



# Inferring species interactions using Granger causality and convergent cross mapping

Frédéric Barraquand<sup>1,2</sup> · Coralie Picoche<sup>1,2</sup> · Matteo Dettò<sup>3</sup> · Florian Hartig<sup>4</sup>

Received: 13 March 2020 / Accepted: 4 September 2020 / Published online: 16 November 2020  
© Springer Nature B.V. 2020

## Abstract

Identifying directed interactions between species from time series of their population densities has many uses in ecology. This key statistical task is equivalent to causal time series inference, which connects to the Granger causality (GC) concept:  $x$  causes  $y$  if  $x$  improves the prediction of  $y$  in a dynamic model. However, the entangled nature of nonlinear ecological systems has led to question the appropriateness of Granger causality, especially in its classical linear multivariate autoregressive (MAR) model form. Convergent cross mapping (CCM), a nonparametric method developed for deterministic dynamical systems, has been suggested as an alternative. Here, we show that linear GC and CCM are able to uncover interactions with surprisingly similar performance, for predator-prey cycles, 2-species deterministic (chaotic), or stochastic competition, as well as 10- and 20-species interaction networks. We found no correspondence between the degree of nonlinearity of the dynamics and which method performs best. Our results therefore imply that Granger causality, even in its linear MAR( $p$ ) formulation, is a valid method for inferring interactions in nonlinear ecological networks; using GC or CCM (or both) can instead be decided based on the aims and specifics of the analysis.

**Keywords** Time series · Interaction network · Causal inference · Feedback · Food web · Community dynamics

## Introduction

Inferring links between different species' population dynamics is a statistical endeavor with profound implications for understanding coexistence mechanisms (Adler et al. 2010, 2018), food web structure and functioning (Berlow et al. 2004; Wootton and Emmerson 2005), as

well as management and conservation at the ecosystem level (Link 2002; Pikitch et al. 2004). However, statistically detecting such dependencies using correlative approaches can be extremely challenging (Coenen and Weitz 2018; Carr et al. 2019). Outside of the usual limitations induced by sample sizes, spatial or temporal co-occurrence (Cazelles et al. 2016) or co-abundance patterns (Stone and Roberts 1991; Loreau and de Mazancourt 2008) do not directly indicate interactions between species (Dormann et al. 2018; Blanchet et al. 2020). For instance, strongly competitive communities usually show a large amount of positive associations between species abundances, not only because abiotic forcing makes synchrony the general rule (Loreau and de Mazancourt 2008), but also because indirect interactions make the enemy of my enemy a friend (Stone and Roberts 1991).

To infer dependencies between species' population dynamics, it is therefore often useful to build on a dynamic ecological and statistical theory. A first step towards inferring interactions between species is to agree on a definition of an interaction (Berlow et al. 2004). For the purpose of this paper, two species  $i$  and  $j$  are deemed to interact if species  $i$ 's population growth rate is affected

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12080-020-00482-7>) contains supplementary material, which is available to authorized users.

✉ Frédéric Barraquand  
frederic.barraquand@u-bordeaux.fr

<sup>1</sup> Institute of Mathematics of Bordeaux, CNRS & University of Bordeaux, Talence, France

<sup>2</sup> Integrative and Theoretical Ecology, LabEx COTE, University of Bordeaux, Pessac, France

<sup>3</sup> Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

<sup>4</sup> Theoretical Ecology, University of Regensburg, Regensburg, Germany

by the population density of species  $j$  or vice versa. This definition maps well to theoretical ecology, where communities are modelled as variations of the generalized Lotka-Volterra equations (e.g., May 1973; Yodzis 1998; Coyte et al. 2015; Eq. 1):

$$\frac{dN_i}{dt} = r_i N_i + \sum_{j=1}^S g_{ij}(N_i, N_j) N_j \quad (1)$$

Interestingly, this definition also matches with that of statistical time series models (Ives et al. 2003; Mutshinda et al. 2009, 2011; Hampton et al. 2013). Embracing that ecological systems are inevitably stochastic, an interaction can therefore be defined as a link from species  $j$ 's density to species  $i$ 's per capita growth rate in a stochastic dynamical system. This has also been referred to as local dependence (Schweder 1970), dynamic causation (Aalen 1987; Aalen et al. 2012; Sugihara et al. 2012), and Granger-Wiener causality (Granger 1969; Geweke 1982; Detto et al. 2012) in the statistics and theoretical ecology literatures.

To infer such dynamic causation from multiple time series, ecologists have used a range of statistical models, most notably multivariate autoregressive models of order one, or MAR(1) models (also called VAR(1)—vector autoregressive models—in the econometrics and neuroscience literatures). When the state variables are species densities, these are multispecies generalizations of the discrete-time Gompertz population growth (Ives et al. 2003; Mutshinda et al. 2009), including an interaction coefficient for each species pair. MAR( $p$ ) models, with a maximum time lag of order  $p \geq 1$ , generalize the MAR(1) framework to more complex dependencies over time, and have been shown to map more exactly to the celebrated Granger-Wiener causality concept (Granger 1969; Sims 1980; Ding et al. 2006; Chen et al. 2006; Barnett et al. 2009, 2014; Detto et al. 2012). Granger-Wiener causality (usually referred to as Granger causality or GC for short) is strongly tied to the physical notion that the cause must precede the effect. Using the temporal order of events for inferring the direction of causality matches the intuition of many biologists (Mayr 1961) and especially ecologists, familiar with predators lagging behind their prey population dynamics (May 1973).

Granger causality combines this idea of temporal precedence of the cause with statistical prediction. If a dynamical model for time series  $y$  has its in-sample predictive ability of future  $y$  values improved by inclusion of time series  $x$  in the predictors, we say that  $x$  Granger causes  $y$ . This is a purely operational definition of causality, yet it is rather general and does not specify any particular model framework. It can in principle be applied to

phenomenological and mechanistic frameworks alike, and it can be extended to nonparametric and spectral settings (Detto et al. 2012). However, the simpler parametric and linear MAR( $p$ ) models are often preferred for Granger causality testing (Lütkepohl 2005). Ecologists have in fact been using the Granger causality concept implicitly many times in the form of MAR(1) models (reviewed in Hampton et al. 2013).

In the last decade, new methods such as convergent cross mapping (CCM; Sugihara et al. 2012), using nonlinear dynamical systems theory and attractor reconstruction, have been introduced to infer interactions between species. Sugihara et al. (2012) criticized the application of Granger causality concept to nonlinear dynamical systems. They deemed GC best suited for linear systems dominated by stochasticity, and unfit to model systems with a highly nonlinear (chaotic) deterministic skeleton, as according to Takens' theorem, lagged values of each variable (e.g., species density) contain information about all other linked variables in the dynamical system. This viewpoint has been subsequently adopted by many ecology studies using CCM (e.g., Ye et al. 2015, 2016; Deyle et al. 2016b; Monster et al. 2017; Harford et al. 2017; Grziwotz et al. 2018). However, while it is correct that the information contained in deterministic dynamical systems cannot be ascribed to a single component of the dynamical system (referred to as “nonseparability” by Sugihara et al. 2012, which is actually true for nonlinear and linear dynamical systems alike, Granger 1969; Runge 2014), the addition of process noise may in fact allow separating the predictive abilities of  $y$  vs.  $(x, y)$  (Runge 2014, p. 19). Given that both nonlinearity and process noise are ubiquitous in ecology, Granger causality could be a potent approach to infer interactions. Moreover, although it can seem intuitive that the linear MAR( $p$ ) implementation of Granger causality will have difficulties with nonlinear time series, it should be noted that MAR( $p$ ) models are usually applied to log(population sizes) in ecological settings (Ives et al. 2003). This log-linear scale (i) allows transforming the log-normal distribution of abundance usually found in data into a normal one and more importantly (ii) transforms multiplicative growth processes into an additive model structure. MAR( $p$ ) models on the log-scale are therefore essentially power-law models when transformed back into the original scale, a flexible way to model monotonic nonlinearities, which is used to approximate nonlinear dynamical systems (Ives et al. 2003).

Modelling has indeed shown that even the simplest MAR(1) models can be surprisingly robust to nonlinearities (Ives 1995; Certain et al. 2018), correctly inferring the sign of interactions in the case of stochastic nonlinear competition with a fixed point and multiple predator-prey systems, including limit cycles. Further evidence that linear

GC can be robust to nonlinearities comes from studies that used linear GC, nonlinear GC, and CCM, and found consistent causal answers with all three (Hannisdal et al. 2017, 2018). These studies provide hints that Granger causality, even in its linear MAR( $p$ ) formulation, may apply well to stochastic and nonlinear ecological dynamical systems.

MAR( $p$ ) models have obvious strengths for causal inference: confidence intervals for coefficients, model selection, and other inferential tools are well understood (Lütkepohl 2005) based on decades of development in time series analysis. By contrast, CCM is relatively new and does not benefit (yet) from the same support from statistical and probability theory. Given the advantages stemming from the great conceptual and practical simplicity of MAR( $p$ ) models, there is a need to better understand in which ecological scenarios linear GC can be a good approximation for interaction inference, and in which cases more sophisticated techniques are needed, such as CCM (Sugihara et al. 2012) or entropy-based methods (e.g., Amblard and Michel 2013; Hannisdal and Liow 2018). With new monitoring tools like metabarcoding making community time series increasingly available, GC methods may become even more interesting for ecologists. For example, Granger causality techniques are currently gaining traction in the rapidly evolving microbiome field that attempts at inferring interactions from metabarcoding data on microorganisms (Gibbons et al. 2017; Mainali et al. 2019; Carr et al. 2019).

In this article, we evaluate the performance of linear MAR( $p$ ) models and compare it to CCM on a number of ecological examples for which CCM is currently thought to be more appropriate. We demonstrate that criticism of the Granger causality concept by Sugihara et al. (2012) may have been induced by nonstandard model selection and evaluation techniques. Using simpler model selection techniques, routinely used by statisticians to infer the lag order  $p$  of MAR( $p$ ) models as well as their parametric structure (Lütkepohl 2005), we show that Granger causality techniques can infer interactions in nonlinear time series surprisingly well. Granger causality and CCM either both work well or both fail to some degree for most case studies, which suggests that seemingly different causality concepts might in fact share hidden similarities. Throughout our analysis, we take care to consider both statistical significance and effect sizes of causal inferences. We then demonstrate that MAR( $p$ ) modelling can be scaled up to large interaction networks using either appropriate model regularization techniques (based on a structured version of the LASSO) or pairwise inference with an appropriate false discovery rate correction. A comparison to CCM is provided in the latter case.

## Methods and models

To start, we recall the basics of Granger causality concepts and MAR( $p$ ) modelling—multivariate autoregressive modelling of order  $p$ —which is the most common way to assess Granger causality (though by no means the only one, see, e.g., Detto et al. (2012) for a nonparametric and spectral Granger approach, Barnett and Seth (2014) for parametric and spectral approaches). We describe shortly thereafter convergent cross mapping (Sugihara et al. 2012), which takes a different approach to causal inference, based on dynamical systems theory and state-space reconstruction. We then describe the real datasets and numerical simulations that will be used for evaluating causal inference methods.

### Causality concepts

#### Granger causality and MAR( $p$ ) implementation

Formally, time series  $\mathbf{x} = (x_t)_{t \in [1, T]}$  Granger causes time series  $\mathbf{y} = (y_t)_{t \in [1, T]}$   $\Leftrightarrow$  including  $x$  in a time series model for  $y$  improves in-sample prediction of  $y$ . In the MAR( $p$ ) framework, this translates into performing two autoregressive model fits to explain time series  $\mathbf{y}$ , one with only  $y$  values and one with both  $y$  and  $x$  values:

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \quad (2)$$

$$y_t = \sum_{i=1}^p \beta_{1i} x_{t-i} + \sum_{i=1}^p \beta_{2i} y_{t-i} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2). \quad (3)$$

Granger causality is inferred if  $\sigma_\epsilon^2 < \sigma_\eta^2$  and such difference is statistically significant. A simple measure of effect size is therefore the log ratio of the sum of squared residuals  $G_{x \rightarrow y} = \ln\left(\frac{\sigma_\eta^2}{\sigma_\epsilon^2}\right)$  (Geweke 1982; Detto et al. 2012). When more than two variables are considered, pairwise GC has to be differentiated from conditional GC (Geweke 1984; Barnett and Seth 2014). Conditional GC occurs whenever a third variable  $z$  is considered and corrected for. When fitting a MAR( $p$ ) model to more than two species, we would typically be interested in conditional GC rather than pairwise GC, with conditional GC correcting for the effects of non-focal species and abiotic covariates. For instance, let us consider a MAR(1) model (4) with 3 species in the classic form of Hampton et al. (2013), where  $\mathbf{N}_t$  is the vector of population densities:

$$\mathbf{x}_t = \ln(\mathbf{N}_t), \quad \mathbf{x}_{t+1} = \mathbf{a} + \mathbf{B}\mathbf{x}_t + \mathbf{C}\mathbf{u}_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim \mathcal{N}_3(\mathbf{0}, \Sigma). \quad (4)$$

Here, the entries in the interaction matrix  $\mathbf{B}$  are defined by:

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \quad (5)$$

and  $\mathbf{C}$  is a matrix representing the effect of environmental covariates  $\mathbf{u}_t$  (Ives et al. 2003; Hampton et al. 2013). Whenever  $b_{12}$  is significantly different from zero, we have a causal influence  $x_2 \rightarrow x_1 | (x_3, \mathbf{u})$ , that is, an influence of  $x_2$  on  $x_1$  conditional to the population density  $x_3$  of species 3 and all the control environmental variables in the vector  $\mathbf{u}$ .

Using centered data so that the intercept disappears, the MAR( $p$ ) model is defined as:

$$\mathbf{y}_{t+1} = \sum_{q=1}^p \mathbf{B}^{(q)} \mathbf{y}_{t-q+1} + \mathbf{e}_t, \quad \mathbf{e}_t \sim \mathcal{N}_d(\mathbf{0}, \Sigma) \quad (6)$$

where  $d$  is the number of system components (individual time series). For a general definition of causal effects, we dropped  $\mathbf{u}_t$  from Eq. 6, as it corresponds to a special case where a subset  $\mathbf{u}_t$  of the variables  $\mathbf{y}_t = (\mathbf{x}_t, \mathbf{u}_t)'$  has a one-way causal impact (i.e.,  $\mathbf{u}_t$  affects  $\mathbf{x}_{t+1}$  but not the other way around, which can be specified as well by forcing the  $\mathbf{B}^{(q)}$  matrices to contain some zeroes). The condition for an interaction from system component  $j$  to system component  $i$  given all other system components (either species densities or environmental variables) then becomes, in a general MAR( $p$ ) setting (according to Eq. 6):

$$\exists b_{ij}^{(q)} \neq 0 \Leftrightarrow y_j \rightarrow y_i | (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_d) \quad (7)$$

where each time lag is indexed by  $q$ .

Conversely, pairwise GC testing between  $y_i$  and  $y_j$  is assessed through a bivariate autoregressive model for each  $(i, j)$  pair, and therefore uses a considerably lower dimensional model, although it may require a false discovery correction to attain meaningful statistical significance (see the next section).

To implement these concepts in practice, we fitted MAR( $p$ ) models using the package `vars` in R (version 3.4.4), which uses ordinary least squares for estimation. We mainly used the BIC as a default for lag order selection, although we also considered other information criteria (see below). The presence of Granger causality was assessed by the statistical significance and magnitude of the interaction matrix coefficients, and more directly using parametric significance tests for nested models. For pairwise Granger causality testing, we used the function `grangertest` in the R package `lmtest` (Zeileis and Hothorn 2002, v0.9-36) which performs a Wald test for nested models (based on the statistical significance of MAR model coefficients). For conditional Granger causality testing, we used the

function `causality` in package `vars` (Pfaff 2008, v1.5-3) which provides F-tests for nested models. Both tests and implementations provided similar answers when compared.

### Granger causality in high-dimensional models

If we have a large number of time series, corresponding to many species, fitting full MAR( $p$ ) models (i.e., models that account for all possible interactions without additional constraints) becomes impractical, unless those time series are extremely long (Michailidis and D’alché Buc 2013). For  $d$  species and  $p$  time lags, a  $d \times d \times p$  dimensional model needs to be fitted to the data. For instance, 10 species with  $p = 2$  yields  $2 \times 10 \times 10 = 200$  parameters in the interaction  $\mathbf{B}^{(q)}$  matrices only. Even a simpler MAR(1) model would be impossible to fit properly without a set of time series of length above 100 (or some added regularization). Preliminary simulations (Certain et al. 2018) suggest that a nonlinear, stochastic ecological system of dimension 10 or 12 requires approximately time series of length 500 to 800 to be fitted properly without implementing additional constraints. To deal with high dimensionality, we considered two solutions:

- Pairwise Granger causality testing with false discovery rate (FDR) correction (Benjamini-Hochberg), with a philosophy similar to Mukhopadhyay and Chatterjee (2006). This is done by fitting bivariate MAR( $p$ ) models, testing for Granger causality in both directions, and then re-adjusting the  $p$  values obtained through the Benjamini and Hochberg (1995) correction.
- LASSO-penalized MAR(1) models with structured penalties, using the R package `SIMoNe` (Chiquet et al. 2008 and Charbonnier et al. 2010, v1.0-3). This allows estimating (through non-zero interaction coefficients) conditional Granger causality. A naive idea would be to use the classic LASSO (Least Absolute Shrinkage and Selection Operator, Tibshirani et al. 2015) to set some of the coefficients to zero. Unfortunately, this approach is known to yield substantial bias whenever there is an important structure (here, modular) in the network (Charbonnier et al. 2010). The technique that we used explicitly accounts for network structure in addition to selecting coefficients with the LASSO, and is described in Electronic Supplementary Material Appendix S1.1.

### Convergent cross mapping

Convergent cross mapping (CCM) was proposed by Sugihara et al. (2012) as an alternative nonparametric method to detect dependencies between time series. CCM relies on state-space reconstruction. We assume two time series  $\mathbf{x} = (x_t)_{t \in [1, T]}$  and  $\mathbf{y} = (y_t)_{t \in [1, T]}$  as previously.

The attractor manifold  $M_X$  is constructed as a set of  $E$ -dimensional vectors  $\tilde{\mathbf{x}}(t) = (x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(E-1)\tau})$  for  $t = 1 + (E-1)\tau$  to  $t = T$ .  $E$  is the embedding dimension, denoting how many time lags one counts back in time. This set of vectors  $\{\tilde{\mathbf{x}}(t)\}$  constitutes the reconstructed manifold. We now find the  $E+1$  nearest neighbors of each  $\tilde{\mathbf{x}}(t)$  in  $M_X$ . Their time indices are denoted  $t_1, \dots, t_{E+1}$ . The reconstruction of  $y_t$  from  $M_X$  proceeds as follows:

$$\hat{y}(t)|M_X = \sum_{i=1}^{E+1} w_i y(t_i) \quad (8)$$

with  $w_i = u_i / \sum_{j=1}^{E+1} u_j$ , and  $u_j = \exp\left(\frac{-d(\tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(t_j))}{d(\tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(t_1))}\right)$  where  $d(\tilde{\mathbf{x}}(t), \tilde{\mathbf{x}}(t_1))$  is the minimal distance between  $\tilde{\mathbf{x}}(t)$  and all other embedded points.

The cross-map skill from  $X$  to  $Y$  is then measured by the correlation coefficient  $\rho(\mathbf{y}, \hat{\mathbf{y}}|M_X) > 0$ , which increases with the size  $L$  of the library of points used to reconstruct the manifold  $M_X$  if  $Y$  causes  $X$ . The surprising thing here is that predicting  $Y$  by  $M_X$  is equivalent to  $Y$  causing  $X$  and not the other way around (Sugihara et al. 2012). Hence, to know if  $X$  causes  $Y$ , we look at  $\rho(\mathbf{x}, \hat{\mathbf{x}}|M_Y)$ .

Due to the absence of a parametric model, there is no formula for the  $p$  value related to the CCM skill  $\rho$ . Several  $p$  value formulations have been proposed under the null hypothesis of no causality from  $X$  to  $Y$ :

- Cobey and Baskerville (2016) suggested  $p(X \leftrightarrow Y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_i (\rho(\mathbf{x}_{b,i}, \hat{\mathbf{x}}_{b,i}|M_{Y,\text{Lmax}}) < \rho(\mathbf{x}_{b,i}, \hat{\mathbf{x}}_{b,i}|M_{Y,\text{Lmin}}))$  where  $n$  is the number of libraries of size  $L$  that were used to build  $M_Y$  and  $\mathbf{x}_{b,i}$  are resampled (bootstrapped) values of the vector  $\mathbf{x}$ .  $M_{Y,\text{Lmax}}$  (respectively,  $M_{Y,\text{Lmin}}$ ) is the manifold constructed with the maximum (respectively, minimum) library size. Two versions of this  $p$  value can be computed depending on whether one samples with replacement (the bootstrap) for the libraries or without replacement (in which case  $M_{Y,\text{Lmin}}$  varies but not  $M_{Y,\text{Lmax}}$ ).
- When two species are forced by a shared forcing driver (e.g., seasonal temperature), spurious causality can emerge. This can be corrected by computing  $n$  surrogate time series  $\mathbf{x}_{\text{surr},i}$ , that keep the periodicity of the signal but shuffle its residuals, so that cross-correlations containing causal information are “erased.” Cross mapping is then computed on the surrogates and compared with the real value (Deyle et al. 2016a). In this case,  $p(X \leftrightarrow Y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_i (\rho(\mathbf{x}_{\text{real}}, \hat{\mathbf{x}}_{\text{real}}|M_Y) < \rho(\mathbf{x}_{\text{surr},i}, \hat{\mathbf{x}}_{\text{surr},i}|M_Y))$ . In fact, the more exact formula is  $p(X \leftrightarrow Y) = \frac{r+1}{n+1}$  where  $r = \sum_{i=1}^n \mathbb{1}_i (\rho(\mathbf{x}_{\text{real}}, \hat{\mathbf{x}}_{\text{real}}|M_Y) < \rho(\mathbf{x}_{\text{surr},i}, \hat{\mathbf{x}}_{\text{surr},i}|M_Y))$ , following North et al. (2002), which we employed here.

- Given that surrogate-based  $p$  values, required in the shared abiotic driver case (item above), were found to perform better than alternative  $p$  values in most contexts, we computed those systematically for all simulations. For model simulations where there was no confounding abiotic driver, surrogates were only computed by permutation of the time series. This was found to be simpler and more efficient than other techniques to provide statistical significance for CCM (ESM Appendix S2.1).

The analyses have been performed using the package rEDM (Ye et al. 2018, v0.7.1). For each time series, we retrieved the best embedding dimension (which maximizes the forecast skill of the simplex, Sugihara and May 1990) and used it in the cross-mapping function, with 100 different libraries for each library size and maximum library size depending on the length of the time series (300 timesteps if not mentioned otherwise). The libraries were obtained with random draws without replacement from the original time series. In high-dimensional cases, we used the same Benjamini-Hochberg correction as for GC.

## Evaluating GC and CCM

GC was evaluated using a criterion of  $p$  value below 0.1 (0.2 in a high-dimensional setting) or an effect size  $G_{x \rightarrow y} > 0.04$  (threshold found and justified in Appendix S2.1), or both criteria simultaneously. Our philosophy here has been to evaluate causality based on both statistical significance and effect sizes, to avoid the well-known drawbacks of considering solely statistical significance. The same philosophy is used for evaluating CCM, where we considered two thresholds for correlation coefficient  $\rho$  of the cross mapping (measuring the effect size), 0.1 and 0.2. These thresholds also originate from preliminary analyses presented in Appendix S2.1.

For each case study, we compared the values of classical scores such as recall or sensitivity (fraction of true interactions  $TP$  found over the total number of true interactions,  $\frac{TP}{TP+FN}$ , where  $FN$  are false negatives) and the specificity (fraction of true negatives  $TN$  over the total number of negatives,  $\frac{TN}{TN+FP}$ , where  $FP$  are false positives). Additionally, we measured similarity between GC- and CCM-detected causalities at the level of individual time series—within a single parameter set and model—to see if they detect matching causalities or have some degree of complementarity. We used the Sokal Michener index  $ISM = \frac{\sum \mathbb{1}_{11} + \sum \mathbb{1}_{00}}{\sum \mathbb{1}_{11} + \sum \mathbb{1}_{10} + \sum \mathbb{1}_{01} + \sum \mathbb{1}_{00}}$  where  $\sum \mathbb{1}_{11}$  indicates the number of simulations for which GC and CCM both indicate causality,  $\sum \mathbb{1}_{10}$  the number of simulations for which GC indicates causality and CCM does not, etc.

## Simulated and real case studies of interacting species population dynamics

### Real data: Veilleux's predator-prey cycles

The first two datasets that we consider are taken from Veilleux (1979) and have been analyzed by other authors with mechanistic models that demonstrated two-way coupling (Jost and Ellner 2000), plausibly with limit cycle behavior (Fig. 1a, b). We additionally created 500 simulated time series from MAR( $p$ ) models that best fit this dataset, to provide a “linear” dynamical version of this empirical system.

### Deterministic chaos in two-species competition models

Our second case study is the two-species discrete-time logistic competition model (Fig. 1c) used in Sugihara et al. (2012) to evaluate the performance of CCM:

$$x_{t+1} = x_t(3.8 - 3.8x_t - 0.02y_t) \quad (9)$$

$$y_{t+1} = y_t(3.5 - 3.5y_t - 0.1x_t) \quad (10)$$

Model parameters are identical to Sugihara et al. (2012), which places this model in the chaotic regime (Lyapunov exponent LE = +0.41). This case study therefore constitutes a strong test of the log-linear MAR( $p$ ) framework. The only setting that was modified compared with Sugihara et al. (2012) is the initial condition, which was randomly drawn from a Uniform(0,1) distribution 500 times. Although we acknowledge that “mirage correlations” can occur in some

datasets, we aimed at reproducing the full distribution of what this model can provide, as there are no justifications to favor one specific set of initial conditions (outside of illustration purposes). The sample size is taken to be  $T = 300$  as in Sugihara et al. (2012), after an initial run of 500 time steps that are discarded to remove transients.

Because a method that finds no interactions whenever absent (i.e., no false positives) is as important as one that finds interactions whenever they are present, we additionally created simulations without interactions:

$$x_{t+1} = x_t(3.8 - 3.8x_t - 0 \times y_t) \quad (11)$$

$$y_{t+1} = y_t(3.5 - 3.5y_t - 0 \times x_t) \quad (12)$$

We evaluated both GC and CCM’s ability to find no interactions between these time series.

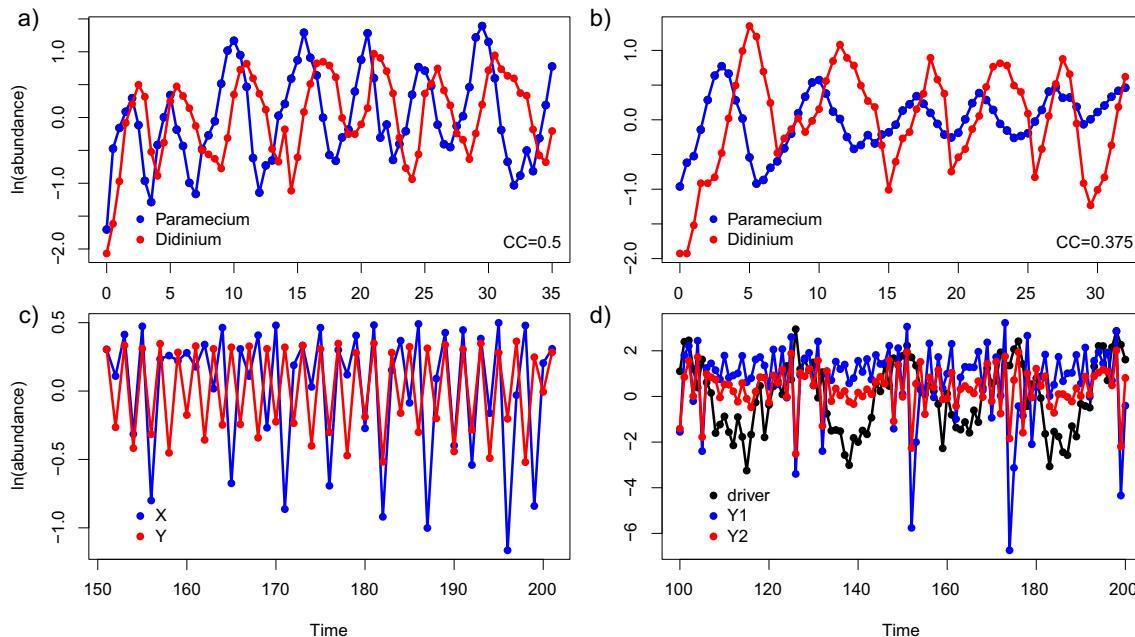
### Two-species stochastic and nonlinear dynamics, including environmental drivers

We consider here a stochastic two-species competition model, with Lotka-Volterra interactions in discrete time and a Ricker type of multispecies density dependence:

$$N_{1,t+1} = N_{1,t} \exp(3 - 4N_{1,t} - 2N_{2,t} + \epsilon_{1,t}) \quad (13)$$

$$N_{2,t+1} = N_{2,t} \exp(2.1 - 0.31N_{1,t} - 3.1N_{2,t} + \epsilon_{2,t}). \quad (14)$$

An illustration is provided in ESM S2.1. This case was already investigated in Certain et al. (2018), including as well an environmental driver on species 1 (but not species 2). The model of Eqs. 13 and 14 has a stochastic Lyapunov



**Fig. 1** Time series of small-community models. Veilleux’s predator-prey data are shown in **a** (dataset CC05) and **b** (dataset CC0375); an example simulation for the 2-species chaotic model is shown in panel **c**

and a simulation of the competition model including an environmental driver is illustrated in panel **d**

exponent (SLE) of  $-0.18$ , and therefore is not sensitive to initial conditions when perturbed by noise (Ellner and Turchin 2005). The stochastic Lyapunov exponent, as elsewhere in this manuscript, was computed following Dennis et al. (2001). The dynamics of the corresponding deterministic skeleton are a two-cycle when there are interactions between species, while without interactions, species 1 is chaotic and species 2 has a two-cycle.

As a fourth case study, we created a scenario to investigate the effect of environmental drivers on the estimation of species interactions (Fig. 1d). This is done with a variant of Eqs. 13 and 14 by adding an environmental driver  $u_t$  that has the same effect on both species, which constitutes a challenge for any causal method ( $u_t$  is a confounding variable):

$$N_{1,t+1} = N_{1,t} \exp(3 + 0.5u_t - 4N_{1,t} - 2N_{2,t} + \epsilon_{1,t}) \quad (15)$$

$$N_{2,t+1} = N_{2,t} \exp(2.1 + 0.5u_t - 0.31N_{1,t} - 3.1N_{2,t} + \epsilon_{2,t}). \quad (16)$$

We considered, as in the deterministic case, the counterparts of the above models where the interspecific interactions are set to zero, i.e.,

$$N_{1,t+1} = N_{1,t} \exp(3 + 0.5u_t - 4N_{1,t} - 0 \times N_{2,t} + \epsilon_{1,t}) \quad (17)$$

$$\mathbf{A} = \begin{pmatrix} -4 & -2 & -0.4 & 0 & 0 & 0 & 0 \\ -0.31 & -3.1 & -0.93 & 0 & 0 & 0 & 0 \\ 0.636 & 0.636 & -2.12 & 0 & 0 & 0 & 0 \\ -0.111 & -0.111 & 0.131 & -3.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & -2 & -2 & -0.4 \\ 0 & 0 & 0 & 0 & -0.31 & -3.1 & -0.93 \\ 0 & 0 & 0 & 0 & 0.636 & -0.636 & -2.12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (20)$$

This Lotka-Volterra model has a stochastic Lyapunov exponent (SLE) of  $+0.33$ . This positive SLE clearly places the model in a noisy chaotic regime (Ellner and Turchin 2005). In addition, we used the Jacobian matrix (of the model in Eq. 19) as the interaction matrix of a MAR(1) model, which has therefore comparable interaction strengths but non-chaotic dynamics. In this case, the dynamical equation is written as:

$$\mathbf{x}_{t+1} = \mathbf{J}\mathbf{x}_t \quad (21)$$

$$\text{with } J_{ij} = \delta_{ij} + a_{ij}N_j^* \quad (22)$$

where  $\mathbf{x}_t = \mathbf{n}_t - \mathbf{n}^*$ , with  $n_t = \ln(N_t)$  and  $\mathbf{n}^*$  being the equilibrium on log-scale,  $\delta_{ij} = 1$  if  $i = j$  and  $\delta_{ij} = 0$  otherwise (see derivation in the ESM Appendix S1.3). By definition, such MAR(1) models have a single fixed point forced by stochasticity when stable (Ives et al. 2003): they

$$N_{2,t+1} = N_{2,t} \exp(2.1 + 0.5u_t - 0 \times N_{1,t} - 3.1N_{2,t} + \epsilon_{2,t}). \quad (18)$$

We ran 500 simulations for each model. The noise was set so that  $\epsilon_{i,t} \sim \mathcal{N}(0, \sigma^2)$  i.i.d. with  $\sigma^2 = 0.01$ , but one should keep in mind that the addition of the environmental drivers increases the level of noise in this system. This slightly increases the Lyapunov exponent ( $\approx -0.15$  for the coupled system).

### Ten-and twenty-species interaction webs

We simulated a ten-species model, which generalizes the two-species Ricker competition to more species and more interaction types, with added stochasticity ( $\sigma^2 = 0.1$ ). This model therefore represents a considerable challenge to interaction inference, due to the large quantity of potential false positives (many zero interactions) combined to both nonlinear dynamics and stronger stochasticity. The dynamical equation can be written as:

$$\mathbf{N}_{t+1} = \mathbf{N}_t \circ \exp(\mathbf{r} + \mathbf{AN}_t + \mathbf{e}_t), \mathbf{e}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \quad (19)$$

where  $\mathbf{N}$  is the abundance vector,  $\sigma^2 = 0.1$  is the process noise variance, and the interaction matrix  $\mathbf{A}$  is defined to be:

$$\text{cannot exhibit chaos and therefore exhibit negative SLEs. We ran 25 simulations over 500 time steps with different initial conditions, for both the chaotic LV model and its log-linear MAR(1) counterpart. We analyzed the last 300 time steps. We slightly modified this model to scale it up to 20 species, with a structure that is still very modular (Eq. S11 in ESM S1.4). For the 20-species model, we also compared Ricker and MAR(1) dynamics for 25 different simulations over 1000 time steps and analyzed the last 700. In the 20-species case, coefficients were drawn from a probability distribution (Eqs. 23 and 24) and therefore differ from one simulation to the next, although we have taken care to avoid coefficients too close to zero by imposing a lower bound:}$$

$$a_{ij} = \chi_{ij} [a_{min} + (a_{max} - a_{min})\text{Beta}(2, 2)] \quad (23)$$

with the bounds of the interaction coefficient selected as:

$$(a_{min}, a_{max}) = \begin{cases} (0.05, 0.1) & \forall i \neq j, \text{ with probability 0.2 (positive interaction)} \\ (-0.2, -0.1) & \forall i \neq j, \text{ with probability 0.8 (negative interaction)} \\ (-0.8, -0.3) & \forall i = j \end{cases}. \quad (24)$$

This construction of the interaction coefficients allows having some realistically strong dominance of the diagonal coefficients, a certain percentage of weak facilitation (20%), and marked competition between species whenever interactions are allowed by the network structure. The 20-species Ricker models thus constructed have SLEs slightly below zero (mean =  $-0.05$ , SD = 0.04), and are therefore less nonlinear (sensu sensitivity to initial conditions) than the 10-species models considered above. The deterministic skeleton of the model has additionally always a stable fixed point. For all datasets, real and simulated alike, the data have been log-transformed and centered before analysis, which is required for GC and does not change performance for CCM (ESM Appendix S2.2). We used a FDR of 20% in all pairwise high-dimensional analyses.

## Results

In the following, we first report both GC/MAR( $p$ ) and CCM results for each dataset or model. The results are then summarized in Fig. 6.

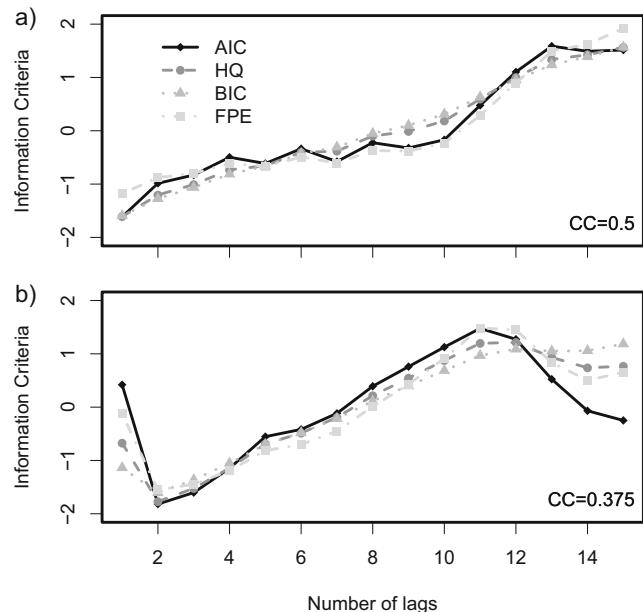
### Real data: Veilleux's predator-prey cycles

On those two datasets, both GC and CCM correctly identify the two-way predator-prey coupling. Surprisingly, CCM also identifies reciprocal causal influences in the linear MAR( $p$ ) approximation. Model selection of MAR( $p$ ) model by all information criteria selected a lag  $p = 1$  for the CC0.5 dataset and a lag of 2 for the CC0.375 dataset (Fig. 2). The  $p$  values for the GC test (null hypothesis: “no GC”) and associated effect sizes demonstrate convincingly that the “no GC” hypothesis can be rejected, for both datasets (Table 1).

CCM also demonstrates bi-directional causality, as demonstrated by the substantial increase in  $\rho(\mathbf{x}, \hat{\mathbf{x}}|M_Y)$  with library size  $L$  in both directions (Fig. 3a and c). This is true for the real data (with or without log transformation, ESM Fig. S5), but also many MAR(1)-simulated dataset using the fitted MAR(1) as the data-generating model (Fig. 3b and d).

### Deterministic chaos in two-species competition models

In the two-species chaotic competition model, high-order temporal lags tend to be selected (ESM Fig. S6) despite the single time lag considered in the simulation model (i.e., higher nonlinearity is expressed as high-order lags). The optimal lag is  $p = 7$  for the model with interactions ( $p = 3$  without interactions), for which we report the results in Table 2. Despite this potential overparameterization, the GC tests show that causality is detected for most lag orders (including  $p = 7$ ) whenever causality is present (ESM Fig. S7). Furthermore, the tests are not able to reject the null hypothesis of no GC when GC is not present (Table 2, Fig. S7 in Appendix for  $p < 7$ ), and the false positive rate is close to the assumed 10%. GC performs therefore surprisingly well in this chaotic context. CCM performs well when considering a simulation model with interactions,



**Fig. 2** Model information criteria as a function of lag order for the predator-prey data, for the two datasets. **a** CC = 0.5 and **b** CC = 0.375

**Table 1**  $p$  values for  $H_0$ : { No Granger causality between  $x$  and  $y$  } and effect sizes of GC

Dataset	CC = 0.5		CC = 0.375	
Lag $p$ in MAR( $p$ )	1		2	
Metrics	p-val	$G_{x \rightarrow y}$	p-val	$G_{x \rightarrow y}$
1 → 2	$2.79 \times 10^{-11}$	0.76	0.0409	0.09
2 → 1	$1.76 \times 10^{-14}$	1.02	0.0464	0.10

but not substantially better than GC concerning the weak causal effect  $2 \rightarrow 1$  (Table 2). The false positives (type I error) are somewhat higher in one causal direction for CCM (Table 2), up to 27% (instead of 10%) unless all  $\rho$  values below 0.2 are discarded (i.e., thresholding based on effect sizes). This is because a large number of simulations still show an increase of  $\rho$  with the library size  $L$  even though there is no causality (ESM Fig. S8). This may have been missed in Sugihara et al. (2012) because specific sets of initial conditions were selected, instead of drawing 500 at random as done here.

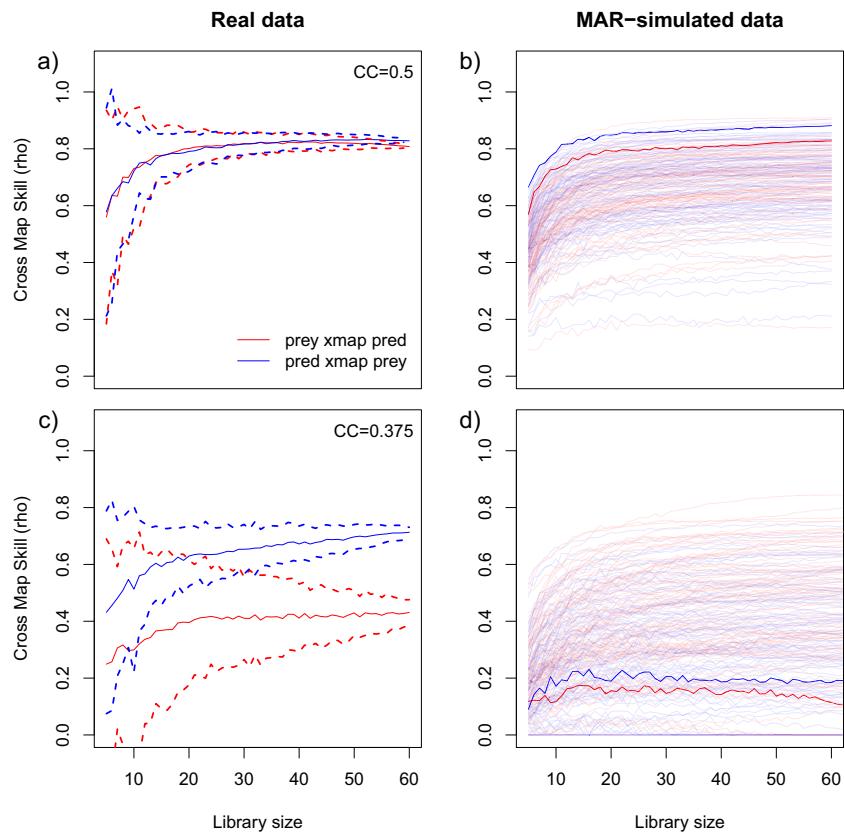
CCM and GC are in general in agreement for specific simulations (i.e., specific initial conditions) corresponding to this model and parameter set: the  $I_{SM}$  similarity index is close to 1, except for the weak interaction  $2 \rightarrow 1$ .

## Two-species stochastic and nonlinear dynamics

### Without environmental driver

In our case study with two-species nonlinear competition and noise, we see that GC and CCM perform quite similarly (Table 3), with both methods able to select properly causality in most cases (>95%). CCM has slightly better rates of interactions found (no false negatives), while GC is a little more conservative, especially when considering a threshold for  $G_{x \rightarrow y}$ , the logarithm of the sum of squares ratio (Table 3). The false positive rate is close to the required 10% level for both methods. Similarity indices are very close to 1, so that the two methods yield essentially similar conclusions when applied to the same time series.

**Fig. 3** Convergent cross mapping for Veilleux's CC0.5 dataset (**a** and **b**) and CC0.375 dataset (**c** and **d**). Dotted lines represent the confidence bands (+/- 2 SD), obtained by bootstrapping. **b** and **d** present CCM analyses on data that were simulated using the best-fitting MAR( $p$ ) models to the Veilleux datasets



**Table 2** Percentages of simulations with Granger causality or CCM between  $x$  (species 1) and  $y$  (species 2) over 500 simulations, for the chaotic 2-species competition model, with interactions (top rows) and without (bottom rows)

Method	Granger	Causality	CCM					$I_{SM}$
Thresholds	pval<0.1	$G_{x \rightarrow y} > 0.04$	Both	pval<0.1	$\rho > 0.1$	$\rho > 0.2$	Both0.1	Both0.2
With inter.								Both
1 → 2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1
2 → 1	50.6	69.8	50.6	56.6	54.8	29.4	54.2	29.4
Without								
1 → 2	11.2	100.0	11.2	17.4	10.0	0.6	10.0	0.6
2 → 1	11.2	2.0	2.0	27.4	27.2	12.2	26.4	12.2

The label Both0.1 (resp. Both0.2) corresponds to a combined detection criterion of  $p$  value  $< 0.1$  and  $\rho > 0.1$  (resp.  $\rho > 0.2$ ). Similarity of causality estimates is indicated by the Sokal-Michener ( $I_{SM}$ ) index, with both significance and effect sizes taken into account for GC and CCM ( $p$  value  $< 0.1$  and  $G_{x \rightarrow y} > 0.04$  or  $\rho > 0.1$ , respectively)

The MAR( $p$ ) model selected by BIC had a lag  $p = 2$  to 6 timesteps (mean  $p = 4$ ) whenever interactions were present,  $p = 2$  whenever interactions were absent, confirming that small lags should be used in such models.

### With an environmental driver

The two-species model with a shared environmental driver (e.g., temperature) is considerably more complex and yields less clear-cut results than stochastic two-species competition. Regarding interactions, CCM is better at uncovering interactions that are present, as GC has a good performance for the strong interaction  $2 \rightarrow 1$  but not the reverse interaction  $1 \rightarrow 2$ . Both GC and CCM have difficulties indicating non-causality (when there are no interactions), and indicate false positives three to four times above the 10% level of the test (Table 4). Thresholding small effect sizes works for GC but does not solve the issue for CCM. Conditional vs. pairwise GC has overall similar performance, conditional GC performs even a little worse: there is little gain in conditioning for temperature in this

case. Additional examination of temperature coefficients (ESM Appendix S2.4) shows that poor estimation of temperature effects is to blame. We use seasonal surrogate time series to assess the significance of CCM, which clearly improves its power to detect interactions, but we still have spurious causalities in CCM when no interactions are present (ESM Fig. S12). This is therefore a scenario where avoiding false causalities is more difficult for CCM—though we should not forget that approximately 67–86% of absent interactions (Table 4) are still being discovered as such by CCM, using both significance and effect size.

### Larger interaction webs

Here, we report the results of analyses for 10- and 20-species modular interaction webs. Lag order selection revealed that low-order MAR( $p$ ) models were selected (ESM Fig. S13), with the BIC indicating  $p = 1$  as the most parsimonious choice. Hence, we have focused on MAR(1) models. The high-dimensional  $S \times S$  MAR(1) models include clustering (see Methods and Appendix S1.1)

**Table 3** Percentages of simulations with Granger causality or CCM between species 1 and species 2 over 500 simulations, for the stochastic 2-species competition model without environmental driver, with interactions (top rows) and without (bottom rows)

Method	Granger	Causality	CCM					$I_{SM}$	
Thresholds	pval<0.1	$G_{x \rightarrow y} > 0.04$	Both	pval<0.1	$\rho > 0.1$	$\rho > 0.2$	Both0.1	Both0.2	Both
With inter.									
1 → 2	98.4	94.0	94.0	100.0	100.0	100.0	100.0	100.0	0.94
2 → 1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1.00
Without									
1 → 2	12.6	0.2	0.2	12.6	11.0	0.2	10.4	0.2	0.89
2 → 1	8.2	0.6	0.6	12.4	10.8	1.4	10.4	1.4	0.89

The label Both0.1 (resp. Both0.2) corresponds to a combined detection criterion of  $p$  value  $< 0.1$  and  $\rho > 0.1$  (resp.  $\rho > 0.2$ ). Similarity of causality estimates is indicated by the Sokal-Michener index ( $I_{SM}$ , between 0 and 1), with both significance and effect sizes taken into account for GC and CCM ( $p$  value  $< 0.1$  and  $G_{x \rightarrow y} > 0.04$  or  $\rho > 0.1$ , respectively)

**Table 4** Percentages of simulations with Granger causality or CCM between species 1 and species 2 over 500 simulations for a model with 2 species and a driver (temperature), with biotic interactions (top rows) and without (bottom rows)

Method	GC pval<0.1	Pairwise $G_{x \rightarrow y} > 0.04$	Both	GC pval<0.1	Conditional $G_{x \rightarrow y} > 0.04$	Both	CCM pval<0.1	$\rho > 0.2$	Both	$I_{SM}$ Both
Thresholds										
With										
1 → 2	39.4	13.6	13.6	45.0	17.8	17.8	99.6	98.4	98.4	0.19
2 → 1	97.8	92.2	92.2	97.8	90.8	90.8	100.0	99.8	99.8	0.91
Without										
1 → 2	36.2	10.2	10.2	42.2	14.4	14.4	41.8	36.2	33.0	0.68
2 → 1	35.2	8.8	8.8	36.2	9.8	9.8	38.6	14.0	14.0	0.82

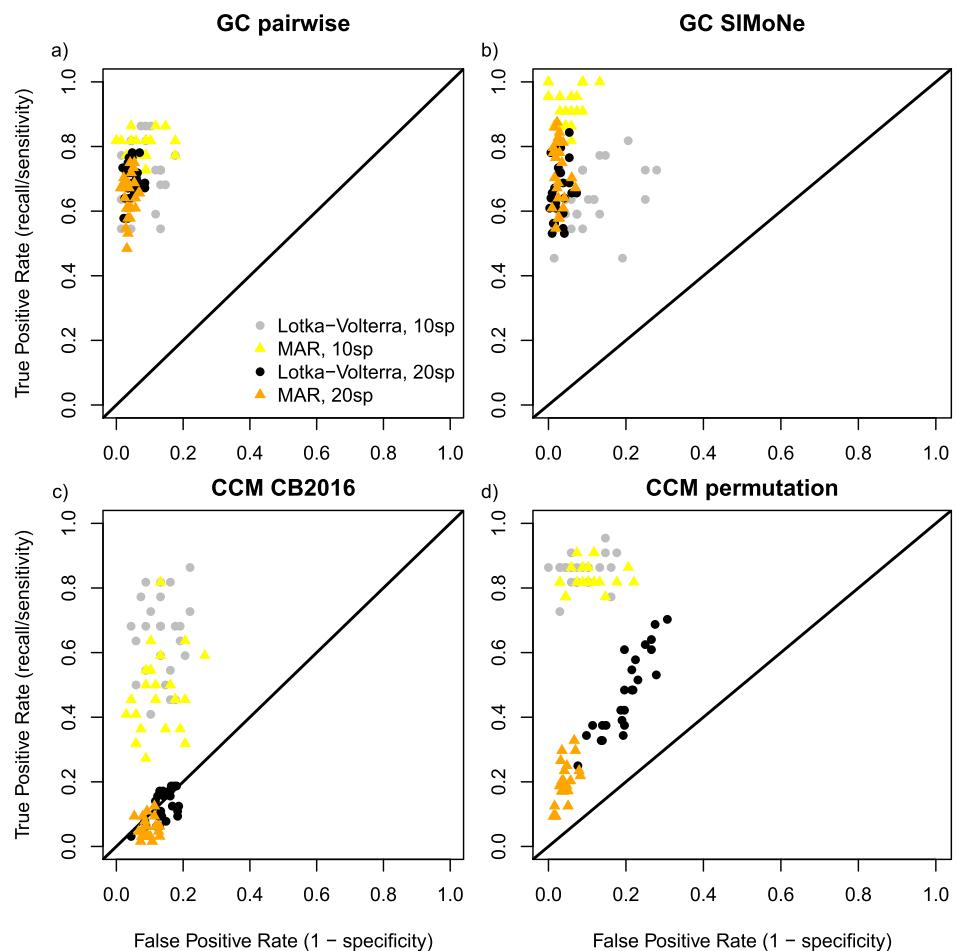
Similarity of causal links is indicated by the Sokal-Michener index ( $I_{SM}$ , between 0 and 1), with both significance and effect sizes taken into account for GC and CCM ( $p$  value  $< 0.1$  and  $G_{x \rightarrow y} > 0.04$  or  $\rho > 0.2$ , respectively). Causalities related to the temperature—not interactions—for CCM are shown in ESM Appendix S2.6

because the basic LASSO-penalized MAR(1) models poorly identify modular interaction webs (Charbonnier et al. 2010).

The recall, or true positive rate, that records how many actual interactions are identified as such, is on average above 70% for the structured LASSO (SiMoNe, Chiquet et al. 2008; Charbonnier et al. 2010) and goes up to 80% (90% for the MAR model with 10 species, where it

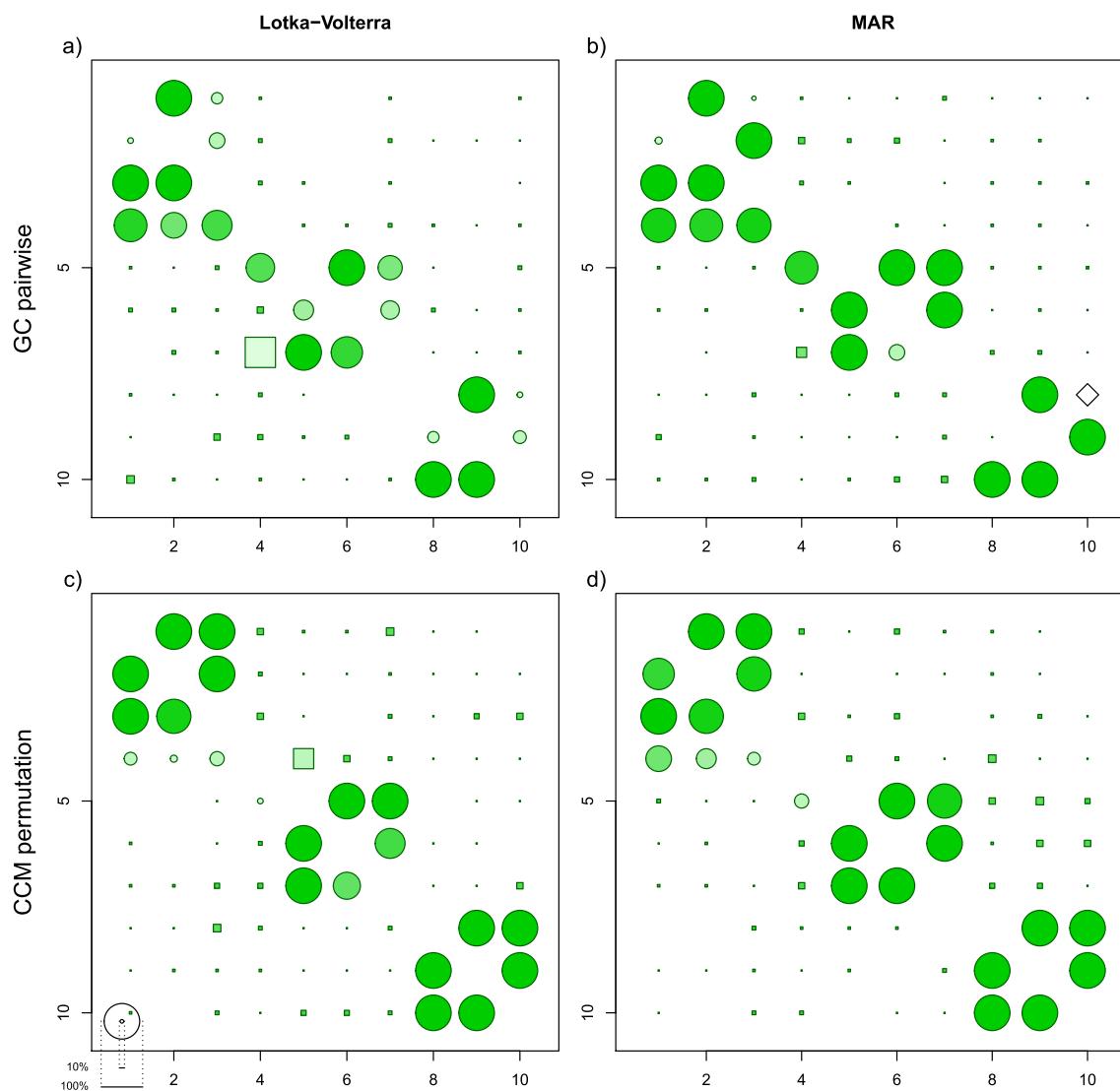
performs better than pairwise Granger causality testing). For the chaotic Lotka-Volterra model with 10-species, the method yields a little more false positives for some simulations than pairwise Granger causality (Fig. 4a and b). For the 20-species models, where interactions are weaker and the dynamics are not chaotic, we have overall a similar performance of pairwise Granger causality testing and the

**Fig. 4** ROC point clouds for the 10- and 20-species model using Granger causality (top) and CCM (bottom), with different ways of computing causality. For GC, **a** corresponds to results from pairwise GC; **b** is computed with structured-LASSO (SiMoNe). For CCM, we used both  $p$  values computed as **c**  $\Pr(\rho(L_{min}) < \rho(L_{max}))$  as suggested by Cobey and Baskerville (2016) and **d** surrogate-based  $p$  values. In the 10-species system, one chaotic reference parameter set is considered with many initial conditions, while in the 20-species model parameters vary for each simulation. The 20-species model is a perturbed fixed point, with negative SLE. The MAR-simulated model is always the MAR(1) model obtained using the Jacobian of the Lotka-Volterra model as an interaction matrix, hence a linearization in log-scale



structured LASSO, with a little higher average sensitivity but also a little more variance for the latter. Both methods show low false positive rates in the 20-species case, below 10%.

Comparing GC and CCM in “ideal” conditions, with the best-performing algorithms for each method (pairwise GC with a Benjamini-Hochberg correction and CCM with surrogate-based  $p$  values), reveals that they reconstruct similar networks for the 10-species case (Fig. 5), both for the chaotic Lotka-Volterra models and the linear (MAR) approximation derived from the Jacobian matrix. Note that the MAR models have milder dynamics (no chaos), but interactions are still strong since their Jacobian matrices match those of the 10-species Lotka-Volterra, where strong interactions have been modelled.

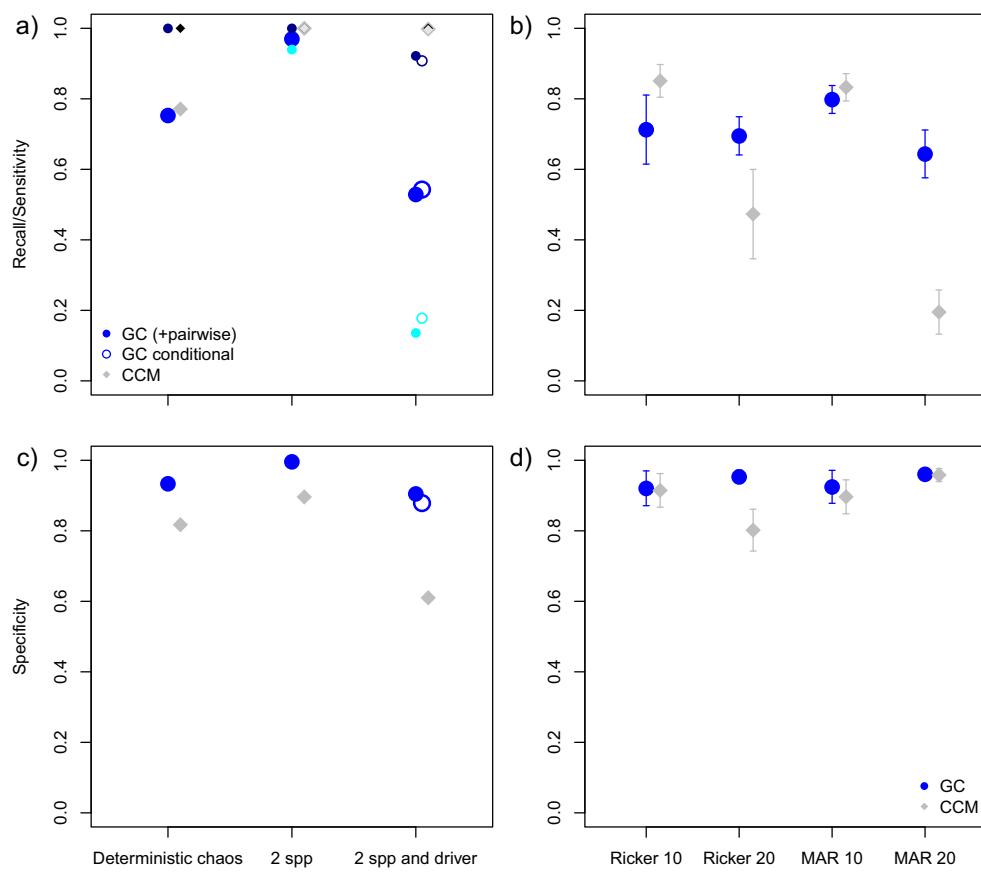


**Fig. 5** Interaction matrices obtained from pairwise GC (top, **a** and **b**) or pairwise CCM (bottom, **c** and **d**) for 10-species communities. Green circles represent true positives, green squares false positives, and empty diamonds false negatives. For true and false positives, the

size of circles and squares is proportional to the proportion of detection over 25 simulations. The color scale is set so that a darker green corresponds to a better performance

### Summary of the results for all simulated case studies

In Fig. 6, we present the recall or sensitivity (fraction of true positives among all positives) and the specificity (fraction



**Fig. 6** Recall and specificity for Granger causality (GC, blue dots) and convergent cross mapping (CCM, gray diamonds), ranked from most to least nonlinear dynamics. Causalities are considered only when they are significant at the 0.1 level (small dimension, **a** and **c** or 0.2 level (large dimension, **b** and **d**). For small dimensionality (**a** and **c**), an additional threshold on effect sizes is considered ( $G_{x \rightarrow y} > 0.04$  for GC,  $\rho > 0.1$  for CCM). Large symbols represent the value of recall and specificity over 500 simulations for two interactions at a time for

small communities. In **a**, the smaller symbols represent the stronger (respectively weaker) interaction in darker (resp. lighter) shade. For the 2 spp. and driver simulation, unfilled circles in **a** and **c** are obtained through conditional Granger causality testing. The smaller circles correspond to the weak interaction (light blue) and the strong interaction (dark blue). For large dimensionality, error bars represent the mean (+/− SD) value of recall and specificity over 25 simulations, since the estimated network varies across simulations

of true negatives among all negatives) for all simulated case studies present in this article. The case studies are ranked by descending order of nonlinearity, from the most nonlinear model (chaotic) to the most linear, for both the low-dimensional systems considered (2 species) and high-dimensional systems (10 and 20 species). Both metrics should be close to 1 for model performance to be high; a high recall is important when the objective of a study is to find all interactions present, and a high specificity is paramount when false positives are costly.

While we would have expected an increase in performance of GC as the dynamics are less nonlinear, combined to a decrease in performance of CCM, Fig. 6 shows a broad overlap in the performances of both methods, both in high- and low-dimensional systems. CCM has slightly higher false positives rates—lower specificity than the required  $\alpha$  level—and higher recall, except for the nonlinear 20 species systems, on which it performs worse. On the other hand,

GC has difficulty finding weak interactions when there is a confounding abiotic driver. The big picture, however, is that both methods display reasonable performances in most situations. A relatively high specificity, which is a key requirement of any interaction-finding method (otherwise, the method just outputs false positives), is found in nearly all cases.

## Discussion

The purpose of this paper was to evaluate the performance of linear and parametric methods for detecting Granger causality (GC) between time series, when the underlying community dynamics are nonlinear, and to compare such performances to a nonparametric and nonlinear popular alternative, convergent cross mapping (CCM). Our main results are that linear GC, implemented using MAR( $p$ )

models, is fairly robust to nonlinearities in ecological dynamics, when applied on the appropriate logarithmic-abundance scale and combined with model selection by information criteria. This was found to be true for all considered nonlinear simulation models, including those demonstrating chaos (Fig. 6). This confirms and extends findings from an investigation of the robustness to nonlinearities of log-linear MAR(1) models (with  $p$  restricted to 1 lag; Certain et al. 2018).

Comparison with the CCM framework by Sugihara et al. (2012) further revealed that CCM and MAR( $p$ ) / Granger causal modelling can in fact—surprisingly—yield relatively similar results in nonlinear and stochastic dynamical systems of interacting species. Evidence for this comes *both* from highly nonlinear systems for which CCM and GC infer the same interactions (deterministic chaos, stochastic competition) and from cases where both methods seem to fail to some degree, such as two competing species forced similarly by a shared environmental driver, where false positives are frequent (between 15 and 40%) for CCM and false negatives frequent for GC (one causal direction missed). Therefore, an important conclusion from our study is that both Granger causality and CCM yield mostly similar inferences on similar datasets, with only a couple of exceptions.

Moreover, we used here false discovery rate corrections and regularized models (i.e., LASSO-penalized models for modular interaction networks; Charbonnier et al. 2010) to tackle relatively high-dimensional systems (10 and 20 species). This allows to better infer Granger causality in lifelike contexts that, we surmise, will be most exciting to ecologists working on interacting species using community-level data. Our results demonstrated that simple pairwise Granger causality (i.e., using  $2 \times 2$  MAR( $p$ ) models many times with a correction for multiple testing) can be as good as the penalized MAR(1) models in finding the interaction network. We elaborate on these results and possible explanations below.

### **Can Granger causality be applied to highly nonlinear coupled dynamical systems?**

Our results showed that Granger causality, in its log-linear form, is robust to the presence of nonlinearities in the underlying multivariate dynamical system. Nonlinear variants of Granger causality (Marinazzo et al. 2008; Yang et al. 2017; Hannisdal and Liow 2018) can also be used to infer interactions in nonlinear and stochastic dynamical systems. These conclusions are further supported by the neuroscience literature (Ding et al. 2006; Chen et al. 2006; Marinazzo et al. 2008; Papana et al. 2013; Barnett and Seth 2014) which, unlike ecology, commonly uses Granger causality on nonlinear (and stochastic) dynamical systems.

Ecologist views over Granger causality have likely been shaped by the influential paper of Sugihara et al. (2012), who suggested that Granger causality would work well for simulated (log)-linear systems (which they referred to as “stochastic”) while CCM would work well for near-deterministic nonlinear dynamical systems. Given the history of both techniques, this makes intuitive sense. However, our tests on simulated data revealed that the domains of applicability of both techniques overlap to a great extent. Several differences between our analyses and those performed by Sugihara et al. (2012) (in their Supplementary Material) allow explaining this greater overlap, which we develop here.

First, Granger causality analyses performed by Sugihara et al. (2012) on the Veilleux and other datasets rely on a slightly dated model selection procedure (pre-information criteria) which produced overparameterized autoregressive models with very long lags (e.g.,  $p > 10$ ). Here, re-analyzing the data with a more classic, information-criteria motivated lag order selection, we have shown that GC is perfectly able to find causality in the classic Veilleux *Paramecium-Didinium* predator-prey datasets that were used as a key nonlinear example by Sugihara et al. (2012).

Second, we sampled many chaotic simulated datasets, corresponding to many initial conditions. Although some chaotic datasets may be difficult to identify for GC techniques, these are very few, as MAR( $p$ ) models and GC inference detected above 95% of the strong true interactions (where CCM found 100%), while both methods detected 50% of weak true interactions, in two-species chaotic models. This result was quite unexpected, as we thought that CCM would completely dominate the scores. Thus, GC testing can be useful for highly nonlinear systems, even chaotic ones, and it additionally tends to produce correct rates of false positives when the null hypothesis of no interactions is true (an important aspect as well).

Third, we found that data simulated with log-linear autoregressive models can also be well-identified by CCM, in both low and high dimensions, even though CCM relies upon the possibility to reconstruct an attractor in state space. This is further proof of the overlap between the domains of applicability of linear GC and CCM. Incidentally, it is also proof that the assumption of a deterministic dynamical system, often seen as a pre-requisite for CCM (Runge et al. 2019a; Langendorf and Doak 2019), may in fact not be needed.

### **How can Granger causality and convergent cross mapping yield similar inferences, in spite of seemingly opposite assumptions?**

Here, we would like to go back to the heart of the issue that Sugihara et al. (2012) highlighted, i.e., “causality reversion”

in nonlinear dynamical systems. Note that while we offer some suggestions as to how GC and CCM performances can overlap to a large extent, we have no definite answer as to why (mathematically speaking); more research is needed on that point.

The standard Granger causality concept holds that whenever a probabilistic model  $Y_{t+h}|(Y_k, X_k)_{k \in [t-p+1, t]}$  better predicts the observed time series  $\mathbf{y} = (y_t)$  than a model  $Y_{t+h}|(Y_k)_{k \in [t-p+1, t]}$ , then “ $x$  is causal for  $y$ .” Most often the time horizon for prediction considered is  $h = 1$ , which is the perspective adopted in this article. CCM instead holds that causality flows from  $x$  to  $y$  whenever  $\rho(\mathbf{x}, \hat{\mathbf{x}}|M_Y)$  increases strongly with the library size  $L$  used to reconstruct  $\mathbf{x}$  from the shadow manifold  $M_Y$ . It seems that in the latter method,  $x$  causes  $y$  whenever knowledge about  $y$  can be used to reconstruct  $x$ . However, verbal reasoning is treacherous there. To determine whether  $x$  causes  $y$ :

- GC compares knowledge about  $Y_t$  vs. knowledge about  $X_t, Y_t$  in prediction of  $Y_{t+1}$
- CCM compares knowledge about  $M_Y$  vs. no knowledge about  $M_Y$  in prediction of  $X_t$ .

There is no direct conditionality upon past  $X_t$  values in the prediction step of the algorithm for CCM. Thus, there seems to be no causality reversion that is intrinsic to nonlinear dynamic testing: GC and CCM are simply two different types of causal inference that are based upon *different assumptions on the conditioning set and ways to select models*.

Finally, GC and CCM both try, as do other approaches based on continuous-time stochastic processes and martingale theory (Commenges and Gégout-Petit 2009; Aalen et al. 2012), to reconstruct a stochastic dynamical system where interactions are defined as influences of state variables upon the rates of change of the system. No matter how different their historical origins are, GC and CCM are bound to exhibit some similarities because they define interactions in a similar manner.

### **Inferring interactions under environmental forcing**

We observed a lowered performance of linear GC in detecting interactions (whenever these were present) in the case of competition with a shared abiotic driver (temperature, which was both autocorrelated and seasonal). Notably, the weak interaction was entirely missed. This seems to be partly due to poor temperature effect estimates. Indeed, there was a bias in the temperature effect estimates (underestimation, ESM Appendix S2.4), that was still present when taking only statistically significant estimates. Additional analyses (in our `2species_driver` code folder; Barraquand and Picoche 2020) have demonstrated that better temperature effect estimates could be obtained if

the temperature consisted of white noise or if the underlying deterministic skeleton of the model was more linear (stable fixed point obtained by lowering intrinsic growth rates). This mirrors results of Certain et al. (2018) who found a good estimation of the temperature effect on growth rates (on one species only) for a stable fixed point in competition models. The fact that we have a good estimation of causal effects with nonlinear models forced by a white noise driver is consistent with our study of the two-species Ricker competition model without the environmental driver: adding a white noise driver amounts to increasing the level of background stochasticity, not its nature. Hence, it seems to be the interaction between a strongly autocorrelated forcing noise and a nonlinear (albeit nonchaotic) stochastic dynamical system that made the estimation challenging for linear GC here. This was true with and without interactions between the species, leading to false negatives, as stated above, but also false positives (although these remained around 10% when using cutoffs on effect size). Because the temperature was seasonal, using a seasonal dummy variable helped improve temperature estimates, but they still were quite variable between simulations. This should not be overly surprising: combined estimation of nonlinear or delayed density-dependence and environmental effects under autocorrelated environmental forcing is a known challenge for parametric methods (Jönsson et al. 2002; Jiang and Shao 2003; Lindén et al. 2013). Non-parametric CCM was found in this particular scenario to avoid well false negatives. However, it did introduce a sizeable number of false positives (between 40 and 15%) which suggests that the method could be better calibrated, a topic that we tackle below.

### **Issues in calculating $p$ values and confidence intervals**

So far, we mainly discussed the performance of CCM and GC in terms of sensitivity and specificity. Such unit-less interpretation of the model outputs dominates benchmarks and tests in the ecology and physics literature (e.g., see recently Krakovská et al. 2018). This implicitly requires setting cutoffs for  $p$  values and effect sizes to decide when an interaction is present. This is possible in a simulation context, but in practice,  $p$  values, Bayes factors, and confidence intervals are the quantities that are typically reported. Therefore, a question of interest is: do GC and CCM consistently produce precise  $p$  values and confidence intervals? Our results show that while overall GC and CCM produce sound results, statistical indicators for both methods are not always very well calibrated. This is exemplified by all the tables in our manuscript in that, in the case where the null hypothesis of no interactions is true, the percentage of  $p$  values below 10%, for both GC and

CCM, does not always match exactly the 10% level of the test employed (although GC fares usually a bit better than CCM in this regard).

The absence of proper calibration for GC is easily explained by the fact that the model that generated the data (nonlinear) and the model used to analyze it (log-linear) are not the same, and thus there is no reason to expect perfectly calibrated  $p$  values. Our results support other findings that confidence intervals for MAR(1) models, when fitted to the data generated by more nonlinear models, tend to be “too narrow” (Certain et al. 2018), in the sense that there is poor coverage of the point estimate. Nonlinear Granger causality methods (Schreiber 2000, see Paluš 2008; Amblard and Michel 2013; Papana et al. 2013, for reviews) could be of use to improve causality detection by obtaining more exact  $p$  values. Transfer entropy (ESM Appendix S1.5), in particular, admits linear GC as a special case (Barnett and Bossomaier 2012), and therefore provides an interesting bridge to classical MAR( $p$ ) modelling.

Why  $p$  values for CCM were also imperfectly calibrated is unclear. While the original CCM article (Sugihara et al. 2012) method did not directly calculate  $p$  values, further work has recommended using surrogate time series to do so (e.g., Deyle et al. 2016a; Ye et al. 2018), a suggestion with which we concur. Cobey and Baskerville (2016) proposed another method to calculate  $p$  values for CCM based on the increase of  $\rho$  with library size  $L$ : although this technique made sense in theory, it was not found to work well in practice (see ESM Section S2.1). We therefore considered several surrogate-based  $p$  values and chose the best-performing ones (Figs. S3, S4 in ESM), but in some cases—with a confounding abiotic driver or with 20 species—this was not completely satisfactory. Formal statistical inference for CCM could therefore be improved. Another idea could be to combine both worlds and perform surrogate-based nonlinear Granger causality inference (Schreiber 2000; Schreiber and Schmitz 2000; Paluš 2008).

### **The specific challenges of high-dimensional, many-species interacting systems**

We found here that both GC and CCM were scalable to large interaction networks (10 or 20 species) for relatively long time series by ecological standards (i.e., 300 to 700 time steps). We used both false discovery rate corrections and regularized models (i.e., LASSO-penalized MAR(1) models developed for modular interaction networks; Charbonnier et al. 2010).

One surprising find, for Granger analyses, was that the structured LASSO did not massively outperform the FDR-corrected pairwise analyses (since the underlying true network is specified from conditional not pairwise

interactions). One way to interpret these results is in terms of correcting for confounders vs. collider bias. Fitting a high-dimensional model, even with regularization through the LASSO, has the benefits of including in the estimation of an interaction  $j \rightarrow i$  the other (potentially many) interacting species. This can be construed as correcting for potential confounders when estimating an interaction. However, any incorrectly included species in the network can generate what is known in causal inference as collider bias (Pearl 2009): if species 6 does not truly affect species 1, but is included in the dynamical model for species 1, then the effect  $2 \rightarrow 1$  might be poorly estimated. Therefore, there is a trade-off between accounting for confounding factors and avoiding collider bias. Pairwise FDR-corrected analyses seemed to realize the best-trade off for the chaotic 10 species Lotka-Volterra models, though not for the 10 species MAR models for which the LASSO was better. This could suggest that when the model functional forms are exact (in the MAR case), avoiding collider bias is easy, but when these are poorly known and the dynamics are highly nonlinear, avoiding collider bias is harder.

However, the performances of the structured LASSO MAR(1) models may be diminished by other choices: (1) we used MAR(1) not MAR( $p$ ), which limits the ability of the autoregressive model to mimic the nonlinear system and (2) we did not use iterative model fitting, e.g., using the first inferred network as prior for the latent structure or as initial condition for further estimates. Both of these ideas may improve the network inference. One reason why we used MAR(1) modelling for the high-dimensional systems, outside of just simplicity, was that  $p = 1$  was selected in the 20-species case based on BIC, ESM Section S2.7. But this selection of the lag itself did not use regularization. Selecting both the interaction matrix sparsity and the lag order through regularization in high-dimensional MAR( $p$ ) modelling is extremely challenging because there are many ways to connect the number of time lags  $p$  to the LASSO penalties (Michailidis and D’alché Buc 2013; Nicholson et al. 2017). Mainali et al. (2019) recently used the R package BigVAR (Nicholson et al. 2017), a promising technique for penalized MAR( $p$ ) fitting, but few interactions were found and model performance was not evaluated with simulations; we have found here in contrast that without a latent network structure, unstructured LASSO-based methods perform poorly on large networks. Hence, our choice of SIMoNe (Charbonnier et al. 2010; Chiquet et al. 2008) which sticks to  $p = 1$  but allows the latent structure to be specified as a stochastic block model (Daudin et al. 2008; see ESM Appendix S1.1 for details). Combining structured LASSO modelling with models more sophisticated than MAR(1) remains an area where development is needed.

## Causal inference for nonlinear and stochastic ecological systems: going further

Overall, both linear Granger causality and convergent cross mapping can show good recall (sensitivity) and specificity for highly nonlinear and stochastic dynamical systems. Their domains of applicability overlap to a great degree. Rather than choosing one of these frameworks based on the supposed degree of nonlinearity or stochasticity of the ecological system studied (e.g., Runge et al. 2019a), we suggest that which one to use may be decided based on the goal and constraints of the analysis.

For instance, (log)-linear GC, being a fully parametric framework, can easily be extended to situations where we have small counts that preclude data transformation. This requires using a log-link function rather than an actual log transformation, as in Poisson Log-Normal models (e.g., Chiquet et al. 2018) and other flavors of latent variable modelling (e.g., Warton et al. 2015; Ovaskainen et al. 2017). It may likewise be very useful when one wants to introduce compositionality constraints in microbiome studies (Bjork et al. 2017). Conversely, convergent cross mapping or nonlinear Granger causality techniques allow for a much finer reconstruction of the attractor shape, which can be very useful to compare to the attractor shape of candidate mechanistic models (e.g., coupled differential equation models), if those exist. In some cases, using both frameworks, to increase the robustness of the interaction inference, is another idea (Hannisdal et al. 2017).

Looking at the various implementations of GC and CCM, it seems that the most critical methodological choices are rarely located along a “linear vs. nonlinear model” gradient, but instead boil down to two characteristics. First, details matter: faulty selection of the lag order  $p$  of autoregressive models results in nonsensical GC inference, and yet proper  $p$  selection yields causal inferences fairly robust to nonlinearities. Likewise, versions of CCM including significance testing are quite sensitive to the  $p$  value definition, and surrogate-based  $p$  values should be preferred. In other words, the devil is always in the details of the test or model selection, for Granger-based or CCM-based methods alike. Second, a key choice to make is what constitutes the “conditioning set,” i.e., the variables that are known to be important confounders and are de facto included in the time series model (Eichler 2013). For instance, an unknown confounder such as seasonal temperature or an invading species can massively thwart any attempt at interaction inference if not corrected for. And even when corrected for (i.e., adding the confounder to the autoregression or considering surrogate time series), this is the scenario where we observed the largest proportion of false positives for both GC and CCM. Strategies to better understand how to choose and handle this conditioning set

when performing causal inference will be, we believe, a very important feature of ecological interaction inference for the years to come. Several algorithms have been already put forward (Eichler 2013; Runge 2018, 2019a; Runge et al. 2019b), and much remains to be done to better incrementally select variables in order to assemble networks.

**Acknowledgments** FB thank Julien Chiquet and Camille Charbonnier for advice on the structured LASSO and Grégoire Certain for discussions on MAR modelling. FB and CP were supported by the French ANR through LabEx COTE (ANR-10-LABX-45). We thank Ethan Deyle, Adam Clark and Hao Ye for feedback, notably regarding surrogate time series testing for CCM. Constructive referee suggestions improved the manuscript, notably the figures.

**Author contributions** All authors contributed to the project design. FB and CP constructed the case studies, wrote the computer code, and analyzed the real and simulated data. All authors contributed to the interpretation of the results. FB wrote a first draft of the manuscript, which was then edited by all authors.

**Data availability** Codes for the analyses presented in this paper are available at <https://github.com/fbarraquand/GCausality> and are published at Zenodo with DOI 10.5281/zenodo.3967591 (Barraquand and Picoche, 2020).

## References

- Aalen OO (1987) Dynamic modelling and causality. *Scand Actuar J* 1987:177–190
- Aalen OO, Røysland K, Gran JM, Ledermann B (2012) Causality, mediation and time: a dynamic viewpoint. *J R Stat Soc Ser A (Stat Soc)* 175:831–861
- Adler P, Ellner S, Levine J (2010) Coexistence of perennial plants: an embarrassment of niches. *Ecol Lett* 13:1019–1029
- Adler P, Smill D, Beard K, Choi R, Furniss T, Kulmatiski A, Meiners J, Tredennick A, Veblen K (2018) Competition and coexistence in plant communities: intraspecific competition is stronger than interspecific competition. *Ecol Lett* 21:1319–1329
- Amblard PO, Michel O (2013) The relation between Granger causality and directed information theory: a review. *Entropy* 15:113–143
- Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys Rev Lett* 103:238701
- Barnett L, Bossomaier T (2012) Transfer entropy as a log-likelihood ratio. *Phys Rev Lett* 109:138105
- Barnett L, Seth AK (2014) The MVGC, multivariate Granger causality toolbox: A new approach to Granger-causal inference. *J Neurosci Methods* 223:50–68
- Barraquand F, Picoche C (2020) Code for Granger causality and CCM analyses. Zenodo, <https://doi.org/10.5281/zenodo.3967591>
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)* 57:289–300
- Berlow EL, Neutel AM, Cohen JE, De Ruiter PC, Ebenman B, Emmerson M, Fox JW, Jansen VA, Iwan Jones J, Kokkoris GD et al (2004) Interaction strengths in food webs: issues and opportunities. *J Animal Ecol* 73:585–598

- Bjork JR, O’Hara RB, Ribes M, Coma R, Montoya JM (2017) The dynamic core microbiome: structure, stability and resistance. *bioRxiv*, pp 137885
- Blanchet FG, Cazelles K, Gravel D (2020) Co-occurrence is not evidence of ecological interactions. *Ecol Lett* 23:1050–1063
- Carr A, Diener C, Baliga NS, Gibbons SM (2019) Use and abuse of correlation analyses in microbial ecology. *ISME J* 13:2647–2655
- Cazelles K, Araújo MB, Mouquet N, Gravel D (2016) A theory for species co-occurrence in interaction networks. *Theor Ecol* 9:39–48
- Certain G, Barraquand F, Gårdmark A (2018) How do MAR(1) models cope with hidden nonlinearities in ecological dynamics? *Methods Ecol Evol* 9:1975–1995
- Charbonnier C, Chiquet J, Ambroise C (2010) Weighted-lasso for structured network inference from time course data. *Stat Appl Genet Mol Biol*:9
- Chen Y, Bressler SL, Ding M (2006) Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *J Neurosci Methods* 150:228–37
- Chiquet J, Smith A, Grasseau G, Matias C, Ambroise C (2008) Simone: statistical inference for modular networks. *Bioinformatics* 25:417–418
- Chiquet J, Mariadassou M, Robin S et al (2018) Variational inference for probabilistic poisson PCA. *Ann Appl Stat* 12:2674–2698
- Cobey S, Baskerville EB (2016) Limits to causal inference with state-space reconstruction for infectious disease. *PloS one* 11:e0169050
- Coenen AR, Weitz JS (2018) Limitations of correlation-based inference in complex virus-microbe communities. *mSystems*:3
- Commenges D, Gégout-Petit A (2009) A general dynamical statistical model with causal interpretation. *J R Stat Soc Ser B (Stat Methodol)* 71:719–736
- Coyte KZ, Schlüter J, Foster KR (2015) The ecology of the microbiome: networks, competition, and stability. *Science* 350:663–666
- Daudin JJ, Picard F, Robin S (2008) A mixture model for random graphs. *Stat Comput* 18:173–183
- Dennis B, Desharnais RA, Cushing JM, Henson SM, Costantino RF (2001) Estimating chaos and complex dynamics in an insect population. *Ecol Monograph* 71:277–303
- Dettlo M, Molini A, Katul G, Stoy P, Palmroth S, Baldocchi D (2012) Causality and persistence in ecological systems: a nonparametric spectral Granger causality approach. *Amer Natur* 179:524–535
- Deyle E, Maher MC, Hernandez RD, Basu S, Sugihara G (2016a) Global environmental drivers of influenza. *Proc Natl Acad Sci* 113:13081–13086
- Deyle E, May R, Munch SB, Sugihara G (2016b) Tracking and forecasting ecosystem interactions in real time. *Proc R Soc B Biol Sci* 283:20152258
- Ding M, Chen Y, Bressler S (2006) Granger causality: basic theory and application to neuroscience. *Handbook of time series analysis*, pp. 437–460
- Dormann CF, Bobrowski M, Dehling DM, Harris DJ, Hartig F, Lischke H, Moretti MD, Pagel J, Pinkert S, Schleuning M et al (2018) Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Glob Ecol Biogeogr* 27:1004–1016
- Eichler M (2013) Causal inference with multiple time series: principles and problems. *Philosophical Transactions of the Royal Society of London A: Mathematical, Phys Eng Sci* 371:20110613
- Ellner S, Turchin P (2005) When can noise induce chaos and why does it matter: a critique. *Oikos* 111:620–631
- Geweke J (1982) Measurement of linear dependence and feedback between multiple time series. *J Am Stat Assoc* 77:304–313
- Geweke JF (1984) Measures of conditional linear dependence and feedback between time series. *J Am Stat Assoc* 79:907–915
- Gibbons SM, Kearney SM, Smillie CS, Alm EJ (2017) Two dynamic regimes in the human gut microbiome. *PLoS Comput Biol* 13:e1005364
- Granger C (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37:424–438
- Grziwotz F, Strauß JF, Hsieh Ch, Telschow A (2018) Empirical dynamic modelling identifies different responses of *Aedes Polynesiensis* subpopulations to natural environmental variables. *Sci Rep* 8:16768
- Hampton SE, Holmes EE, Scheef LP, Scheuerell MD, Katz SL, Pendleton DE, Ward EJ (2013) Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology* 94:2663–2669
- Hannisdal B, Haaga KA, Reitan T, Diego D, Liow LH (2017) Common species link global ecosystems to climate change: dynamical evidence in the planktonic fossil record. *Proc R Soc B Biol Sci* 284:20170722
- Hannisdal B, Liow LH (2018) Causality from palaeontological time series. *Palaeontology* 61:495–509
- Harford WJ, Karnauskas M, Walter JF, Liu H (2017) Non-parametric modeling reveals environmental effects on bluefin tuna recruitment in Atlantic, Pacific, and Southern Oceans. *Fisher Oceanogr* 26:396–412
- Ives AR (1995) Predicting the response of populations to environmental change. *Ecol* 76:926–941
- Ives A, Dennis B, Cunningham K, Carpenter S (2003) Estimating community stability and ecological interactions from time-series data. *Ecol Monogr* 73:301–330
- Jiang L, Shao N (2003) Autocorrelated exogenous factors and the detection of delayed density dependence. *Ecology* 84:2208–2213
- Jonzén N, Lundberg P, Ranta E, Kaitala V (2002) The irreducible uncertainty of the demography-environment interaction in ecology. *Proc R Soc Lond Ser B: Biol Sci* 269:221–225
- Jost C, Ellner SP (2000) Testing for predator dependence in predator-prey dynamics: a non-parametric approach. *Proc R Soc Lond B Biol Sci* 267:1611–1620
- Krakovská A, Jakubík J, Chvosteková M, Coufal D, Jajcay N, Paluš M (2018) Comparison of six methods for the detection of causality in a bivariate time series. *Phys Rev E* 97:042207
- Langendorf RE, Doak DF (2019) Can community structure causally determine dynamics of constituent species? A test using a host-parasite community. *The American Naturalist* 194:E66–E80
- Lindén A, Fowler MS, Jonzén N (2013) Mischaracterising density dependence biases estimated effects of coloured covariates on population dynamics. *Popul Ecol* 55:183–192
- Link JS (2002) What does ecosystem-based fisheries management mean. *Fisheries* 27:18–21
- Loreau M, de Mazancourt C (2008) Species synchrony and its drivers: neutral and nonneutral community dynamics in fluctuating environments. *Amer Natur* 172:E48–E66
- Lütkepohl H (2005) New Introduction to Multiple Time Series Analysis. Springer
- Mainali K, Bewick S, Vecchio-Pagan B, Karig D, Fagan WF (2019) Detecting interaction networks in the human microbiome with conditional Granger causality. *PLoS Comput Biol* 15:e1007037
- Marinazzo D, Pellicoro M, Stramaglia S (2008) Kernel-Granger causality and the analysis of dynamical networks. *Phys Rev E* 77:1–9
- May R (1973) Stability and complexity in model ecosystems. Princeton University Press, Princeton
- Mayr E (1961) Cause and effect in biology. *Science* 134:1501–1506
- Michailidis G, D’alché Buc F (2013) Autoregressive models for gene regulatory network inference: sparsity, stability and causality issues. *Math Biosci* 246:326–334

- Mønster D, Fusaroli R, Tylén K, Roepstorff A, Sherson JF (2017) Causal inference from noisy time-series data—testing the convergent cross-mapping algorithm in the presence of noise and external influence. *Futur Gener Comput Syst* 73:52–62
- Mukhopadhyay ND, Chatterjee S (2006) Causality and pathway search in microarray time series experiment. *Bioinformatics* 23:442–449
- Mutshinda CM, O'Hara RB, Woiwod IP (2009) What drives community dynamics? *Proc R Soc B: Biol Sci* 276:2923–2929
- Mutshinda CM, O'Hara RB, Woiwod IP (2011) A multispecies perspective on ecological impacts of climatic forcing. *J Anim Ecol* 80:101–107
- Nicholson W, Matteson D, Bien J (2017) BigVar: Tools for modeling sparse high-dimensional multivariate time series. *arXiv:1702.07094*
- North BV, Curtis D, Sham PC (2002) A note on the calculation of empirical p values from Monte Carlo procedures. *Amer J Human Gen* 71:439–441
- Ovaskainen O, Tikhonov G, Norberg A, Guillaume Blanchet F, Duan L, Dunson D, Roslin T, Abrego N (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol Lett* 20:561–576
- Paluš M (2008) From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics. *ContempPhys* 48:307–348
- Papana A, Kyrtsov C, Kugiumtzis D, Diks C (2013) Simulation study of direct causality measures in multivariate time series. *Entropy* 15:2635–2661
- Pearl J (2009) Causal inference in statistics: an overview. *Stat Surv* 3:96–146
- Pfaff B (2008) VAR, SVAR and SVEC Models: implementation within R package vars. *J Stat Softw*:27
- Pikitch E, Santora E, Babcock A, Bakun A, Bonfil R, Conover D, Dayton P, Doukakis P, Fluharty D, Heheman B et al (2004) Ecosystem-based fishery management. *Science* 305:346–347
- Runge J (2014) Detecting and quantifying causality from time series of complex systems. Ph.D thesis, Humboldt-Universitätzu, Berlin. Mathematisch-Naturwissenschaftliche Fakultät
- Runge J (2018) Causal network reconstruction from time series: from theoretical assumptions to practical estimation. *Chaos: Interdiscip J Nonlinear Sci* 28:075310
- Runge J, Bathiany S, Boltt E, Camps-Valls G, Coumou D, Deyle E, Glymour C, Kretschmer M, Mahecha MD, Muñoz-marí J et al (2019a) Inferring causation from time series in earth system sciences. *Nat Commun* 10:2553
- Runge J, Nowack P, Kretschmer M, Flaxman S, Sejdinovic D (2019b) Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci Adv* 5:eaau4996
- Schreiber T (2000) Measuring information transfer. *Phys Rev Lett* 85:461
- Schreiber T, Schmitz A (2000) Surrogate time series. *Physica D: Nonlinear Phenom* 142:346–382
- Schweder T (1970) Composable Markov processes. *J Appl Probab* 7:400–410
- Sims C (1980) Macroeconomics and reality. *Econometrica* 48:1–48
- Stone L, Roberts A (1991) Conditions for a species to gain advantage from the presence of competitors. *Ecology* 72:1964–1972
- Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344:734
- Sugihara G, May R, Ye H, Hsieh Ch, Deyle E, Fogarty M, Munch S (2012) Detecting causality in complex ecosystems. *Science* 338:496–500
- Tibshirani R, Wainwright M, Hastie T (2015) Statistical learning with sparsity: the Lasso and generalizations. Chapman and Hall/CRC
- Veilleux BG (1979) An analysis of the predatory interaction between *Paramecium* and *Didinium*. *J Anim Ecol* 48:787–803
- Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC, Hui FK (2015) So many variables: joint modeling in community ecology. *Trends Ecol Evol* 30:766–779
- Wootton J, Emmerson M (2005) Measurement of interaction strength in nature. *Annual Review of Ecology. Evol Syst* 36:419–444
- Yang G, Wang L, Wang X (2017) Reconstruction of complex directional networks with group lasso nonlinear conditional Granger causality. *Sci Rep* 7:2991
- Ye H, Deyle E, Gilarranz LJ, Sugihara G (2015) Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific Reports* 5
- Ye H, Sugihara G (2016) Information leverage in interconnected ecosystems: overcoming the curse of dimensionality. *Science* 353:922–925
- Ye H, Clark A, Deyle E, Munch S, Cai J, Cowles J, Daon Y, Edwards A, Keyes O, Stagge J, Ushio M, White E, Sugihara G (2018) rEDM: applications of empirical dynamic modeling from time series. R package version 0.7.1
- Yodzis P (1998) Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *J Anim Ecol* 67:635–658
- Zeileis A, Hothorn T (2002) Diagnostic checking in regression relationships. *R News* 2:7–10

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.