

SYMPOSIUM

CAUSALITY FROM PALAEOONTOLOGICAL TIME SERIES

by BJARTE HANNISDAL^{1,2,3}  and LEE HSIANG LIOW^{4,5} ¹Department of Earth Science, University of Bergen, PO Box 7803, 5020, Bergen, Norway; bjarte.hannisdal@uib.no²K.G. Jebsen Centre for Deep Sea Research, University of Bergen, Norway³Bjerknes Centre for Climate Research, University of Bergen, Norway⁴Centre for Ecological & Evolutionary Synthesis, Department of Biosciences, University of Oslo, PO Box 1066 Blindern, 0316, Oslo, Norway; l.h.liow@ibv.uio.no⁵Natural History Museum, University of Oslo, Norway

Typescript received 29 January 2018; accepted in revised form 5 April 2018

Abstract: As custodians of deep time, palaeontologists have an obligation to seek the causes and consequences of long-term evolutionary trajectories and the processes of ecosystem assembly and collapse. Building explicit process models on the relevant scales can be fraught with difficulties, and causal inference is typically limited to patterns of association. In this review, we discuss some of the ways in which causal connections can be extracted from palaeontological time series and provide an overview of three recently developed analytical frameworks that have been applied to palaeontological

questions, namely linear stochastic differential equations, convergent cross mapping and transfer entropy. We outline how these methods differ conceptually, and in practice, and point to available software and worked examples. We end by discussing why a paradigm of dynamical causality is needed to decipher the messages encrypted in palaeontological patterns.

Key words: time series, dynamical systems, Granger causality, stochastic differential equations, convergent cross mapping, transfer entropy.

WE INHABIT a world fundamentally shaped by the deep history of geobiological interactions. Long-term ecological and evolutionary dynamics coupled with geological processes have made our planet a vast reservoir of living and fossil biomass that is literally fuelling our modern civilization, and the rapid depletion of this biomass holds dire consequences for humankind (Schramski *et al.* 2015). As modern ecosystems head towards uncharted territories, geohistorical data are our only record of ecosystems undisturbed by human activities and of biotic responses to global change in the past (Dietl *et al.* 2015). Palaeontologists thus have an obligation not only to document the rich complexity of life's history, but also to extract from this complexity the causal structures that govern its dynamics (Fig. 1).

The recent growth of large-scale global and regional palaeontological data compilations, as well as high-resolution examination of local stratigraphic sections, has provided momentum for palaeontologists to rise to this challenge. Increasingly comprehensive records of taxonomic richness, abundance and phenotypic variability, combined with improved chronological and phylogenetic constraints, yield time series that can capture the

dynamics of underlying biological processes, and be meaningfully compared with other deep-time records of environmental change. We will argue that these time series beg for causal explanations that are not restricted to unique events, but accommodate dynamical, temporally extensive modes of causality.

Although a strong interest in using quantitative approaches to understand temporal trends in palaeontology can be traced back to Simpson (1944) and earlier (Brinkmann 1929), we believe that the first person to introduce our field to formal time series analyses was David Raup. While his collaboration with Gould, Schopf and Simberloff on stochastic evolution was ongoing (Raup *et al.* 1973), he proposed a more formalized use of stochastic time-series models (random walks) to study evolutionary time series (Raup 1977a). These models became the canonical null hypotheses against which presumed evolutionary patterns were judged (Raup & Crick 1981; Bookstein 1987; Roopnarine *et al.* 1999; Hannisdal 2006). Today, the use of stochastic time series models of phenotypic evolution has shifted from null hypothesis testing to parameter estimation and model selection (Hunt 2006; Hannisdal 2007; Hunt & Rabosky 2014). If a time series of

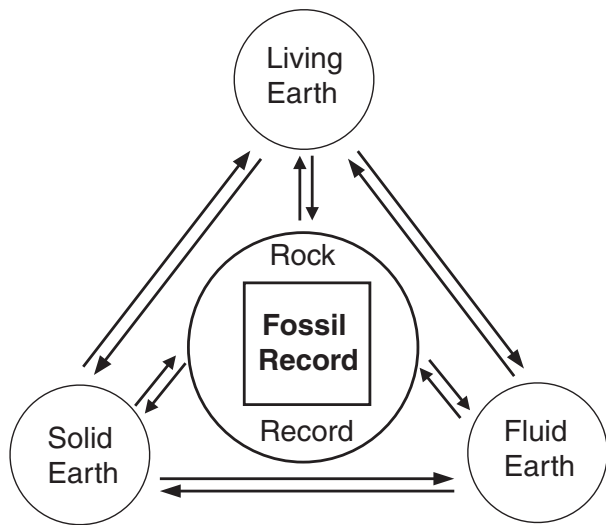


FIG. 1. The multiple causes of the fossil record. Schematic representation of the multi-directional relationships between living, fluid and solid earth components that combine to create the fossil record. Fluid earth includes atmosphere and hydrosphere. Note that while the rock record is viewed as a physical object with mass and chemical composition that interacts with the other active components (denoted by circles), the fossil record is here viewed as a subset of the rock record. Adapted from Peters & Heim (2011).

phenotypic variability in fossil organisms is compared to a group of different models, then a better fit to a specific model is usually interpreted as statistical support for the (implicitly causal) mechanism implied by the preferred model.

Raup & Sepkoski (1984) also pioneered the use of Fourier analysis to study the frequency power spectrum of extinction intensities, concluding that there is a *c.* 26 million-year periodicity in extinction at the family level. They speculated on astronomical and astrophysical causes of the periodicity (and therefore of the extinctions), because ‘purely biological or earthbound physical causes seem incredible’. Their paper was promptly followed by Kitchell & Pena (1984) who used autoregressive integrated moving average (ARIMA) models to infer a dominant periodicity of about 31 million years with substantial uncertainty. Kitchell & Pena stressed that ‘periodicity in a time series is insufficient evidence that a periodic external force is causally responsible’, not least because in natural systems, (quasi-)periodicity can arise through internal dynamics without periodicity in the external forcing, like in the case of predator–prey cycles. The existence of a significant periodicity in the Phanerozoic extinction record and the extent to which matching frequencies in other processes (astronomical or Earth bound) imply a causal connection remain topics of continued debate in the current palaeontological literature

(Melott & Bambach 2017; Erlykin *et al.* 2018). Similar controversies surrounding the causal implications of matching periodicities are simmering in related fields, such as the existence of Milankovitch cycles in Palaeozoic sedimentary sequences (Hinnov *et al.* 2016; Smith *et al.* 2016) or in seafloor bathymetry (Huybers *et al.* 2016; Olive *et al.* 2016). One way to break the impasse, is to confront our time series with methods that seek to distinguish causative from correlative relationships.

From prediction to causality

Causality is a slippery concept that is widely discussed across all scientific disciplines (e.g. Hill 1965; Winship & Morgan 1999; Pearl 2009; Illari *et al.* 2011). For instance, the search term ‘causality and science’ returned more than 300 articles in the period 2008–2017 from the journal *Philosophy of Science* alone (averaging six articles per issue). We will not attempt to summarize the history of the causality concept here, but we note that the criteria for identifying causal relationships have been continuously subject to revision in the sciences. For instance, Karl Pearson simply equated causation with correlation in the form of contingency tables (Pearson 1911). Hill (1965) famously discussed nine minimal criteria for empirically identifying causal agents of disease that were the benchmark for epidemiologists for decades. Today, only one of these criteria, temporal precedence (i.e. the disease-causing agent needs to occur before the disease), is generally accepted in its original form (Fedak *et al.* 2015).

Controlled experiments, in which the effect of a single factor is isolated and tested by direct manipulation, are considered to be the gold standard of scientific causal assessment. For a system of the kind outlined in Figure 1, however, experiments are confined to numerical representation of processes *in silico*. For example, Earth system models used in near-term climate projections are arguably the most sophisticated tools available for numerical experiments (Taylor *et al.* 2012). Yet, despite the vast complexity of these models, the inclusion of key biological processes is still in its infancy (e.g. Wieder *et al.* 2015) and experiments are currently limited to millennial time scales or shorter. Models of intermediate complexity are better suited for exploring mechanistic hypotheses in deep time (e.g. Meyer *et al.* 2016). Explicit process models, including highly idealized mathematical representations, will undoubtedly play an increasingly important role in causal assessment in palaeontology.

Ultimately, however, process models need to specify a model structure assumed to plausibly represent causal mechanisms, which creates a vulnerability to model misspecification that can be difficult to assess (Wood & Thomas 1999; Babbie *et al.* 2014). On the spectrum between

explicit process models and statistical analysis of patterns of correlation, there is room for concepts that use either stochastic parameterizations or an entirely equation-free technique to detect causal connections directly from observed time series. In the following, we present three such methods that are amenable to palaeontological time series analyses, namely: (1) linear stochastic differential equations; (2) convergent cross mapping; and (3) transfer entropy. Before we present the nuts and bolts of these techniques, it is necessary to provide a brief introduction to two underlying concepts from the field of time series analysis: Granger causality and dynamical system reconstruction.

Granger causality rests on the idea that a cause both precedes and helps predict its effect. Inspired by earlier theories of prediction, Granger (1963, 1969) implemented this idea in the context of econometric time series analysis. His original formulation was in terms of linear vector autoregressive models, in which the value of a variable at a given time is modelled as a linear weighted sum of a set of values from its own past, and from the past of a set of other variables. Each variable is a time series representing a stochastic process, and the task is to find the weights that minimize the prediction error. In the classical sense, a variable X is said to ‘Granger cause’ Y if our ability to predict Y declines when X is removed from the set of predictive variables. Granger causality is a special case of transfer entropy (see Transfer Entropy, below) and the two are equivalent for Gaussian variables (Barnett *et al.* 2009).

It is not uncommon to explore the ‘cause precedes effect’ criterion by performing time lagged cross-correlation analyses on a pair of time series, where the lagged series is interpreted as the effect if a significant cross correlation is found between the two offset time-series. Such an approach may have levels of false positives and false negatives that are unacceptable (Liow *et al.* 2015) due to mismatches between the underlying processes and the implicit assumptions inherent in the lagged analyses performed.

In the context of time series analysis, (lagged) correlations and associated regression analysis need to correct for serial (or auto-)correlation, which is one of the problems Granger helped elucidate with his work on cointegration, unit root tests and causality. We do not intend to review the use of classical methods like regression analysis here (see Brown *et al.* 2011 for a review of their use in the climate change ecology literature). We note, however, that the development of many standard statistical methods, including regression, was motivated by a classical experimental setting in which the cause–effect mechanism is known and regression coefficients are intended to measure (typically linear) effect size. The methods reviewed here are meant to expand our toolkit to handle more general systems where mechanisms may be unknown and experiments are not feasible.

The criterion of improved prediction also has its limitations. For example, if watching the weather report helps us to predict the weather, then it Granger-causes the weather, even if it is obviously not a true cause. In addition, Granger’s definition makes the unrealistic assumption that all relevant variables are included in the analysis. A variety of extensions and modifications of the original definition have since been proposed (e.g. Chen *et al.* 2004; Eichler 2005; Barrett *et al.* 2010; Barnett & Seth 2015) including the much more flexible linear stochastic differential equation framework described below, which permits the identification of unobserved (latent) confounding factors and feedback.

Linear systems have the distinct advantage that they can be broken down into parts that can be analysed separately and then recombined. Matters become worse if there is strong nonlinearity, which is ubiquitous in nature and in our everyday experience. In a system where components are interdependent, such as ecosystems or Earth systems (Fig. 1), variables can interact in such a way that any linear correlations between them can change as the state of the system changes. Small perturbations at tipping points can be associated with sudden regime shifts in the response of ecosystems to environmental change (Burkett *et al.* 2005) and in the dynamics of speciation (Nosil *et al.* 2017). May (1976) alerted biologists to the fact that even the simplest mathematical relations between two species can yield highly complex, nonlinear and unpredictable dynamics. Most nonlinear dynamical systems cannot be solved analytically, but a geometric approach can be used to characterize their properties. From a dynamical systems perspective, predictive causality is derived from the geometry of the dynamics, which can be reconstructed without accounting for all the component variables constituting the whole system. To provide some context for the two equation-free methods described herein, we need to highlight a few key concepts.

In dynamical systems theory, the state of the system is described by coordinates in a state space (or phase space for continuous-time systems). The number of interacting components in the system determines the dimensionality of its state space, and the time evolution of the system forms trajectories in this space. In an open, dissipative system that exchanges energy and matter with its environment, any dynamics will cease unless there is some driving force. Over a sufficient period of time, the driving force and the dissipation will tend to balance, and the system will settle on a typical behaviour. This typical behaviour, towards which the system ‘gravitates’, is confined to a subset of the state space known as an attractor. Examples of attractors can be stable fixed points (e.g. a pendulum with friction will converge on a position of rest), limit cycles (e.g. predator–prey oscillations) or more complex subsets such as the strange attractors of chaotic

systems (e.g. the famous ‘butterfly’-shaped attractor of the Lorenz system). The attractor is itself an invariant subset of the dynamics: if we pick any point on the attractor and follow its evolution under the dynamics of the system, the resulting trajectory lies entirely on the attractor. This implies that to study the long-term behaviour of the system we can seek to reconstruct the attractor, rather than the entire set of unobserved (latent) variables constituting the whole system. State-space (attractor) reconstruction is a starting point for the two equation-free time series analysis approaches outlined below.

Despite their conceptual and practical differences, all three methods discussed in this review are ultimately rooted in the notion of prediction improvement as a statistical way of making inroads to causality. They all begin with measurements (assumed to represent variables that capture unobserved processes of interest) which are arranged as time series data that are then subject to statistical analyses (Fig. 2). Other approaches exist that are based on a stronger form of causality. The general framework of structural equation modelling (Shipley 2016) pioneered by Sewall Wright (1921) has been formalized by Pearl (2009) as model-based causal inference on graphs. This approach rests on philosophical connections between causality and probability (Spirtes *et al.* 2000) and on the notion of intervention (Woodward 2003), and involves the mathematical machinery of a causal calculus on directed acyclic graphs and Bayesian inference (Pearl 2009).

The causal graph formalism is widely used for modelling causality, especially in the social sciences, and puts special emphasis on interventions. For example, watching the weather report may Granger-cause the weather, but through intervention we could actively manipulate the weather forecast to discover that it does not cause the weather. Concerns have been raised, however, that the causal graph formalism is less well-suited for explaining spatially and structurally complex biological phenomena (Kaiser 2016), and that the interventionist approach is not applicable to dynamical systems in general (Weber 2016). Further discussion of causal graph methods is outside the scope of this review.

Some of the key features of the methods described below are summarized in Table 1. These features are subject to change because all three methods are topics of active research. We invite the reader to explore the methods on dedicated websites that provide software, detailed user manuals and worked examples cited below.

TIME-SERIES APPROACHES FOR CAUSAL INFERENCE

Linear stochastic differential equations

Linear stochastic differential equations (SDEs) are applied to (palaeo)biological questions most commonly in the form of an Ornstein–Uhlenbeck (OU) process for

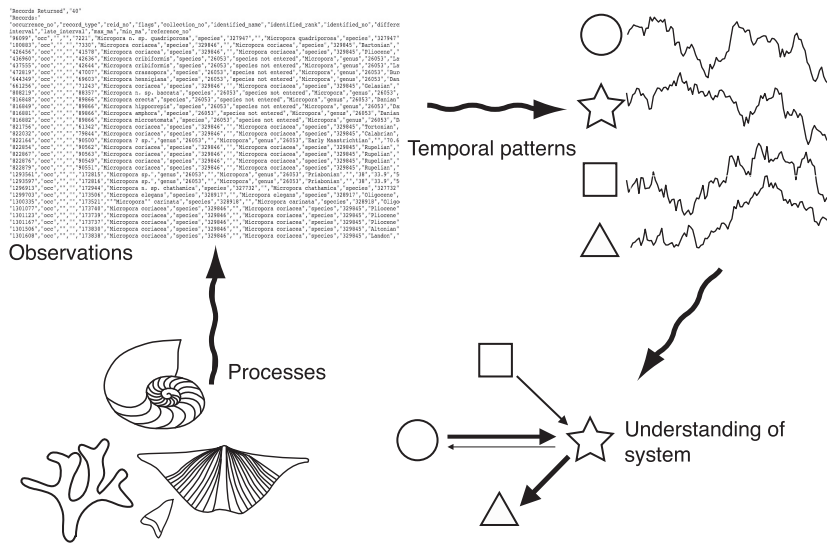


FIG. 2. Extracting causal structures from palaeontological time series. Ecological and evolutionary processes interact with environmental dynamics to produce a fossil record (represented in the diagram by iconic fossils), which we then sample. Selected measurements (shown as output from a database), together with chosen environmental data are then arranged in temporal sequence, as time series. To avoid any subliminal ranking, we denote variables by generic symbols. If these time series are sufficiently informative, we can use this inherent dynamical information to characterize the strength and directionality of causal connections between the variables, represented as a causal diagram. Straight arrows between the symbols indicate causal directionality and the thickness of the arrows indicates the strength of the relationships.

TABLE 1. Key features of the three causal time series analysis methods described here.

	General approach	Unobserved processes	Causality	Statistical inference	Temporal sampling	Uncertainty	Non-stationarity	Number of variables
SDE	Parametric modelling	Latent process modelling	Granger; explicit time	Bayesian	(Ir)regular (Not) contemporaneous	Explicitly modelled	Explicitly modelled	Extendible to multiple variables
CCM	Dynamical system reconstruction	Implicit (embedding theorem)	Reconstructed states of response predict states of driver	Surrogate testing	(Ir)regular Contemporaneous	Monte Carlo resampling	Implicit	Bivariate, pairwise analyses
TE	Dynamical system reconstruction	Implicit (embedding theorem)	Directed information flow	Surrogate testing	(Ir)regular Contemporaneous	Monte Carlo resampling	Transfer operator approximation*	Generalizable to multiple variables

*Applies to the specific TE implementation discussed here.

describing time series of phenotypes (e.g. Lande 1976; Hunt 2006). In the last few years, SDEs have also been used in palaeontology to model systems within an extended Granger causal framework.

To explain how SDEs can be used as a causal inference tool, we first briefly describe the structure of a basic linear SDE. An SDE is simply a differential equation in which at least one term is a stochastic process. We can write a basic linear SDE in this form:

$$dX(t) = -\alpha(X(t) - \mu)dt + \sigma dB(t) \quad (1)$$

where $X(t)$ is the process of interest (dependent variable) that changes with the independent variable time t . The first part of the right-hand side of the equation, $\alpha(X(t) - \mu)dt$, describes the deterministic part of how X changes with time. The second part of that, $\sigma dB(t)$ is the stochastic component. α and μ are parameters that characterize the deterministic part of the process while σ describes the stochastic part. α , μ and σ are often assumed to be constant in biological applications, but they need not be. If the deterministic component is absent, such that $dX(t) = \sigma dB(t)$, and σ is constant, the expression is reduced to a random walk (commonly referred to as Brownian motion, or a Wiener process). Where α , μ and σ are all constant, Equation 1 represents an OU model which, roughly speaking, describes a random walk with a tendency to move towards a long-term average (μ). For instance, if X represents body size, and the left-hand side of Equation 1 reflects how fast body size is changing, then α indicates how strongly body size is ‘pulled’ towards the average body size μ , while σ is the standard deviation of the stochastic changes in body size.

To illustrate how we can use SDEs to examine whether there is a causal (Granger) relationship between two time-series, let’s consider a hypothesis. Say that we would like to test if sea surface temperature (SST) has influenced body size in a taxon of interest. We need to have time series of both a SST proxy (SST) and size measurements (Size) for our taxon that overlap substantially in time (but they don’t necessarily have to be measured at the same time points).

We could describe SST as

$$dSST(t) = -\alpha_{SST}(SST(t) - \mu_{SST})dt + \sigma_{SST}dB_{SST}(t) \quad (2)$$

and Size as

$$dSize(t) = -\alpha_{Size}(Size(t) - \mu_{Size})dt + \sigma_{Size}dB_{Size}(t) \quad (3)$$

by simply substituting X in Equation 1 by SST or Size, and putting SST and Size subscripts on these equations to signify that these parameters are specific to SST and Size.

As such, there is no modelled relationship between SST and Size. However, we can imagine that SST and Size are

correlated or even causally related. To model a correlative relationship, we can write the pair of equations

$$dSST(t) = -\alpha_{SST}(SST(t) - \mu_{SST})dt + \sigma_{SST}dB_{SST}(t)$$

(which is exactly the same as Equation 2)

and

$$dSize(t) = -\alpha_{Size}(Size(t) - \mu_S)dt + \sigma_{Size}(1 - \rho)^{0.5}dB_{Size}(t) + \rho\sigma_{Size}dB_{SST}(t) \quad (4)$$

In Equation 4, there is an extra term ρ which describes the temporal correlation between SST and Size in the stochastic part of the equation. Note also that we could have written the same for Equation 2 and the described relationship between SST and Size would be the same.

To explicitly model a Granger causal relationship between SST and Size where SST drives Size we retain Equation 2 and write for Size,

$$dSize(t) = -\alpha_{Size}(Size(t) - \mu_{Size} - \beta[SST(t) - \mu_{SST}])dt + \sigma_{Size}dB_{Size}(t) \quad (5)$$

The deterministic part of the SST equation has entered the deterministic part of the Size equation such that Size now depends on SST. The stochastic part of the equation is unchanged. If we wanted to include the reversed directionality of causality, we would need to write an equivalent equation for SST (which, unlike in the correlative equation, is not symmetrical). Note that no causal link from another time series can be modelled if the putative ‘response’ time series is purely stochastic. We can then compare the different models for SST and Size using a chosen model comparison approach.

Some advantages of SDE for palaeontological data are the following. First, because continuous processes are modelled, there is no need for observed time series to be measured at equidistant or the same time points. Second, there is no need to detrend the data series given that stationarity is not assumed in SDE (e.g. random walks are not strictly stationary). Third, SDEs allow for the modelling of latent variables that are not observed. Fourth, feedback loops, that is causal arrows in both directions, are permissible. These last two features explicitly relax two assumptions of the original formulation of Granger causality. Fifth, uncertainty in measurements can be included in parameter estimation. In addition, because SDE is parametric, it allows us to characterize the properties of stochastic processes and the timescale on which one time-series responds to another.

However, linear SDEs are unsuited to systems that are highly non-linear. They are also in general not amenable to studying counts (e.g. species richness) unless changes

in species richness can be justified as proxies of rates of diversification, extinction or other such continuous-valued quantities that reflect underlying processes. Observed time series need to have enough data points to be informative, but we refrain from specifying the minimum data requirements for any of the methods, because these requirements depend on the ratio of signal to noise, and on the type of dynamics of the underlying processes. Even though it is possible to go beyond pairwise studies of time series in SDE, analysing systems with more than three time-series can be quite expensive computationally. In addition, like any other approach that utilizes model comparison, but especially so because causal hypotheses are involved, the relative best fit of models needs to be interpreted with caution, as even the best model in a limited pool of models may have poor absolute fit to the data (Pennell *et al.* 2015; Voje *et al.* 2018). SDEs, as currently implemented, cannot accommodate substantial changes in the causal relationship between variables through time. If there is evidence to suggest that relationships do not remain the same over time, then this hypothesis could be tested with SDEs by dividing the time series into two or more segments.

The linear stochastic differential equation approach to causal inference has already seen a number of palaeontological applications. SDE has been used to show that coccolithophore size evolution in the Cenozoic was not driven by global temperature changes but was characterized by local spatial dynamics (Reitan *et al.* 2012). The SDE method has also been applied to the classic ‘ships passing in the night’ question of bivalve versus brachiopod diversification dynamics in the Phanerozoic, where it was shown that bivalve extinction rates drove brachiopod origination rates (Liow *et al.* 2015; Reitan & Liow 2017). Recently, it has also been used in parallel with the equation-free methods described in the following sections to show that the observed global occupancy of planktonic foraminifera has been dynamically coupled to past oceanographic changes captured in deep-ocean temperature reconstructions for the Cenozoic (Hannisdal *et al.* 2017). Code written in R and C to run linear SDEs in causal inference is currently available at <http://folk.uio.no/trondr/layered/> but a fully documented R package is forthcoming.

Convergent cross mapping

As noted above, to study the long-term behaviour of a deterministic dynamical system we may characterize the state space (attractor), rather than the entire set of unobserved variables constituting the whole system. Attractor reconstruction is often the first step in non-linear time series analysis, estimation of invariants (see Transfer

Entropy, below) and forecasting/prediction. It was shown in 1980 that for a dynamical system of multiple coupled components, a time series of any single component can be used to obtain a faithful representation of the dynamics of the whole system, by combining the time series with lagged instances of itself (Packard *et al.* 1980; Takens 1981).

If a time series $X = \{x(t_1), \dots, x(t_n)\}$ is a component of a dynamical system, we can reconstruct the state space of the system by constructing m -dimensional time-delay embedding vectors

$$E_X = \{[x(t_i), x(t_i - \tau), x(t_i - 2\tau), \dots, x(t_i - m\tau)]\} \quad (6)$$

for all $i \in [1, \dots, n - m\tau]$

where $x(t_i)$ is the scalar value of the time series X at time t_i , and τ is the embedding lag. The embedding lag is only relevant for computational purposes and is unrelated to any true causal delay in the dynamics. For a dynamical system of unknown dimensionality, the optimal embedding dimension m needs to be estimated (Sauer *et al.* 1991). Loosely speaking, the Takens embedding theorem says that there is a one-to-one mapping between the reconstructed states in E_X and the states of the whole system (Takens 1981; Sauer *et al.* 1991). The reconstructed attractor thus preserves the dynamics of the system. If two variables X and Y are coupled components of the same dynamical system (i.e. they are causally connected) then the one-to-one mapping between a reconstruction E_Y and the system attractor is also true for a time series of Y . Consequently, E_X and E_Y will be in one-to-one correspondence with each other, which can serve as a criterion for causal detection.

Attractor reconstruction using delay embeddings has found many applications in nonlinear time series analysis (Kantz & Schreiber 2003). Sugihara *et al.* (2012) introduced convergent cross mapping (CCM) as a non-parametric statistical approach to detecting causality between two observed time series. The CCM method estimates dynamical coupling strength by quantifying the extent to which the dynamics of the driver time series are encoded in a time-delay embedding of the putative response variable.

The CCM algorithm iteratively draws a subset of a time series Y (the prediction set) and finds, for each point P_i in the prediction set, the corresponding subset L_i in the reconstructed state space E_X (the library set). Next, it locates the nearest neighbours of L_i in E_X and uses a geometric projection technique (Sugihara & May 1990) to compute a predicted value P_i^* . The closeness of prediction, or cross-mapping skill, is measured by the Pearson correlation between the predicted value P_i^* and the observed value P_i . The convergence part of CCM comes from the expectation that, if the variables belong to the

same dynamical system, then the cross-mapping skill should increase and converge with increasing size of the library set used in the reconstruction.

In contrast to Granger causality approaches, which use past values of the forcing variable to predict future values of the response variables, CCM asks whether the reconstructed states of the response variable can be used to predict the states of the forcing variable. Hence, the notation ‘ X xmap Y ’ refers to estimating $y(t_i)$ from the corresponding delay-embedding reconstruction of $x(t_i)$, which can be interpreted as ‘ Y is causally influencing X ’. In CCM, causal directionality is detected by assessing whether prediction skill from X to Y is greater than vice versa.

State-space reconstruction requires a sufficient embedding dimension m (the number of lagged instances included in the embedding vector), and the CCM algorithm selects the embedding parameters by optimizing self-prediction using a geometric projection technique (Sugihara & May 1990). See Haaga *et al.* (2018) for more details.

In systems with strong unidirectional forcing, the response variable can be fully encoded in the forcing variable (‘synchronization’), and cross mapping will be observed in both directions, even if true causality is unidirectional. To address this problem, Ye *et al.* (2015a) introduced a lagged CCM to distinguish synchronization from true two-way causality. In practice, beyond the sign of the peak lag, current implementations of CCM cannot reliably detect the absolute magnitude of a causal time delay, and the peak lag may not be unique if the forcing variable is strongly periodic. Instead, Haaga *et al.* (2018) proposed to integrate CCM skill over a window of positive and negative lags, thus sacrificing absolute time delay information for the sake of obtaining a more robust detection of causal directionality. Indeed, several aspects of the CCM approach to causality are topics of lively discussion, including different ways of subsampling the data to construct libraries (Luo *et al.* 2015; Ye *et al.* 2015b), the appropriate surrogate data for significance assessment (Baskerville & Cobey 2017; Sugihara *et al.* 2017; Haaga *et al.* 2018) and the ability to infer absolute causal time delays (van Nes *et al.* 2015; Ye *et al.* 2015a; Coufal *et al.* 2017; Haaga *et al.* 2018).

CCM has been applied to records in which samples are not evenly spaced in time (e.g. van Nes *et al.* 2015) but the state-space cross-mapping concept demands that the variables being compared are sampled at contemporaneous points, which might necessitate bin-averaging, interpolation, or imputation of data. CCM does not assume stationarity, hence there is no need to detrend the observed time series. Measurement errors, sampling noise, or chronological uncertainty have to be dealt with by Monte Carlo iterative analyses, which can be computationally costly. CCM is only defined for the bivariate case,

but, in principle, transitive multivariate causal chains can also be resolved through pairwise analyses (Ye *et al.* 2015a). For dynamical systems in general, the relationship between the coupled components will vary depending on the state of the system (think of the two ‘wings’ of the Lorenz ‘butterfly’ attractor). In the CCM approach, two variables are causally coupled if they can be shown to belong to the same dynamical system, but that does not imply that the relationship between them is fixed through time. With sufficient data, a moving window analysis can be used to test for temporal changes in the strength and directionality of coupling.

The convergent cross-mapping approach to causal inference has already found a few palaeontological applications, including time series of Cenozoic planktonic foraminifera (Hannisdal *et al.* 2017), Phanerozoic marine invertebrate diversity (Cermeño *et al.* 2017) and Holocene reef urchins (Cramer *et al.* 2018).

Software for running CCM is available as an R package (rEDM) at <https://github.com/ha0ye>, and wrapper functions that simplify lagged CCM analyses with surrogate testing are available at <https://www.earthsystemevolution.com/page/tools/>.

Transfer entropy

Granger causality was originally inspired by ideas from the theory of prediction, and Granger himself noted early on its linkages to information theory Granger (1963). A predictive notion of causality is closely related to the notion of a directional flow of information, and information-theoretic techniques are now widely used to measure causal dependence in dynamical systems (Hlaváčková-Schindler *et al.* 2007; Amblard & Michel 2013).

One of the most popular information-theoretic measures of directional information flow is Schreiber’s (2000) transfer entropy, which starts conventionally with the Markov property of conditional independence: If two processes X and Y are independent, then knowing the past l states of Y has no influence on the state transition probability of X , from $x(t_i)$ to $x(t_{i+1})$, beyond knowing the past k states of X alone. To simplify notation, we denote $x(t_{i+1})$ as x_{i+1} , and a vector $x(t_i, \dots, t_{i-k})$ of k past states is written as $x_i^{(k)}$:

$$p(x_{i+1}|x_i^{(k)}, y_i^{(l)}) = p(x_{i+1}|x_i^{(k)}) \quad (7)$$

Schreiber proposed to use the Kullback–Leibler divergence (Kullback & Leibler 1951; also called Kullback entropy, or relative entropy) to quantify the incorrectness of assuming independence, by estimating how much information is lost if we use the probability on the right hand side to approximate the probability on the left hand side of Equation 7:

$$TE_{Y \rightarrow X} = \sum_{x,y} p(x_{i+1}, x_i^{(k)}, y_i^{(l)}) \log \frac{p(x_{i+1}|x_i^{(k)}, y_i^{(l)})}{p(x_{i+1}|x_i^{(k)})} \quad (8)$$

Transfer entropy (TE) is a non-symmetric measure of information flow, which is equivalent to the conditional mutual information of Paluš *et al.* (2001; Vejmelka & Paluš 2008). Note that there is no natural scale for entropy, hence TE as defined above only yields a relative measure of whether Y is more relevant for predicting X than vice versa. Many modifications and different implementations of TE have subsequently been proposed (e.g. Chen *et al.* 2004; Staniek & Lehnertz 2008; Runge *et al.* 2012; Yu *et al.* 2017).

Verdes (2005) described a generalized TE conditioned on a third variable (i.e. a ‘partial’ TE), which he named information transfer (IT), and suggested a modified estimation algorithm that seemed to maintain statistical power even for relatively short and noisy time series. The IT has been adapted and applied to palaeontological and geological time series analysis (Hannisdal & Peters 2010; Hannisdal 2011a) to test the common-cause and redundancy hypotheses for explaining associations between rock and fossil records (Hannisdal 2011b; Hannisdal & Peters 2011; Dunhill *et al.* 2014, 2017), and to detect causal interactions between long-term climate change and global planktonic ecosystems (Hannisdal *et al.* 2012, 2017) (Fig. 3). The IT method has been found to be consistent with other measures of causal dependence (Hannisdal 2011a; Hannisdal *et al.* 2017). On the other hand, the Verdes algorithm is a heuristic approach, and the IT as a directionality measure can be biased if the time series have very different levels of non-stationarity, or trend. In practice, the appropriate level of detrending or other data pre-processing (Hannisdal 2011a; Hannisdal *et al.* 2012) has to be tested on a case-by-case basis.

To address these shortcomings, a new approach to estimating transfer entropy is being developed by D. Diego, K. Haaga and BH (see <https://www.earthsystemevolution.com/page/tools/>). This approach is also based on the concept of an attractor as a subset of a system’s state space, where each state consists of a set of current and past values from the different time series being analysed. Not all points on an attractor will be visited with the same frequency, and the frequency with which a typical trajectory will visit some portion of the state space is related to the density of the attractor, which is invariant under the dynamics. If we assume that two time-series X and Y represent variables that belong to the same dynamical system, and that the system has settled on its attractor, then it is possible to interpret the invariant density of a given region of the attractor as the probability for the system to be in any of the states in that region. The usual method for estimating the probability distributions required for

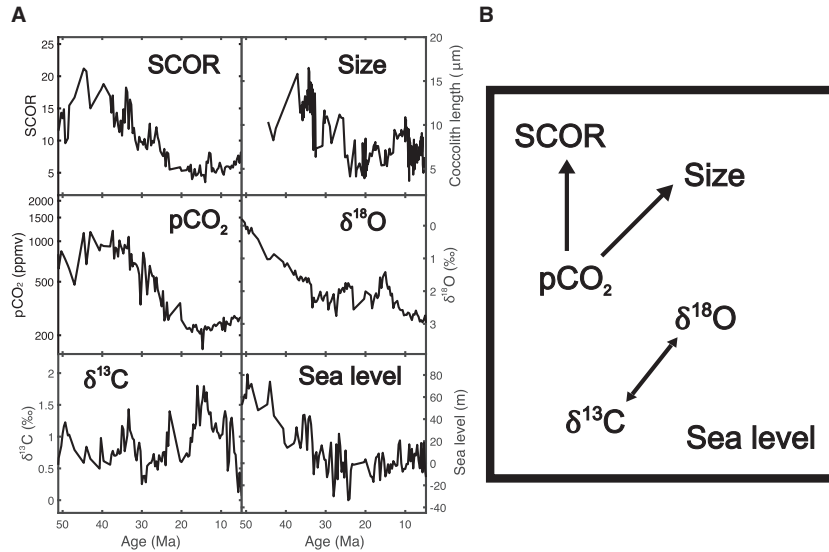


FIG. 3. Detecting a long-term planktonic response to Cenozoic changes in atmospheric CO_2 . A, six time-series representing coccolithophore Summed Common Species Occurrence Rate (SCOR; a proxy for relative changes in global abundance) and assemblage coccolithophore size (Hannisdal *et al.* 2012), reconstructed atmospheric $p\text{CO}_2$ based on the $\delta^{13}\text{C}$ of alkenones and $\delta^{11}\text{B}$ of carbonate (Pearson & Palmer 2000; Pagani *et al.* 2005, 2011; Pearson *et al.* 2009), deep-sea carbonate $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ (Zachos *et al.* 2001, 2008) and reconstructed sea level (Miller *et al.* 2005). These time series are redrawn from the original data sources used in Hannisdal *et al.* (2012). B, the causal relationships inferred by IT analysis of directional information flow among these observed time series (Hannisdal *et al.* 2012), where arrows point from the driver to the response. The reconstructed $p\text{CO}_2$ record captures a signal of processes that have influenced coccolithophore size and SCOR over the period 50–5 Ma. Note that sea level has no inferred relationship with any of the other time series whilst the relationship between $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ indicates a two-way interaction. See Hannisdal *et al.* (2012) for more details, and Hannisdal *et al.* (2017) for a study of Cenozoic planktonic foraminifera SCOR.

computing transfer entropy is the method of nearest neighbours (e.g. Kraskov *et al.* 2004), which is very efficient but may lose accuracy unless the number of points in an embedding reconstruction of the attractor is very large. Instead, Diego *et al.* take a different approach inspired by the work of Froyland (1997), by discretizing the state space into simple volume partitions (simplices) and tracking how these volumes change between time steps. The movement and deformation of volumes through time approximate the action of the underlying governing processes. The dynamics are approximated by the so-called transfer operator, in the form of a Markov matrix of transition probabilities, which is used to estimate the invariant probability distribution needed to compute TE. The aim of this approach is to improve accuracy for short time series, and because the transfer operator is stationary by construction, there is no need to pre-process non-stationary data.

Both CCM and TE are referred to as ‘model-free’, or ‘equation-free’ methods, because they take a geometric approach to detecting dynamical coupling without using model equations to specify causal mechanisms. TE allows time series to be unevenly sampled in time, but different variables need to be compared at contemporaneous points for the calculation of TE to be meaningful. Unlike the original TE (Schreiber 2000) and IT (Verdes 2005), the transfer

operator TE does not assume stationarity and obviates the need to detrend the observed time series. Like CCM, the TE currently requires a Monte Carlo approach to account for noise or chronological uncertainty. Unlike the CCM, however, the TE can be directly generalized to three or more variables to test for confounding, or common-cause, interactions. That said, with more variables added in a conditional TE, longer time series are needed to ensure sufficient information content. TE allows for the relationship between coupled variables to vary depending on the state of the system and, with enough data, temporal changes in the strength and directionality of coupling can be assessed with a moving window analysis.

Software for computing the transfer operator TE using the high-performance computing language Julia is available at <https://www.earthsystemevolution.com/page/tools/>. An R package of wrapper functions is forthcoming. This website also features worked examples and interactive applications with examples of dynamical systems.

FROM EVENT-BASED REASONING TO DYNAMICAL CAUSALITY

It is a quirk of human cognition that as we receive a continuous stream of sensory input, we tend to segment this

dynamic information into discrete events with a beginning and an end (Michotte 1963; Zacks & Tversky 2001; Shipley & Zacks 2008). Temporally extended or ongoing changes tend to be perceived as a background, or status quo, against which events are reified and subsequently 'filled in' with a cause-effect interpretation (Strickland & Keil 2011). Event-based reasoning can impede our understanding of dynamic interactions in complex natural systems, such as ecosystems or climate systems (Raia 2008; Grotzer *et al.* 2013). In historical geology and palaeontology, an event-based causal focus is enhanced by the nature of the stratigraphic record, which, 'like the life of a soldier, consists of long periods of boredom and short periods of terror' in Derek Ager's famous aphorism (Ager 1973). Indeed, event stratigraphy is central to chronostratigraphic correlation and the refinement of the geological time scale itself. We certainly do not wish to downplay the importance of singular events in macroevolution, from key evolutionary innovations to mass extinctions, nor the need to understand the role of unreplicated events in macroevolutionary theory or phylogenetic comparative biology (Jablonski 2017*a, b*; Uyeda *et al.* 2017). We do argue, however, that causal attribution in the context of palaeontological time series needs a dynamical, process-like notion of causality to complement our research on events. Raup himself lamented the tendency in our field of 'finding specific causes for specific evolutionary events' rather than seeking generalizations across groups of events (Raup 1977*b*). The same concern has been raised for the interpretation of carbon isotope excursions across the Phanerozoic, where underlying dynamics may explain commonalities among a large number of isotope excursions that are typically treated as unique events (Bachan *et al.* 2017).

Palaeontological and palaeoenvironmental records provide access to only a very small subset of the variables that may have interacted to generate the patterns observed in those records. Moreover, most deep-time records are themselves phenomenological proxies, which are typically a compounding of different underlying processes (e.g. the oxygen isotope record from marine carbonates captures past changes in temperature, ice volume, salinity and pH). But even if we cannot know all the interacting components (or 'detail complexity') of the relevant system, observed time series can capture important aspects of its dynamical complexity. If so, then the dynamical information inherent in the observed time series can be used to detect causal connections. As noted above in the method descriptions, this data-driven approach places certain demands on the information content of the data that are not always met in relatively short, noisy and irregular palaeontological time series. In addition, we note that there are other important empirical considerations that we have not discussed in this

review, including the appropriateness of measurements (Houle *et al.* 2011) and the perennial problems associated with disentangling temporal process from stratigraphic architecture (Holland 1995; Hannisdal 2006, 2007; Patzkowsky & Holland 2012). On the other hand, the quality and resolution of deep-time records are steadily improving, including innovative approaches such as continuous-time age models (e.g. Nelsen *et al.* 2016; Husson & Peters 2017).

Thinking about the innumerable interactions that conspire to generate our observed palaeontological patterns can be quite dizzying, ranging from the gene regulatory networks involved in development (Peter & Davidson 2015), and the vast complexity of ecosystems (e.g. Lima-Mendez *et al.* 2015), to biotic feedbacks on global climate on million-year time scales (e.g. Beerling & Berner 2005). To illustrate how a dynamical notion of causality might help us see the proverbial forest for the trees, we make an analogy with another system of immense complexity that allows us to think about such things in the first place (and get a headache): the human brain. While the basic causal mechanisms behind brain activity at the micro-level of a single neuron have essentially been understood since the 1950s, macro-level activities such as movement, cognition and perception involve interactions of massive numbers of neurons across large systems of the brain. Recent technologies such as functional magnetic resonance imaging and electroencephalography provide time series of this system-level behaviour, which can be fruitfully analysed using concepts from dynamical systems theory and surrogate data testing (Breakspear 2017) as well as Granger causality (Bressler & Seth 2011). Causal hypotheses can thus be tested on observed time series without recourse to modelling the underlying complexity. A time-series analysis approach to causality in neuroscience has been criticized on the grounds that it can fail to resolve the full structure of the underlying mechanism (e.g. Stokes & Purdon 2017). However, this particular critique fails to make the distinction between quantification of directional information flow and explicit modelling of physical causal mechanisms, both of which are legitimate tasks, but the latter is beyond the purview of the time-series analysis approach. Instead, a data-driven causal analysis can uncover dynamical relationships between components of complex and poorly constrained systems (e.g. Fig. 3), and thus help guide efforts to build models of the underlying mechanisms.

Another tentative insight from the study of large-scale brain activity that could resonate with palaeontologists is that macro-scale dynamics are not simply the sum of micro-level components, and that causal emergence allows for top-down causality (e.g. Hoel *et al.* 2013). Although fossils are not neurons, and the question still remains as to what extent macroevolutionary processes

are emergent or generated bottom-up from microevolutionary processes (Erwin 2000; Simons 2002; Jablonski 2017a), we note that new theoretical results have raised the tantalizing prospect of a fundamental link between the processes of evolution and the principles of learning (Watson & Szathmáry 2016).

The types of linear thinking that we typically use in standard data analysis (e.g. variance partitioning, regression coefficients) assume that components can be analysed separately and then added back together. However, this approximation often fails in both living and Earth systems, where components are intertwined and may interact in such a way that any linear correlations are transient and may even switch sign (referred to as ‘mirage correlations’ by Sugihara *et al.* 2012). Although there is no substitute for a well-posed question that can be answered with limited data and simple methods, we must acknowledge that the systems from which we obtain time series (and the causal questions we ask of them) may be far from simple, and may require analyses that are conceptually and computationally involved. We want to know what has ‘driven’, ‘tracked’, ‘interacted with’ or ‘controlled’ the intriguing patterns of temporal change observed in deep-time records, but the way we pose causal questions belies the epistemic challenges involved in answering them. The methods discussed in this review invite us to take a step back and think carefully about what it means to formulate and test causal hypotheses on time series.

A defining topic of palaeontology is the interplay of ecological, evolutionary and environmental processes responsible for the long-term maintenance and turnover of Earth’s biota. Palaeontological data provide pre-anthropogenic baselines and deep-time lessons critical for understanding current systems as well as their recent, near-term and long-term future responses to change (National Research Council (US) 2011; Hannisdal *et al.* 2012, 2017; Harnik *et al.* 2012; Dietl *et al.* 2015; Finnegan *et al.* 2015; Kidwell 2015; Schmidt 2017). With the looming threat of large-scale disruption of global ecosystems, there is a need for palaeontologists to be bifocal: we need to zoom in to uncover the historical particulars of individual events, but also zoom out to search for underlying causal structures that may hold general lessons across multiple events. A dynamical, process-like concept of causality may be a Rosetta stone that can assist in deciphering the messages encrypted in palaeontological time series.

Acknowledgements. We thank Rachel Warnock, Alex Dunhill, Erin Saupé and Mark Sutton for their invitations to present parts of this review at the Geological Society of America Annual Meeting 2017 topical session ‘Advances in Computational Paleobiology: Reshaping Our Understanding of the Fossil Record’ (LHL), the 2017 Annual Symposium of the Palaeontological

Association titled ‘Evolutionary Modelling in Palaeontology’ (LHL), and the European Geosciences Union 2018 General Assembly technical session ‘Advances in palaeobiological modelling’ (BH). We thank the Palaeontological Association for sponsoring all three symposia and Andrew Smith for inviting us to write this review. LHL thanks Trond Reitan for discussions and his contributions to method development, and The Research Council of Norway grant number 235073 for funding. BH thanks Kristian Haaga and David Diego for their invaluable contributions. BH is funded by the Research Council of Norway grant 231259, and by the Bergen Research Foundation. We thank two anonymous reviewers for their helpful comments and suggestions on an earlier version of this paper.

Editor. Andrew Smith

REFERENCES

- AGER, D. V. 1973. *The nature of the stratigraphical record*. Macmillan, London.
- AMBLARD, P. O. and MICHEL, O. J. J. 2013. The relation between Granger causality and directed information theory: a review. *Entropy*, **15**, 113–143.
- BABTIE, A. C., KIRK, P. and STUMPF, M. P. H. 2014. Topological sensitivity analysis for systems biology. *Proceedings of the National Academy of Sciences*, **111**, 18507–18512.
- BACHAN, A., LAU, K. V., SALTZMAN, M. R., THOMAS, E., KUMP, L. R. and PAYNE, J. L. 2017. A model for the decrease in amplitude of carbon isotope excursions across the Phanerozoic. *American Journal of Science*, **317**, 641–676.
- BARNETT, L. and SETH, A. K. 2015. Granger causality for state-space models. *Physical Review E*, **91**, 40101.
- BARRETT, A. B. and SETH, A. K. 2009. Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical Review Letters*, **103**, 238701.
- BARRETT, A. B., BARNETT, L. and SETH, A. K. 2010. Multivariate Granger causality and generalized variance. *Physical Review E*, **81**, 41907.
- BASKERVILLE, E. B. and COBEY, S. 2017. Does influenza drive absolute humidity? *Proceedings of the National Academy of Sciences*, **114**, E2270–E2271.
- BEERLING, D. J. and BERNER, R. A. 2005. Feedbacks and the coevolution of plants and atmospheric CO₂. *Proceedings of the National Academy of Sciences*, **102**, 1302–1305.
- BOOKSTEIN, F. L. 1987. Random walk and the existence of evolutionary rates. *Paleobiology*, **13**, 446–464.
- BREAKSPEAR, M. 2017. Dynamic models of large-scale brain activity. *Nature Neuroscience*, **20**, 340–352.
- BRESSLER, S. L. and SETH, A. K. 2011. Wiener–Granger causality: a well established methodology. *NeuroImage*, **58**, 323–329.
- BRINKMANN, R. 1929. Statistisch-biostratigraphische Untersuchungen an mitteljurassischen Ammoniten über Artbegriff und Stammesentwicklung. *Abhandlungen der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, N.F.*, **13**, 1–249.

- BROWN, C. J., SCHOEMAN, D. S., SYDEMAN, W. J., BRANDER, K., BUCKLEY, L. B., BURROWS, M., DUARTE, C. M., MOORE, P. J., PANDOLFI, J. M., POLOCZANSKA, E., VENABLES, W. and RICHARDSON, A. J. 2011. Quantitative approaches in climate change ecology. *Global Change Biology*, **17**, 3697–3713.
- BURKETT, V. R., WILCOX, D. A., STOTTLEMYER, R., BARROW, W., FAGRE, D., BARON, J., PRICE, J., NIELSEN, J. L., ALLEN, C. D., PETERSON, D. L., RUGGERONE, G. and DOYLE, T. 2005. Nonlinear dynamics in ecosystem response to climatic change: case studies and policy implications. *Ecological Complexity*, **2**, 357–394.
- CERMEÑO, P., BENTON, M. J., PAZ, Ó. and VÉRARD, C. 2017. Trophic and tectonic limits to the global increase of marine invertebrate diversity. *Scientific Reports*, **7**, 15969.
- CHEN, Y., RANGARAJAN, G., FENG, J. and DING, M. 2004. Analyzing multiple nonlinear time series with extended Granger causality. *Physics Letters A*, **324**, 26–35.
- COUFAL, D., JAKUBÍK, J., JAJCAY, N., HLINKA, J., KRAKOVSKÁ, A. and PALUŠ, M. 2017. Detection of coupling delay: a problem not yet solved. *Chaos*, **27**, 83109.
- CRAMER, K. L., O'DEA, A., CARPENTER, C. and NORRIS, R. D. 2018. A 3000 year record of Caribbean reef urchin communities reveals causes and consequences of long-term decline in *Diadema antillarum*. *Ecography*, **41**, 164–173.
- DIETL, G. P., KIDWELL, S. M., BRENNER, M., BURNEY, D. A., FLESSA, K. W., JACKSON, S. T. and KOCH, P. L. 2015. Conservation paleobiology: leveraging knowledge of the past to inform conservation and restoration. *Annual Review of Earth & Planetary Sciences*, **43**, 79–103.
- DUNHILL, A. M., HANNISDAL, B. and BENTON, M. J. 2014. Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nature Communications*, **5**, 4818.
- BROCKLEHURST, N. and BENTON, M. J. 2017. On formation-based sampling proxies and why they should not be used to correct the fossil record. *Palaentology*, **61**, 119–132.
- EICHLER, M. 2005. A graphical approach for evaluating effective connectivity in neural systems. *Philosophical Transactions of the Royal Society B*, **360**, 953–967.
- ERLYKIN, A. D., HARPER, D. A. T., SLOAN, T. and WOLFENDALE, A. W. 2018. Periodicity in extinction rates. *Palaentology*, **61**, 149–158.
- ERWIN, D. H. 2000. Macroevolution is more than repeated rounds of microevolution. *Evolution & Development*, **2**, 78–84.
- FEDAK, K. M., BERNAL, A., CAPSHAW, Z. A. and GROSS, S. 2015. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerging Themes in Epidemiology*, **12**, 14.
- FINNEGAN, S., ANDERSON, S. C., HARNIK, P. G., SIMPSON, C., TITTENSOR, D. P., BYRNES, J. E., FINKEL, Z. V., LINDBERG, D. R., LIOW, L. H., LOCKWOOD, R., LOTZE, H. K., McCLAIN, C. R., McGUIRE, J. L., O'DEA, A. and PANDOLFI, J. M. 2015. Paleontological baselines for evaluating extinction risk in the modern oceans. *Science*, **348**, 567–570.
- FROYLAND, G. 1997. Estimating physical invariant measures and space averages of dynamical systems indicators. *Bulletin of the Australian Mathematical Society*, **56**, 157–159.
- GRANGER, C. W. J. 1963. Economic processes involving feedback. *Information & Control*, **6**, 28–48.
- 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424.
- GROTZER, T. A., KAMARAINEN, A. M., TUTWILER, M. S., METCALF, S. and DEDE, C. 2013. Learning to reason about ecosystems dynamics over time: the challenges of an event-based causal focus. *BioScience*, **63**, 288–296.
- HAAGA, K. A., BRENDRYEN, J., DIEGO, D. and HANNISDAL, B. 2018. Forcing of Late Pleistocene ice volume by spatially variable summer energy. *Open Science Framework*. <https://doi.org/10.17605/OSF.IO/C2DZV>
- HANNISDAL, B. 2006. Phenotypic evolution in the fossil record: numerical experiments. *The Journal of Geology*, **114**, 133–153.
- 2007. Inferring phenotypic evolution in the fossil record by Bayesian inversion. *Paleobiology*, **33**, 98–115.
- 2011a. Non-parametric inference of causal interactions from geological records. *American Journal of Science*, **311**, 315–334.
- 2011b. Detecting common-cause relationships with directional information transfer. *Geological Society, London, Special Publications*, **358**, 19–29.
- and PETERS, S. E. 2010. On the relationship between macrostratigraphy and geological processes: quantitative information capture and sampling robustness. *The Journal of Geology*, **118**, 111–130.
- 2011. Phanerozoic earth system evolution and marine biodiversity. *Science*, **334**, 1121–1124.
- HENDERIKS, J. and LIOW, L. H. 2012. Long-term evolutionary and ecological responses of calcifying phytoplankton to changes in atmospheric CO₂. *Global Change Biology*, **18**, 3504–3516.
- HAAGA, K. A., REITAN, T., DIEGO, D. and LIOW, L. H. 2017. Common species link global ecosystems to climate change: dynamical evidence in the planktonic fossil record. *Proceedings of the Royal Society B*, **284**, 20170722.
- HARNIK, P. G., LOTZE, H. K., ANDERSON, S. C., FINKEL, Z. V., FINNEGAN, S., LINDBERG, D. R., LIOW, L. H., LOCKWOOD, R., McCLAIN, C. R., McGUIRE, J. L., O'DEA, A., PANDOLFI, J. M., SIMPSON, C. and TITTENSOR, D. P. 2012. Extinctions in ancient and modern seas. *Trends in Ecology & Evolution*, **27**, 608–617.
- HILL, A. B. 1965. The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, **8**, 295–300.
- HINNOV, L. A., WU, H. and FANG, Q. 2016. Reply to the comment on “Geologic evidence for chaotic behavior of the planets and its constraints on the third-order eustatic sequences at the end of the Late Paleozoic Ice Age” by Fang, Q., Wu, H., Hinnov, L. A., Jing, X., Wang, X., & Jiang, Q. [Palaeogeography Palaeoclimatology Palaeoecology, Volume 400, 15 December 2015, 848–859 (2015)]. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **461**, 475–480.

- HLAVÁČKOVÁ-SCHINDLER, K., PALUŠ, M., VEJ-MELKA, M. and BHATTACHARYA, J. 2007. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, **441**, 1–46.
- HOEL, E. P., ALBANTAKIS, L. and TONONI, G. 2013. Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences*, **110**, 19790–19795.
- HOLLAND, S. M. 1995. The stratigraphic distribution of fossils. *Paleobiology*, **21**, 92–109.
- HOULE, D., PÉLABON, C., WAGNER, G. P. and HANSEN, T. F. 2011. Measurement and meaning in biology. *The Quarterly Review of Biology*, **86**, 3–34.
- HUNT, G. 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology*, **32**, 578–601.
- and RABOSKY, D. L. 2014. Phenotypic evolution in fossil species: pattern and process. *Annual Review of Earth & Planetary Sciences*, **42**, 421–441.
- HUSSON, J. M. and PETERS, S. E. 2017. Atmospheric oxygenation driven by unsteady growth of the continental sedimentary reservoir. *Earth & Planetary Science Letters*, **460**, 68–75.
- HUYBERS, P., LANGMUIR, C., KATZ, R. F., FERGUSON, D., PROISTOESCU, C. and CARBOTTE, S. 2016. Comment on “Sensitivity of seafloor bathymetry to climate-driven fluctuations in mid-ocean ridge magma supply”. *Science*, **352**, 1405.
- ILLARI, P. M., RUSSO, F. and WILLIAMSON, J. 2011. *Causality in the sciences*. Oxford University Press.
- JABLONSKI, D. 2017a. Approaches to macroevolution: 1. General concepts and origin of variation. *Evolutionary Biology*, **44**, 427–450.
- 2017b. Approaches to macroevolution: 2. Sorting of variation, some overarching issues, and general conclusions. *Evolutionary Biology*, **44**, 451–475.
- KAISER, M. I. 2016. On the limits of causal modeling: spatially-structurally complex biological phenomena. *Philosophy of Science*, **83**, 921–933.
- KANTZ, H. and SCHREIBER, T. 2003. *Nonlinear time series analysis*. Cambridge University Press.
- KIDWELL, S. M. 2015. Biology in the Anthropocene: challenges and insights from young fossil records. *Proceedings of the National Academy of Sciences*, **112**, 4922–4929.
- KITCHELL, J. A. and PENA, D. 1984. Periodicity of extinctions in the geologic past: deterministic versus stochastic explanations. *Science*, **226**, 689–692.
- KRASKOV, A., STÖGBAUER, H. and GRASSBERGER, P. 2004. Estimating mutual information. *Physical Review E*, **69**, 16.
- KULLBACK, S. and LEIBLER, R. A. 1951. On information and sufficiency. *The Annals of Mathematical Statistics*, **22**, 79–86.
- LANDE, R. 1976. Natural selection and random genetic drift in phenotype evolution. *Evolution*, **30**, 314–334.
- LIMA-MENDEZ, G., FAUST, K., HENRY, N., DECELLE, J., COLIN, S., CARCILLO, F., CHAFFRON, S., IGNACIO-ESPINOSA, J. C., ROUX, S., VINCENT, F., BITTNER, L., DARZI, Y., WANG, J., AUDIC, S., BERLINE, L., BONTEMPI, G., CABELLO, A. M., COPPOLA, L., CORNEJO-CASTILLO, F. M., D’OVIDIO, F., MEESTER, L. DE, FERRERA, I., GARET-DELMAS, M. J., GUIDI, L., LARA, E., PESANT, S., ROYO-LLONCH, M., SALAZAR, G., SÁNCHEZ, P., SEBASTIAN, M., SOUFFREAU, C., DIMIER, C., PICHERAL, M., SEARSON, S., KANDELS-LEWIS, S., GORSKY, G., NOT, F., OGATA, H., SPEICH, S., STEMMANN, L., WEISENBACH, J., WINCKER, P., ACINAS, S. G., SUNAGAWA, S., BORK, P., SULLIVAN, M. B., KARSENTI, E., BOWLER, C., VARGAS, C. DE, RAES, J., BOSS, E., FOLLOWS, M., GRIMSLEY, N., HINGAMP, P., IUDICONE, D., JAILLON, O., KARP-BOSS, L., KRZIC, U., REYNAUD, E. G., SARDET, C., SIERACKI, M. and VELAYOUDON, D. 2015. Determinants of community structure in the global plankton interactome. *Science*, **348**, 1262073.
- LIOW, L. H., REITAN, T. and HARNIK, P. G. 2015. Ecological interactions on macroevolutionary time scales: clams and brachiopods are more than ships that pass in the night. *Ecology Letters*, **18**, 1030–1039.
- LUO, M., KANTZ, H., LAU, N.-C., HUANG, W. and ZHOU, Y. 2015. Questionable dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature: Fig. 1. *Proceedings of the National Academy of Sciences*, **112**, E4638–E4639.
- MAY, R. M. 1976. Simple mathematical models with very complicated dynamics. *Nature*, **261**, 459–467.
- MELOTT, A. L. and BAMBACH, R. K. 2017. Periodicity in the extinction rate and possible astronomical causes – comment on mass extinctions over the last 500 myr: an astronomical cause? (Erlykin *et al.*). *Palaeontology*, **60**, 911–920.
- MEYER, K. M., RIDGWELL, A. and PAYNE, J. L. 2016. The influence of the biological pump on ocean chemistry: implications for long-term trends in marine redox chemistry, the global carbon cycle, and marine animal ecosystems. *Geobiology*, **14**, 207–219.
- MICHOTTE, A. 1963. *The perception of causality*. [Translated by T.R. Miles and E. Miles] Methuen, London.
- MILLER, K. G., KOMINZ, M. A., BROWNING, J. V., WRIGHT, J. D., MOUNTAIN, G. S., KATZ, M. E., SUGARMAN, P. J., CRAMER, B. S., CHRISTIE-BLICK, N. and PEKAR, S. F. 2005. The Phanerozoic record of global sea-level change. *Science*, **310**, 1293–1298.
- NATIONAL RESEARCH COUNCIL (US) 2011. *Understanding earth’s deep past: lessons for our climate future*. National Academies Press.
- NELSEN, M. P., DIMICHELE, W. A., PETERS, S. E. and BOYCE, C. K. 2016. Delayed fungal evolution did not cause the Paleozoic peak in coal production. *Proceedings of the National Academy of Sciences*, **113**, 2442–2447.
- NES, E. H. VAN, SCHEFFER, M., BROVKIN, V., LENTON, T. M., YE, H., DEYLE, E. and SUGIHARA, G. 2015. Causal feedbacks in climate change. *Nature Climate Change*, **5**, 445–448.
- NOSIL, P., FEDER, J. L., FLAXMAN, S. M. and GOMPERT, Z. 2017. Tipping points in the dynamics of speciation. *Nature Ecology & Evolution*, **1**, 1.

- OLIVE, J.-A., BEHN, M. D., ITO, G., BUCK, W. R., ESCARTÍN, J. and HOWELL, S. 2016. Response to Comment on “Sensitivity of seafloor bathymetry to climate-driven fluctuations in mid-ocean ridge magma supply”. *Science*, **352**, 1405.
- PACKARD, N. H., CRUTCHFIELD, J. P., FARMER, J. D. and SHAW, R. S. 1980. Geometry from a time series. *Physical Review Letters*, **45**, 712–716.
- PAGANI, M., ZACHOS, J. C., FREEMAN, K. H., TRIPLE, B. and BOHATY, S. 2005. Marked decline in atmospheric carbon dioxide concentrations during the Paleocene. *Science*, **309**, 600–603.
- HUBER, M., LIU, Z., BOHATY, S. M., HENDERIKS, J., SIJP, W., KRISHNAN, S. and DECONTO, R. M. 2011. The role of carbon dioxide during the onset of antarctic glaciation. *Science*, **334**, 1261–1264.
- PALUŠ, M., KOMÁREK, V., HRNČÍŘ, Z. and ŠTĚRBOVÁ, K. 2001. Synchronization as adjustment of information rates: detection from bivariate time series. *Physical Review E*, **63**, 46211.
- PATZKOWSKY, M. E. and HOLLAND, S. M. 2012. *Stratigraphic paleobiology: understanding the distribution of fossil taxa in time and space*. University of Chicago Press.
- PEARL, J. 2009. *Causality*. Cambridge University Press.
- PEARSON, K. 1911. *The grammar of science*. Walter Scott.
- PEARSON, P. N. and PALMER, M. R. 2000. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature*, **406**, 695–699.
- FOSTER, G. L. and WADE, B. S. 2009. Atmospheric carbon dioxide through the Eocene–Oligocene climate transition. *Nature*, **461**, 1110–1113.
- PENNELL, M. W., FITZJOHN, R. G., CORNWELL, W. K. and HARMON, L. J. 2015. Model adequacy and the macroevolution of angiosperm functional traits. *The American Naturalist*, **186**, E33–E50.
- PETER, I. S. and DAVIDSON, E. H. 2015. *Genomic control process: development and evolution*. Academic Press.
- PETERS, S. E. and HEIM, N. A. 2011. Macrostratigraphy and macroevolution in marine environments: testing the common-cause hypothesis. *Geological Society, London, Special Publications*, **358**, 95–104.
- RAIA, F. 2008. Causality in complex dynamic systems: a challenge in Earth Systems Science education. *Journal of Geoscience Education*, **56**, 81.
- RAUP, D. M. 1977a. Stochastic models in evolutionary palaeontology. *Developments in Palaeontology & Stratigraphy*, **5**, 59–78.
- 1977b. Probabilistic Models in Evolutionary Paleobiology: a random walk through the fossil record produces some surprising results. *American Scientist*, **65**, 50–57.
- RAUP, D. and CRICK, R. 1981. Evolution of single characters in the Jurassic Ammonite *Kosmoceras*. *Paleobiology*, **7**, 200–215.
- RAUP, D. M. and SEPKOSKI, J. J. 1984. Periodicity of extinctions in the geologic past. *Proceedings of the National Academy of Sciences*, **81**, 801–805.
- GOULD, S. J., SCHOPF, T. J. M. and SIMBERLOFF, D. S. 1973. Stochastic models of phylogeny and the evolution of diversity. *The Journal of Geology*, **81**, 525–542.
- REITAN, T. and LIOW, L. H. 2017. An unknown Phanerozoic driver of brachiopod extinction rates unveiled by multivariate linear stochastic differential equations. *Paleobiology*, **43**, 537–549.
- SCHWEDER, T. and HENDERIKS, J. 2012. Phenotypic evolution studied by layered stochastic differential equations. *Annals of Applied Statistics*, **6**, 1531–1551.
- ROOPNARINE, P. D., BYARS, G. and FITZGERALD, P. 1999. Anagenetic evolution, stratophenetic patterns, and random walk models. *Paleobiology*, **25**, 41–57.
- RUNGE, J., HEITZIG, J., PETOUKHOV, V. and KURTHS, J. 2012. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical Review Letters*, **108**, 258701.
- SAUER, T., YORKE, J. A. and CASDAGLI, M. 1991. Embedology. *Journal of Statistical Physics*, **65**, 579–616.
- SCHMIDT, D. N. 2017. Determining climate change impacts on ecosystems: the role of palaeontology. *Palaeontology*, **61**, 1–12.
- SCHRAMSKI, J. R., GATTIE, D. K. and BROWN, J. H. 2015. Human domination of the biosphere: rapid discharge of the earth-space battery foretells the future of humankind. *Proceedings of the National Academy of Sciences*, **112**, 9511–9517.
- SCHREIBER, T. 2000. Measuring information transfer. *Physical Review Letters*, **85**, 461–464.
- SHIPLEY, B. 2016. *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference with R*. Cambridge University Press.
- SHIPLEY, T. F. and ZACKS, J. M. 2008. *Understanding events: from perception to action*. Oxford University Press.
- SIMONS, A. M. 2002. The continuity of microevolution and macroevolution. *Journal of Evolutionary Biology*, **15**, 688–701.
- SIMPSON, G. G. 1944. *Tempo and mode in evolution*. Columbia University Press.
- SMITH, D. G., BAILEY, R. J. and VAUGHAN, S. 2016. Comment on ‘Geologic evidence for chaotic behavior of the planets and its constraints on the third-order eustatic sequences at the end of the Late Paleozoic Ice Age’ by Fang, Q., Wu, H., Hinnov, L. A., Jing, X., Wang, X., & Jiang, Q. [Palaeogeography Palaeoclimatology Palaeoecology 400 (2015) 848–859]. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **461**, 472–474.
- SPIRITES, P., GLYMOUR, C. N. and SCHEINES, R. 2000. *Causation, prediction, and search*. MIT Press.
- STANIEK, M. and LEHNERTZ, K. 2008. Symbolic transfer entropy. *Physical Review Letters*, **100**, 158101.
- STOKES, P. A. and PURDON, P. L. 2017. A study of problems encountered in Granger causality analysis from a neuroscience perspective. *Proceedings of the National Academy of Sciences*, **114**, E7063–E7072.
- STRICKLAND, B. and KEIL, F. 2011. Event completion: event based inferences distort memory in a matter of seconds. *Cognition*, **121**, 409–415.
- SUGIHARA, G. and MAY, R. M. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, **344**, 734–741.
- MAY, R., YE, H., HSIEH, C., DEYLE, E., FOGARTY, M. and MUNCH, S. 2012. Detecting causality in complex ecosystems. *Science*, **338**, 496–500.

- DEYLE, E. R. and YE, H. 2017. Reply to Baskerville and Cobey: misconceptions about causation with synchrony and seasonal drivers. *Proceedings of the National Academy of Sciences*, **114**, E2272–E2274.
- TAKENS, F. 1981. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, **898**, 366–381.
- TAYLOR, K. E., STOUFFER, R. J. and MEEHL, G. A. 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**, 485–498.
- UYEDA, J. C., ZENIL-FERGUSON, R. and PENNELL, M. W. 2017. Rethinking phylogenetic comparative methods. *BioRxiv*. <https://doi.org/10.1101/222729>
- VEJMEĽKA, M. and PALUŠ, M. 2008. Inferring the directionality of coupling with conditional mutual information. *Physical Review E*, **77**, 26214.
- VERDES, P. F. 2005. Assessing causality from multivariate time series. *Physical Review E*, **72**, 26222.
- VOJE, K. L., STARRFELT, J. and LIOW, L. H. 2018. Model adequacy and microevolutionary explanations for stasis in the fossil record. *The American Naturalist*, **191**, 509–523.
- WATSON, R. A. and SZATHMÁRY, E. 2016. How can evolution learn? *Trends in Ecology & Evolution*, **31**, 147–157.
- WEBER, M. 2016. On the incompatibility of dynamical biological mechanisms and causal graphs. *Philosophy of Science*, **83**, 959–971.
- WIEDER, W. R., GRANDY, A. S., KALLENBACH, C. M., TAYLOR, P. G. and BONAN, G. B. 2015. Representing life in the earth system with soil microbial functional traits in the MIMICS model. *Geoscientific Model Development*, **8**, 1789–1808.
- WINSHIP, C. and MORGAN, S. L. 1999. The estimation of causal effects from observational data. *Annual Review of Sociology*, **25**, 659–706.
- WOOD, S. N. and THOMAS, M. B. 1999. Super-sensitivity to structure in biological models. *Proceedings of the Royal Society B*, **266**, 565–570.
- WOODWARD, J. 2003. *Making things happen: a causal theory of explanation*. Oxford University Press.
- WRIGHT, S. 1921. Correlation and causation. *Journal of Agricultural Research*, **20**, 557–585.
- YE, H., DEYLE, E. R., GILARRANZ, L. J. and SUGIHARA, G. 2015a. Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific Reports*, **5**, 14750.
- SUGIHARA, G., DEYLE, E. R., MAY, R. M., SWANSON, K. and TSONIS, A. A. 2015b. Reply to Luo et al.: robustness of causal effects of galactic cosmic rays on interannual variation in global temperature. *Proceedings of the National Academy of Sciences*, **112**, E4640–E4641.
- YU, M., HILLEBRAND, A., GOUW, A. A. and STAM, C. J. 2017. Horizontal visibility graph transfer entropy (HVGTE): a novel metric to characterize directed connectivity in large-scale brain networks. *NeuroImage*, **156**, 249–264.
- ZACHOS, J., PAGANI, M., SLOAN, L., THOMAS, E. and BILLUPS, K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**, 686–693.
- ZACHOS, J. C., DICKENS, G. R. and ZEEBE, R. E. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, **451**, 279–283.
- ZACKS, J. M. and TVERSKY, B. 2001. Event structure in perception and conception. *Psychological Bulletin*, **127**, 3–21.