2,..., \tau_5$. The subscripts $_{mo}$ and $_{obs}$ abbreviate model and observation, respectively. T_{ref} is set to 15°C. The term ξ, termed "intrinsic" or "dynamical

15 noise", here represents values from uncorrelated, normally distributed noise. Having dynamic noise is essential for a method utilizing conditional independence tests. It is based on the assumption that a process or state is never fully described by its deterministic part because there are unresolved intrinsic processes, summarized as ξ.

The model was fitted to real observational data (using radiation, temperature and land-atmosphere fluxes) of daily time resolution, measured by eddy-covariance method (Baldocchi et al., 1988; Baldocchi, 2003) from FLUXNET, by minimizing

20 the sum of squared residuals using the gradient descent implemented in the `Optim.jl` package (Mogensen and Riseth, 2018). We fitted the model to 72 sites listed in Table B1 given in the Supplementary Material section. The value range for the coupling coefficients c_1 to c_4 and c_5 were set to [0.2,1] and [1,2.5], respectively. The lags were limited to integer values in the range [0,25]. For the fitting, the dynamical noise ξ was set to 0. The distributions of obtained lags and coupling coefficients are given in the Supplementary Material Fig. A1 to D1. The fitting process thus generated 72 sets of parameters, containing coupling

25 coefficients and lag values, which were used for the time series generation.

From each of the 72 sets of parameters we generated four sets of time series each having a length of 500 years. The time series generation was initiated using two types of data: first, uncorrelated, normally distributed noise, and second, unprocessed radiation data as used during the fitting (the available radiation data was repeated to 500 years). The resulting datasets are called baseline dataset and seasonality dataset, respectively. In both cases, the model was run twice, once with homoscedastic

30 (constant variance of ξ), once with heteroscedastic dynamical noise ξ. To induce heteroscedasticity, ξ was multiplied with a mean daily variance that was extracted for each variable at each FLUXNET site. In Supplementary Material Fig. E1 a five year

time series excerpt from Hainich site (Knohl et al., 2003b) is shown. A third dataset is generated by anomalization (subtraction of smoothed seasonal mean) of the seasonality dataset.

2.2.2 Eddy Covariance Data - Majadas de Tiètar Experimental site

Data from three towers located in Majadas de Tiètar, (ES-LMa, ES-LM1, ES-LM2), a Mediterranean Savanna in central
5 Spain, are used (coordinates of central tower: 39°56'25"N 5°46'29"W). Meaurements include the exchange of CO_2 between the land surface and atmosphere at half-hourly resolution using the eddy covariance method. The three tower footprints received different fertilisation treatments in spring 2015 (El-Madany et al., 2018). We consider data from before the fertilization from April 2014 to March 2015 of shortwave downward radiation (Rg), air temperature (T), net ecosystem exchange (NEE), vapour pressure deficit (VPD), sensible heat (H) and latent heat (LE). The average temperature within this period was 17.3°C, with a
10 total precipitation of 765mm. Most precipitation fell between October and April.

We expect the causal imprints in the data to vary between seasons and during the course of the day. To satisfy causal stationarity, we estimate networks separately for each month and consider only samples for which the potential radiation was above $\frac{4}{5}$ of the potential daily maximum, which corresponds to midday samples. We used a mask type that limits only the receiver variable to the respective month and day time values (cf. Table A1 for PCMCI parameter settings). This setting causes
15 time series lengths ranging from 239 datapoints in December to 372 datapoints in July. Minimal and maximal lags were set to 0 and 8, respectively. Due to the limited number of years, we left the data unprocessed, i.e. we did not subtract a seasonal mean for anomalisation. Constraining the samples to separate month and midday values reduces the effect of seasonality as a common driver that would lead to spurious links. Furthermore, to correct for multiple testing we applied the Benjamini-Hochberg false discovery rate correction (Benjamini and Hochberg, 1995). Thereby, the p-values for the whole graph obtained from the MCI
20 step are adjusted to control the number of false discoveries (Runge et al., 2018). We chose a two-sided significance level of 0.01.

2.2.3 Gridded global data set

The second observational case study was performed on a global data set. We used data with 0.5° spatial and monthly temporal resolution from 1982 to 2008. The dataset is composed of three climatic variables, global radiation (Rg), temperature (T) and
25 precipitation (P), and one vegetation state index, the Normalized Difference Vegetation Index (NDVI). Both temperature and precipitation datasets were taken from the Climate Research Unit (CRU), version TS3.10 (Harris et al., 2014). The radiation data stems from the Climate Research Unit and National Centers for Environmental Prediction dataset (CRUNCEP, Viovy, 2016). The used NDVI data stems from the Global Inventory Monitoring and Modeling Systems (GIMMS) in version 3g_v1 (Pinzon and Tucker, 2014).
30 To examine the influence of radiation, temperature, and precipitation on NDVI by means of PCMCI we used the following settings. We compute the anomalies by subtracting a smoothed seasonal mean. A maximal time lag of three months was chosen based on the largest lag with significant partial correlation among all pairs of variables, partialling out only the autocorrelation of each variable. The receiver variable was limited to the growing season defined by T>0 and NDVI>0.2, which allows good

comparison to Wu et al. (2015). The significance level (α_{pc}) in the condition selection phase (cf. Sect. 2.1.2) was chosen based on the AIC selection criterion. A concise list of PCMCI parameters that were altered from default settings is given in Table A1.
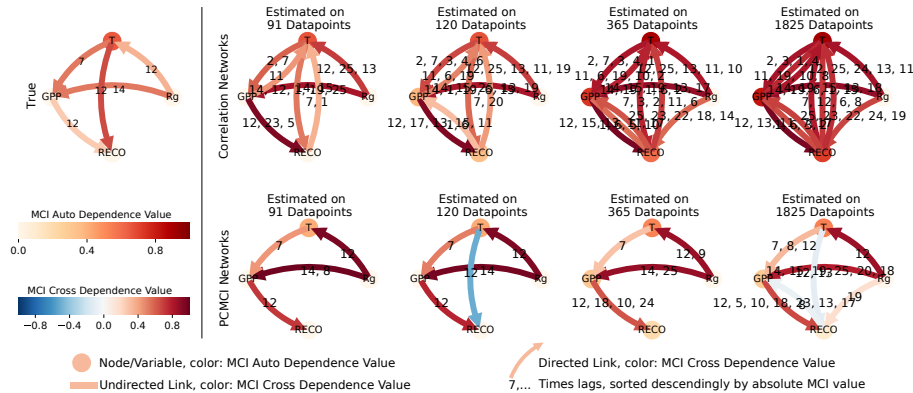
Figure 1. The artificial datasets are generated with a prescribed interaction structure (True Network), which is obtained by fitting the test model to the FLUXNET sites. Here we show for four time series length the process graphs estimated via both lagged correlation and PCMCI. The data used stems from the homoscedastic realisation of the seasonality dataset of the Hainich site. The significance level was set to 0.01. The number of time lag labels were limited to five in the correlation networks. But for the longest time series typically the whole range of lags (0-25) was significant.

3 Results

3.1 Test Model

As motivating example, in Fig. 1 we show PCMCI and lagged correlation networks in the form of process graphs. It is clearly visible, that many more spurious links pass the significance threshold of 0.01 using lagged correlation as compared to using
5   PCMCI. Those spurious links can complicate the analysis or lead to false assumptions and misleading hypotheses. We examined four cases of different time series lengths: 91 [183<=doy<=274] and 120 [153<=doy<=274] days, 1 and 5 years for daily data (doy: day of the year). For each time series length and each parameter set, the causal network structure was estimated for 100 realisations of the model (each based on a realization of intrinsic noise), which allowed the estimation of false positive (FPR) and true positive (TPR) detection rates. The detection rates are calculated for each tower, FPR in general and
10 TPR link-wise. The TPR for each link is its sum of detections among 100 realisations divided by 100. The FPR is the number of falsely detected links divided by the number of all possible false links and 100. The summary of the experiments i.e. the overall false positive rate (FPR) and the distributions of the link's true positive rate (TPR) across sites are given in Fig. 2 and 3, respectively. The blue violin plots always report the case of normal distributed (non heteroscedastic) intrinsic noise and the corresponding orange violin plots summarize the case of heteroscedastic noise. The effect of heteroscedasticity and seasonality
15 is then assessed by comparing the distributions obtained from the baseline dataset to the results of the seasonality dataset.

The FPR of homoscedastic time series in the baseline dataset is in the expected range of 0.01, the chosen significance level, indicating a well calibrated test due to fulfilled assumptions. The assumption of stationarity is violated as soon as heteroscedas-
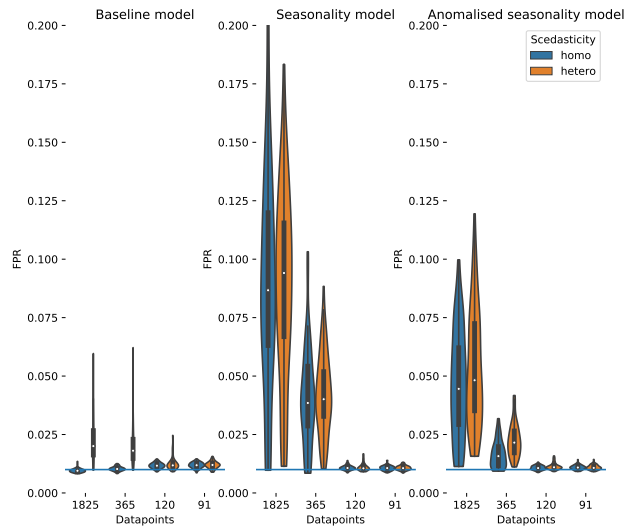
Figure 2. The distribution of false positive detection rates estimated for the baseline dataset, the seasonality dataset and the anomalised seasonality dataset (mean seasonal cycle subtracted). The distributions are given for different time series length (number of datapoints). Additionally, the distributions are split to show the impact of heteroscedastic noise (orange) compared to normal distributed noise (blue). The significance level of 0.01 is given by a blue horizontal line.

ticity or seasonality is present. The effect on the FPR is an increase above 0.01 for time series length of 1 and 5 years with a much stronger increase due to seasonality (factor of 4 and 8, respectively) than for heteroscedasticity (factor of 2).

The effect of non-stationarities on the TPR differs among the links. The detection of linear links ($\mathcal{R}g \rightarrow \mathcal{T}$ and $\mathcal{T} \rightarrow \mathcal{T}$) is not affected by seasonality and slightly improves for heteroscedastic dynamical noise. The detection of non-linear links is
5   improved by seasonality with the strongest effects in the link $\mathcal{T} \rightarrow \mathcal{GPP}$. The link $\mathcal{T} \rightarrow \mathcal{Reco}$ has a stonger non-linearity and therefore the detection rate shows a weaker effect on seasonality. Furthermore, the coupling coefficient c_5, the base of T, can be close to or be exactly one (cf. Fig. B1). This would actually cause on the one hand the effect of $\mathcal{T} \rightarrow \mathcal{Reco}$ to vanish, rendering a detection impossible and on the other hand result in a linear dependence of $\mathcal{GPP}$ on $\mathcal{Reco}$ which improves its detection. Heteroscedasticity seems to have a slight negative effect on non-linear links. In general, the TPRs in the seasonality dataset are
10 quite high, even for non-linear links, and predominantly above 80% and often reaching 100%.

Comparing the TPRs of the non-linear links, shows some disparity. The links $\mathcal{T} \rightarrow \mathcal{GPP}$ and $\mathcal{T} \rightarrow \mathcal{Reco}$ experience zero detection in the baseline dataset but partially considerable rates in the seasonality dataset with a strong dependence on the time series length. On the contrary, the median of the TPR of the links $\mathcal{R}g \rightarrow \mathcal{GPP}$ and $\mathcal{GPP} \rightarrow \mathcal{Reco}$ is above 95% in the seasonality dataset, even for time series length as short as 91 days, but remains high in the baseline dataset. The removal of
15 the seasonal cycle keeps the TPRs largely unaffected, but reduces the FPR. Nevertheless, it still remains above the significance level by a factor of four and two for five and one year time series length, respectively.
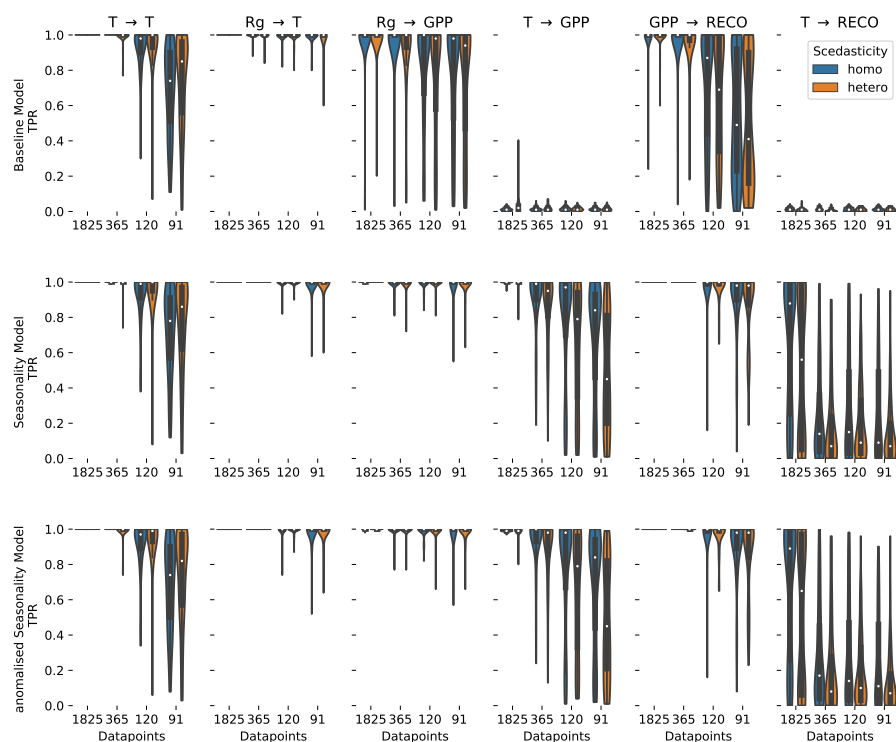
Figure 3. The distribution of the true positive detection rates for each link in the test model estimated for the baseline dataset, the seasonality dataset, and the anomalised seasonality dataset. The distributions are given for different time series length (number of datapoints). Additionally the distributions are split to show the impact of heteroscedastic noise (orange) compared to normally distributed noise (blue).

In summary, the seasonality dataset exhibits high TPR even for non-linear links. Compared to stationary time series, the detection of non-linear links actually benefits from seasonality. The high detection, though, comes at the cost of a high false positive rate for time series length of and above one year. To a certain degree, the increase in FPR can be counteracted by anomalization.

3.2 Majadas de Tiètar dataset

At first we look at the link consistency by comparing networks that were obtained for each tower within a month. The comparison is done for two months with strongly differing climate conditions: April and August. In Fig. 4 we compare the estimated link strengths (effect size estimated via partial correlation) as long as the corresponding links are significant in at least one network. The confidence intervals are overlapping for the majority of links, suggesting that the uncertainty of the fluxes is much smaller than the observed effects (El-Madany et al., 2018). Exceptions are found for only a few links ($Rg\overset{0}{\rightarrow}T$, $Rg\overset{1}{\rightarrow}LE$, $VPD\overset{1}{\rightarrow}VPD$, $H\overset{2}{\rightarrow}H$, $NEE\overset{0}{\rightarrow}LE$, $H\overset{2}{\rightarrow}NEE$, $NEE\overset{4}{\rightarrow}NEE$, number above the arrow indicates the lag) where the detection rates do not or barely overlap. Cross links (a link from one variable to another) with two or more significant appearances are predomi-
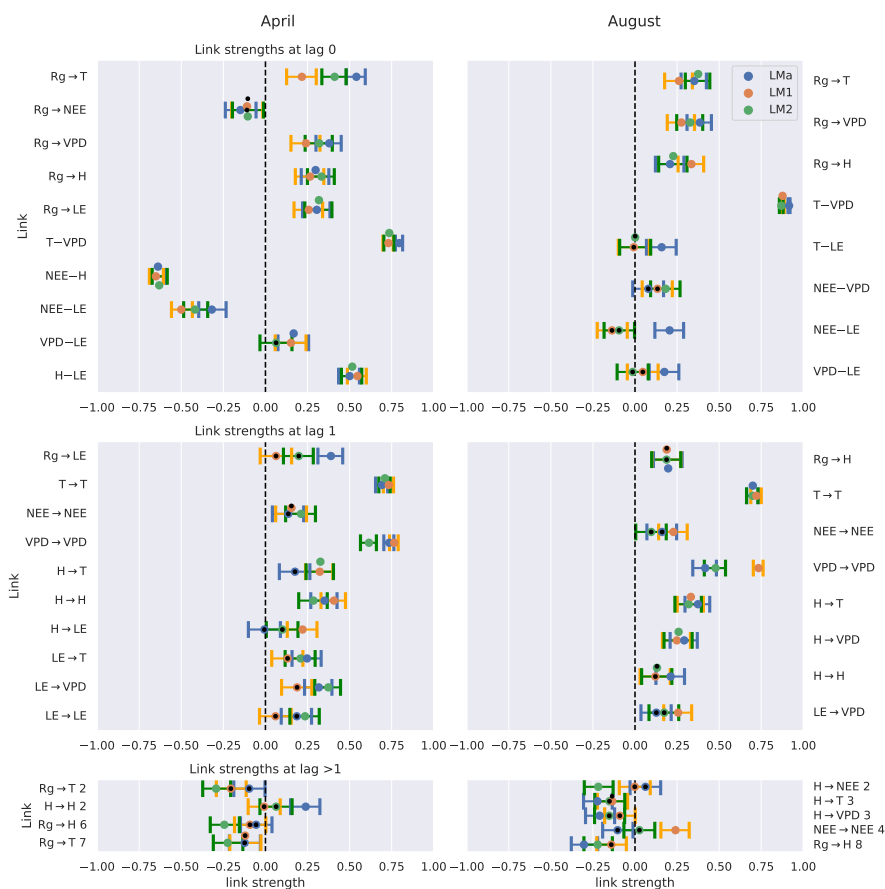
Figure 4. Comparison of the networks of three eddy covariance measurement stations (LMa, LM1, LM2) located in Majadas (Spain). Links that are found to be significant in one of the three networks are included. For each link, the calculated strength of all three networks is plotted with its 90% confidence interval. The colors blue, orange, and green correspond to the towers LMa, LM1, and LM2, respectively. The significance threshold is 0.01. If a link does not pass the significance, it is marked by a black dot. The links are grouped into lag 0 (top), lag 1 (middle) and all lags greater than 1 (bottom). Negative NEE is associated with carbon uptake by the ecosystem. Links at lag 0 are left undirected (−), yet as Rg is set as main driver, links incorporating Rg at lag 0 are directed (→).

nantly at zero lag. Approximately half of the links with lag one are auto-dependent links (a link from the past of a variable to its present). Comparing the links between the month April and August, distinct differences can be noticed. First, August has slightly fewer significant links compared to April. Second, the only links remaining that are significant in two or three towers are between atmospheric variables. Third, the remaining link strengths tend to be weaker in August than in April.

5   The difference among the seasons is further investigated in Fig. 5 which shows process graphs for each month of the year. We combine the networks of the three towers to one process graph by plotting every link that is significant in at least one tower. The process graphs in Fig. 5 visualize clearly gradual changes within the interaction structure of the biosphere-atmosphere system during the course of a year. During the main growing season from February to May, NEE is coupled strongly to the energy fluxes latent (LE) and sensible heat (H). These connections weaken, disappear or even switch sign with start and course

10   of the dry season. Less regularly NEE also shows connections to radiation (Rg) and temperature (T). Between the atmospheric variables, a basic network between VPD, T, Rg and H remains intact and relatively constant in strength. The dominance of contemporaneous links is found as well, as seen already in Fig. 4. Besides the decoupling of NEE from any variable in the dry period, there are additional interesting patterns. For example, the positive reappearance of the link between NEE and LE in September. Here, the onset of precipitation events (cf. Fig. G1) occurred that lead to strong respiration peaks (Ma et al.,

15   2012). Creating such a network via lagged correlation would result in much more significant links (causing the network to be not interpretable as opposed to PCMCI) and NEE does not decouple from the atmosphere in August (cf. Fig. F1).

The above results demonstrate that PCMCI is sensitive enough to capture seasonal differences and certain physiological reasonable biosphere behaviour. Moreover, PCMCI yields a better interpretable network structure than pure correlation approaches.

20   **3.3   Global**

Subject of inspection were the significant lags and MCI values of each climatic variable on NDVI. In Fig. 6 the maximal MCI value and the corresponding lag are plotted for the links $X_{t-\tau} \rightarrow \text{NDVI} : \tau \in \{0, 1, 2, 3\}$ with X being one of the climatic drivers radiation, temperature, or precipitation. The chosen significance threshold was set to 0.05. Fig. 7 shows the climatic driver with largest MCI per grid point. PCMCI detects a regionally varying influence of climatic drivers. As expected, the boreal

25   regions are strongly driven by temperature instantaneously, while (semi-) arid regions, which correspond predominantly to grass or prairie dominated areas, respond strongest to precipitation at a time lag of one month. Radiation is found to have a comparatively low spatial effect with hot spots in south and east China, central Russia and east Canada.

The dominant lags are found to be zero and one. Just a very small fraction of the total area shows a maximal MCI value at a higher lag of two or three months. The lags are also not equally distributed among the climatic drivers. Radiation and

30   temperature are predominantly strongest at lag zero, while precipitation has a much larger fraction of area showing the strongest response at lag one. Regions where the impact of Rg on NDVI is strongest at lag 1 tend to respond negatively to Rg but positively to precipitation at lag one. On the other hand, a large part of regions with the strongest impact of precipitation at lag zero respond negatively to it but positively to radiation.
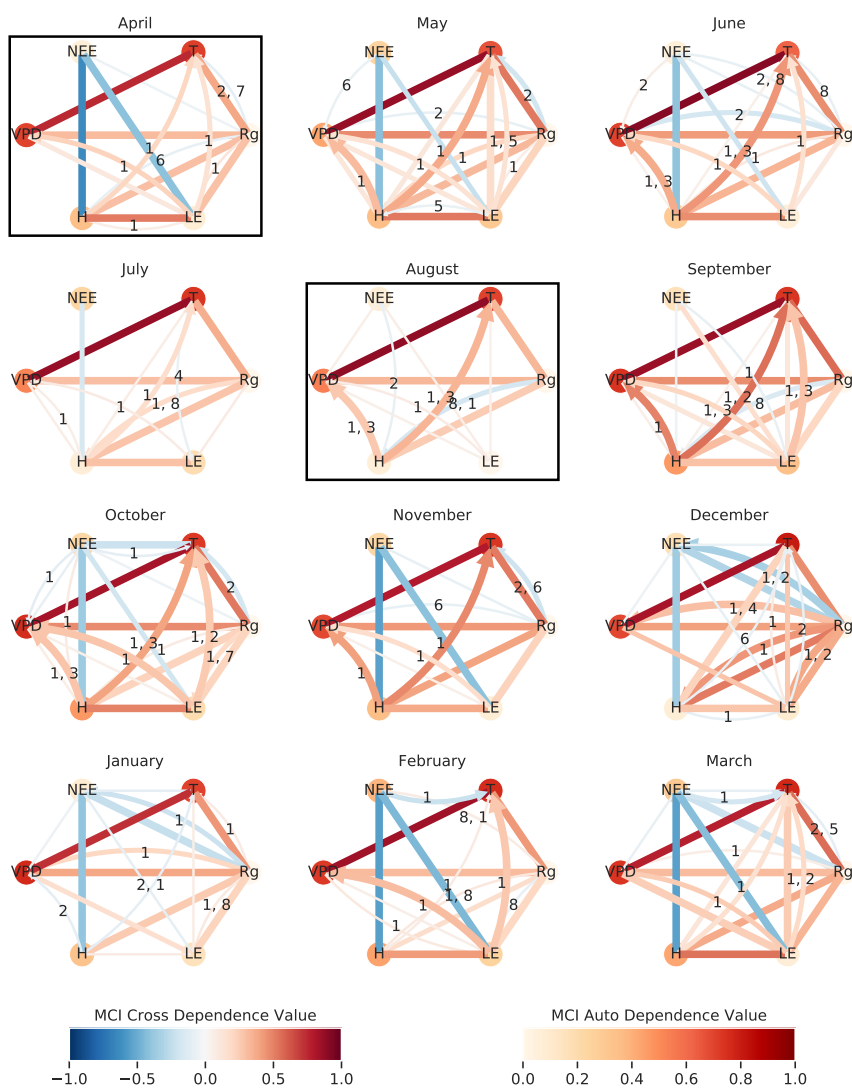
Figure 5. To visualize the gradual changes in interaction structure the networks of the three towers are combined for each month. The number of significant occurences of a link is given by its width. The link strength, given by the link color, is calculated by averaging the significant links of the towers. The link's lag is shown in the centre of each arrow, sorted in descending order of link strength. The resulting graphs are shown for April 2014 till March 2015. The significance threshold is 0.01. The networks of April and August, illustrated in Fig. 4, are highlighted by a box.
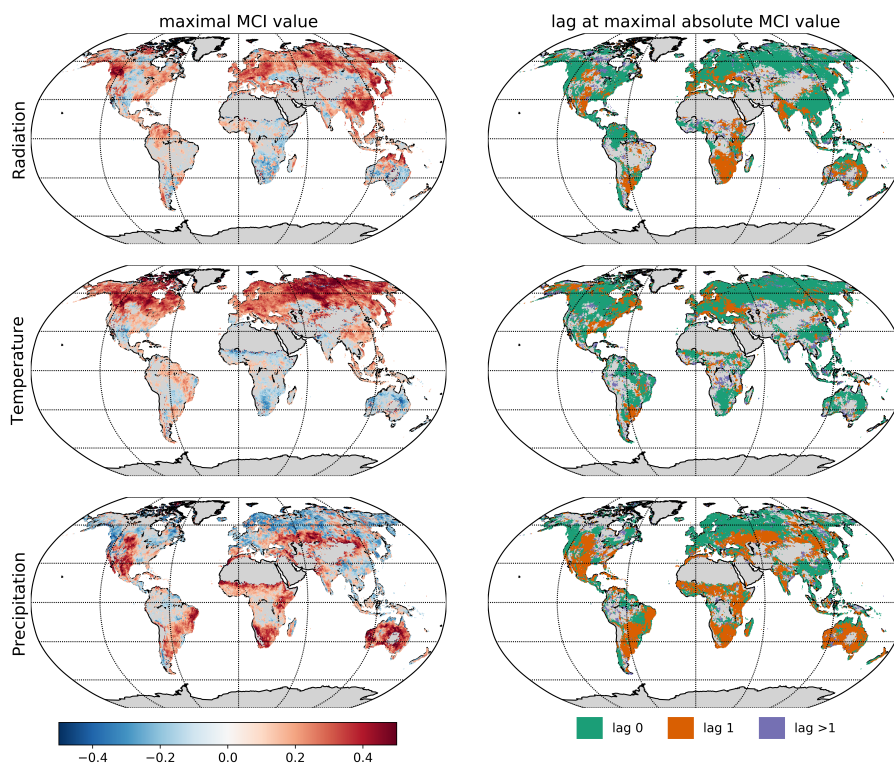
Figure 6. Influence of climatic drivers on NDVI as calculated by PCMCI. The first column shows the estimated causal influence given as maximal absolute MCI value of climatic drivers on NDVI. The second column gives the time lag at which the maximal absolute MCI value occures (in month).
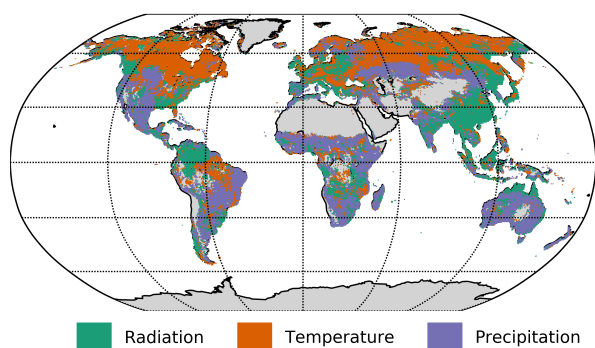
Figure 7. Map of the strongest climatic driver (largest absolute MCI value) per grid point.

In summary, PCMCI estimates coherent interaction patterns which match well with anticipated behaviour based on vegetation type and prevailing climatic conditions.

4 Discussion

Causal discovery methods promise an improved understanding and can help to come up with new hypotheses about the interaction between biosphere and atmosphere (Christiansen and Peters, 2018; Runge et al., 2019). But the underlying assumptions need to be properly taken into account. The coupled biosphere-atmosphere system possesses several challenges that potentially

5 violate the underlying assumptions of causal discovery in general and the employed method's assumptions in particular. Here, we investigate the effect of a violation of assumptions on PCMCI network estimates.

With regard to expected non-linearities in biosphere-atmosphere interactions, using a linear independence test within the PCMCI framework may not be adequate. We motivate our choice with the following arguments: first, non-linearities are often approximated linearly. Second, a linear regression based test has a much higher power for detecting linear links than a non

10 parametric test (Runge, 2018a) and can, hence, detect links already at smaller sample sizes. Third, linear partial correlation is easily interpretable, for example, positive and negative MCI values. The results of the test model support our choices. In general, the derived results show a high detection power with a strong consistency in calculated effect strengths on eddy covariance data and global, regularly gridded reanalysis data that leads to well interpretable patterns. Observed drawbacks are a high FPR in case of violated assumptions, especially strong periodicity, as well as the appearance of contemporaneous lags

15 in measurement datasets.

4.1 Lessons learned from the test model

The probability to detect a link with PCMCI depends strongly on a link's MCI effect size, which is larger for strong variance in the driver and a low variance in the receiver (Runge et al., 2018). Several results can be explained by this observation. First, the variance of three out of five drivers of cross dependencies in the test model are either directly or indirectly (via $\mathcal{GPP}$)

20 influenced by $\mathcal{R}g$, which has the highest variance of all variables. Consequently, the detection power of the three links is large, almost 100%. In comparison, the other variables' variances are weaker, since they are influenced by $\mathcal{T}$, which results in a lower detection power. This is the origin of the disparity in detection rates of the non-linear links. Second, also the partially strong increase in TPR of non-linear links (influenced in a multiplicative way by $\mathcal{T}$) from the baseline dataset to the seasonality dataset can be explained by this increase in variance. A multiplicative link is actually not generally expected to be found by

25 ParCorr (Runge, 2018a), but the value of the multiplicative factor is dominated by the seasonal value, and not the dynamical noise, which might cause rather a scaling of the dynamical noise terms rather than a random distortion. Third, the dependence on the variance ratio can also explain the difference in TPR between homoscedastic (equal error variance) and heteroscedastic (error variance changing over time) time series, i.e., the variance of $\mathcal{R}g$ and $\mathcal{GPP}$ exhibits a strong seasonality with its peak in summer, while the variance of $\mathcal{T}$ is rather constant. This explains, for example, the strong decrease in TPR for the link

30 $\mathcal{T} \rightarrow \mathcal{GPP}$ at 91 days time series length when comparing homoscedasticity to heteroscedasticity. The decrease in TPR is less pronounced when another season, implying a different variance, is chosen for this comparison. As links with weak driver variance and strong response variance are more likely to be missed, one may ask which effect this will have on the detection

of feedback loops where one variable has low and the other high variance. Here lies a limitation of the test model where no feedback loops were implemented.

Seasonality and heteroscedasticity constitute violations of the stationarity assumption underlying the independence test ParCorr. As shown in Runge (2018a), including the cause of the non-stationarity as an exogenous driver in the analysis allows

5 PCMCI to regress out its influence on the other variables. However, for ParCorr this is only valid if the dependence on the non-stationary driver is linear. Therefore, the regression on $\mathcal{R}g$ fails for $\mathcal{GPP}$ and $\mathcal{R}eco$ in the test model. With this ill-posed setting, the probability to detect false links increases with increasing time series length or when more periods are included. Stationarity in mean is obviously also not fully guaranteed when subtracting the seasonal mean. Here we observe that the FPR stays above the significance level for the anomalised seasonality dataset. One can ask whether the FPR stays above the

10 significance threshold because subtracting the seasonal mean does not remove the heteroscedasticity. However, we attribute this high FPR to a not fully removed seasonality since the FPR of both homoscedastic and heteroscedastic time series decreases by roughly the same amount in the anomalised seasonality dataset and the effect of heteroscedasticity is rather weak in the baseline dataset. Much of the influence of heteroscedasticity is also removed when limiting the analysis to a specific time period, i.e. season, which makes the data causally stationarity (cf. Sect. 2.1). For example, the link from radiation to GPP vanishes in

15 winter as there is mostly no active plant material left. One can argue, as it is done in Peters et al. (2017), that the causality of a system is invariant even between seasons, the physical mechanism is the same in all seasons. For example, a functional dependence $f(x)$ might be 'flat' for small values of x and linearly increasing for larger values. If only small values occur in the winter season, then the link is absent, while it 'appears' only in the summer season. Across all seasons, this can be considered as a nonlinear functional dependence $f(x)$. In practice, restricting an analysis to different seasons can help in interpreting the

20 mechanism, here in a linear framework.

Summarizing the results of the test model, the different detection rates, disparity among non-linear links, and the detection of multiplicative links are largely explainable via the effect of the variance on the link detection. Yet, the discussion revealed the need for further research in several aspects. On the one hand, feedback loops are not included in the test model yet are an important aspect in natural systems. On the other hand, removing non-stationarities is essential to keep the false positive rate

25 in the expected range, but standard procedures of subtracting the mean seasonal cycle are not sufficient. Further, the effect of non-stationarity on the causal network structure needs to be investigated.

4.2 Causal interpretation of estimated networks from observational data

In both the half-hourly time resolved eddy covariance data and the monthly global dataset the predominant type of dependence found is contemporaneous. PCMCI leaves these undirected since no time order indicating the flow of causal information

30 is available. Further, as discussed in Sect. 2.1, contemporaneous common drivers or mediators are not accounted for. The consequence is that both spurious contemporaneous and spurious lagged links can appear, if they are due to contemporaneous variables. For interactions that are contemporaneous in nature since they occur on considerably shorter time scales than the time resolution, therefore, PCMCI is not the optimal choice regarding a causal interpretation and other methods should be applied in conjunction (Runge et al., 2019). Further, we faced a trade off between fulfilling causal assumptions and detection power.

In practice, accounting for causal stationarity (by limiting the analysis to certain periods of the dataset) means decreasing the number of available data points while accounting for causal sufficiency leads to an increase in dimensionality by adding variables and increasing the maximal lag. Both will lead to a decrease in detection power, which can affect the network structure. PCMCI alleviates the curse of dimensionality by applying a condition selection step, but still one cannot indefinetly

5   add more variables. Another important factor that affects detection power and dimensionality is the time resolution. There are several points in favour and against increasing time resolution. On the one hand, increasing time resolution can resolve contemporaneous links and potentially increases the detection power due to an increased number of datapoints. On the other hand, the dimensionality increases if the maximal lag is adapted. Further, causal information might be split apart and distributed over more lags, rendering the links at each individual lag less detectable. This can cause links to disappear, but links can also

10 appear if new processes are resolved at a higher time scale. At last, observational noise (measurement errors) might be larger in higher resolution data than in aggregated data, as it is averaged out in the latter and thus affects link detection less. Consequently, when comparing network structures based on different settings, i.e. maximal lag, included variables, time resolution, and considered time period, the (dis-)appearance of single links among specific variables can stem from several factors, i.e. a change in detection power, a changed (conditional) dependency, or due to a common driver. These factors together with a

15 non-zero false positive detection rate are challenging for a causal interpretation. Therefore, detected links should be interpreted with care and can give rise to new hypotheses and analyses involving further variables. Generally, a causal interpretation is more robust regarding the absence of links (cf. Sect.2.1). In particular it does not require that all common drivers are observed.

Nevertheless, robust patterns were identified in our studies that are also consistent with other studies. Furthermore, a causal analysis has the advantage of an enhanced interpretability compared to correlative approaches. First of all, we could show

20 that the networks' estimated link strengths are consistent for observational data, even though measurement error affects the data. The dataset used was suitable for this analysis, as the measurement stations are located in a reasonably homogeneous ecosystem that shows only little spatial variation (El-Madany et al., 2018). Thus, also the interaction between biosphere and atmosphere is expected to change only marginally across space. Second, the gradual changes in plant activity that are taking place in the ecosystem of Majadas throughout the year do very well emerge in the coupling strength of daytime NEE to the

25 atmospheric variables. The observed decoupling during the dry season is in accordance with the one of a soybean field during drought conditions observed by Ruddell and Kumar (2009). The gradual changes in ecosystem activity are not visible in a pure (lagged) correlation analysis or are only visible in color or density changes but the large number of significant links prevents any detailed interpretation on the physical mechanisms and changes thereof. The large number of significant links compared to the PCMCI networks stems solely from the absence of conditioning on common drivers or mediating variables, which often

30 further leads to an overestimation of the link strength in correlation networks. As a result, processes, such as the decoupling of NEE during the dry period, stay hidden. To reduce the effect of confounding, often analyses utilize partial correlation (see e.g. Buermann et al., 2018). However, a partial correlation can introduce new dependencies (as opposed to removing them) if one conditions on causal effects of the variables under consideration (the 'marrying parents' effect). This issue is avoided in PCMCI by only conditioning on past variables. Additionally, PCMCI chooses only relevant variables as conditions by applying

the PC condition selection step which is especially valuable in high dimensional study cases and improves detection power and computation time (Runge, 2018a).

The global study of climatic drivers of vegetation shows a general pattern of lags and dependence strengths of vegetation on climatic variables that is easily-interpretable. The boreal regions appear energy limited and especially driven by temperature (cf. Fig 6c), while the strongest dependence of (semi-)arid regions on precipitation reflects their limitation in water supply. Two recent studies performed a similar analysis. Both Wu et al. (2015) and Papagiannopoulou et al. (2017b) investigated lagged effects and dependence strengths of NDVI on precipitation, temperature and radiation. Wu et al. (2015) estimated the lags of the strongest effects via an univariate regression of the climatic drivers on NDVI and subsequently used those lags to fit a multivariate regression model of the climatic drivers on NDVI and determined their relative effects. Papagiannopoulou et al. (2017b) applied a non-linear Granger causality framework utilising a random forest predictive model; the method was presented by Papagiannopoulou et al. (2017a). We recognize that similar patterns are observed in Wu et al. (2015), but the lags at the maximal MCI value are usually lower than the one found in Wu et al. (2015), which stems from the methodical differences. Besides having used anomaly values, PCMCI regresses both NDVI and the climatic drivers on their parents before calculating the MCI value (cf. Sect. 2.1.3). This especially removes the influence of autocorrelation. Runge et al. (2014) shows how autocorrelation affects the correlative lag causing it to be larger for stronger autocorrelation; thereby the correlative lag may become larger than the causal lag. Therefore, according to Fig. 6b, the causal information embedded in monthly resolution is predominantly received within one month. Finding the strongest causal links at a time lag up to one month appear in agreement with Papagiannopoulou et al. (2017b). Also the spatial distribution of the strongest climatic influences compares well. But there are certain noteworthy differences which not necessarily stem from masking differences, i.e. that we took only values belonging to the growing season while Papagiannopoulou et al. (2017b) took the whole time series. First, there is little significant Granger causality of water availability found in boreal regions while there are significant negative causal dependencies detected via PCMCI. Second, NDVI in arid regions is not or barely Granger caused by radiation and temperature, but in parts shows a negative PCMCI value on those variables. There might be physiological reasons that can explain the PCMCI patterns, i.e. water logging or too high temperatures. To explain the differences though, we could identify two possible reasons. First, Papagiannopoulou et al. (2017b) masked out negative influences of radiation arguing that radiation is not negatively affecting NDVI. They found, that negative influences of Rg are usually a consequence of poor conditioning on other variables. Second, a precipitation event in boreal regions coincides with a reduction in radiation and temperature. Boreal regions usually do not suffer from water shortages. Thus they respond stronger to the reduction of radiation and temperature than precipitation. As precipitation is coupled negatively to radiation and temperature at lag zero, the effect of precipitation on NDVI is found to be negative. Thus, the link $P \xrightarrow{-} NDVI$ might be an effect of the contemporaneous common driver scheme $P \xleftarrow{-} Rg \xrightarrow{+} NDVI$ and therefore would not be causal. In fact, a similar argumentation can be given for the negative impact of temperature and radiation on NDVI in arid regions.

In summary, we pointed out the need for careful interpretations in applying causal discovery methods and especially highlighted the challenges linked to the study of biosphere-atmosphere interaction via PCMCI. We demonstrated that the network structures estimated from observational data are explainable with respect to plant physiology and climatic effects. Finally, our

study shows that causal methods can deliver better interpretability and a much improved process understanding in comparison to correlation and bivariate Granger causality analyses that are ambiguous to interpret since they do not account for common drivers.

4.3 Outlook

5   The preceding discussion has shed light on the merits of PCMCI as well as the challenges of applying causal discovery methods. Runge et al. (2019) discuss further challenges and methods and give an outlook how multiple methods can be combined to alleviate limitations.

5 Conclusions

Here we tested PCMCI, an algorithm that estimates causal graphs from empirical time-series. We specifically explored two types of data sets that are highly relevant in biogeosciences: eddy covariance measurements of land-atmosphere fluxes and global satellite remote sensing of vegetation greenness. The causal graphs estimated from the eddy covariance data collected

5 in a Mediterranean site confirm patterns we would expect in these ecosystems: During the dry season's plants senescence, for instance, the ecosystem's carbon cycle (NEE) decouples from meteorological variability. On the contrary during the main growing season with warm and humid conditions strong links between NEE, LE and H characterise the graph. Not only the strongly contrasting states emerge in the graph structure using the causal framework, but also the gradual transitions that relate to minor changes like the connectivity of sensible heat to temperature with progressing dryness. A purely correlative analy-

10 sis, instead, is not able to resolve these patterns. PCMCI allows us to identify and focus on much fewer, but highly relevant dependencies only. Applying the approach to three replicated eddy covariance systems shows the robustness of the method to random errors in the fluxes measurements and confirm one of the assumption of eddy covariance: above a relatively homogeneous terrain the fluxes measured should be spatially invariant, and so the underlying causal relationship between climate and fluxes. The global analysis of NDVI in relation to climatic drivers confirms the known patterns of dependence strengths of veg-

15 etation on climatic variables: boreal regions are energy limited and especially driven by temperature and secondarily radiation, while in semi-arid regions vegetation dynamics are strongly dependent on water supply. However, obtained response times of vegetation to climatic variations are lower using PCMCI than correlation which can be attributed to a better treatment of the autocorrelation in the time-series and cross-relations among climate variables. Compared to merely correlative approaches, this leads to a interpretable pattern of driver-response relationships. In short, the new developments achieved in causal inference

20 allow to gain well constrained insights on processes, that would otherwise be drowning in the correlation chaos. Therefore we hope that this study fosters usage of causal inference in analysing interactions and feedbacks of the biosphere-atmosphere system and furthermore exhibits our demand of further developments.

Table A1. PCMCI parameters that were used differently from default settings.

Dataset	significance	α_{pc}	tau_min	tau_max	selected_variables	mask_type	fdr_method
Test Model	0.01	0.4	0	25	[1,2,3]	'none'	'none'
Majadas Dataset	0.01	None	0	8	[1,2,3,4,5]	'y'	'fdr_bh'
Gridded global data set	0.05	None	0	3	[1,2,3]	'y'	'fdr_bh'

Code and data availability. The eddy covariance data of the FLUXNET sites can be downloaded from the official webpage (https://fluxnet.fluxdata.org/).
CRU temperature and precipitation data is available at http://badc.nerc.ac.uk/data/cru/.
CRUNCEP radiation data can be downloaded via ftp://nacp.ornl.gov/synthesis/2009/frescati/temp/land_use_change/original/readme.htm.
The NDVI dataset is available at http://ecocast.arc.nasa.gov/data/pub/gimms/3g/.

5 The TIGRAMITE software package that includes PCMCI can be found on github https://github.com/jakobrunge/tigramite/. All other code
will be made available upon request.

FLUXNET-ID	start year	end year	Data Reference	FLUXNET-ID	start year	end year	Data Reference
AT-Neu	2002	2012	Wohlfahrt et al.	DK-ZaH	2000	2014	Lund et al. (2012)
AU-Cpr	2010	2014	Meyer et al. (2015)	FI-Hyy	1996	2014	Suni et al. (2003)
AU-DaP	2007	2013	Beringer et al. (2011a)	FI-Sod	2001	2014	Thum et al. (2007)
AU-DaS	2008	2014	Hutley et al. (2011)	FR-Fon	2005	2014	Delpierre et al. (2016)
AU-Dry	2008	2014	Cernusak et al. (2011)	FR-LBr	1996	2008	Berbigier et al. (2001)
AU-How	2001	2014	Beringer et al. (2007)	FR-Pue	2000	2014	Rambal et al. (2004)
AU-Stp	2008	2014	Beringer et al. (2011b)	GF-Guy	2004	2014	Bonal et al. (2008)
AU-Tum	2001	2014	Leuning et al. (2005)	IT-BCi	2004	2014	Vitale et al. (2016)
BE-Lon	2004	2014	Moureaux et al. (2006)	IT-Col	1996	2014	Valentini et al. (1996)
BE-Vie	1996	2014	Aubinet et al. (2001)	IT-Lav	2003	2014	Marcolla et al. (2003)
BR-Sa3	2000	2004	Saleska et al. (2003)	IT-MBo	2003	2013	Marcolla et al. (2011)
CA-Man	1994	2008	Brooks et al. (1997)	IT-Noe	2004	2014	Spano et al.
CA-NS2	2001	2005	Bond-Lamberty et al. (2004)	IT-Ro1	2000	2008	Rey et al. (2002)
CA-NS3	2001	2005	Wang et al. (2002a)	IT-SRo	1999	2012	Chiesi et al. (2005)
CA-NS5	2001	2005	Wang et al. (2002b)	IT-Tor	2008	2014	Galvagno et al. (2013)
CA-NS6	2001	2005	Wang et al. (2002c)	JP-SMF	2002	2006	Matsumoto et al. (2008)
CA-Qfo	2003	2010	Chen et al. (2006)	NL-Hor	2004	2011	Jacobs et al. (2007)
CA-SF2	2001	2005	Rayment and Jarvis (1999a)	RU-Fyo	1998	2014	Kurbatova et al. (2008)
CA-SF3	2001	2006	Rayment and Jarvis (1999b)	US-ARM	2003	2012	Fischer et al. (2007)
CH-Cha	2005	2014	Merbold et al. (2014)	US-Blo	1997	2007	Schade et al.
CH-Dav	1997	2014	Zielis et al. (2014)	US-Ha1	1991	2012	Wofsy et al. (1993)
CH-Fru	2005	2014	Imer et al. (2013)	US-Me2	2002	2014	McDowell et al. (2004)
CH-Lae	2004	2014	Etzold et al. (2011)	US-Me6	2010	2014	Ruehr et al. (2012a)
CH-Oe1	2002	2008	Ammann et al. (2009)	US-MMS	1999	2014	Pryor et al. (1999)
CH-Oe2	2004	2014	Dietiker et al. (2010)	US-Ne1	2001	2013	Gitelson et al. (2003)
CZ-wet	2006	2014	Dušek et al. (2012)	US-Ne2	2001	2013	Cassman et al. (2003a)
DE-Akm	2009	2014	Bernhofer et al. (a)	US-Ne3	2001	2013	Cassman et al. (2003b)
DE-Geb	2001	2014	Anthoni et al. (2004)	US-SRG	2008	2014	Ruehr et al. (2012b)
DE-Gri	2004	2014	Prescher et al. (2010a)	US-SRM	2004	2014	Scott et al. (2008)
DE-Hai	2000	2012	Knohl et al. (2003a)	US-Ton	2001	2014	Tang et al. (2003)
DE-Kli	2004	2014	Prescher et al. (2010b)	US-Twt	2009	2014	Hatala et al. (2012)
DE-Lkb	2009	2013	Lindauer et al. (2014)	US-UMB	2000	2014	Rothstein et al. (2000)
DE-Obe	2008	2014	Bernhofer et al. (b)	US-UMd	2007	2014	Nave et al. (2011)
DE-Spw	2010	2014	Bernhofer et al. (c)	US-Var	2000	2014	Xu et al. (2004)
DE-Tha	1996	2014	Grünwald and Bernhofer (2007)	US-Whs	2007	2014	Scott et al. (2006)
DK-NuF	2008	2014	Westergaard-Nielsen et al. (2013)	US-Wkg	2004	2014	Emmerich (2003)
DK-Sor	1996	2014	Pilegaard et al. (2011)				

Table B1. List of FLUXNET sites used for the generation of artificial datasets and the time periode used.
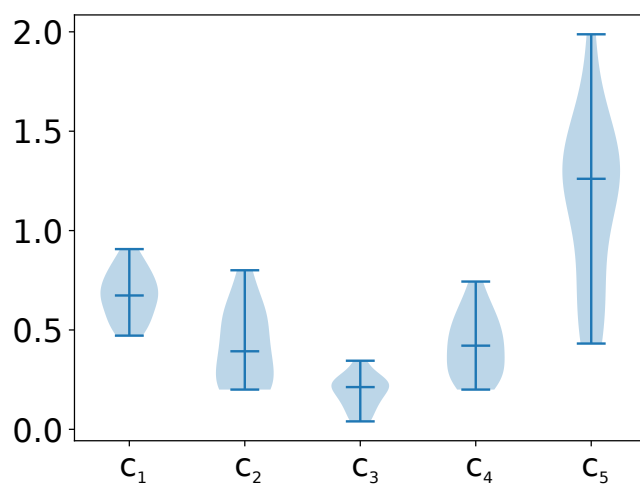
Figure A1. Distribution of coupling coefficients obtained after fitting the test model to the Fluxnet sites. Here shown are the distributions used for generation of heteroscedastic time series.
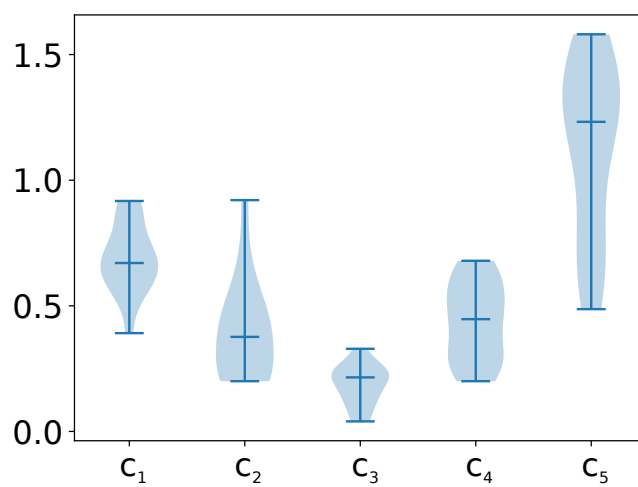
Figure B1. Distribution of coupling coefficients obtained after fitting the test model to the Fluxnet sites. Here shown are the distributions used for generation of homoscedastic time series.
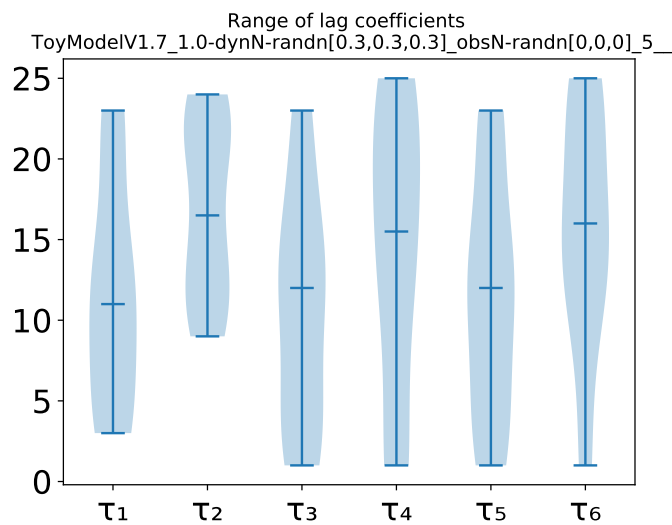
Figure C1. Distribution of time lags obtained after fitting the test model to the Fluxnet sites. Here shown are the distributions used for generation of heteroscedastic time series.
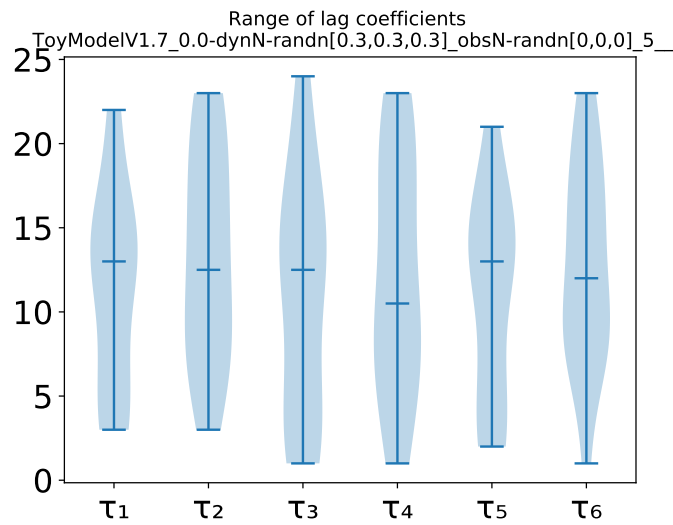
Figure D1. Distribution of time lags obtained after fitting the test model to the Fluxnet sites. Here shown are the distributions used for generation of homoscedastic time series.
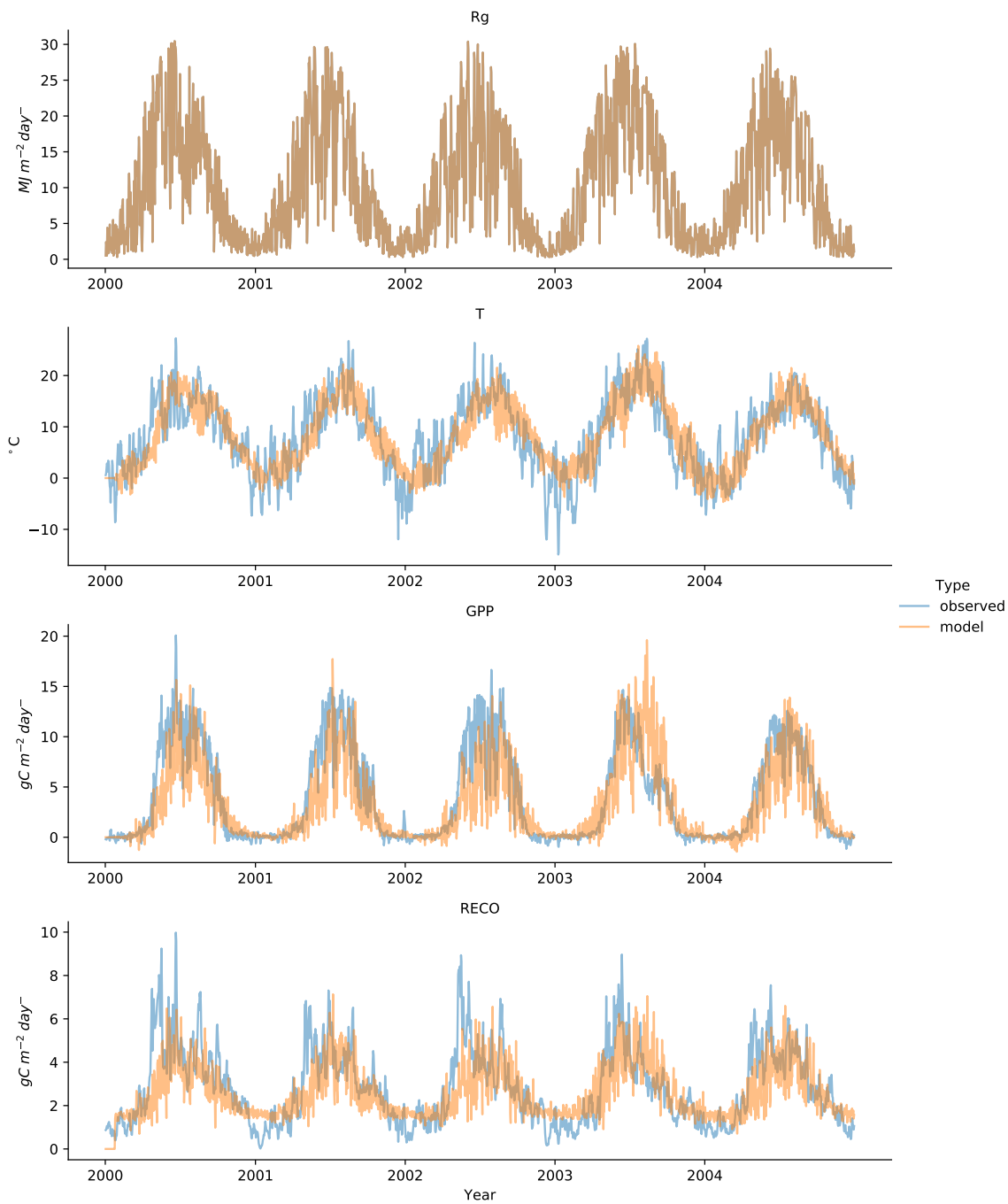
Figure E1. Observed (blue) and model (orange) time-series for Hainich Fluxnet site. The model data was produced with heteroscedastic noise.
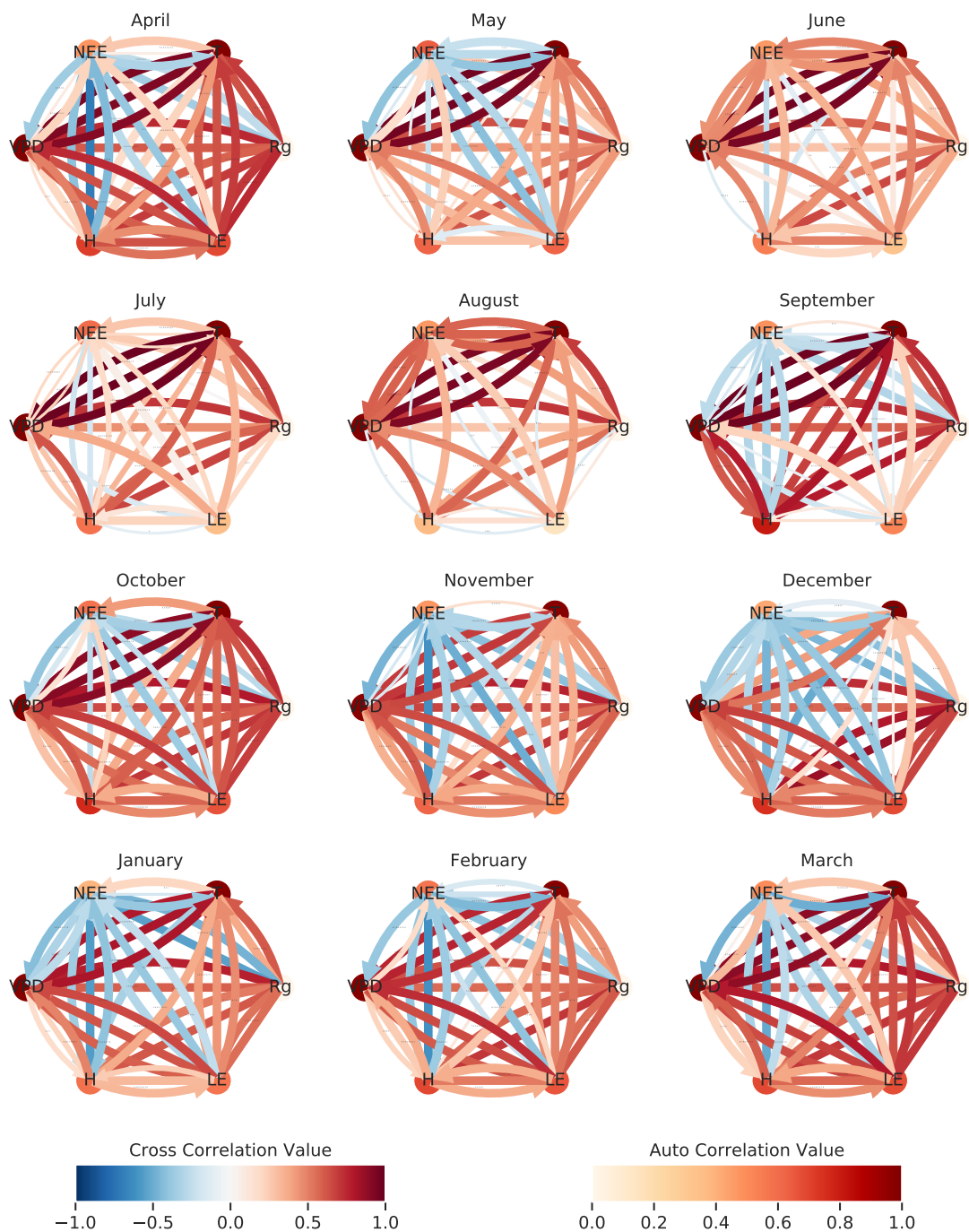
Figure F1. Same as Fig. 5 but using simple correlation analysis to estimate the graph structures. The number of significant occurences of a link is given by its width. The link strength, given by the link color, is calculated by averaging the significant links of the towers. Link labels indicating the lag were removed to improve link visibility. They typically ranged from 1 to 8 (full range of possible lags). The resulting graphs are shown for April 2014 till March 2015. The significance threshold is 0.01.

Figure G1. Precipitation in Majadas Majadas de Tiètar measured at the three tower sites from April 2014 to March 2015. Missing values are set to -1.

Author contributions. CK and MDM designed the study with contributions from JR and DGM. CK conducted the analysis and wrote the manuscript. All authors helped to impove the manuscript. AC, MM, TEM, OPP conducted field experiments in Majadas, processed and provided its data as well helped with interpretation.

Competing interests. The authors declare that they have no competing interests.

5 *Acknowledgements.* We thank Maha Shadaydeh, Rune Christiansen, Jonas Peters, Milan Flach and Markus Reichstein for useful comments. CK thanks the Max Planck Research School for global Biogeochemical Cycles for supporting his PhD project. The authors affiliated with 1. thank the European Space Agency for funding the "Earth System Data Lab" project.

   This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC,

10 OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices.

References

Ammann, C., Spirig, C., Leifeld, J., and Neftel, A.: Assessment of the Nitrogen and Carbon Budget of Two Managed Temperate Grassland Fields, Agriculture, Ecosystems & Environment, 133, 150–162, https://doi.org/10.1016/j.agee.2009.05.006, 2009.

Anthoni, P. M., Knohl, A., Rebmann, C., Freibauer, A., Mund, M., Ziegler, W., Kolle, O., and Schulze, E.-D.: Forest and Agricultural Land-Use-Dependent CO2 Exchange in Thuringia, Germany, Global Change Biology, 10, 2005–2019, https://doi.org/10.1111/j.1365-2486.2004.00863.x, 2004.

Attanasio, A.: Testing for linear Granger causality from natural/anthropogenic forcings to global temperature anomalies, Theoretical and Applied Climatology, 110, 281–289, https://doi.org/10.1007/s00704-012-0634-x, https://doi.org/10.1007/s00704-012-0634-x, 2012.

Attanasio, A., Pasini, A., and Triacca, U.: A contribution to attribution of recent global warming by out-of-sample Granger causality analysis, Atmospheric Science Letters, 13, 67–72, https://doi.org/10.1002/asl.365, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.365, 2012.

Aubinet, M., Chermanne, B., Vandenhaute, M., Longdoz, B., Yernaux, M., and Laitat, E.: Long Term Carbon Dioxide Exchange above a Mixed Forest in the Belgian Ardennes, Agricultural and Forest Meteorology, 108, 293–315, https://doi.org/https://doi.org/10.1016/S0168-1923(01)00244-1, 2001.

Baldocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere – the state and future of the eddy covariance method, Global Change Biology, 20, 3600–3609, https://doi.org/10.1111/gcb.12649, http://dx.doi.org/10.1111/gcb.12649, 2014.

Baldocchi, D., Ryu, Y., and Keenan, T.: Terrestrial Carbon Cycle Variability [version 1; peer review: 2 approved, F1000Research, 5(F1000 Faculty Rev), https://doi.org/(https://doi.org/10.12688/f1000research.8962.1), 2016.

Baldocchi, D. D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, Global Change Biology, 9, 479–492, https://doi.org/10.1046/j.1365-2486.2003.00629.x, https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2486.2003.00629.x, 2003.

Baldocchi, D. D., Hincks, B. B., and Meyers, T. P.: Measuring Biosphere-Atmosphere Exchanges of Biologically Related Gases with Micrometeorological Methods, Ecology, 69, 1331–1340, https://doi.org/10.2307/1941631, https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1941631, 1988.

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society: Series B (Methodological), 57, 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x, https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x, 1995.

Berbigier, P., Bonnefond, J.-M., and Mellmann, P.: CO2 and Water Vapour Fluxes for 2 Years above Euroflux Forest Site, Agricultural and Forest Meteorology, 108, 183–197, https://doi.org/10.1016/S0168-1923(01)00240-4, 2001.

Beringer, J., Hutley, L. B., Tapper, N. J., and Cernusak, L. A.: Savanna Fires and Their Impact on Net Ecosystem Productivity in North Australia, Global Change Biology, 13, {990–1004}, https://doi.org/{10.1111/j.1365-2486.2007.01334.x}, 2007.

Beringer, J., Hutley, L. B., Hacker, J. M., Neininger, B., and U, K. T. P.: Patterns and Processes of Carbon, Water and Energy Cycles across Northern Australian Landscapes: From Point to Region, Agricultural and Forest Meteorology, 151, {1409–1416}, https://doi.org/{10.1016/j.agrformet.2011.05.003}, 2011a.

Beringer, J., Hutley, L. B., Hacker, J. M., Neininger, B., and U, K. T. P.: Patterns and Processes of Carbon, Water and Energy Cycles across Northern Australian Landscapes: From Point to Region, Agricultural and Forest Meteorology, 151, 1409–1416,

https://doi.org/https://doi.org/10.1016/j.agrformet.2011.05.003, savanna Patterns of Energy and Carbon Integrated Across the Landscape (SPECIAL), 2011b.

Bernhofer, C., Grünwald, T., Moderow, U., Hehn, M., Eichelmann, U., Prasse, H., and Postel, U.: FLUXNET2015 DE-Akm Anklam, https://doi.org/10.18140/FLX/1440213, a.

5 Bernhofer, C., Grünwald, T., Moderow, U., Hehn, M., Eichelmann, U., Prasse, H., and Postel, U.: FLUXNET2015 DE-Obe Oberbärenburg, https://doi.org/10.18140/FLX/1440151, b.

Bernhofer, C., Grünwald, T., Moderow, U., Hehn, M., Eichelmann, U., Prasse, H., and Postel, U.: FLUXNET2015 DE-Spw Spreewald, https://doi.org/10.18140/FLX/1440220, c.

Bonal, D., Bosc, A., Ponton, S., Goret, J.-Y., Burban, B., Gross, P., Bonnefond, J.-M., Elbers, J., Longdoz, B., Epron, D., Guehl, J.-M., and

10 Granier, A.: Impact of Severe Dry Season on Net Ecosystem Exchange in the Neotropical Rainforest of French Guiana, Global Change Biology, 14, 1917–1933, https://doi.org/10.1111/j.1365-2486.2008.01610.x, 2008.

Bonan, G.: Ecological Climatology: Concepts and Applications, Cambridge University Press, 3 edn., https://doi.org/10.1017/CBO9781107339200, 2015.

Bond-Lamberty, B., Wang, C., and Gower, S. T.: Net Primary Production and Net Ecosystem Production of a Boreal Black Spruce Wildfire

15 Chronosequence, Global Change Biology, 10, 473–487, https://doi.org/10.1111/j.1529-8817.2003.0742.x, 2004.

Brooks, J. R., Flanagan, L. B., Varney, G. T., and Ehleringer, J. R.: Vertical gradients in photosynthetic gas exchange characteristics and refixation of respired CO2 within boreal forest canopies, Tree Physiology, 17, 1–12, https://doi.org/10.1093/treephys/17.1.1, https://doi.org/10.1093/treephys/17.1.1, 1997.

Buermann, W., Forkel, M., O'Sullivan, M., Sitch, S., Friedlingstein, P., Haverd, V., K. Jain, A., Kato, E., Kautz, M., Lienert, S., Lombardozzi,

20 D., E. M. S. Nabel, J., Tian, H., J. Wiltshire, A., Zhu, D., Smith, W., and D. Richardson, A.: Widespread seasonal compensation effects of spring warming on northern plant productivity, Nature, 562, https://doi.org/10.1038/s41586-018-0555-7, 2018.

Cassman, K. G., Dobermann, A., Walters, D. T., and Yang, H.: Meeting Cereal Demand While Protecting Natural Resources and Improving Environmental Quality, Annual Review of Environment and Resources, 28, 315–358, https://doi.org/10.1146/annurev.energy.28.040202.122858, 2003a.

25 Cassman, K. G., Dobermann, A., Walters, D. T., and Yang, H.: Meeting Cereal Demand While Protecting Natural Resources and Improving Environmental Quality, Annual Review of Environment and Resources, 28, 315–358, https://doi.org/10.1146/annurev.energy.28.040202.122858, 2003b.

Cernusak, L. A., Hutley, L. B., Beringer, J., Holtum, J. A. M., and Turner, B. L.: Photosynthetic Physiology of Eucalypts along a Sub-Continental Rainfall Gradient in Northern Australia, Agricultural and Forest Meteorology, 151, {1462–1470},

30 https://doi.org/{10.1016/j.agrformet.2011.01.006}, 2011.

Chen, J. M., Govind, A., Sonnentag, O., Zhang, Y., Barr, A., and Amiro, B.: Leaf Area Index Measurements at Fluxnet-Canada Forest Sites, Agricultural and Forest Meteorology, 140, 257–268, https://doi.org/10.1016/j.agrformet.2006.08.005, 2006.

Chiesi, M., Maselli, F., Bindi, M., Fibbi, L., Cherubini, P., Arlotta, E., Tirone, G., Matteucci, G., and Seufert, G.: Modelling Carbon Budget of Mediterranean Forests Using Ground and Remote Sensing Measurements, Agricultural and Forest Meteorology, 135, 22–34,

35 https://doi.org/10.1016/j.agrformet.2005.09.011, 2005.

Christiansen, R. and Peters, J.: Switching Regression Models and Causal Inference in the Presence of Latent Variables, arXiv preprint arXiv:1808.05541, 2018.

Delpierre, N., Berveiller, D., Granda, E., and Dufrêne, E.: Wood Phenology, Not Carbon Input, Controls the Interannual Variability of Wood Growth in a Temperate Oak Forest, New Phytologist, 210, 459–470, https://doi.org/10.1111/nph.13771, 2016.

Detto, M., Molini, A., Katul, G., Stoy, P., Palmroth, S., and Baldocchi, D.: Causality and Persistence in Ecological Systems: A Nonparametric Spectral Granger Causality Approach., The American Naturalist, 179, 524–535, https://doi.org/10.1086/664628, https://doi.org/10.1086/664628, pMID: 22437181, 2012.

Dietiker, D., Buchmann, N., and Eugster, W.: Testing the Ability of the DNDC Model to Predict CO2 and Water Vapour Fluxes of a Swiss Cropland Site, Agriculture, Ecosystems & Environment, 139, 396–401, https://doi.org/10.1016/j.agee.2010.09.002, 2010.

Dušek, J., Čížková, H., Stellner, S., Czerný, R., and Květ, J.: Fluctuating Water Table Affects Gross Ecosystem Production and Gross Radiation Use Efficiency in a Sedge-Grass Marsh, Hydrobiologia, 692, 57–66, https://doi.org/10.1007/s10750-012-0998-z, 2012.

Ebert-Uphoff, I. and Deng, Y.: Causal Discovery for Climate Research Using Graphical Models, Journal of Climate, 25, 5648–5665, https://doi.org/10.1175/JCLI-D-11-00387.1, https://doi.org/10.1175/JCLI-D-11-00387.1, 2012.

El-Madany, T. S., Reichstein, M., Perez-Priego, O., Carrara, A., Moreno, G., Martín, M. P., Pacheco-Labrador, J., Wohlfahrt, G., Nieto, H., Weber, U., Kolle, O., Luo, Y.-P., Carvalhais, N., and Migliavacca, M.: Drivers of spatio-temporal variability of carbon dioxide and energy fluxes in a Mediterranean savanna ecosystem, Agricultural and Forest Meteorology, 262, 258 – 278, https://doi.org/https://doi.org/10.1016/j.agrformet.2018.07.010, http://www.sciencedirect.com/science/article/pii/S0168192318302314, 2018.

Elsner, J. B.: Evidence in support of the climate change–Atlantic hurricane hypothesis, Geophysical Research Letters, 33, https://doi.org/10.1029/2006GL026869, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL026869, 2006.

Elsner, J. B.: Granger causality and Atlantic hurricanes, Tellus A: Dynamic Meteorology and Oceanography, 59, 476–485, https://doi.org/10.1111/j.1600-0870.2007.00244.x, https://doi.org/10.1111/j.1600-0870.2007.00244.x, 2007.

Emmerich, W. E.: Carbon Dioxide Fluxes in a Semiarid Environment with High Carbonate Soils, Agricultural and Forest Meteorology, 116, 91–102, https://doi.org/10.1016/S0168-1923(02)00231-9, 2003.

Etzold, S., Ruehr, N. K., Zweifel, R., Dobbertin, M., Zingg, A., Pluess, P., Häsler, R., Eugster, W., and Buchmann, N.: The Carbon Balance of Two Contrasting Mountain Forest Ecosystems in Switzerland: Similar Annual Trends, but Seasonal Differences, Ecosystems, 14, 1289–1309, https://doi.org/10.1007/s10021-011-9481-3, 2011.

Fischer, M. L., Billesbach, D. P., Berry, J. A., Riley, W. J., and Torn, M. S.: Spatiotemporal Variations in Growing Season Exchanges of CO2, H2O, and Sensible Heat in Agricultural Fields of the Southern Great Plains, Earth Interactions, 11, 1–21, https://doi.org/10.1175/EI231.1, 2007.

Galvagno, M., Wohlfahrt, G., Cremonese, E., Rossini, M., Colombo, R., Filippa, G., Julitta, T., Manca, G., Siniscalco, C., di Cella, U. M., and Migliavacca, M.: Phenology and Carbon Dioxide Source/Sink Strength of a Subalpine Grassland in Response to an Exceptionally Short Snow Season, Environmental Research Letters, 8, 025 008, https://doi.org/10.1088/1748-9326/8/2/025008, 2013.

Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., and Leavitt, B.: Remote Estimation of Leaf Area Index and Green Leaf Biomass in Maize Canopies, Geophysical Research Letters, 30, https://doi.org/10.1029/2002GL016450, 2003.

Granger, C. W. J.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods, Econometrica, 37, 424–438, http://www.jstor.org/stable/1912791, 1969.

Green, J., G. Konings, A., Alemohammad, S. H., Berry, J., Entekhabi, D., Kolassa, J., Lee, J.-E., and Gentine, P.: Regionally strong feedbacks between the atmosphere and terrestrial biosphere, NATURE GEOSCIENCE, 10, 410–414, https://doi.org/http://dx.doi.org/10.1038/ngeo2957, https://www.nature.com/ngeo/journal/v10/n6/pdf/ngeo2957.pdf, 2017.

Grünwald, T. and Bernhofer, C.: A Decade of Carbon, Water and Energy Flux Measurements of an Old Spruce Forest at the Anchor Station Tharandt, Tellus B: Chemical and Physical Meteorology, 59, 387–396, https://doi.org/10.1111/j.1600-0889.2007.00259.x, 2007.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., Van der Linden, S., Leitão, P. J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., and Sang, B.: The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation, Remote Sensing, 7, 8830–8857, https://doi.org/10.3390/rs70708830, https://www.mdpi.com/2072-4292/7/7/8830, 2015.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, https://doi.org/10.1002/joc.3711, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.3711, 2014.

Hatala, J. A., Detto, M., and Baldocchi, D. D.: Gross Ecosystem Photosynthesis Causes a Diurnal Pattern in Methane Emission from Rice, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL051303, 2012.

Hutley, L. B., Beringer, J., Isaac, P. R., Hacker, J. M., and Cernusak, L. A.: A Sub-Continental Scale Living Laboratory: Spatial Patterns of Savanna Vegetation over a Rainfall Gradient in Northern Australia, Agricultural and Forest Meteorology, 151, {1417–1428}, https://doi.org/{10.1016/j.agrformet.2011.03.002}, 2011.

Imer, D., Merbold, L., Eugster, W., and Buchmann, N.: Temporal and Spatial Variations of Soil CO_2, CH_4 and N_2O Fluxes at Three Differently Managed Grasslands, Biogeosciences, 10, 5931–5945, https://doi.org/https://doi.org/10.5194/bg-10-5931-2013, 2013.

Jacobs, C. M. J., Jacobs, A. F. G., Bosveld, F. C., Hendriks, D. M. D., Hensen, A., Kroon, P. S., Moors, E. J., Nol, L., Schrier-Uijl, A., and Veenendaal, E. M.: Variability of Annual CO_2 Exchange from Dutch Grasslands, Biogeosciences, 4, 803–816, https://doi.org/https://doi.org/10.5194/bg-4-803-2007, 2007.

Justice, C., Townshend, J., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D., and Morisette, J.: An overview of MODIS Land data processing and product status, Remote Sensing of Environment, 83, 3 – 15, https://doi.org/https://doi.org/10.1016/S0034-4257(02)00084-6, http://www.sciencedirect.com/science/article/pii/S0034425702000846, the Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring, 2002.

Knohl, A., Schulze, E.-D., Kolle, O., and Buchmann, N.: Large Carbon Uptake by an Unmanaged 250-Year-Old Deciduous Forest in Central Germany, Agricultural and Forest Meteorology, 118, 151–167, https://doi.org/10.1016/S0168-1923(03)00115-1, 2003a.

Knohl, A., Schulze, E.-D., Kolle, O., and Buchmann, N.: Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany, Agricultural and Forest Meteorology, 118, 151 – 167, https://doi.org/https://doi.org/10.1016/S0168-1923(03)00115-1, http://www.sciencedirect.com/science/article/pii/S0168192303001151, 2003b.

Kodra, E., Chatterjee, S., and Ganguly, A. R.: Exploring Granger causality between global average observed time series of carbon dioxide and temperature, Theoretical and Applied Climatology, 104, 325–335, https://doi.org/10.1007/s00704-010-0342-3, https://doi.org/10.1007/s00704-010-0342-3, 2011.

Kretschmer, M., Coumou, D., Donges, J. F., and Runge, J.: Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation, Journal of Climate, 29, 4069–4081, https://doi.org/10.1175/JCLI-D-15-0654.1, https://doi.org/10.1175/JCLI-D-15-0654.1, 2016.

Kurbatova, J., Li, C., Varlagin, A., Xiao, X., and Vygodskaya, N.: Modeling Carbon Dynamics in Two Adjacent Spruce Forests with Different Soil Conditions in Russia, Biogeosciences, 5, 969–980, https://doi.org/https://doi.org/10.5194/bg-5-969-2008, 2008.

Leuning, R., Cleugh, H. A., Zegelin, S. J., and Hughes, D.: Carbon and Water Fluxes over a Temperate Eucalyptus Forest and a Tropical Wet/Dry Savanna in Australia: Measurements and Comparison with MODIS Remote Sensing Estimates, Agricultural and Forest Meteorology, 129, 151–173, https://doi.org/https://doi.org/10.1016/j.agrformet.2004.12.004, 2005.

Lindauer, M., Schmid, H. P., Grote, R., Mauder, M., Steinbrecher, R., and Wolpert, B.: Net Ecosystem Exchange over a Non-Cleared
5   Wind-Throw-Disturbed Upland Spruce Forest—Measurements and Simulations, Agricultural and Forest Meteorology, 197, 219–234, https://doi.org/10.1016/j.agrformet.2014.07.005, 2014.

Lund, M., Falk, J. M., Friborg, T., Mbufong, H. N., Sigsgaard, C., Soegaard, H., and Tamstorf, M. P.: Trends in CO2 Exchange in a High Arctic Tundra Heath, 2000–2010, Journal of Geophysical Research: Biogeosciences, 117, https://doi.org/10.1029/2011JG001901, 2012.

Ma, S., Baldocchi, D. D., Hatala, J. A., Detto, M., and Yuste, J. C.: Are rain-induced ecosystem respiration pulses
10   enhanced by legacies of antecedent photodegradation in semi-arid environments?, Agricultural and Forest Meteorology, 154-155, 203 – 213, https://doi.org/https://doi.org/10.1016/j.agrformet.2011.11.007, http://www.sciencedirect.com/science/article/pii/S0168192311003364, 2012.

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M. E., García-Santos, G., Fernandes, R., and Berger, M.: Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land, Remote Sensing of Environment, 120, 91 –
15   101, https://doi.org/https://doi.org/10.1016/j.rse.2011.09.026, http://www.sciencedirect.com/science/article/pii/S0034425712000648, the Sentinel Missions - New Opportunities for Science, 2012.

Marcolla, B., Pitacco, A., and Cescatti, A.: Canopy Architecture and Turbulence Structure in a Coniferous Forest, Boundary-Layer Meteorology, 108, 39–59, https://doi.org/10.1023/A:1023027709805, 2003.

Marcolla, B., Cescatti, A., Manca, G., Zorer, R., Cavagna, M., Fiora, A., Gianelle, D., Rodeghiero, M., Sottocornola, M., and Zampedri, R.:
20   Climatic Controls and Ecosystem Responses Drive the Inter-Annual Variability of the Net Ecosystem Exchange of an Alpine Meadow, Agricultural and Forest Meteorology, 151, 1233–1243, https://doi.org/10.1016/j.agrformet.2011.04.015, 2011.

Matsumoto, K., Ohta, T., Nakai, T., Kuwada, T., Daikoku, K., Iida, S., Yabuki, H., Kononov, A. V., van der Molen, M. K., Kodama, Y., Maximov, T. C., Dolman, A. J., and Hattori, S.: Energy Consumption and Evapotranspiration at Several Boreal and Temperate Forests in the Far East, Agricultural and Forest Meteorology, 148, 1978–1989, https://doi.org/10.1016/j.agrformet.2008.09.008, 2008.

25 McDowell, N. G., Bowling, D. R., Bond, B. J., Irvine, J., Law, B. E., Anthoni, P., and Ehleringer, J. R.: Response of the Carbon Isotopic Content of Ecosystem, Leaf, and Soil Respiration to Meteorological and Physiological Driving Factors in a Pinus Ponderosa Ecosystem, Global Biogeochemical Cycles, 18, https://doi.org/10.1029/2003GB002049, 2004.

McPherson, R. A.: A review of vegetation—atmosphere interactions and their influences on mesoscale phenomena, Progress in Physical Geography, 31, 261–285, https://doi.org/10.1177/0309133307079055, https://doi.org/10.1177/0309133307079055, 2007.

30 Merbold, L., Eugster, W., Stieger, J., Zahniser, M., Nelson, D., and Buchmann, N.: Greenhouse Gas Budget (CO2, CH4 and N2O) of Intensively Managed Grassland Following Restoration, Global Change Biology, 20, 1913–1928, https://doi.org/10.1111/gcb.12518, 2014.

Meyer, W. S., Kondrlova, E., and Koerber, G. R.: Evaporation of Perennial Semi-Arid Woodland in Southeastern Australia Is Adapted for Irregular but Common Dry Periods, Hydrological Processes, 29, {3714–3726}, https://doi.org/{10.1002/hyp.10467}, 2015.

Miralles, D. G., Gentine, P., Seneviratne, S. I., and Teuling, A. J.: Land–atmospheric feedbacks during droughts and heatwaves: state of the
35   science and current challenges, Annals of the New York Academy of Sciences, 0, https://doi.org/10.1111/nyas.13912, https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/nyas.13912, 2018.

Mogensen, P. K. and Riseth, A. N.: Optim: A mathematical optimization package for Julia, Journal of Open Source Software, 3, 615, https://doi.org/10.21105/joss.00615, 2018.

Monson, R. and Baldocchi, D.: Terrestrial Biosphere-Atmosphere Fluxes, Cambridge University Press, https://doi.org/10.1017/CBO9781139629218, 2014.

Moureaux, C., Debacq, A., Bodson, B., Heinesch, B., and Aubinet, M.: Annual Net Ecosystem Carbon Exchange by a Sugar Beet Crop, Agricultural and Forest Meteorology, 139, 25–39, https://doi.org/https://doi.org/10.1016/j.agrformet.2006.05.009, 2006.

5 Nave, L. E., Gough, C. M., Maurer, K. D., Bohrer, G., Hardiman, B. S., Moine, J. L., Munoz, A. B., Nadelhoffer, K. J., Sparks, J. P., Strahm, B. D., Vogel, C. S., and Curtis, P. S.: Disturbance and the Resilience of Coupled Carbon and Nitrogen Cycling in a North Temperate Forest, Journal of Geophysical Research: Biogeosciences, 116, https://doi.org/10.1029/2011JG001758, 2011.

Papagiannopoulou, C., Miralles, D. G., Decubber, S., Demuzere, M., Verhoest, N. E. C., Dorigo, W. A., and Waegeman, W.: A non-linear Granger-causality framework to investigate climate–vegetation dynamics, Geoscientific Model Development, 10, 1945–1960,

10   https://doi.org/10.5194/gmd-10-1945-2017, https://www.geosci-model-dev.net/10/1945/2017/, 2017a.

Papagiannopoulou, C., Miralles, D. G., Dorigo, W. A., Verhoest, N. E. C., Depoorter, M., and Waegeman, W.: Vegetation anomalies caused by antecedent precipitation in most of the world, Environmental Research Letters, 12, 074 016, http://stacks.iop.org/1748-9326/12/i=7/a=074016, 2017b.

Pearl, J.: Causality, Cambridge University Press, https://doi.org/10.1017/CBO9780511803161, 2009.

15 Peters, J., Janzing, D., and Schölkopf, B.: Elements of Causal Inference: Foundations and Learning Algorithms, MIT Press, Cambridge, MA, USA, 2017.

Pilegaard, K., Ibrom, A., Courtney, M. S., Hummelshøj, P., and Jensen, N. O.: Increasing Net CO2 Uptake by a Danish Beech Forest during the Period from 1996 to 2009, Agricultural and Forest Meteorology, 151, 934–946, https://doi.org/10.1016/j.agrformet.2011.02.013, 2011.

Pinzon, J. E. and Tucker, C. J.: A Non-Stationary 1981–2012 AVHRR NDVI3g Time Series, Remote Sensing, 6, 6929–6960,

20   https://doi.org/10.3390/rs6086929, http://www.mdpi.com/2072-4292/6/8/6929, 2014.

Prescher, A.-K., Grünwald, T., and Bernhofer, C.: Land Use Regulates Carbon Budgets in Eastern Germany: From NEE to NBP, Agricultural and Forest Meteorology, 150, 1016–1025, https://doi.org/10.1016/j.agrformet.2010.03.008, 2010a.

Prescher, A.-K., Grünwald, T., and Bernhofer, C.: Land Use Regulates Carbon Budgets in Eastern Germany: From NEE to NBP, Agricultural and Forest Meteorology, 150, 1016–1025, https://doi.org/10.1016/j.agrformet.2010.03.008, 2010b.

25 Pryor, S. C., Barthelmie, R. J., and Jensen, B.: Nitrogen Dry Deposition at an AmeriFlux Site in a Hardwood Forest in the Midwest, Geophysical Research Letters, 26, 691–694, https://doi.org/10.1029/1999GL900066, 1999.

Qi, W. and Dubayah, R. O.: Combining Tandem-X InSAR and simulated GEDI lidar observations for forest structure mapping, Remote Sensing of Environment, 187, 253 – 266, https://doi.org/https://doi.org/10.1016/j.rse.2016.10.018, http://www.sciencedirect.com/science/article/pii/S003442571630390X, 2016.

30 Rambal, S., Joffre, R., Ourcival, J. M., Cavender-Bares, J., and Rocheteau, A.: The Growth Respiration Component in Eddy CO2 Flux from a Quercus Ilex Mediterranean Forest, Global Change Biology, 10, 1460–1469, https://doi.org/10.1111/j.1365-2486.2004.00819.x, 2004.

Rayment, M. B. and Jarvis, P. G.: Seasonal Gas Exchange of Black Spruce Using an Automatic Branch Bag System, Canadian Journal of Forest Research, 29, 1528–1538, https://doi.org/10.1139/x99-130, 1999a.

Rayment, M. B. and Jarvis, P. G.: Seasonal Gas Exchange of Black Spruce Using an Automatic Branch Bag System, Canadian Journal of

35   Forest Research, 29, 1528–1538, https://doi.org/10.1139/x99-130, 1999b.

Rey, A., Pegoraro, E., Tedeschi, V., Parri, I. D., Jarvis, P. G., and Valentini, R.: Annual Variation in Soil Respiration and Its Components in a Coppice Oak Forest in Central Italy, Global Change Biology, 8, 851–866, https://doi.org/10.1046/j.1365-2486.2002.00521.x, 2002.

Rothstein, D. E., Zak, D. R., Pregitzer, K. S., and Curtis, P. S.: Kinetics of Nitrogen Uptake by Populus Tremuloides in Relation to Atmospheric CO2 and Soil Nitrogen Availability, Tree Physiology, 20, 265–270, https://doi.org/10.1093/treephys/20.4.265, 2000.

Ruddell, B. L. and Kumar, P.: Ecohydrologic process networks: 1. Identification, Water Resources Research, 45, n/a–n/a, https://doi.org/10.1029/2008WR007279, http://dx.doi.org/10.1029/2008WR007279, w03419, 2009.

5   Ruehr, N. K., Martin, J. G., and Law, B. E.: Effects of Water Availability on Carbon and Water Exchange in a Young Ponderosa Pine Forest: Above- and Belowground Responses, Agricultural and Forest Meteorology, 164, 136–148, https://doi.org/10.1016/j.agrformet.2012.05.015, 2012a.

Ruehr, N. K., Martin, J. G., and Law, B. E.: Effects of Water Availability on Carbon and Water Exchange in a Young Ponderosa Pine Forest: Above- and Belowground Responses, Agricultural and Forest Meteorology, 164, 136–148,
10   https://doi.org/10.1016/j.agrformet.2012.05.015, 2012b.

Runge, J.: Causal network reconstruction from time series: From theoretical assumptions to practical estimation, Chaos: An Interdisciplinary Journal of Nonlinear Science, 28, 075 310, https://doi.org/10.1063/1.5025050, https://doi.org/10.1063/1.5025050, 2018a.

Runge, J.: TIGRAMITE-Causal discovery for time series datasets, https://github.com/jakobrunge/tigramite/, 2018b.

Runge, J., Heitzig, J., Petoukhov, V., and Kurths, J.: Escaping the Curse of Dimensionality in Estimating Multivariate Transfer En-
15   tropy, Phys. Rev. Lett., 108, 258 701, https://doi.org/10.1103/PhysRevLett.108.258701, https://link.aps.org/doi/10.1103/PhysRevLett.108.258701, 2012.

Runge, J., Petoukhov, V., and Kurths, J.: Quantifying the Strength and Delay of Climatic Interactions: The Ambiguities of Cross Correlation and a Novel Measure Based on Graphical Models, Journal of Climate, 27, 720–739, https://doi.org/10.1175/JCLI-D-13-00159.1, https://doi.org/10.1175/JCLI-D-13-00159.1, 2014.

20   Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D.: Detecting causal associations in large nonlinear time series datasets, arXiv e-prints, arXiv:1702.07007v2, https://arxiv.org/abs/1702.07007v2, 2018.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., et al.: Inferring causation from time series in Earth system sciences, Nature communications, 10, 2553, 2019.

Saleska, S. R., Miller, S. D., Matross, D. M., Goulden, M. L., Wofsy, S. C., da Rocha, H. R., de Camargo, P. B., Crill, P.,
25   Daube, B. C., de Freitas, H. C., Hutyra, L., Keller, M., Kirchhoff, V., Menton, M., Munger, J. W., Pyle, E. H., Rice, A. H., and Silva, H.: Carbon in Amazon Forests: Unexpected Seasonal Fluxes and Disturbance-Induced Losses, Science, 302, 1554–1557, https://doi.org/10.1126/science.1091165, 2003.

Schade, G. W., Goldstein, A. H., and Lamanna, M. S.: Are Monoterpene Emissions Influenced by Humidity?, Geophysical Research Letters, 26, 2187–2190, https://doi.org/10.1029/1999GL900444.

30   Scott, R. L., Huxman, T. E., Cable, W. L., and Emmerich, W. E.: Partitioning of Evapotranspiration and Its Relation to Carbon Dioxide Exchange in a Chihuahuan Desert Shrubland, Hydrological Processes, 20, 3227–3243, https://doi.org/10.1002/hyp.6329, 2006.

Scott, R. L., Cable, W. L., and Hultine, K. R.: The Ecohydrologic Significance of Hydraulic Redistribution in a Semiarid Savanna, Water Resources Research, 44, https://doi.org/10.1029/2007WR006149, 2008.

Shadaydeh, M., Garcia, Y. G., Mahecha, M., Reichstein, M., and Denzler, J.: Analyzing the Time Variant Causality in Ecological Time
35   Series: A Time-Frequency Approach, in: International Conference on Ecological Informatics (ICEI), pp. 151–152, 2018.

Spano, D., Duce, P., Marras, S., Sirca, C., Arca, A., Zara, P., Ventura, A., Mereu, S., and Sanna, L.: FLUXNET2015 IT-Noe Arca di Noe - Le Prigionette, https://doi.org/10.18140/FLX/1440171.

Spirtes, P. and Glymour, C.: An Algorithm for Fast Recovery of Sparse Causal Graphs, Social Science Computer Review, 9, 62–72, https://doi.org/10.1177/089443939100900106, https://doi.org/10.1177/089443939100900106, 1991.

Spirtes, P., Glymour, C., and Scheines, R.: Causation, prediction, and search, MIT Press, 2001.

Suni, T., Rinne, J., Reissell, A., Altimir, N., Keronen, P., Rannik, U., Dal Maso, M., Kulmala, M., and Vesala, T.: Long-Term Measurements
of Surface Fluxes above a Scots Pine Forest in Hyytiala, Southern Finland, 1996-2001, BOREAL ENVIRONMENT RESEARCH, 8,
{287–301}, 2003.

Tang, J., Baldocchi, D. D., Qi, Y., and Xu, L.: Assessing Soil CO2 Efflux Using Continuous Measurements of CO2 Profiles in Soils with
Small Solid-State Sensors, Agricultural and Forest Meteorology, 118, 207–220, https://doi.org/10.1016/S0168-1923(03)00112-6, 2003.

Thum, T., Aalto, T., Laurila, T., Aurela, M., Kolari, P., and Hari, P.: Parametrization of Two Photosynthesis Models at the Canopy Scale in a
Northern Boreal Scots Pine Forest, Tellus B, 59, 874–890, https://doi.org/10.1111/j.1600-0889.2007.00305.x, 2007.

Valentini, R., Angelis, P. D., Matteucci, G., Monaco, R., Dore, S., and Mucnozza, G. E. S.: Seasonal Net Carbon Dioxide Exchange of a
Beech Forest with the Atmosphere, Global Change Biology, 2, 199–207, https://doi.org/10.1111/j.1365-2486.1996.tb00072.x, 1996.

Viovy, N.: CRUNCEP data set, availableat:ftp://nacp.ornl.gov/synthesis/2009/frescati/temp/land_use_change/original/readme.htm,
lastaccess:June2016, 2016.

Vitale, L., Di Tommasi, P., D'Urso, G., and Magliulo, V.: The Response of Ecosystem Carbon Fluxes to LAI and Environmental Drivers in a
Maize Crop Grown in Two Contrasting Seasons, International Journal of Biometeorology, 60, 411–420, https://doi.org/10.1007/s00484-
015-1038-2, 2016.

Wang, C., Bond-Lamberty, B., and Gower, S. T.: Environmental Controls on Carbon Dioxide Flux from Black Spruce Coarse Woody Debris,
Oecologia, 132, 374–381, https://doi.org/10.1007/s00442-002-0987-4, 2002a.

Wang, C., Bond-Lamberty, B., and Gower, S. T.: Environmental Controls on Carbon Dioxide Flux from Black Spruce Coarse Woody Debris,
Oecologia, 132, 374–381, https://doi.org/10.1007/s00442-002-0987-4, 2002b.

Wang, C., Bond-Lamberty, B., and Gower, S. T.: Environmental Controls on Carbon Dioxide Flux from Black Spruce Coarse Woody Debris,
Oecologia, 132, 374–381, https://doi.org/10.1007/s00442-002-0987-4, 2002c.

Westergaard-Nielsen, A., Lund, M., Hansen, B. U., and Tamstorf, M. P.: Camera Derived Vegetation Greenness Index as Proxy for
Gross Primary Production in a Low Arctic Wetland Area, ISPRS Journal of Photogrammetry and Remote Sensing, 86, 89–99,
https://doi.org/10.1016/j.isprsjprs.2013.09.006, 2013.

Wofsy, S. C., Goulden, M. L., Munger, J. W., Fan, S.-M., Bakwin, P. S., Daube, B. C., Bassow, S. L., and Bazzaz, F. A.: Net Exchange of
CO2 in a Mid-Latitude Forest, Science, 260, 1314–1317, https://doi.org/10.1126/science.260.5112.1314, 1993.

Wohlfahrt, G., Hammerle, A., Haslwanter, A., Bahn, M., Tappeiner, U., and Cernusca, A.: Seasonal and inter-annual variability of the net
ecosystem CO2 exchange of a temperate mountain grassland: Effects of weather and management, Journal of Geophysical Research:
Atmospheres, 113, https://doi.org/10.1029/2007JD009286, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009286.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., Helmer, E.,
Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P. S., Vermote, E. F., Vogelmann, J., Wulder, M. A., and Wynne, R.: Free Access to
Landsat Imagery, Science, 320, 1011–1011, https://doi.org/10.1126/science.320.5879.1011a, https://science.sciencemag.org/content/320/
5879/1011.1, 2008.

Wright, S.: Correlation and causation, J. Agric. Res., 20, 557–580, https://ci.nii.ac.jp/naid/10010273333/en/, 1921.

Wu, D., Zhao, X., Liang, S., Zhou, T., Huang, K., Tang, B., and Zhao, W.: Time-lag effects of global vegetation responses to climate change, Global Change Biology, 21, 3520–3531, https://doi.org/10.1111/gcb.12945, https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.12945, 2015.

Xu, L., Baldocchi, D. D., and Tang, J.: How Soil Moisture, Rain Pulses, and Growth Alter the Response of Ecosystem Respiration to Temperature, Global Biogeochemical Cycles, 18, https://doi.org/10.1029/2004GB002281, 2004.

Zielis, S., Etzold, S., Zweifel, R., Eugster, W., Haeni, M., and Buchmann, N.: NEP of a Swiss Subalpine Forest Is Significantly Driven Not Only by Current but Also by Previous Year's Weather, Biogeosciences, 11, 1627–1635, https://doi.org/https://doi.org/10.5194/bg-11-1627-2014, 2014.