

Causality and Persistence in Ecological Systems: A Nonparametric Spectral Granger Causality Approach

Matteo Detto,^{1,2,*} Annalisa Molini,^{3,4,5} Gabriel Katul,^{4,5} Paul Stoy,⁶ Sari Palmroth,⁴ and Dennis Baldocchi²

1. Smithsonian Tropical Resource Institute, Apartado Postal 0843-03092 Balboa, Ancon, Panama; 2. Department of Environmental Science, Policy, and Management, University of California, Berkeley, California 94720; 3. Masdar Institute Science and Technology, Abu Dhabi, United Arab Emirates; 4. Nicholas School of the Environment, Duke University, Durham, North Carolina 27708; 5. Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, North Carolina 27708; 6. Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, Montana 59717

Submitted December 20, 2010; Accepted December 12, 2011; Electronically published February 20, 2012

Online enhancements: appendixes, zip file.

ABSTRACT: Directionality in coupling, defined as the linkage relating causes to their effects at a later time, can be used to explain the core dynamics of ecological systems by untangling direct and feedback relationships between the different components of the systems. Inferring causality from measured ecological variables sampled through time remains a formidable challenge further made difficult by the action of periodic drivers overlapping the natural dynamics of the system. Periodicity in the drivers can often mask the self-sustained oscillations originating from the autonomous dynamics. While linear and direct causal relationships are commonly addressed in the time domain, using the well-established machinery of Granger causality (G-causality), the presence of periodic forcing requires frequency-based statistics (e.g., the Fourier transform), able to distinguish coupling induced by oscillations in external drivers from genuine endogenous interactions. Recent nonparametric spectral extensions of G-causality to the frequency domain pave the way for the scale-by-scale decomposition of causality, which can improve our ability to link oscillatory behaviors of ecological networks to causal mechanisms. The performance of both spectral G-causality and its conditional extension for multivariate systems is explored in quantifying causal interactions within ecological networks. Through two case studies involving synthetic and actual time series, it is demonstrated that conditional G-causality outperforms standard G-causality in identifying causal links and their concomitant timescales.

Keywords: spectral Granger causality, complex systems, periodic forcing, photosynthesis-respiration coupling.

Introduction

Systems exhibiting mutualisms and competition or other forms of two-way interactions are ubiquitous in ecology. Some examples taken from population and community

ecology include Lotka-Volterra predator-prey systems (Roughgarden 1975 and references therein; Chesson 1994; Holmes et al. 1994), nutrient recycling (Loreau 2001), and human-biosphere interactions (e.g., Brander and Taylor 1998; Raupach 2007). Some populations, such as insects, that display nonlinear demographic dynamics (Costantino et al. 1997) exhibit complex endogenous behaviors, even in the absence of environmental variability. For others, the interplay with environmental drivers is crucial in determining their physical state, as in the case of vegetation and soil moisture in water-limited regions (Detto et al. 2006).

Common to all of these examples is some understanding of the dynamical nature of the core ecological system, which is often encoded in a set of autonomous differential equations (e.g., a Lotka-Volterra-like system). These systems are causal in that their dynamics can be explained in terms of conditionals in the form of “if A had not occurred, B would not occur.” When such dynamics lead to self-sustained oscillations and, at the same time, some external time-dependent forcing is acting on this system, it becomes difficult to distinguish the signature of autonomous oscillatory dynamics from the effects of the external forcing. In a similar context, any coupling among ecological variables induced by external drivers is hard to distinguish from the endogenous correlation structure of the core ecological system. Consequently, it is not surprising that in ecological time series analysis, one of the most challenging goals is to determine how one state variable causes the future evolution of another.

The standard approach to confront this challenge is through controlled manipulative experiments. This empirical approach provides strong evidence because it can reveal unambiguous mechanistic causal relationships. However, in many natural settings, such manipulative ex-

* Corresponding author; e-mail: dettom@si.edu.

Am. Nat. 2012. Vol. 179, pp. 524–535. © 2012 by The University of Chicago. 0003-0147/2012/17904-52705\$15.00. All rights reserved.

DOI: 10.1086/664628

periments are difficult to implement. For example, experimental manipulation of both air temperature and CO₂ concentration are difficult at the ecosystem scale (Shaw et al. 2002), and findings from such experiments may be scale dependent or constrained by the limited spatial scale. Even well-designed manipulative experiments are not free from spurious results, as may occur when covariates are inadvertently changed by the manipulation.

Another less direct approach to infer causal interactions (if possible) relies on observational data (Shipley 2000), for example, using structural equation modeling (SEM). Path analysis, a special case of SEM, can be formulated as a multiple regression problem aimed at decomposing different sources of correlation within an ecological system rather than explicitly accounting for directional coupling in time (i.e., causes preceding effects). However, the existence of correlation itself does not necessarily entail causation as implied by an action and a subsequent reaction. Correlation metrics are essentially symmetric and cannot, by definition, measure time asymmetry—often considered the signature of causality (Hlavackova-Schindler et al. 2007; Lungarella et al. 2007; Paluš and Vejmelka 2007)—nor can they distinguish between unidirectional and bidirectional couplings. Additionally, path analysis handles with difficulty any colinearity (Petraitis et al. 1996).

The idea of causality has been developed and studied in a number of disciplines ranging from philosophy to science and by no means can be covered in a single article (e.g., see Peters 2001, pp. 128–136, for a comprehensive synthesis of definitions and measures in ecology and Pearl 2009 for applications in statistical inference). The scope of this work is a specific yet widely used type of causality known as Granger causality (or G-causality; Granger 1969), where cause and effect are linearly related. This causality metric originated in econometrics but is now proliferating to a number of disciplines, including geosciences (Salvucci et al. 2002; Kaufmann et al. 2008; Smirnov and Mokhov 2009) and neurosciences (Bernasconi and König 1999; Ding et al. 2006). G-causality explains phenomena by showing them to be due to effects originating from prior causes in time. When these prior causes are accounted for, predictions of the phenomena are improved against a null hypothesis that does not account for these prior causes. Hence, the premise of G-causality encodes two of the three cardinal aspects of causality (explanation, prediction, and teleology) singled out by Mayr (1960) in biology. However, due to the predominantly periodic nature of external drivers encountered in ecology, as well as the complex nature of ecological systems, an extension of G-causality to the frequency domain (i.e., a spectrally transformed series) is warranted and proposed here. The basic idea is to decompose and analyze causal

relationships as a function of the frequency at which they occur instead of as a function of time.

Spectral methods have been widely adopted in ecology and are reviewed elsewhere (e.g., Platt and Denman 1975; Priestley 1981). In the multitude of possible ecological applications, the goal remains the same: to move from a framework where variability of the analyzed system is a function of time (time domain) to a new approach where the variance can be decomposed scale by scale, allowing for a clear view of how different frequencies impact the considered process.

The principle of causation (Mayr 1960) and the search for causality in relation to space and timescales within an ecological process have a long history in ecology, preceding even the work by Granger in economics. Kershaw's (1963) seminal work investigated the length scales of vegetation clumping while trying to assess causality by linking the scale of the pattern with its generating mechanism. In this case, short-length scale patterns (~10 cm) suggested a causal connection based on plant morphology, while length scales of tens of meters implied interactions with topography, water drainage networks, or soil composition (i.e., mainly but not completely exogenous mechanisms). A similar approach can be applied to ecological systems displaying a linear causal structure in time. In a linear system, an exogenous forcing at a given period propagates through the system unaltered (except through its phase angle) to produce an effect at the same frequency.

The approach may fail in the presence of nonlinearity. In this case, a driver acting on a specific frequency can produce a response with variability at one or more different frequencies (Pascual and Ellner 2000). Extracting information from nonlinear time series is inherently challenging: “when the analyzed data cannot be distinguished from realizations of a linear process, it is hardly expected to extract useful information by a nonlinear descriptor” (Paluš 2008, p. 320), especially for short, noisy time series. In any case, linearized models can often provide acceptable approximations of the dynamics of ecological systems, allowing for the estimation of main direct interactions by simpler linear causal metrics (Nisbet and Gurney 1982; Royama 1992). In spite of this, nonlinearity tests may establish the extent to which linear analysis is pertinent to ecological data without generating spurious causal relationships (Paluš 2008; Papan et al. 2011).

Here, the main theory describing G-causality in the time domain is briefly reviewed. Next, the analysis is extended to the spectral domain, where systems of reciprocally interacting variables are subjected to stochastic forcing. Finally, a novel approach based on conditional G-causality in the spectral domain is presented. This conditional statistic, hereafter referred to as conditional spectral G-causality, allows for differentiating direct causal linkage from

indirect causal linkage between multiple state variables even under the action of external drivers. This approach builds on the so-called spectral decomposition of G-causality (Chen et al. 2006; Dhamala et al. 2008), where time series are transformed to the frequency domain and the causal relationships between periodic drivers and self-sustained oscillations are assessed scale by scale.

As an illustration, two applications of G-causality are presented: one theoretical (simulation based) and the other empirical. The theoretical case focuses on a system composed of primary producers and decomposers that are allowed to interact in a compartmental model while incorporating stochastic components. The empirical case explores the causal interaction between photosynthesis and soil respiration in forested ecosystems when time series measurements are conditioned on soil temperature and the presence of leaves. This empirical case study was previously analyzed by Stoy et al. (2007), using a lagged correlation, or “pulse response,” analysis. Their analysis failed to establish a strong causal relationship precisely because periodicity occurred at multiple timescales (in the temperature and incident radiation) overlapping with the characteristic scale of endogenous interaction between photosynthesis and respiration. These two applications are intended to provide a methodological basis for the detection of G-causality in ecological systems while illustrating its usefulness as a diagnostic approach.

Methods of Analysis

Granger Causality: A Brief Review

Granger (1969, p. 430) defines causality as follows: “if some other series y_i contains information in past terms that helps in the prediction of x_i and if this information is contained in no other series used in the predictor, then y_i is said to cause x_i .” This connection between causality and predictability led Granger to express linear causality in a parametric form, drawing from established tools in time series analysis (see Box and Jenkins 1970), such as autoregressive models.

If an autoregressive representation of order m is used, the bivariate model assessing the influence of a random variable y on another random variable x (and vice versa) is given by

$$x_n = \sum_{j=1}^m a_{1,j} x_{n-j} + \sum_{j=1}^m a_{2,j} y_{n-j} + \varepsilon_n, \quad (1a)$$

$$y_n = \sum_{j=1}^m b_{1,j} x_{n-j} + \sum_{j=1}^m b_{2,j} y_{n-j} + \xi_n, \quad (1b)$$

where ε and ξ are the prediction errors while a and b are coefficients describing the linear interactions between the

variables, with subscript j indicating time lags. When the above equation is compared with a univariate model, $x_n = \sum_{j=1}^m a_j x_{n-j} + \eta_n$, and when the multivariate model outperforms the univariate case (e.g., $\sigma_\varepsilon^2 < \sigma_\eta^2$), y is said to have a causal effect on x (and similarly for the effect of x on y). This is the statistical interpretation of causality proposed by Granger (1969) and is commonly referred to as Granger causality, or G-causality.

Note here that correlation itself does not necessarily imply an improvement in prediction (Aldrich 1995). Correlation is a measure of coupling strength, which can originate from both causation and dependence on common causes.

G-causality is a measure of coupling, with time directionality being explicit. For this reason, it is based on prediction errors rather than on linear interactions among coefficients. Traditionally, it is expressed as the ratio between the residual variance of the bivariate and the null model (i.e., the univariate case) and is given as

$$G_{y \rightarrow x} = \ln \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}. \quad (2)$$

If the variables x and y do not interact, there will be no improvement in using y to predict x ; that is, $\sigma_\varepsilon^2 \approx \sigma_\eta^2$ and $G_{y \rightarrow x} \approx 0$, even if the two variables are correlated. If otherwise y has a causal influence on x , $\sigma_\varepsilon^2 < \sigma_\eta^2$, so $G_{y \rightarrow x} > 0$. The causal effect of x on y , $G_{x \rightarrow y}$, is defined in a similar manner. G-causality varies between 0 and infinity; it is analogous but not identical to $-\ln(1 - R^2)$, where R^2 is the usual coefficient of determination (the proportion of the variance explained by a linear model). Because of the temporal sequence shown in equations (1), it is clear that G-causality can capture only functional relationships for which cause and effect are sufficiently separated in time.

From Time to Frequency: Spectral G-Causality and the Nonparametric Approach

Ecological processes, especially those connected with the exchanges of mass and energy between biosphere and atmosphere, often display strong oscillations due to diurnal or seasonal cycles (Baldocchi et al. 2001b; Katul et al. 2001). Interannual variability in many natural ecosystems is regulated by large climatic fluctuations (Stenseth et al. 2002), which can be revealed by spectral analysis. Hence, it is convenient to extend the definition of the causality in equation (2) to the frequency domain and introduce the spectral analogue of G-causality I as a function of frequency f (Geweke 1982), using

$$I(f)_{y \rightarrow x} = \ln \left\{ \frac{S_{xx}(f)}{S_{xx}(f) - [\mathbf{\Gamma}_{yy} - (\mathbf{\Gamma}_{xy}^2 / \mathbf{\Gamma}_{xx})] |\mathbf{H}_{xy}(f)|^2} \right\}, \quad (3)$$

where $S_{xx}(f)$ is the value of the spectral density of x (also called the periodogram, or power spectrum) at frequency f , $\mathbf{\Gamma}$ is the error covariance matrix for the system of equations (1) and $\mathbf{H}(f)$ is a spectral transfer function matrix resulting from rewriting system (1) in the Fourier domain.

As in equation (2), if the variables x and y are not interacting, the numerator and denominator of equation (3) are equal and $I(f)_{y \rightarrow x} \approx 0$. If y manifests a causal influence on x at a specific frequency f_s , then $I(f_s)_{y \rightarrow x} > 0$.

The robustness of these statistics must be tested against the null hypothesis that the two series do not exhibit any causal interactions at a particular frequency. In general, the analytical solution for the causality structure of the system is not a priori known, necessitating the use of Monte Carlo simulation methods and surrogate data. The null hypothesis that no linear causal relationships exist among the variables considered is here tested using the so-called iterative amplitude adjusted Fourier transform (IAAFT) surrogates. The IAAFT surrogates, recently adopted in testing rainfall nonlinearity and cross-scale causality in time (Roux et al. 2009; Molini and Katul 2010), simultaneously preserve the probability distribution and the power spectrum of the original time series. Thus, the linear correlation structure of the original time series is conserved by the IAAFT, while any other form of coupling or nonlinear correlation is destroyed.

Since \mathbf{S} , \mathbf{H} , and $\mathbf{\Gamma}$ can be inferred from the time series, equation (3) does not require any assumption regarding the autoregressive order of the model representing the data and the determination of $I(f)$ can be considered non-parametric. The spectral density matrix \mathbf{S} can be estimated directly from the spectral and cross-spectral functions (using Fourier, wavelet transform, or analogous spectral methods). The matrix factorization (Wilson 1972; Dhamala et al. 2008) allows \mathbf{H} and $\mathbf{\Gamma}$ to be computed from the equality $\mathbf{S} = \mathbf{H}\mathbf{\Gamma}\mathbf{H}'$, where the prime indicates the matrix adjoint. Full details of such factorization are provided in appendix A, available online.

This approach is particularly effective for time series exhibiting apparent periodicities, superimposed on long-range memory, which would otherwise require higher-order autoregressive models to reconstruct the observed dynamics. The estimation error of an autoregressive model is known to increase with order. Moreover, order misspecifications may produce spurious causality effects (Hlavackova-Schindler et al. 2007).

The adopted scheme can be extended to the multivariate case, which now involves k stochastic variables (z_1, \dots, z_k). In this extension, it becomes possible to compute the so-

called conditional G-causality $I(f)_{y \rightarrow x | z_1, \dots, z_k}$ (see details in app. A); that is, y causes x , given that z_1, \dots, z_k cause x or y (Chen 2006).

Statistics Performance and Key Applications

Case 1: A Stoichiometric Model

The interactions between primary producers and decomposers are often studied by using a stoichiometric approach. Following Daufresne and Loreau (2001), the cycle of organic and inorganic matter within an ecosystem can be represented by three compartments: primary producers (P), detritus (L), and decomposers (D), along with a pool of inorganic nutrients (N).

The detritus generated by producers constitutes the substrate for the communities of decomposers, which mineralize nutrients and release carbon back into the atmosphere. All the compartments have constant C : N ratios (here denoted by α for P and L and by β for D , with $\alpha > \beta$), which allows formulating explicit mass-balance constraints and expressing stocks and fluxes independently in terms of C or N. The models combine indirect mutualism and competition because decomposers recycle nutrients as well as compete with producers for them. The dynamical equations for a closed system (zero dilution rate) expressed in nutrient stocks are

$$\begin{aligned} \dot{P} &= \Phi_n - eP, \\ \dot{L} &= eP - \Phi_d, \\ \dot{D} &= \Phi_i + \Phi_d - mD, \\ \dot{N} &= mD - \Phi_i - \Phi_n, \end{aligned} \quad (4)$$

where e and m are the primary producers' death rate and the decomposers' mineralization rate, respectively, while Φ_n , Φ_d , and Φ_i represent the nutrient uptake, decomposition, and immobilization fluxes, respectively. These fluxes can be described using a nutrition consumption function of a Lotka-Volterra type (nonlinear interaction) or, alternatively, using a donor-controlled mechanism (linear interactions). For this model, the results of the donor-controlled mechanism approach were shown to be qualitatively the same as the Lotka-Volterra type (Daufresne and Loreau 2001). Namely, the system can be assumed dynamically linear. This schematization is feasible only when decomposers are carbon limited. The immobilization flux can be expressed as $\Phi_i = [(\alpha - \beta)/\beta]\Phi_d$. If a donor-controlled interaction type is assumed, nutrition and decomposition fluxes are $\Phi_n = nN$ and $\Phi_d = dL$, respectively, where n is the plant nutrition rate and d is the detritus decomposition rate. Because the total nutrient

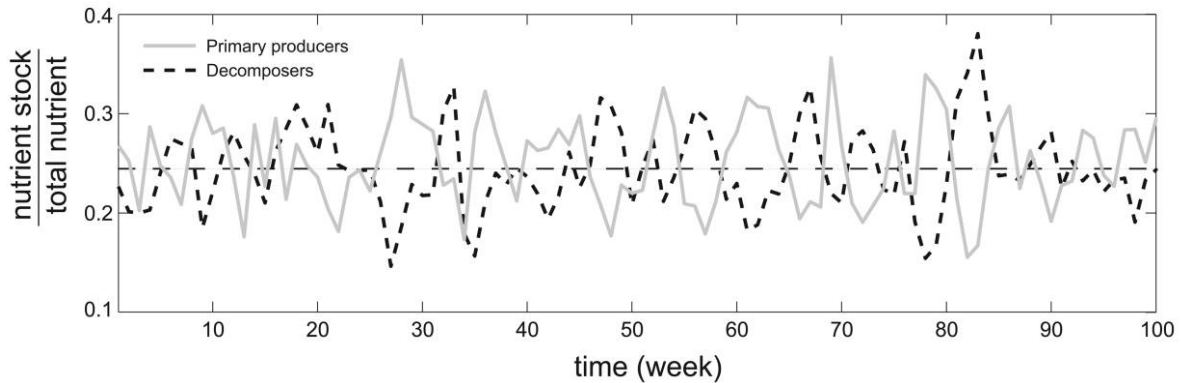


Figure 1: Sample realization of normalized nutrient stocks for primary producers and decomposers, simulated with equation (5) ($P^* = N^* = 0.25N_T$).

stock is constant for a closed system ($N_T = N + P + L + D$), equation (4) reduces to three independent equations:

$$\begin{aligned} \dot{P} &= nN_T - (e + n)P - nL - nD, \\ \dot{L} &= eP + dL, \\ \dot{D} &= \frac{\alpha}{\beta}dL - mD. \end{aligned} \tag{5}$$

The stable steady state, obtained by setting all time derivatives to 0, is given by

$$\begin{aligned} P^* &= \frac{N_T}{\chi}, \\ L^* &= \frac{e}{d}P^*, \\ D^* &= \frac{\alpha e}{\beta m}P^*, \\ N^* &= \frac{e}{n}P^*, \end{aligned} \tag{6}$$

where $\chi = 1 + (e/n) + (e/d) + (\alpha/\beta)(e/m)$.

In the presence of environmental noise, the discrete form of the deterministic system in equation (5) is analogous to the linear first-order autoregressive model in equations (1). The introduction of random terms is assumed to represent the variability of the nutrition, decomposition, mineralization, and mortality rates as may be induced by hurricane, pest or disease, changes in mineralization rate in response to temperature fluctuations, and so on.

In natural (not controlled) ecosystems P , L , D , and N vary over very different timescales (from years to days) and are characterized by different abundances in turn influencing parameters m , e , d , and n . We simulated 50 short realizations, each with 100 weekly steps, to mimic real ecological time series, for the system in equation (5). For simplicity, the total nutrient stock is equally divided at equilibrium among organic and inorganic compartments ($P^* = L^* = D^* = N^* = N_T/4$) and $\alpha = 1$, $\beta = 0.8$, $n = e = d = 0.5$, $m = 0.625$. The three stochastic variables are assumed to be independent Gaussian noises with variance equal to 10^{-3} . Note that this assumption is not required and analogous results would be obtained with a different type of noise (e.g., Poisson) even if the noises were not independent (e.g., autocorrelated or correlated with each other, as in natural ecosystems). As an illustration, figure 1 shows one realization for primary producers and decomposers, expressed in stock of nutrients, while figure 2 depicts the conditional G-causality for different interaction schemes among L , P , and D .

From these simulations, the conditional G-causality correctly detected the dynamical interactions represented in the model; L and D have a direct influence on P (fig. 2*b*, 2*c*), P controls L (fig. 2*d*), but D has no direct effect on L (fig. 2*e*). Finally, D is controlled by the amount of detritus L (fig. 2*h*), as the decomposers are carbon limited. Pairwise G-causalities in figure 2 (dashed lines) fail to determine unambiguously the real direct interactions and their strengths. In fact, the G-causalities in figure 2*f* and 2*g* are not 0. Producers, for example, have an effect on decomposers but only indirectly, as mediated by the detritus pool (fig. 2*g*). The bivariate causality of L on P (fig. 2*b*) is also overestimated. The detritus appears to have more influence

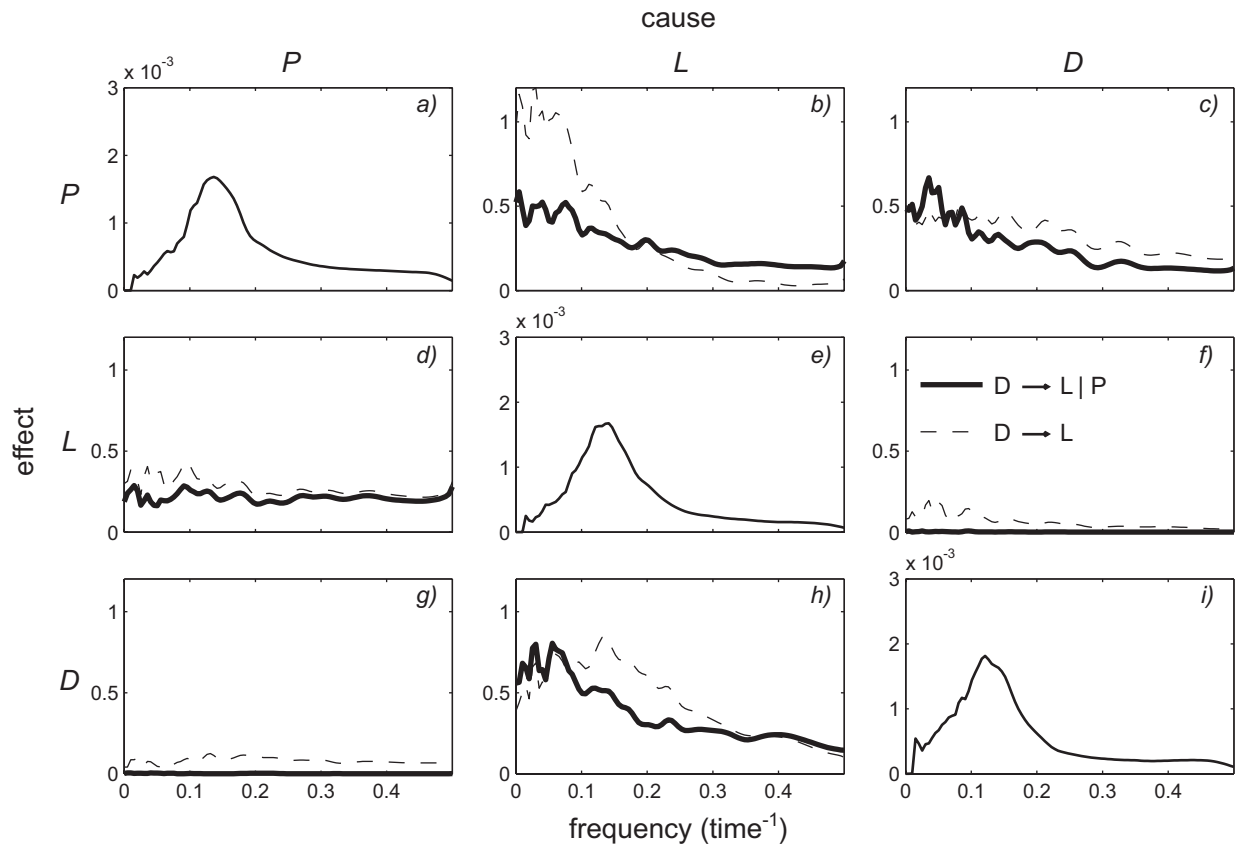


Figure 2: Conditional Granger causalities (G-causalities; bold solid lines) for the model simulations indicate the direct interactions among two variables, given the effect of the third (e.g., the effect of L on P , given D ; b). Spectral estimations are shown in the diagonal panels, and the G-causalities for the bivariate case are shown as dashed lines. The bold lines represent the desired goal of the analysis because they isolate direct effects, while the dashed lines indicate the apparent connections that are due to a combination of direct and indirect effects. Legend in f applies to all panels.

on producers than on decomposers (cf. fig. 2*b*, fig. 2*h*) when actually the model parameters were originally selected to show the opposite ($n = 0.5$, $\alpha/\beta d = 0.625$).

The choice of parameters determines the frequency corresponding to the peak in the power spectra (fig. 2, diagonal panels), which in this example is $f_{\max} = 0.13$ week $^{-1}$.

The performance of the G-causality in detecting real interactions was tested by repeating this numerical experiment 1,000 times using 50 and 10 realizations separately for each simulation. The intent of these different realization sizes is to assess the rapid convergence of the G-causality metric. Because the model structure is known a priori here, analytical solutions can be obtained directly from the model's parameters (Chen et al. 2006). In figure 3, a case of interest is shown that represents the decomposers' carbon limitation hypothesis, that is, the influence of the detritus on decomposers after taking into account

the effect of producers. When the 50 realizations are used, estimates that are much closer to the analytical solution are obtained when compared to using only 10 realizations.

A large number of numerical simulations suggest that unless the variance of the noise becomes too small compared to measuring errors, conditional G-causality always detected the correct directionality in this system. This finding was not altered even when the noises were correlated with each other. For further examples, see appendix B, available online; original Matlab code developed and used in the examples is also available online as a zip file.

Case 2: Ecosystem Application

Based on manipulation experiments, stable and radio isotope techniques, and soil gas phase CO_2 concentration measurements, a number of studies (Horwath et al. 1994; Andrews et al. 1999; Hogberg et al. 2001) have suggested

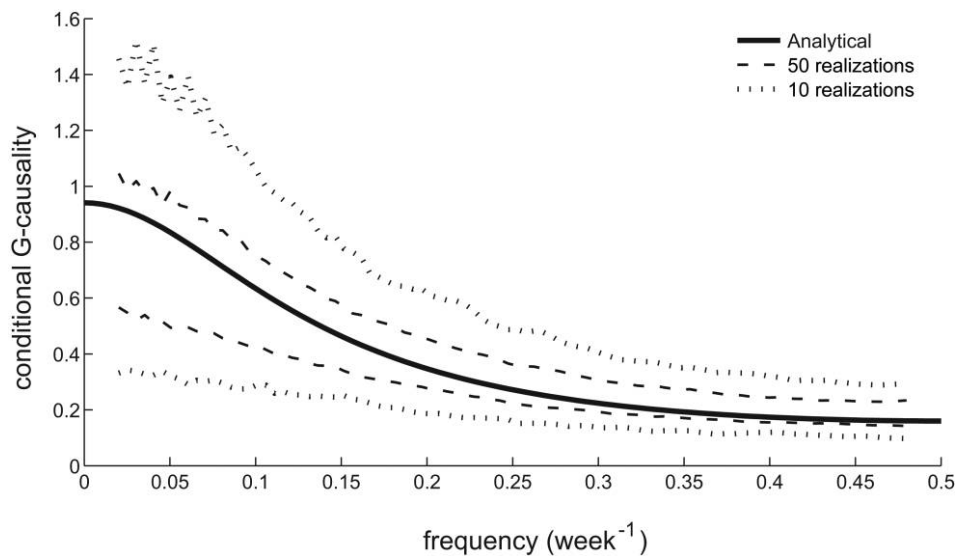


Figure 3: Analytical conditional Granger causality for the influence of the detritus on decomposers, given the effect of producers (carbon limitation hypothesis). The 95% envelopes are obtained from 1,000 simulations of the model with 50 realizations (dashed lines) and 10 realizations (dotted lines).

that canopy assimilation can modulate soil respiration (i.e., a causal link exists between these two variables) with a lag time that varies from a few hours to several days (see Kuzyakov and Gavrichkova 2010 for a review). Representing the relationship between carbon inputs and outputs in ecological models is critical to understand ecosystem carbon cycling. However, detecting causal links between photosynthesis and respiration is complicated by the interacting role of abiotic factors (temperature and water availability), as well as physical and biological CO_2 transport mechanisms. The diversity of primary producers and decomposers implies the coexistence of several compartments with different turnover rates (Loreau 2001), some of which may act in synchrony with these physical drivers.

Stoy et al. (2007) conducted a lagged cross-correlation analysis (the correlation among variables at different offset time lags) between canopy photosynthesis (A_c) estimated from eddy-covariance time series and forest floor respiration (R_s) estimated from automated chamber time series. The study was conducted in adjacent planted pine (PP) and hardwood (HW) forested ecosystems within the Duke Forest, Durham, North Carolina. A statistically significant peak in the cross-correlation function between A_c and R_s emerged at time lags between 1 and 3 days in both forest stands during summer months under finite canopy photosynthesis. The time lag was consistent with timescale arguments that consider the soil transport of CO_2 as strictly diffusive and adds a representative phloem transport time-

scale (the time required for carbon to travel from leaf to phloem to root or endomycorrhizal surface) to complete the pathway from leaf to soil to atmosphere.

Stable isotopic studies also revealed the signature of photosynthesis in soil respiration in PP at timescales of less than 1 week (Andrews et al. 1999). Separate studies have shown that most soil-respired carbon entered the ecosystem via photosynthesis no more than 1 month earlier (Taneva et al. 2006; Drake et al. 2008). Collectively, these independent studies demonstrate causal links between canopy photosynthesis and root respiration (i.e., delayed response following a carbon input) on short timescales.

When the correlation analysis in Stoy et al. (2007) was repeated in the winter months and the hardwood forest had minimal leaf area and negligible photosynthesis, apart from the presence of some evergreen species in the footprint of the tower, surprisingly similar time lags for the maximum cross-correlation between A_c and R_s emerged. It was concluded that the observed wintertime lag is due to multiple forcing functions (mainly radiation and soil temperature) rather than carbon input into the soil. When the effects of soil temperature on respiration were partially filtered (via a nonlinear filter), the maximum cross-correlation between A_c and R_s within the hardwood forest was weaker yet still maintained statistical significance. These findings motivate the development of a robust test of causality at multiple timescales where ecosystem CO_2 assimilation, transport, and respiration occur.

To illustrate the advantages of conditional G-causality in interpreting the input-output dynamics of the photosynthesis–soil respiration system, the same published data set by Stoy et al. (2007) was analyzed. The study comprises 4 years of automated chamber measurements of hourly forest floor fluxes. These measurements were taken at different locations within the stand, along with eddy-covariance measurements of water vapor and CO₂ fluxes above the forest canopy at PP and HW (for a more exhaustive description of the experiment, see also Palmroth et al. 2005; Drake et al. 2008).

The time series of R_s , A_c , and soil temperature (T_s) for each stand were divided into blocks of 30 days. The variables for each block were normalized to have zero mean and unit variance. The ensemble spectra and cospectra were then evaluated as averages in the growing season series (April–August) and non-growing season series (October–February). The analysis was carried out separately for these two periods to demonstrate the causal role of A_c on R_s , which should hold true only when leaves are present.

Figures 4 and 5 represent the results of spectral G-causality analysis for PP and HW, respectively. The ensemble spectra show a consistent peak corresponding to the daily

cycle. At the same frequency, the bivariate G-causalities are all significant. Reference to the causality of R_s on A_c is also shown to be 0 (respiration has no influence on assimilation at these scales). When conditional G-causality is computed (conditioned on T_s), the peak at 1 day⁻¹ is greatly reduced, particularly in the wintertime, when the deciduous forest is dormant.

This analysis clearly demonstrates that conventional correlation analysis or standard G-causality can fail in isolating causal linkages in multivariate systems, while conditional G-causality agrees with logical expectations derived from other experimental techniques such as stable and radio isotopes. In winter, there is little carbon input from a near-leafless HW stand, yet forest floor respiration remains appreciable, primarily due to heterotrophic and autotrophic soil respiration. After conditioning on soil temperature with the conditional G-causality approach, it becomes clear that much of this respiration is not induced by photosynthesis but by T_s , which oscillates on diurnal frequencies. Significance levels were obtained using IAAFT data of A_c series when evaluating G-causality and its conditional counterpart.

Because A_c is derived by flux separation of net ecosystem

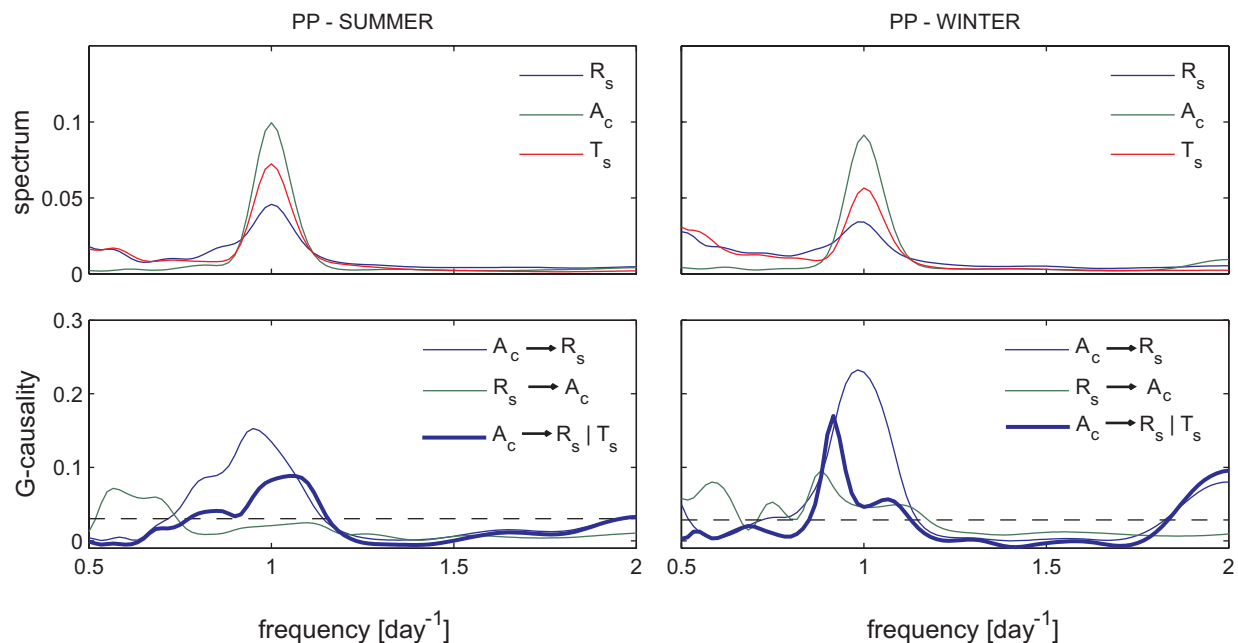


Figure 4: Granger causality (G-causality) and conditional G-causality (conditioned on soil temperature T_s) between soil respiration (R_s) and canopy assimilation (A_c) rate for the evergreen pine plantation (PP) are shown in the bottom panels, respectively. The ensemble spectra for the respective summer and winter series are shown in the top panels for reference. Dashed lines represent 95% confidence intervals around the null hypothesis (no G-causality), obtained via an ensemble of 1,000 iterative amplitude adjusted Fourier transform surrogate time series. Note that the bivariate G-causality overestimates the interaction between A_c and R_s when temperature is not taken into account.

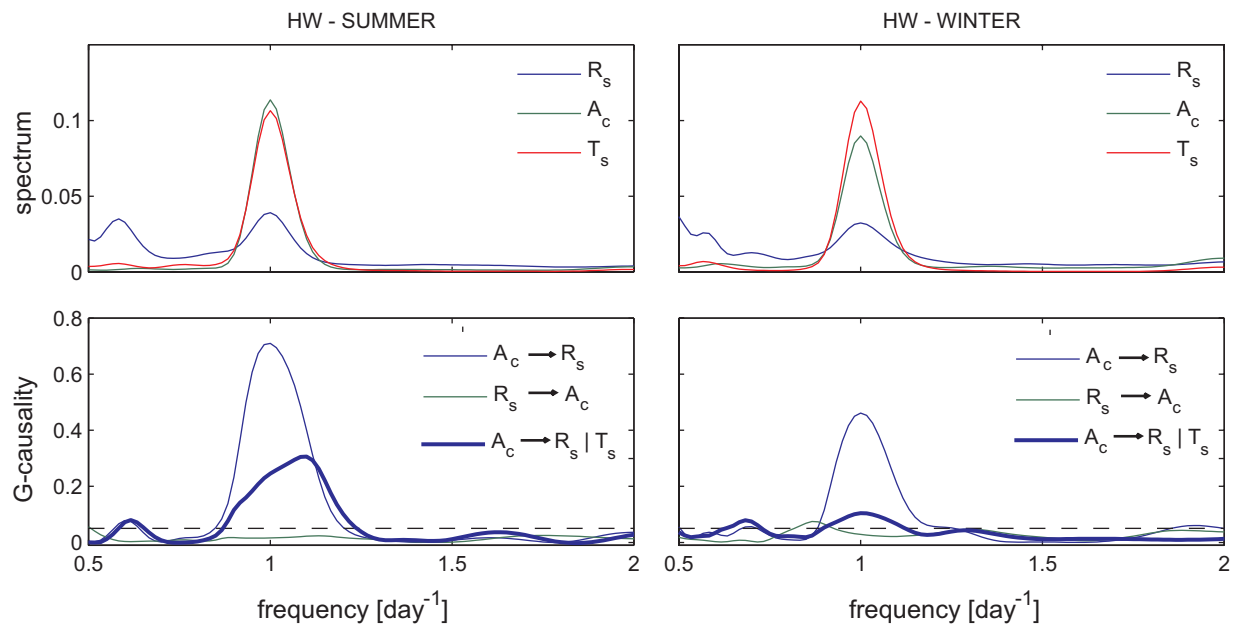


Figure 5: Same as figure 4 but for the broadleaf deciduous (hardwood [HW]) forest adjacent to the evergreen pine plantation. Note the reduced peak in the conditional Granger causality (G-causality) at 1 day^{-1} during winter months, compared to summer periods, or when compared to standard pairwise G-causality.

CO_2 exchange (Stoy et al. 2006), artificial correlations between A_c and ecosystem respiration may bring undesirable effects into their correlation even though the two measurement systems are independent (eddy covariance and automated chambers). In principle, this is not an issue for directional causality because self-correlation should manifest itself only in the interdependence (instantaneous causality). However, to exclude any doubts, the same analysis was repeated using a canopy conductance model with bulk conductance obtained independently using water vapor flux measurements from the eddy-covariance systems and vapor pressure deficit observations. Because photosynthesis is tightly coupled with bulk conductance, and assuming there is minor variation in the effective ratio of intercellular to ambient atmospheric CO_2 , then variations in bulk conductance can be used as surrogates for variation in A_c without resorting to any estimates of ecosystem respiration. The results were similar to those derived using the eddy-covariance CO_2 flux data for A_c (not shown) even though the causalities were slightly reduced. This further confirms the suitability of spectral conditional G-causality while supporting the hypothesis that relationships between inputs and outputs of carbon are causal at short timescales.

Drivers for A_c , namely, photosynthetically active radiation, generally exhibit more temporal variability than soil

temperature on such short timescales. Carbon flows within the plant and into the soil are presumably dampened (and lagged) by the presence of plant carbon pools. Nonetheless, traces of variability connected with the plant activity can fingerprint the soil processes, as evidenced in this study as well as other independent stable isotope analysis.

Discussion and Conclusions

Determining the arrow of causal linkages between two or more time-varying measures is now receiving significant attention in a number of disciplines such as economics, neuroscience, and climate science. The dynamics of many ecological systems share a number of striking similarities with these fields (marked variability over a wide range of timescales and space scales, periodic forcing imposing strong correlation but not causation among variables, etc.). Causality, as defined in Granger's (1969, 1988) work, embodies the dependence between two time series with a nonzero lag time in their cross-correlation. A spectral extension of G-causality was proposed here to determine the scale-dependent interactions among ecological variables. This spectral approach presents a decisive advantage when compared with the time-based scheme since oscillatory forcing can be identified, localized within a specific and

limited number of frequencies (e.g., seasonal, diurnal, interannual, etc.), and removed via conditional analysis. Despite these advantages, the spectral G-causality and the conditional G-causality remain underexploited tools in ecological systems (whether in models or in data). The main motivation here was to illustrate their potentials in untangling and explaining causal relationships between the different components of ecological systems.

Broadly speaking, persistence, long-term memory, or long-range dependence often characterizes ecosystem processes. This memory is encoded in the measured autocorrelation function, in the spectral density function, or in the coefficient matrix of an autoregressive process, depending on whether a parametric or a nonparametric version of G-causality is adopted. Soil respiration, like other processes in the soil system, “remembers” the characteristics of the past states. Changes in soil states are slow when compared to the aboveground system, as the processes are mediated by storage in the plant pools, diffusion in the soil-plant system, and the complex bonding structure of soil organic matter. Respiration depends not only on instantaneous combinations of soil temperature and soil moisture but also on the complex interactions between carbon pools (e.g., standing biomass and microbial biomass). At the Duke Forest sites, both bacteria and fungi are actively present. If this memory, which is clearly beyond the diurnal cycle, is persistent enough, traces of these interactions may be revealed via conditional G-causality. This topic is currently under investigation, and fortunately, the time is ripe for stepping into long-term analysis of ecological systems, given the wealth of ecological measurements now available from the FLUXNET network (Baldocchi et al. 2001a) and other similar databases. This method, as any other statistical method, does not provide any mechanistic explanation of the origin of the interactions. However, it is an invaluable tool that can be exploited to verify ecological hypotheses. Future development of this type of analysis may begin to incorporate some nonlinear causal interactions (e.g., Marinazzo et al. 2008), namely, the situation in which a forcing at a given frequency produces an effect over one or more different frequencies.

Acknowledgments

We thank M. Dhamala for useful discussions. The data collection at the Duke Forest was supported by the U.S. Department of Energy through the Office of Biological and Environmental Research Terrestrial Carbon Processes program (Free-Air CO₂ Enrichment and National Institute for Climatic Change Research grants DE-FG02-95ER62083, DE-FC02-06ER64156). M.D. and D.B. ac-

knowledge funding from the National Science Foundation (grant 0628720). G.K. and A.M. acknowledge funding from the National Science Foundation (grants NSF-EAR-10-13339, NSF-CBET-10-3347, NSF-AGS-11-02227). M.D. also acknowledges the Global Forest Carbon Research Initiative at the Center for Tropical Forest Science.

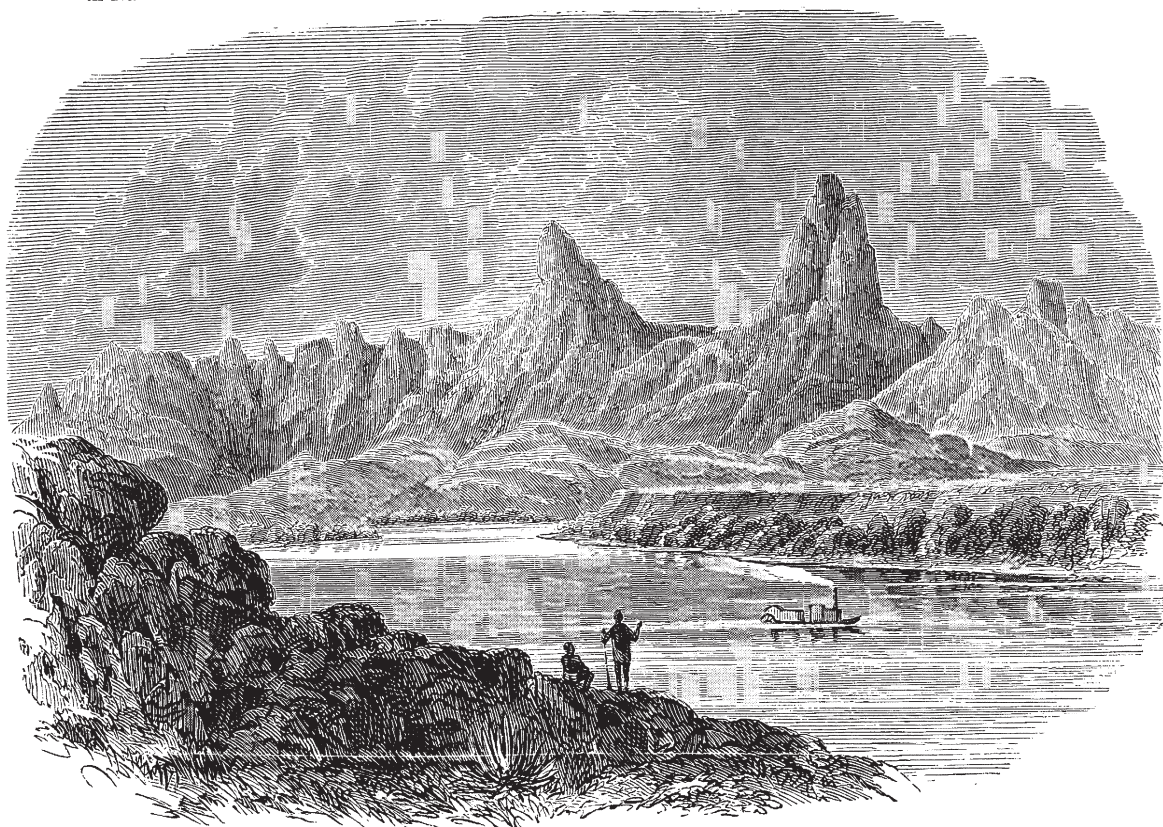
Literature Cited

- Aldrich, J. 1995. Correlations genuine and spurious in Pearson and Yule. *Statistical Science* 10:364–376.
- Andrews, J. A., K. G. Harrison, R. Matamala, and W. H. Schlesinger. 1999. Separation of root respiration from total soil respiration using carbon-13 labeling during free-air carbon dioxide enrichment (FACE). *Soil Science Society of America Journal* 63:1429–1435.
- Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, et al. 2001a. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society* 82:2415–2434.
- Baldocchi, D., E. Falge, and K. Wilson. 2001b. A spectral analysis of biosphere-atmosphere trace gas flux densities and meteorological variables across hour to multi-year timescales. *Agricultural and Forest Meteorology* 107:1–27.
- Bernasconi, C., and P. König. 1999. On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. *Biological Cybernetics* 81:199–210.
- Box, G., and G. Jenkins. 1970. *Time series analysis: forecasting and control*. Holden-Day, San Francisco.
- Brander, J. A., and M. S. Taylor. 1998. The simple economics of Easter Island: a Ricardo-Malthus model of renewable resource use. *American Economic Review* 88:119–138.
- Chen, Y. H., S. L. Bressler, and M. Ding. 2006. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *Journal of Neuroscience Methods* 150:228–237.
- Chesson, P. 1994. Multispecies competition in variable environments. *Theoretical Population Biology* 45:227–276.
- Costantino, R. F., R. A. Desharnais, J. M. Cushing, and B. Dennis. 1997. Chaotic dynamics in an insect population. *Science* 275:389–391.
- Daufresne, T., and M. Loreau. 2001. Ecological stoichiometry, primary producer-decomposer interactions, and ecosystem persistence. *Ecology* 82:3069–3082.
- Detto, M., N. Montaldo, J. D. Albertson, M. Mancini, and G. G. Katul. 2006. Soil moisture and vegetation controls on evapotranspiration in a heterogeneous Mediterranean ecosystem on Sardinia, Italy. *Water Resources Research* 42:W08419.
- Dhamala, M., G. Rangarajan, and M. Ding. 2008. Estimating Granger causality from Fourier and wavelet transforms of time series data. *Physical Review Letters* 100:018701.
- Ding, M., Y. Chen, and S. L. Bressler. 2006. Granger causality: basic theory and application to neuroscience. Pages 437–460 in B. Schelter, M. Winterhalder, and J. Timmer, eds. *Handbook of time series analysis: recent theoretical developments and applications*. Wiley, Weinheim.
- Drake, J. E., P. C. Stoy, R. B. Jackson, and E. H. DeLucia. 2008. Fine-

- root respiration in a loblolly pine (*Pinus taeda* L.) forest exposed to elevated CO₂ and N fertilization. *Plant Cell and Environment* 31:1663–1672.
- Geweke, J. 1982. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association* 77:304–313.
- Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37:424–438.
- . 1988. Some recent developments in the concept of causality. *Journal of Econometrics* 39:199–211.
- Hlavackova-Schindler, K., M. Paluš, M. Vejmelka, and J. Bhattacharya. 2007. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports* 441:1–46.
- Högberg, P., A. Nordgren, N. Buchmann, A. F. S. Taylor, A. Ekblad, M. N. Högberg, G. Nyberg, M. Ottosson-Löfvenius, and D. J. Read. 2001. Large-scale forest girdling shows that current photosynthesis drives soil respiration. *Nature* 411:789–792.
- Holmes, E. E., M. A. Lewis, J. E. Banks, and R. R. Veit. 1994. Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology* 75:17–29.
- Horwath, W. R., K. S. Pregitzer, and E. A. Paul. 1994. ¹⁴C allocation in tree-soil systems. *Tree Physiology* 14:1163–1176.
- Katul, G., C.-T. Lai, K. Schäfer, B. Vidakovich, J. Albertson, D. Ellsworth, and R. Oren. 2001. Multiscale analysis of vegetation surface fluxes: from seconds to years. *Advances in Water Resources* 24:1119–1132.
- Kaufmann, R. K., L. F. Paletta, H. Q. Tian, R. B. Myneni, and R. D. D'Arrigo. 2008. The power of monitoring stations and a CO₂ fertilization effect: evidence from causal relationships between NDVI and carbon dioxide. *Earth Interactions* 12, doi:10.1175/2007EI240.1.
- Kershaw, K. A. 1963. Pattern in vegetation and its causality. *Ecology* 44:377–388.
- Kuzakov, Y., and O. Gavrichkova. 2010. Time lag between photosynthesis and CO₂ efflux from soil: a review. *Global Change Biology* 16:3386–3406.
- Loreau, M. 2001. Microbial diversity, producer-decomposer interactions and ecosystem processes: a theoretical model. *Proceedings of the Royal Society B: Biological Sciences* 268:303–309.
- Lungarella, M., K. Ishiguro, Y. Kuniyoshi, and N. Otsu. 2007. Methods for quantifying the causal structure of bivariate time series. *International Journal of Bifurcation and Chaos* 17:903–921.
- Marinazzo, D., M. Pellicoro, and S. Stramaglia. 2008. Kernel method for nonlinear Granger causality. *Physical Review Letters* 100:144103.
- Mayr, E. 1960. Cause and effect in biology. *Science* 134:1501–1506.
- Molini, A., G. G. Katul, and A. Porporato. 2010. Causality across rainfall timescales revealed by continuous wavelet transforms. *Journal of Geophysical Research* 115:D14123.
- Nisbet, R. M., and W. S. C. Gurney. 1982. *Modelling fluctuating populations*. Wiley, Chichester.
- Palmroth, S., C. A. Maier, H. R. McCarthy, A. C. Oishi, H.-S. Kim, K. H. Johnsen, C. G. Katul, and R. Oren. 2005. Contrasting responses to drought of forest floor CO₂ efflux in a loblolly pine plantation and a nearby oak-hickory forest. *Global Change Biology* 11:421–434.
- Paluš, M. 2008. From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics. *Contemporary Physics* 48:307–348.
- Paluš, M., and M. Vejmelka. 2007. Directionality of coupling from bivariate time series: how to avoid false causalities and missed connections. *Physical Review E* 75:056211.
- Papana, A., D. Kugiumtzis, and P. G. Larsson. 2011. Reducing the bias of causality. *Physical Review E* 83:036207.
- Pascual, M., and S. P. Ellner. 2000. Linking ecological patterns to environmental forcing via nonlinear time series models. *Ecology* 81:2767–2780.
- Pearl, J. 2009. *Causality: models, reasoning, and inference*. 2nd ed. Cambridge University Press, Cambridge.
- Peters, R. H. 2001. *A critique for ecology*. Cambridge University Press, Cambridge.
- Petit, J. R., J. Jouzel, D. Raynaud, N. I. Barkov, J.-M. Barnola, I. Basile, M. Bender, et al. 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* 399:429–436.
- Petratis, P. S., A. E. Dunham, and P. H. Niewiarowski. 1996. Inferring multiple causality: the limitations of path analysis. *Functional Ecology* 10:421–431.
- Platt, T., and K. L. Denman. 1975. Spectral analysis in ecology. *Annual Review of Ecology and Systematics* 6:189–210.
- Plowright, R. K., S. H. Sokolow, M. E. Gorman, P. Daszak, and J. E. Foley. 2008. Causal inference in disease ecology: investigating ecological drivers of disease emergence. *Frontiers in Ecology and the Environment* 6:420–429.
- Priestley, M. B. 1981. *Spectral analysis and time series*. Academic Press, London.
- Ranta, E., V. Kaitala, and P. Lundberg. 1998. Population variability in space and time: the dynamics of synchronous population fluctuations. *Oikos* 83:376–382.
- Raupach, M. R. 2007. Dynamics of resource production and utilization in two-component biosphere-human and terrestrial carbon systems. *Hydrology and Earth System Sciences* 11:875–889.
- Rietkerk, M., S. C. Dekker, P. C. de Ruiter, and J. van de Koppel. 2004. Self-organized patchiness and catastrophic shifts in ecosystems. *Science* 305:1926–1929.
- Roughgarden, J. 1975. Simple model for population dynamics in stochastic environments. *American Naturalist* 109:713–736.
- Roux, S. G., V. Venugopal, K. Fienberg, A. Arneodo, and E. Foufoula-Georgiou. 2009. Evidence for inherent nonlinearity in temporal rainfall. *Advances in Water Resources* 32:41–48.
- Royama, T. 1992. *Analytical population dynamics*. Chapman & Hall, London.
- Salvucci, G. D., J. A. Saleem, and R. Kaufmann. 2002. Investigating soil moisture feedbacks on precipitation with tests of Granger causality. *Advances in Water Resources* 25:1305–1312.
- Schreiber, T., and A. Schmitz. 2000. Surrogate time series. *Physica D* 142:346–382.
- Shaw, M. R., E. S. Zavaleta, N. R. Chiariello, E. E. Cleland, H. A. Mooney, and C. B. Field. 2002. Grassland responses to global environmental changes suppressed by elevated CO₂. *Science* 298:1987–1990.
- Shipley, B. 2000. *Cause and correlation in biology: a user's guide to path analysis, structural equations, and causal inference*. Cambridge University Press, Cambridge.
- Smirnov, D. A., and I. I. Mokhov. 2009. From Granger causality to long-term causality: application to climatic data. *Physical Review E* 80:016208.
- Stenseth, N. C., A. Mysterud, G. Ottersen, J. W. Hurrell, K.-S. Chan, and M. Lima. 2002. Ecological effects of climate fluctuations. *Science* 297:1292–1296.

- Stoy, P. C., G. G. Katul, M. B. S. Siqueira, J.-Y. Juang, K. A. Novick, and R. Oren. 2006. An evaluation of methods for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration. *Agricultural and Forest Meteorology* 141: 2–18.
- Stoy, P. C., S. Palmroth, A. C. Oishi, E. Ward, M. B. S. Siqueira, J.-Y. Juang, K. A. Novick, K. Johnsen, G. G. Katul, and R. Oren. 2007. Are ecosystem carbon inputs and outputs coupled at short timescales? a case study from adjacent pine and hardwood forests using impulse-response analysis. *Plant Cell and Environment* 30: 700–710.
- Taneva, L., J. S. Phippen, W. Schlesinger, and M. A. Gonzalez-Meler. 2006. The turnover of carbon pools contributing to soil CO₂ and soil respiration in a temperate forest exposed to elevated CO₂ concentration. *Global Change Biology* 12:983–994.
- Wilson, G. T. 1972. Factorization of matricial spectral densities. *SIAM Journal on Applied Mathematics* 23:420–426.

Associate Editor: Benjamin M. Bolker
Editor: Ruth G. Shaw



CHIMNEY PEAK.

“The north-western prolongation of the Purple Hills, which form the first three cañons in the river. The two pinnacles of ‘Chimney Peak,’ looming up in the background, are composed of trap. This being much harder than the material of the neighboring rocks has yielded less to the action of the elements,” from “The Chasms of the Colorado,” by A. Hyatt (*American Naturalist*, 1868, 2:359–365).