

Identifying dynamic coupling between interacting populations using Granger causality and convergent cross-mapping: a reconciliation

F. Barraquand, C. Picoche, M. Detto & F. Hartig

April 14, 2019

Abstract

Finding who interacts with whom based on time series data is a key endeavour of statistical ecology. Here, we show that multivariate autoregressive (MAR) modelling of finite but arbitrary order can be used, with appropriate log transformation and model selection based on information criteria, in order to infer causal links between interacting populations. Causality is here understood in the sense of Granger, i.e. time series x Granger causes y if information on present and past values of x helps predicting future values of y . Recent nonlinear approaches such as convergent cross-mapping (CCM) have been put forward to alleviate potential issues of Granger causality (GC), notably due to the assumption of linearity in most implementations of GC. This has led to a relative skepticism about the ability of (log)-linear parametric methods to infer causal links in nonlinear dynamical systems. The merits of nonlinear and nonparametric models notwithstanding, the present results show that the demise of parametric and linear modelling has been somewhat exaggerated. Not only parametric MAR(p) models are able to infer causal links for a number of nonlinear (stochastic) systems, but they can even do so for datasets with highly nonlinear dynamical behaviour (e.g., limit cycles in real data, simulated chaotic dynamics). GC demonstrates, surprisingly, similar performances to CCM in a number of highly nonlinear cases. We further show that the (log)-linear framework can be extended to large interaction networks, provided some assumptions about the interaction structure. Finally, we discuss nonparametric, spectral and nonlinear extensions of Granger causality approaches that could be of use to ecologists.

Keywords: time series, interaction network, interaction strength, causal inference, feedback, food web, community dynamics.

Introduction

There is a diversity of viewpoints regarding the importance of interactions between species for predicting their joint dynamics; a division that can be traced back at least to the Clementsian (integrated) vs Gleasonian (stochastically dominated) views of plant communities (Chase, 2003). Today, whether one should predict population dynamics as nearly isolated populations or integrated communities merges with management concerns, as exemplified by debates on ecosystem-based fisheries management in fisheries science (Link, 2002; Pikitch *et al.*, 2004). One reason for this continued tension within ecology is that most attempts to infer interaction strengths rely on relatively strong assumptions (Berlow *et al.*, 2004; Wootton & Emmerson, 2005). Reliably inferring how populations of different species are able to influence one another’s dynamics is therefore a longstanding quest of ecology.

At one end of the spectrum of beliefs on interaction networks roles in determining community dynamics, a large part of theoretical ecology models communities as variations of the generalized Lotka-Volterra equations (e.g., May, 1973; Yodzis, 1998; Coyte *et al.*, 2015; eq. 1), where the dynamics of food webs or competitive systems are modelled as fairly coupled deterministic dynamical systems, in discrete or continuous time. This implicitly assumes that interactions between populations of different species are many and strong enough (even if interaction strength is asymmetrically distributed) to deserve to be modelled explicitly, and such that interactions rates can be weighted by body mass or metabolic rates (Yodzis & Innes, 1992; Brose *et al.*, 2006; Weitz & Levin, 2006; Rall *et al.*, 2008; Brose, 2010).

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

At the other end of the spectrum, other branches of theoretical ecology have highlighted that many species may have weak dynamical coupling to their competitors, enemies and resources. One reason for this is that intraspecific density-dependence dominates the dynamics (Barabás *et al.*, 2017) and even more so as community size increases. A variant being that as most consumers are generalists rather than specialists, their dynamics can be primarily determined by the intricacies of stage structure within populations (Murdoch *et al.*, 2002; Barraquand *et al.*, 2017) rather than between species interactions. Recent statistical evidence seems to substantiate that view: most empirically-minded trials to infer interspecific interactions between species in large field communities, using time series of counts, have yielded so far weak estimates of net species interactions (Mutshinda *et al.*, 2009, 2011; Barraquand *et al.*, 2018). In contrast to the statistical models though, experimental measures have shown that weak interactors can have strong impact within their communities (but experimental measures of interactions are not equivalent to the interaction concepts of dynamical models, Berlow *et al.*, 2004).

Summarizing the evidence derived from theoretical and statistical models, ecologists are currently in the puzzling position where both theory and data have produced conflicting reports regarding the importance of species interactions for predicting community dynamics. To improve ecological modelling overall, a better statistical inference of dynamic coupling between species is therefore paramount - preferably one able to deal with the numerous nonlinearities present in ecological models.

To quantify dynamic coupling between species, ecologists have used a number of statistical models ranging from mechanistic to purely phenomenological. They include most notably multivariate autoregressive models or order one, or MAR(1) models (also called VAR(1) - vector autoregressive models). These are statistical multispecies generalizations of the Gompertz discrete-time single-species models (Ives *et al.*, 2003; Mutshinda *et al.*, 2009). MAR(1) can to a large extent infer linkages between species population dynamics (Certain *et al.*, 2018). MAR(p) models, with a maximum time lag of order $p \geq 1$, generalize the MAR(1) framework familiar to ecologists and are interestingly linked to one of the most celebrated causality concept, the so-called Granger-Wiener causality (Granger, 1969; Sims, 1980; Ding *et al.*, 2006; Chen *et al.*, 2006; Barnett *et al.*, 2009; Detto *et al.*, 2012; Sugihara *et al.*, 2012; Barnett & Seth, 2014).

Granger-Wiener causality (Granger causality or GC for short) is a causality concept that is strongly tied to the idea that the cause must precede in time the effect. There are other ways to define causality, through interventions notably (e.g. Pearl, 2009) but these are usually less practical in a dynamical setting. Therefore, if a dynamical model for time series y can see its in-sample predictive power improve by inclusion of time series x , we say that x Granger-causes y . This definition is rather general and does not specify, in general, any sort of model framework. It can in principle be applied to phenomenological and mechanistic frameworks alike, as well as in nonparametric settings (Detto *et al.*, 2012). However, GC testing is often used in contexts where the a priori information on community dynamics is limited, so that one needs to reduce the universe of possible model formulations to a handful of contrasted scenarios. In this context, GC is usually tested within the context of statistically friendly MAR(p) models for whom confidence intervals for coefficients, model selection, and other inferential tools are well understood. In an ecological setting, we use mainly log-linear models because the MAR framework applies to log-abundances (Ives *et al.*, 2003).

In the last decade, other methods such as Convergent Cross-Mapping (CCM; Sugihara *et al.*, 2012) have been put forward as a new way to infer interactions between species, and alleviate problems due to the limitation of linear autoregressive models (although, as we say above and shall show later, MAR(p) models are log-linear, which makes them essentially power-law models).

In this article, we first evaluate the performance of 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 somewhat harsh 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 (Lütkepohl, 2005), in order to infer lag order p of MAR(p) models as well as their parametric structure, we show that Granger causality techniques can in fact infer interactions surprisingly well in nonlinear cases. We then highlight intriguing parameter configurations and empirical case studies where Granger causality and Sugihara et al.’s convergent cross-mapping (CCM) either both fail or both work, which suggests that seemingly foreign causality concepts might in fact share some underlying mathematical similarities. We finally demonstrate that Granger causality / MAR(p) modelling can be scaled up to large interaction networks using appropriate model regularization techniques (LASSO-based).

Methods and models

In the following, we recall the basics of Granger causality concepts and MAR(p) - Multivariate AutoRegressive - modelling, which is the most common way (not the only one) to assess Granger causality. For completeness, we also present some common nonlinear alternatives to MAR(p) modelling which stays within the Granger causality purview in Appendix . We describe shortly thereafter convergent-cross mapping (Sugihara *et al.*, 2012), which is implemented as well but 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

$\mathbf{x} = (x_t)_{t \in [1:T]}$ Granger-causes $\mathbf{y} = (y_t)_{t \in [1:T]}$ \iff 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 time series autoregressions to explain time series y , one with only y and one with y and x :

$$y_t = \sum_{i=1}^L a_i y_{t-i} + \eta_t \quad (2)$$

$$y_t = \sum_{i=1}^L a_{1i} x_{t-i} + \sum_{i=1}^L a_{2i} y_{t-i} + \epsilon_t \quad (3)$$

Granger causality then occurs if $\sigma_\epsilon^2 < \sigma_\eta^2$. When more than two variables are considered, simple GC has to be differentiated from conditional GC. Conditional GC occurs whenever a third variable z is considered and corrected for.

When fitting a MAR(p) model, we typically estimate conditional GC. For instance, let us consider a MAR(1) model (eq. 4) with 3 species, which may be familiar to ecologists through the works of Ives *et al.* (2003); Hampton *et al.* (2013):

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

so that its \mathbf{B} (interaction) matrix is defined by

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

Here, 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

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

We drop \mathbf{u}_t which corresponds to a special case where a subset of the variables has a one-way causal impact. The condition for an interaction from species j to species i given all other species densities and variables then becomes, in a general MAR(p) setting (according to eq. 6,):

$$\exists b_{ij}^{(q)} \neq 0 \Leftrightarrow x_j \rightarrow x_i | (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_d) \quad (7)$$

where each time lag is indexed by q . This highlights immediately the high model dimensionality that can be attained with high temporal lag order p and high number of species. Conversely, direct or pairwise GCausality testing between x_i and x_j is assessed through a bivariate autoregressive model for each (i, j) pair, and therefore uses a considerably lower-dimensional model. Direct GC testing requires, however, some false discovery correction to attain meaningful statistical significance (Mukhopadhyay & Chatterjee, 2006). In the following, we use a Benjamini-Hochberg correction (REF).

MAR(p) model fitting has been performed using the package **vars** in **R**, which uses ordinary least squares for estimation. We used the BIC as a default for lag order selection, although we also considered other information criteria for lag order selection (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` which performs a Wald test for nested models. For conditional Granger causality testing, we use the function `causality` in package `vars` which provides F-tests for the nested models. Both tests and implementations provided similar answers when compared.

Convergent-cross mapping

Convergent cross-mapping (Sugihara *et al.*, 2012) relies on state-space reconstruction. We assume two time series $\mathbf{x} = (x_t)_{t \in [1:T]}$ and $\mathbf{y} = (y_t)_{t \in [1:T]}$. The presentation is adapted from Sugihara et al.'s articles and also Palus (2018). The attractor manifold M_X is constructed as a set of E -dimensional vectors $\tilde{\mathbf{x}}(t) = (x(t), x(t-\tau), x(t-2\tau), \dots, x(t-(E-1)\tau), \dots)$ for $t = 1 + (E-1)\tau$ to $t = T$. E is the embedding dimension. This sets of vectors constitutes the reconstructed manifold. We now find the $E+1$ nearest neighbours of $\tilde{\mathbf{x}}(t)$ 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:

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

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.

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

Even though there are no standard ways of computing associated p-values for the convergent cross-mapping skill ρ , several formulations have been proposed:

- except specified otherwise, we used the test used in Cobey & Baskerville, 2016 in which $p(X \rightarrow Y) = \frac{1}{n} \sum \mathbf{1}((\rho(\mathbf{x}, \hat{\mathbf{x}}|M_{Y,max}) < \rho(\mathbf{x}, \hat{\mathbf{x}}|M_{Y,min}))$ where n is the number of libraries of size L that were used to build M_Y and $M_{Y,max}$ (respectively, $M_{Y,min}$) is the manifold constructed with the maximum (respectively, minimum) library size
- for periodic signals, or, more generally, time series that share a common, sometimes unknown, driver, spurious causality can be found between time series. We compute surrogate time series which keep the periodicity of the signal but modify its residuals. Cross-mapping is then computed on the surrogates and compared to the real value. In this case, $p(X \rightarrow Y) = \frac{1}{n} \sum \mathbf{1}((\rho(\mathbf{x}_{real}, \hat{\mathbf{x}}_{real}|M_Y) < \rho(\mathbf{x}_{surr}, \hat{\mathbf{x}}_{surr}|M_Y))$
- for large simulations (10 and 20 species), a method similar to the above was used but surrogates were

only computed by randomly shuffling the causal time series.

The analyses have been performed in R [3.4.4] using the package rEDM [version 0.7.1 Ye *et al.*, 2018]. For each time series, we retrieved the best embedding dimension (which maximizes the forecast skill of the simplex) and used it in the cross-mapping function, with 100 different libraries for each library size and library size depending on the length of the time series [this is not enough: we need to discuss the size we choose and to be consistent in our scripts]. The libraries were obtained with random draws without replacement from the original time series.

Granger causality in high-dimensional models

Full MAR(p) model fitting is highly impractical for high-dimensional models, unless very long time series are considered or special constraints are implemented to reduce the dimensionality (Michailidis & d’Alché Buc, 2013). If there are d species and p timelags, a $d \times d \times p$ dimensional model needs to be fitted to the data. For instance, let us imagine that a system involving 10 species with at least one cycling species is considered (and its cyclicity is probably not induced by the species in the network). To model it properly, we need $p = 2$ (long, regular cyclic behaviour in a one species AR(p) model requires $p > 1$). We then have 200 parameters in the interaction $\mathbf{B}^{(a)}$ matrices only. Even a MAR(1) model will have 100 elements, and therefore would be impossible to fit properly without a set of time series of length above 100. Preliminary simulations (Certain *et al.*, 2018; Barraquand *et al.*, 2018) suggest that a nonlinear, stochastic ecological system of dimension 10 or 12 requires approximately times series of length 500 to 800 to be fitted properly without implementing additional constraints. To deal with high-dimensionality for time series of long yet reasonable length (100 to 300 timesteps), we have considered two solutions:

- Pairwise Granger causality testing with False Discovery Rate (FDR) correction (Benjamini-Hochberg), with a philosophy similar to Mukhopadhyay & Chatterjee (2006). This done by fitting bivariate MAR(p) models, test for Granger causality in both directions, and then re-adjust the p-values obtained through the Benjamini-Hochberg correction.
- LASSO-penalized MAR(1) models with structured penalties, using the R package **simone** (Chiquet *et al.*, 2008; Charbonnier *et al.*, 2010). This allows to estimate (through non-zero interaction coefficients) conditional Granger causality. A naive approach to MAR(1) estimation would simply use the LASSO (Least Absolute Shrinkage and Selection Operator, REFs Hastie et al.) to set some of the coefficients to zero. Unfortunately, this approach is known to yield substantial bias whenever there is an important structure (for instance, modular) in the network. We have therefore used a technique that explicitly

accounts for network structure in addition to selecting coefficients with the LASSO (Charbonnier *et al.*, 2010).

Simulated and real datasets of interacting species population dynamics

Real data: Veilleux’s predator-prey cycles

The two first datasets that we consider are taken from Veilleux (1979) and have been analysed by other authors with mechanistic models which demonstrated two-way coupling (Jost & Ellner, 2000).

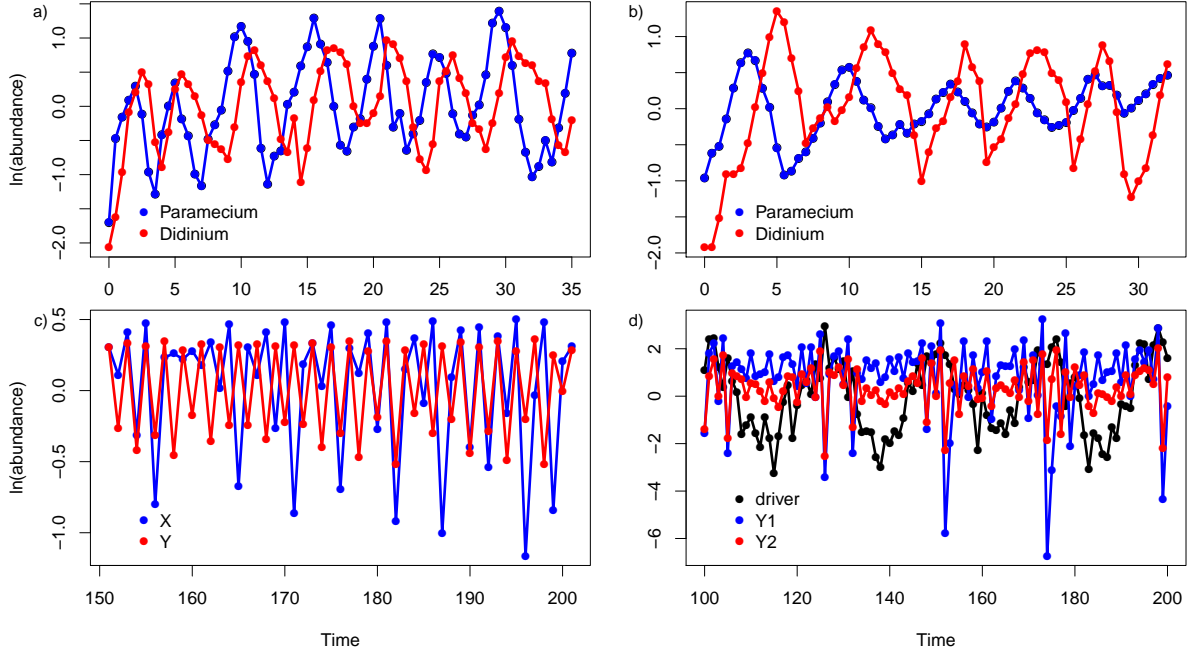


Figure 1: Time series of small-community models. Veilleux’s predator-prey data are shown in (a) (dataset CC05) and (b) (dataset CC0375); an example simulation for the 2-species chaotic model is shown in panel (c) and a simulation of the competition model including an environmental driver is illustrated in panel (d).

Deterministic chaos in two-species competition models

Here we use the same model as in Sugihara *et al.* (2012), a two-species discrete-time logistic competition model

$$x_{t+1} = x_t(3.8 - 3.8x_t - 0.02y_t) \quad (8)$$

$$y_{t+1} = y_t(3.5 - 3.5y_t - 0.1x_t) \quad (9)$$

The models are, as in Sugihara *et al.* (2012), in the chaotic regime, which constitutes a strong test of the log-linear MAR(p) framework. The only thing that is varied is the initial condition, which is taken at random 500 times. Although we acknowledge that “mirage correlations” can occur in some datasets, we aimed at reproducing a realistic sample 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 $t_m = 300$ as in Sugihara *et al.* (2012).

Because a method that finds no interactions whenever absent is as important as one that finds interactions whenever they are present, we also consider the following deterministic “null competition” model:

$$x_{t+1} = x_t(3.8 - 3.8x_t - 0 \times y_t) \quad (10)$$

$$y_{t+1} = y_t(3.5 - 3.5y_t - 0 \times x_t) \quad (11)$$

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

Two-species stochastic and nonlinear dynamics, including environmental drivers

First we consider a stochastic two-competition model, with Lotka-Volterra interactions in discrete time and a Ricker type of multispecies density-dependence.

$$N_{1,t+1} = N_{1,t} \exp(3 - 4N_{t,1} - 2N_{t,2} + \epsilon_{1,t}) \quad (12)$$

$$N_{2,t+1} = N_{2,t} \exp(2.1 - 0.31N_{t,1} - 3.1N_{t,2} + \epsilon_{2,t}) \quad (13)$$

This case was already investigated in Certain *et al.* (2018), but including an environmental driver on species 1. Then we consider a variant of this model 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)

$$N_{1,t+1} = N_{1,t} \exp(3 + 0.5u_t - 4N_{t,1} - 2N_{t,2} + \epsilon_{1,t}) \quad (14)$$

$$N_{2,t+1} = N_{2,t} \exp(2.1 + 0.5u_t - 0.31N_{t,1} - 3.1N_{t,2} + \epsilon_{2,t}) \quad (15)$$

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

$$N_{1,t+1} = N_{1,t} \exp(3 + 0.5u_t - 4N_{t,1} - 0 \times N_{t,2} + \epsilon_{1,t}) \quad (16)$$

$$N_{2,t+1} = N_{2,t} \exp(2.1 + 0.5u_t - 0 \times N_{t,1} - 3.1N_{t,2} + \epsilon_{2,t}) \quad (17)$$

We run 500 simulations for each model.

Five-, ten-, and twenty-species interaction webs

[[We're not actually sure we're gonna do the 5-species model]]

We consider first the 5-species deterministic model considered by Sugihara *et al.* (2012), in a slightly modified Ricker-density-dependence version, which allowed to have non-negative equilibria for all initial conditions (unlike the initial logistic formulation whose trajectories frequently crossed zero, since this is an issue with the logistic map):

$$\begin{aligned} N_{t+1,1} &= N_{t,1} \exp(4 - 4N_{t,1} - 2N_{t,2} - 0.4N_{t,3}) \\ N_{t+1,2} &= N_{t,2} \exp(3.1 - 0.31N_{t,1} - 3.1N_{t,2} - 0.93N_{t,3}) \\ N_{t+1,3} &= N_{t,3} \exp(0.12 + 0.636N_{t,1} + 0.636N_{t,2} - 2.12N_{t,3}) \\ N_{t+1,4} &= N_{t,4} \exp(3.8 - 0.111N_{t,1} - 0.111N_{t,2} + 0.131N_{t,3} - 3.8N_{t,4}) \\ N_{t+1,5} &= N_{t,5} \exp(4.1 - 0.082N_{t,1} - 0.111N_{t,2} - 0.125N_{t,3} - 4.1N_{t,5}) \end{aligned}$$

We consider a 10 species model which generalises the two-species Ricker competition to more species and more interaction types, with added stochasticity, and 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 stochasticity. The dynamical equation can be written as

$$\mathbf{n}_{t+1} = \mathbf{n}_t \circ \exp(\mathbf{r} + \mathbf{A}\mathbf{n}_t + \mathbf{e}_t), \mathbf{e}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \quad (18)$$

where \mathbf{n} is the abundance, the error $\sigma^2 = 0.1$ and the interaction matrix is defined to be

$$A = \begin{pmatrix} -4 & -2 & -0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.31 & -3.1 & -0.93 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.636 & 0.636 & -2.12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.111 & -0.111 & 0.131 & -3.8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & -2 & -2 & -0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.31 & -3.1 & -0.93 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.636 & 0.636 & -2.12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & -2 & -0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.31 & -3.1 & -0.93 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.636 & 0.636 & -2.12 \end{pmatrix} \quad (19)$$

We use the same interaction matrix with Gompertz (MAR(1)) dynamics, and run 25 simulations over 500 time steps with different initial conditions.

We slightly modified this model to scale it up to 20 species. Its structure is still fairly modular (eq. 20)

$$B = \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_{2,1} & b_{2,2} & b_{2,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_{3,1} & b_{3,2} & b_{3,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_{4,1} & b_{4,2} & b_{4,3} & b_{4,4} & b_{4,5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_{5,4} & b_{5,5} & b_{5,6} & b_{5,7} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_{6,5} & b_{6,6} & b_{6,7} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_{7,5} & b_{7,6} & b_{7,7} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{8,8} & b_{8,9} & b_{8,10} & b_{8,11} & b_{8,12} & b_{8,13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{9,8} & b_{9,9} & b_{9,10} & b_{9,11} & b_{9,12} & b_{9,13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{10,8} & b_{10,9} & b_{10,10} & b_{10,11} & b_{10,12} & b_{10,13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{11,8} & b_{11,9} & b_{11,10} & b_{11,11} & b_{11,12} & b_{11,13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{12,8} & b_{12,9} & b_{12,10} & b_{12,11} & b_{12,12} & b_{12,13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{13,8} & b_{13,9} & b_{13,10} & b_{13,11} & b_{13,12} & b_{13,13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{14,11} & b_{14,12} & b_{14,13} & b_{14,14} & b_{14,15} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{15,14} & b_{15,15} & b_{15,16} & b_{15,17} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{16,15} & b_{16,16} & b_{16,17} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{17,15} & b_{17,16} & b_{17,17} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{18,18} & b_{18,19} & b_{18,20} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{19,18} & b_{19,19} & b_{19,20} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{20,18} & b_{20,19} & b_{20,20} \end{pmatrix} \quad (20)$$

For the 20-species model, we also compare Ricker and Gompertz dynamics for 25 different simulations over 1000 time steps, but in this case, coefficients are drawn from a random distribution (eq. 21 and eq. 22).

$$b_{i,j} = b_{min} + (b_{max} - b_{min})\beta(2,2) \quad (21)$$

$$(b_{min}, b_{max}) = \begin{cases} (0.05, 0.1) & \forall i \neq j, \text{ with probability } 0.2 \text{ (positive interaction)} \\ (-0.2, -0.1) & \forall i \neq j, \text{ with probability } 0.8 \text{ (negative interaction)} \\ (-0.8, -0.3) & \forall i = j \end{cases} \quad (22)$$

For all datasets, real and simulated alike, the data are log-transformed before analysis.

Results

In each section, we apply both GC/MAR(p) modelling and CCM, except for high-dimensional datasets.

Real data: Veilleux's predator-prey cycles

Model selection of MAR(p) model by all information criteria selected a lag $p = 1$ for the CC05 dataset and a lag of 2 for the CC0.375 dataset (Fig. 2). The p-values for the GC test (null hypothesis: “no GC”) demonstrate convincingly that the “no GC” hypothesis can be rejected, for both datasets (1).

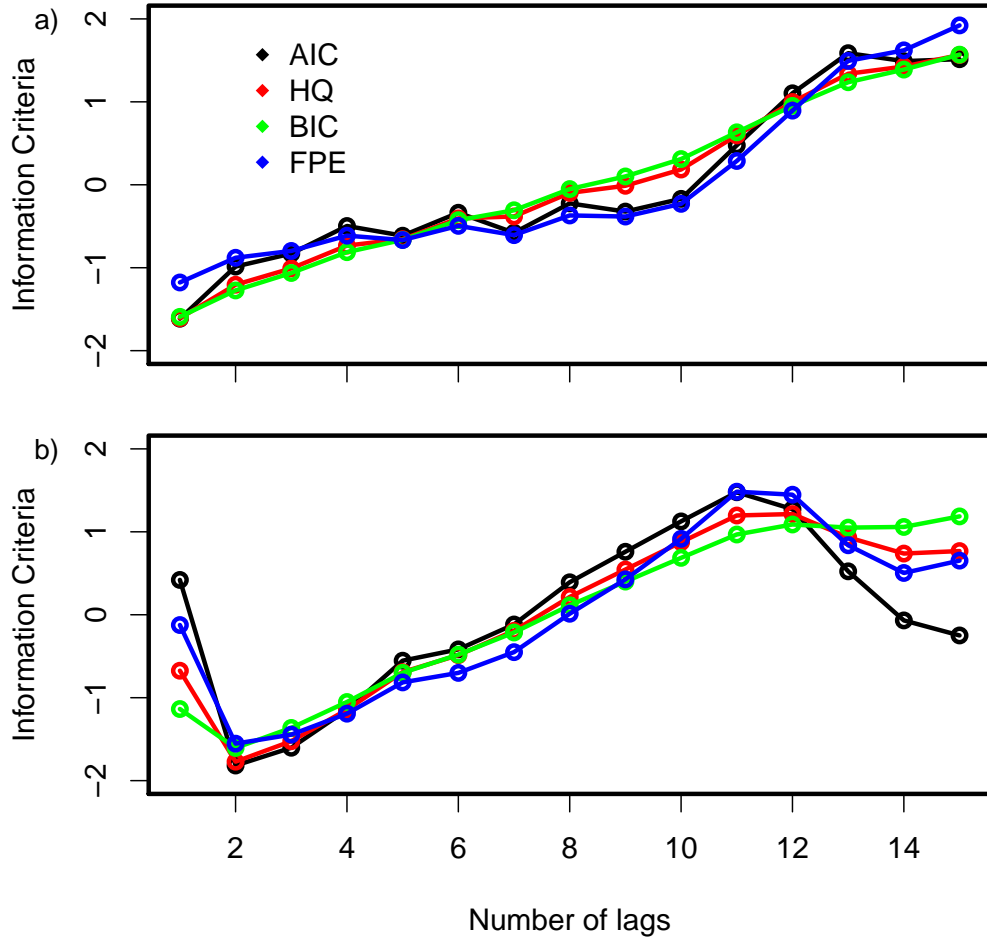


Figure 2: Results of model information criteria vs. lag order for the predator-prey data, for the two datasets.

Dataset	CC = 05	CC = 0.375
Lag p in VAR(p)	1	2
$1 \rightarrow 2$	2.79×10^{-11}	0.0409
$2 \rightarrow 1$	1.76×10^{-14}	0.0464

Table 1: P-values for H₀: No Granger causality between x and y .

CCM also demonstrate bi-directional causality, as demonstrated by the increase in $\rho(X, \hat{X}|M_Y)$ with library size L in both directions (Fig. 3).

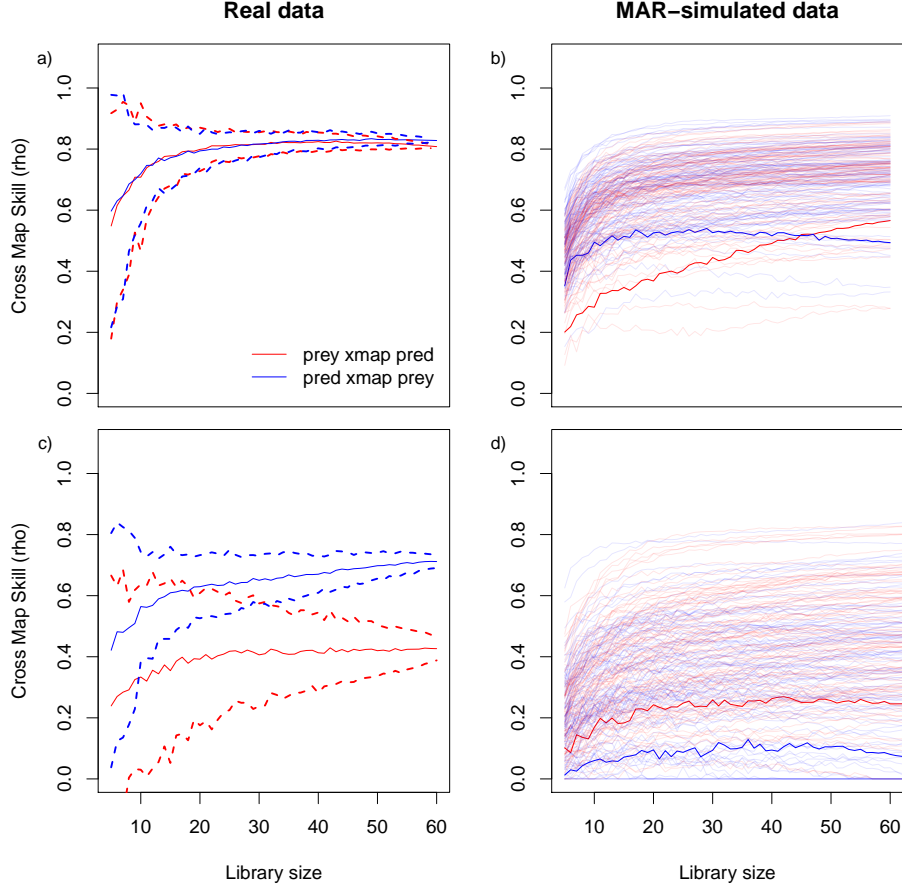


Figure 3: Convergent cross-mapping for Veilleux's CC05 dataset (a and b) and CC0.375 dataset (c and d). Dotted lines represent the confidence bands, obtained by bootstrapping.

Deterministic chaos in two-species competition models

In the 2-species chaotic competition model high-order lags tend to be selected (Fig. 14) despite the single timelag considered in the simulation model (i.e., higher nonlinearity is identified as high-order lags). Despite this potential overparameterization, the GC tests show that causality is detected for most timelags (including the optimal one, $p=7$) whenever causality is present. Further, the tests are not able to reject the null hypothesis of no GC when GC is not present (Fig. 4). GC performs therefore surprisingly well in this chaotic context.

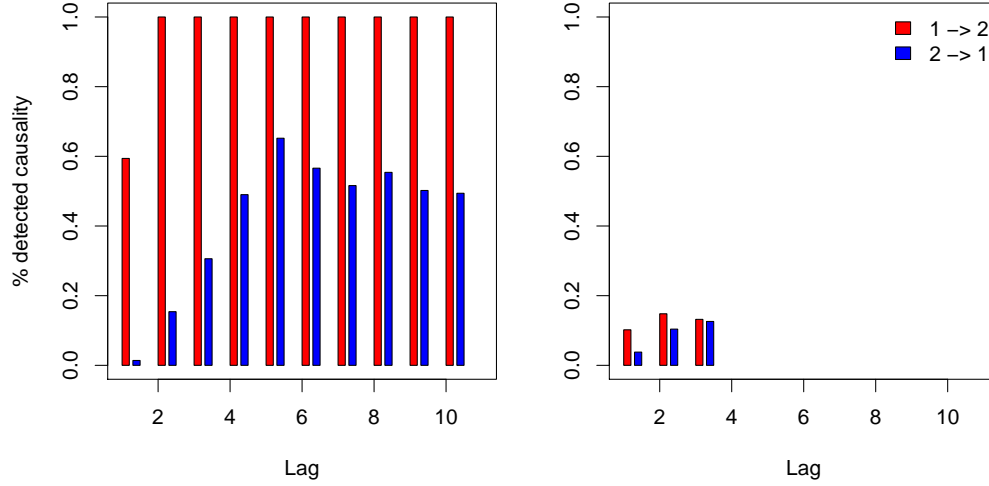


Figure 4: Proportion of detected Granger-causality, at the 10% significance threshold, over 500 chaotic simulations with (left) and without (right) actual interactions between species, depending on the number of time lags taken into account (x-axis). Without interactions, the optimal lag is 3 and the Wald test cannot be performed for $p > 3$.

	With interactions	Without interactions
	GC/CCM	GC/CCM
$1 \rightarrow 2$	100%/100% ($\phi = 1$)	13%/6% ($\phi = 0.11$)
$2 \rightarrow 1$	52%/48% ($\phi = 0.25$)	13%/26% ($\phi = 0.07$)

Table 2: Proportion of simulations with Granger-causality or CCM between x and y over 500 simulations, for the chaotic 2-species competition model. [FB and CP: We're still a bit unclear about the use of the ϕ coefficient for quantifying the agreement between the methods (from two binary time series). Intuitively, ϕ should be high for both high and low percentages for both methods, because they both produces in this case a lot of ones or a lot of zeroes. These results may suggest that ones and zeroes do not have completely symmetrical effects? TODO: Introduce handmade function for computing ϕ CP: Still not ok. What do you do when you have 0s in the denominator?]

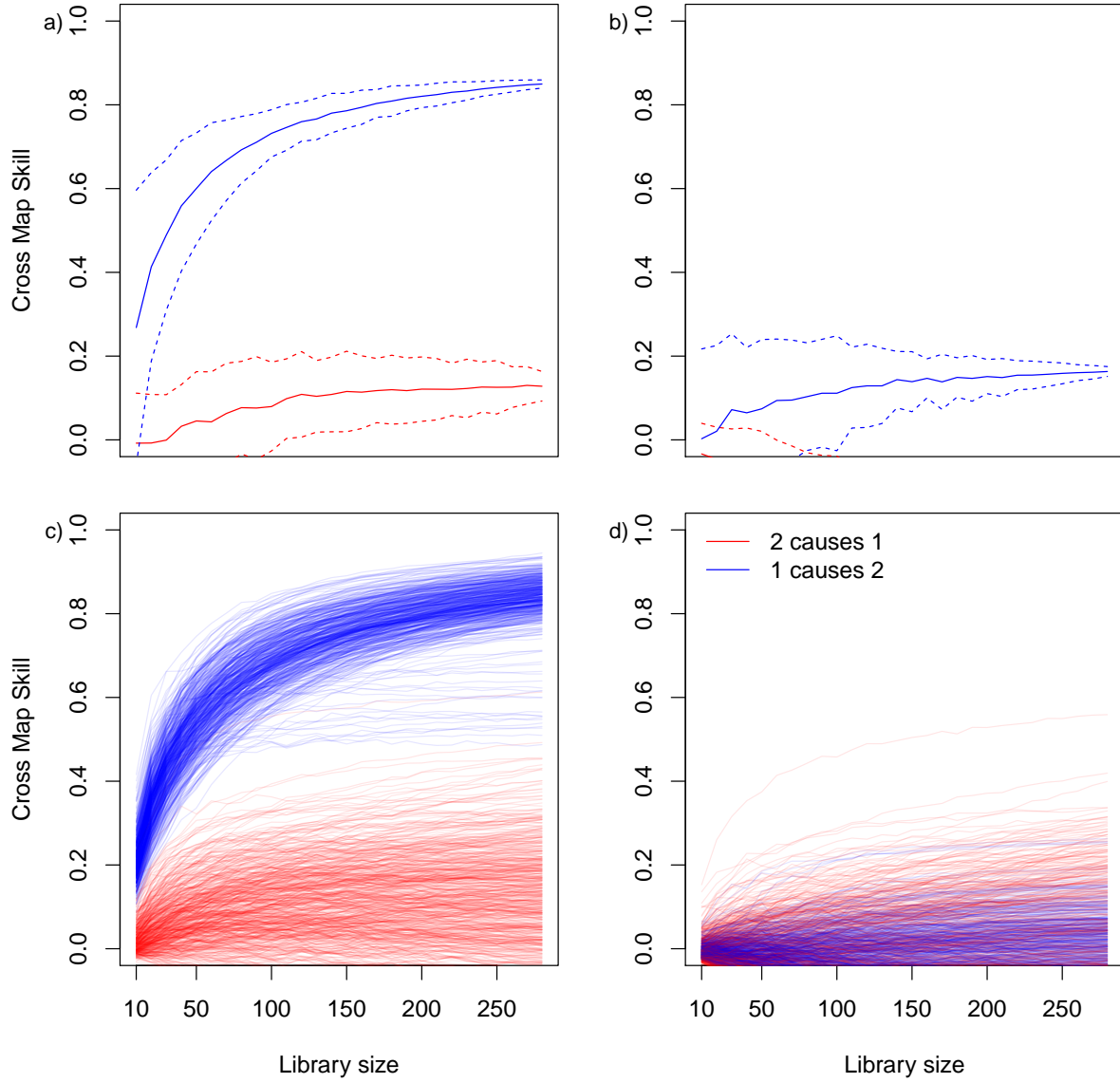


Figure 5: Convergent-cross mapping on simulated deterministic data, with (left) and without (right) competition between the two species

Two-species stochastic and nonlinear dynamics

Without environmental driver

In our case with 2-species nonlinear competition and noise, we see that GC and CCM perform quite similarly (Table 3), with both methods able to select in most cases causality and non-causality.

	With interactions	Without interactions
	GC/CCM	GC/CCM
$1 \rightarrow 2$	97%/28% ($\phi = 0.01$)	0%/0% ($\phi = 1$)
$2 \rightarrow 1$	100%/100% ($\phi = 1$)	0%/0% ($\phi = 1$)

Table 3: Proportion of simulations with Granger-causality or CCM between x and y over 500 simulations, for the stochastic 2-species competition model without environmental driver

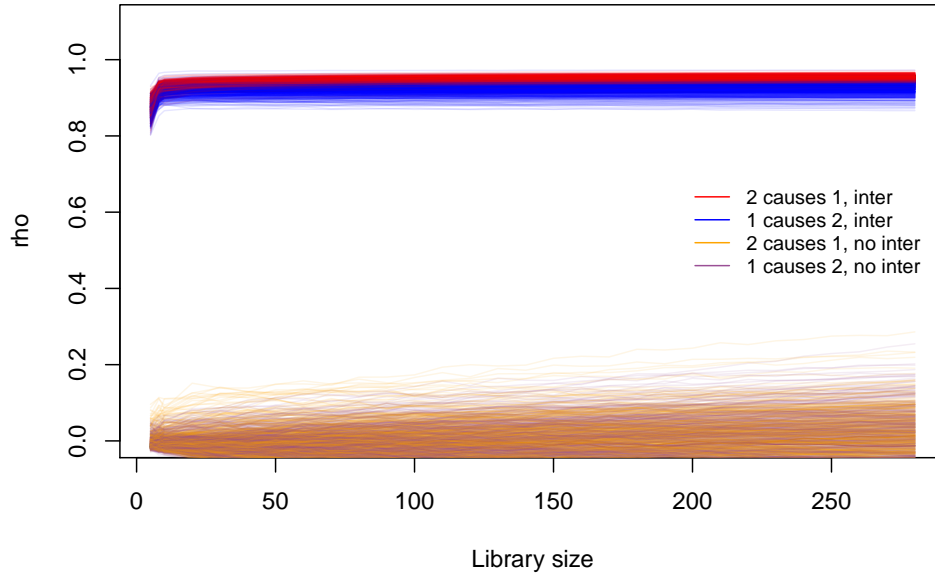


Figure 6: Convergent-cross mapping on simulated stochastic data, with (red, blue) and without (orange, purple) competition between the two species, over 500 simulations.

The MAR(p) model selected by BIC had a lag of $p = 3$ timesteps, confirming that small lags should be used in such models.

With an environmental driver

This considerably more complex case where the two species are influenced by a shared environmental driver (e.g., temperature) yielded less clear cut results than stochastic 2-species competition. Overall, both methods recognize the effect of temperature on temperature on the two species growth, with and without interactions (slightly lower score for CCM for species 2, but overall good performance, Table 4) . CCM was better at uncovering interactions that were present in this case, GC had a good performance for the strong interaction $2 \rightarrow 1$ but not the reserve $1 \rightarrow 2$. Surprisingly, this performance ranking is the reverse of that without the environmental driver. Both GC and CCM had difficulties indicating non-causality in this case (when there

are no interactions), and indicated false positives, averaged over the two possible interactions, about 50% for both methods. It is therefore quite logical to say that both methods fail to detect an absence of interaction in this case. We also asked the “stupid” question of whether species 1 and 2 influenced temperature, assuming that both methods would fail. GC did surprisingly well in this context, showing only between 15 and 20% of false positives.

	With interactions	Without interactions
	GC/CCM	GC/CCM
sp1->temp	21%/98% ($\phi = -0.06$)	19%/99% ($\phi = 0.1$)
temp->sp1	100%/99%	100%/100% ($\phi = 1$)
sp1->sp2	42%/100%	53%/20% ($\phi = -0.02$)
sp2->sp1	97%/100% (-0.01)	48%/72% ($\phi = -0.02$)
sp2->temp	16%/100%	14%/20% ($\phi = -0.01$)
temp->sp2	100%/100%	100%/72%

Table 4: Proportion of simulations with Granger-causality or CCM between x and y over 500 simulations for a model with 2 species and a driver (temperature)

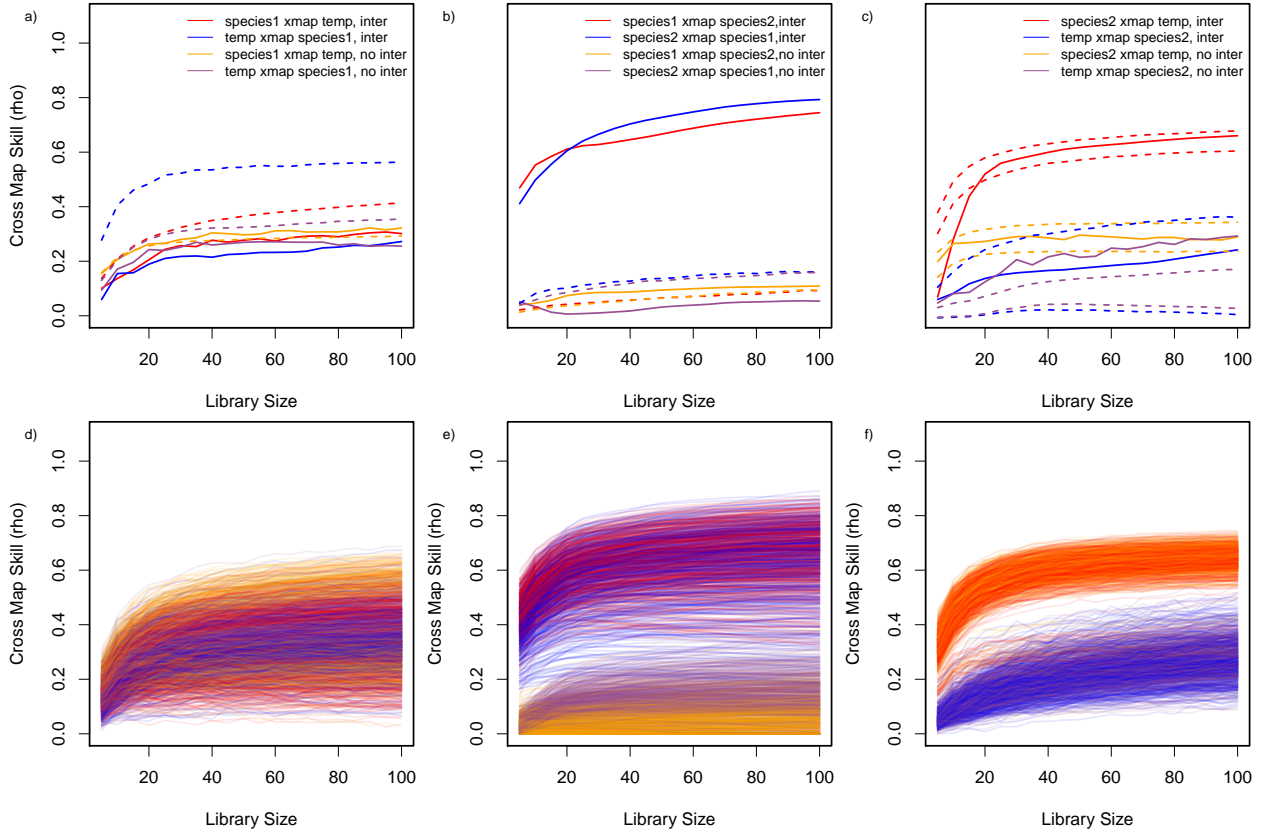


Figure 7: Convergent cross-mapping for the two species forced by an environmental driver (denoted as temp), when interactions are present (blue, red) and when interactions are absent (purple, orange), for 500 simulations. Dashed lines indicate the 10% interval for rho-values obtained from surrogate time series, i.e., time series that have the same seasonal forcing but whose cross-correlations are altered.

Larger interaction webs

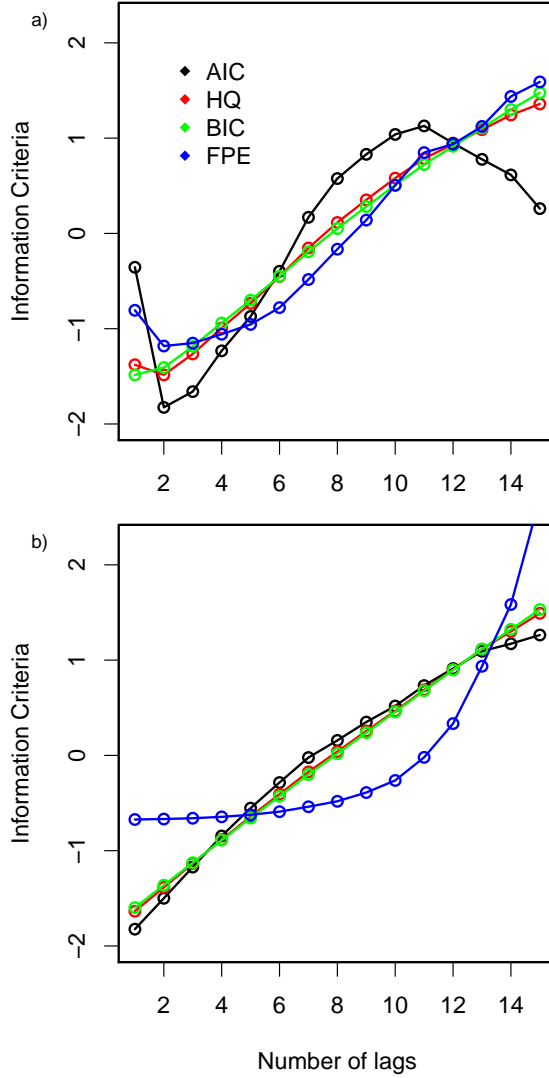


Figure 8: Lag order selection for the 10-species (a) and [one of the] 20-species stochastic community model (b).

Here we report the results of analyses for 10 and 20 species modular interaction webs (5 species as in Sugihara *et al.* done in Supplement SXX [yet to do]). Lag order selection revealed that low-order $MAR(p)$ models were selected (Fig. 8), 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 SXX) because the basic LASSO-penalized $VAR(1)$ models poorly identify modular interactions webs (Charbonnier *et al.*, 2010). The recall, or true positive rate, that records how many true positives are identified as such, is almost always above 60% for the LASSO-based method (siMoNe, Chiquet *et al.*, 2008;

Charbonnier *et al.*, 2010) and up to 80%, which is a relatively good performance. Surprisingly, we found that pairwise (direct) Granger causality testing was quite efficient, in this case more than the LASSO-based method. One issue though is that the level of significance has to be adjusted (we use a global $\alpha = 20\%$), which is always somewhat subjective (but there are also always several tuning parameters in LASSO-based inferences, so some degree of subjectivity is everywhere).

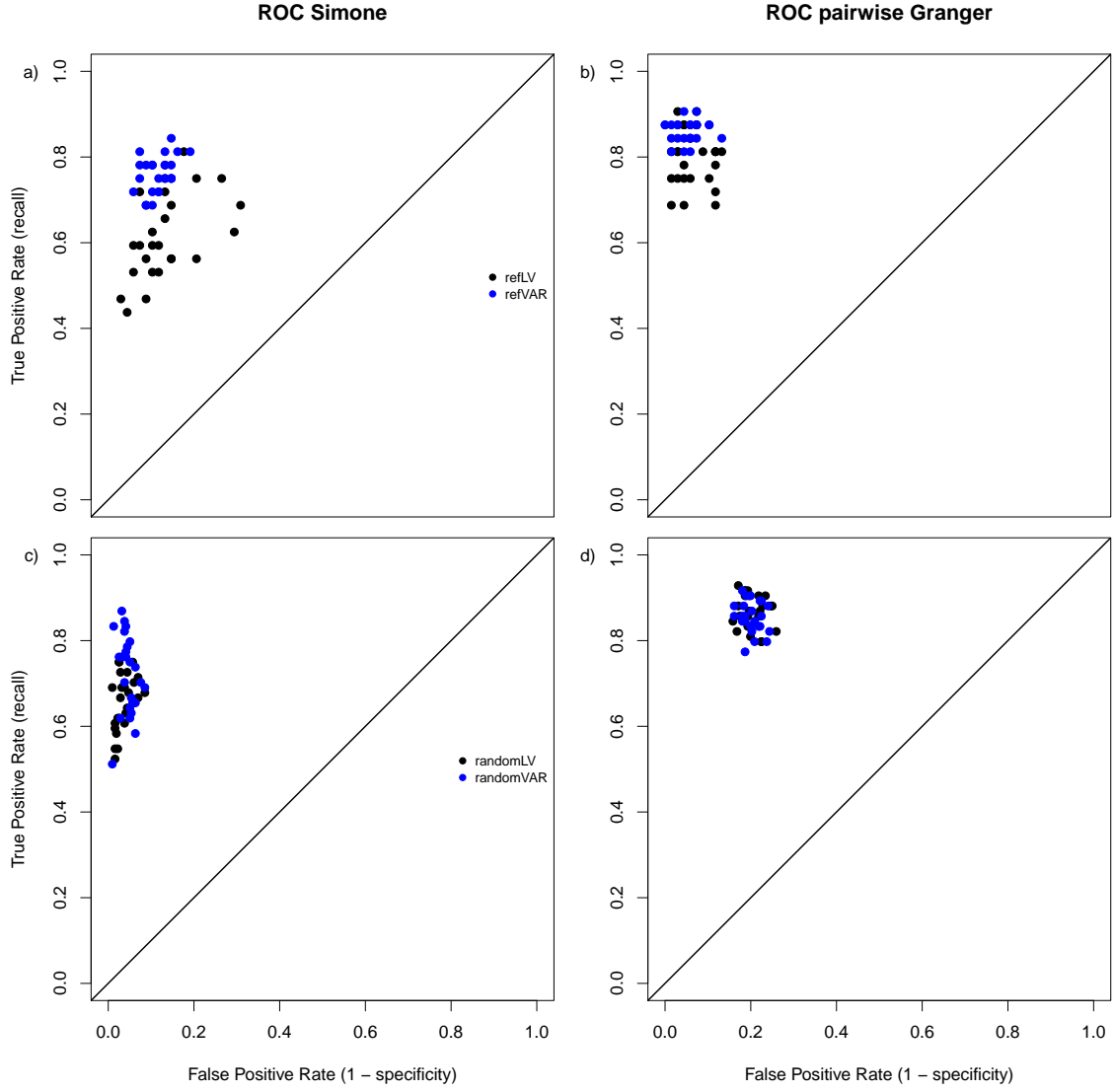


Figure 9: ROC plot for the 10 and 20 species webs, as analysed with a LASSO-based penalized MAR(1) model including clustering. In panel (a) and (b), each point represent multiple simulations of the same parameter set; in (c) and (d), we see multiple parameter sets. **[Perhaps we can add a third column where we show CCM pairwise, if the results are not too ugly when we have the simplex].**

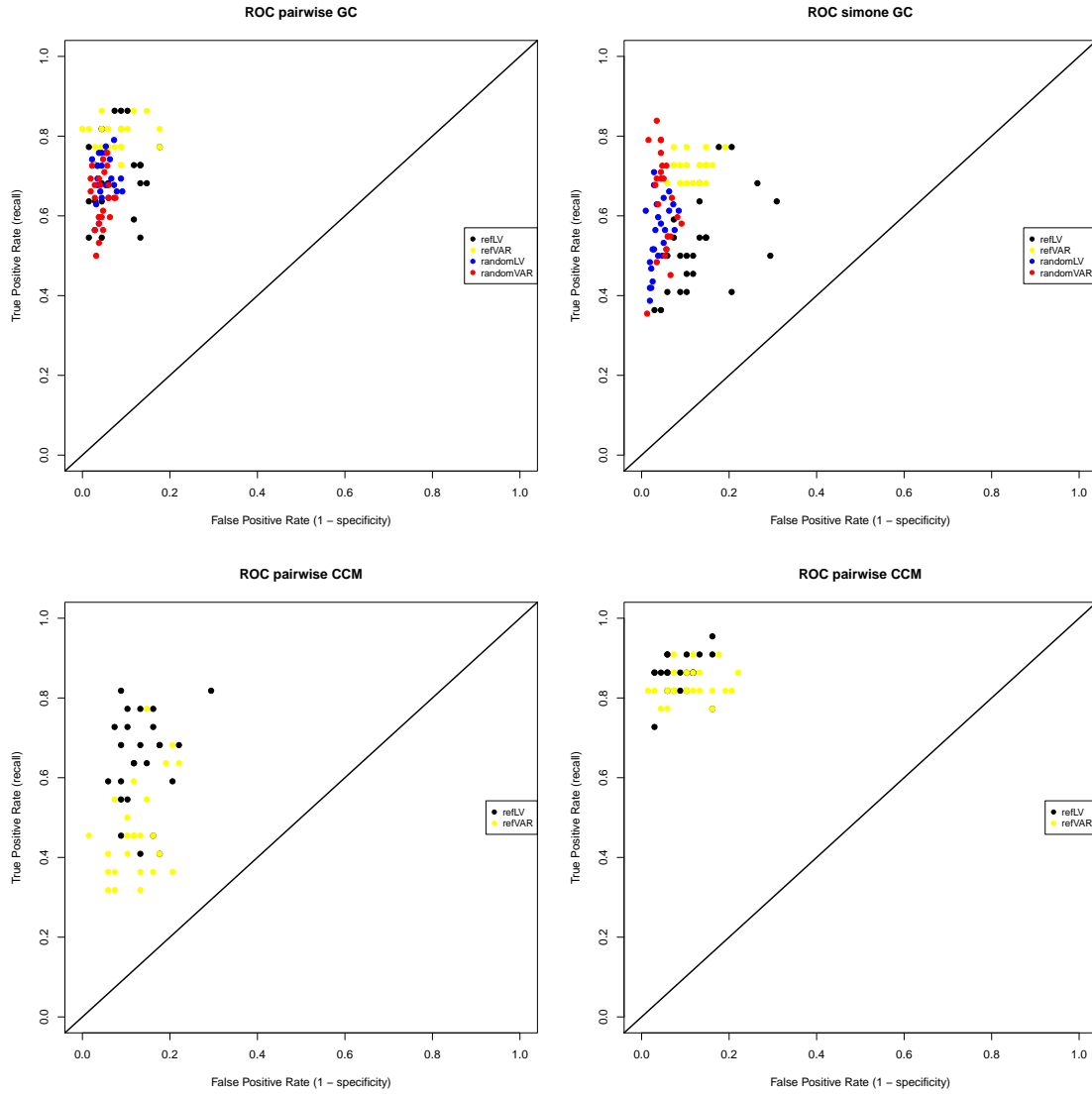


Figure 10: ROC curves for the 10 and 20-species model using the Granger Causality method (top) and the CCM method (bottom), with different ways of computing the causality. For GC, left is pairwise ; right is LASSO. For CCM, left is the Cobey-Baskerville method, right is the “shuffling”-method (no 20-species model yet, still running)

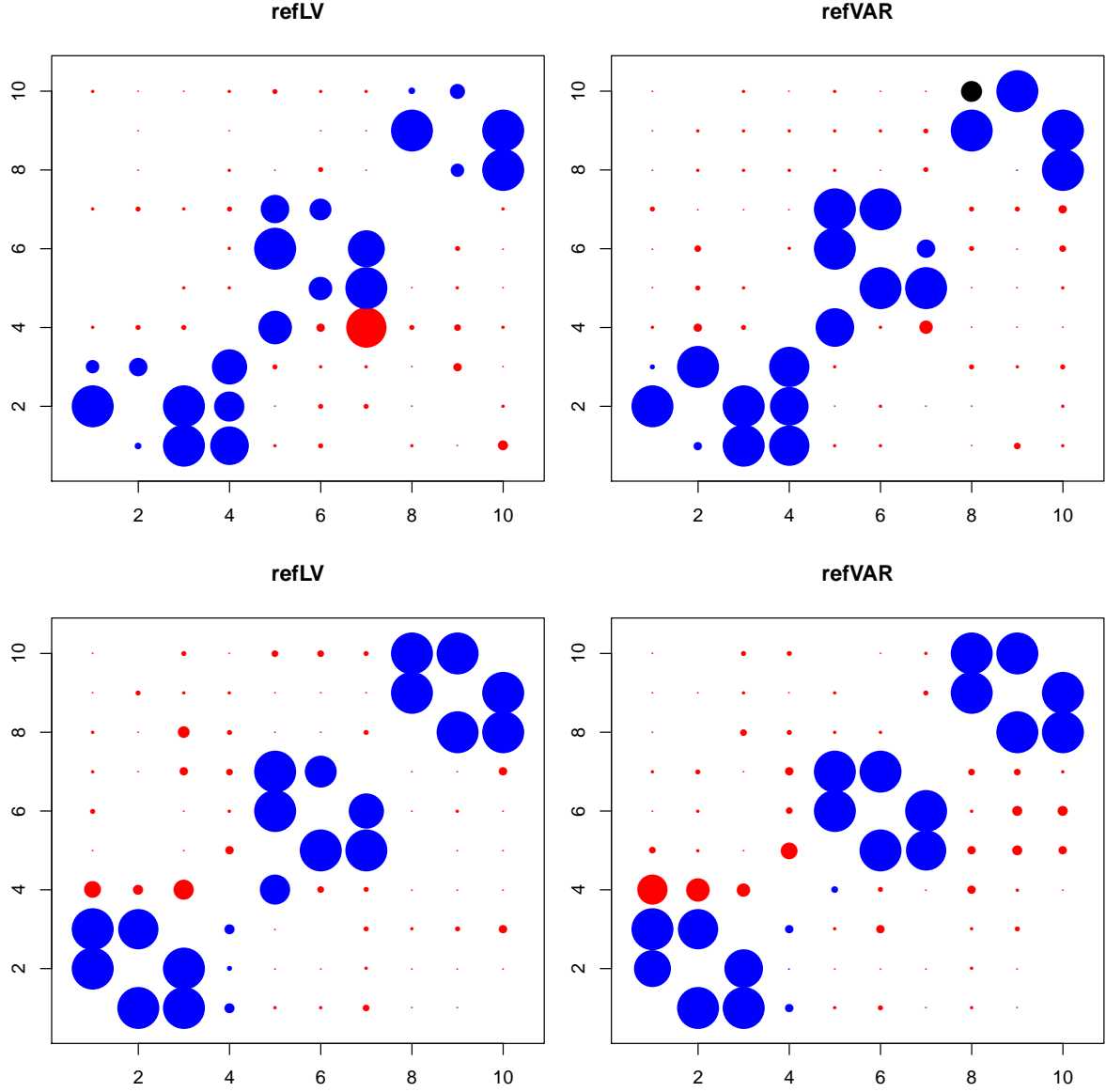


Figure 11: Interaction matrices obtained from pairwise GC (top) or 'shuffling' CCM (bottom). Blue circles are the true positives, red circles are false positives and black circles are false negatives. For the true and false positives, the size of the circles is proportional to the proportion of detection over 25 simulations.

Discussion

We have shown above that MAR(p) modelling and Granger causality testing are fairly robust to nonlinearities in ecological dynamics, when applied on the appropriate logarithmic-abundance scale, and combined with model selection by information criteria. This is true for many nonlinear simulation models, including deterministic chaos. 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 to 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 both infer interactions (stochastic competition) and from cases where both methods seem to fail to some degree (i.e., two competing species forced similarly by a shared environmental driver). Therefore, an important conclusion from our study is that both Granger causality and CCM are able to yield reasonable inferences on similar datasets, unless there is a very high degree of confounding (see also Cobey & Baskerville, 2016, who showed similar results).

Moreover, we use here false discovery rates and regularized models (i.e., LASSO-penalized MAR(1) models developed for modular interaction networks, Charbonnier *et al.*, 2010) to tackle relatively-high dimensional models (10 and 20 species). This allows to better infer Granger causality in these contexts that, we surmise, will be most exciting to ecologists working on interacting species using community-level data. The results demonstrate that simple pairwise Granger causality (i.e., using 2×2 MAR(p) models) are almost as good as the penalized MAR(1) models in finding the interaction network. We elaborate on these results and possible explanations below.

Is Granger causality a useful concept for highly nonlinear dynamical systems, and why?

Sugihara *et al.* (2012) suggested that Granger causality would work well for simulated (log)-linear systems while CCM would work well for nonlinear dynamical systems. Reality, however, is less clear-cut. First, the analyses performed by Sugihara *et al.* (2012) *et al.* on the Veilleux and other datasets seem to rely on a slightly unorthodox model selection procedure. Reanalysing the data with a more classic, information-criteria motivated lag order selection, we show that in fact GC is perfectly able to find causality in the Veilleux predator-prey datasets.

We also reported above on results where even deterministic chaos can, surprisingly, be well identified by MAR(p) models / GC inference (we expected CCM to completely dominate the scores and GC to fail entirely, but GC provided reasonable performances). Thus GC can be useful for highly nonlinear systems. Moreover, the converse is also true: some log-linear cases can also be well-identified by CCM, even though CCM relies upon the possibility to reconstruct an attractor in state space. Of course, as CCM has been developed specifically from nonlinear dynamical systems theory (and relies upon state-space reconstruction of chaotic models), it seems more appropriate for highly nonlinear and weakly stochastic cases, while MAR(p) models are best performers in weakly nonlinear and more stochastic situations. Nonetheless, it was very surprising

to see how well MAR(p) models can perform in deterministic chaos in a number of cases, unlike what could have been expected from the previous study of Sugihara *et al.* (2012). Though we clearly acknowledge that some chaotic scenarios may inevitably lead to a difficulty in GC inference.

Several recent attempts to perform nonlinear inference of interactions support our findings. Suzuki *et al.* (2017) use stochastic GLV models and a model selection algorithm derived from Fisher & Mehta (2014), and even though they use the S-map, because their evaluation is based on prediction, they actually perform a nonlinear Granger inference. . The neuroscience, gene regulation networks and related dynamical systems literature is awash with nonlinear GC inferences, often confirmed by simulations of nonlinear stochastic systems (Marinazzo *et al.*, 2008; Yang *et al.*, 2017) [some more examples here]. Hence there is fairly good evidence that either log-linear (power-law) or nonlinear GC inference is possible for nonlinear stochastic systems of interacting species.

As shown by our analyses for large interaction webs, GC can be extended to quite large webs, both in a pairwise GC context and in a conditional GC (full model fitting) context. **[More discussion of these results here and more on CCM results when we have those]**

How can Granger causality and convergent cross-mapping yield similar inferences?

Here, we would like to go back to the heart of the issue that Sugihara *et al.* (2012) highlighted, causality reversion in nonlinear dynamical systems. The standard Granger causality holds that whenever a model $Y_{t+h}|(Y_t, X_{t \in A_t})$ better predicts the observed time series (y_t) than a model $Y_{t+h}|(Y_{t \in A_t})$, then x is causal for y . CCM instead holds that whenever $\rho(x, \hat{x}|M_Y)$ increases strongly with the library size L used to reconstruct 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+h}
- CCM compares knowledge about Y_t vs. no knowledge about Y_t in prediction of X_t .

Using a standard autoregressive model, the equivalent of a CCM test $x \rightarrow y$ would be predicting X_t by a model $X_t|Y_t$ vs X_t . There is no conditionality upon past X_t values in the prediction step of the algorithm. Thus there is no causality reversion that is intrinsic to nonlinear dynamic testing: GC and CCM are simply two different types of causality testing that are based upon *different assumptions on the conditioning set and ways to select models*. We therefore conclude for these methods and concepts to work relatively similarly,

they must share some underlying similarities that not yet evident to theoretical and statistical ecologists. More mathematical research on the possible connections and differences between these methods is obviously needed to better see in which scenarios each should be favored.

Going further

Based on our simulations, complemented with those of the neuroscience literature (Ding *et al.*, 2006; Chen *et al.*, 2006; Barnett & Seth, 2014; Marinazzo *et al.*, 2008), we suggest that Granger causality, in its log-linear (this paper) or nonlinear varieties (Marinazzo *et al.*, 2008; Suzuki *et al.*, 2017), is appropriate to infer ecological interactions. CCM is another interesting method to infer interactions. Although it has been shown that some scenarios like seasonal forcing might render interaction inference difficult (Cobey & Baskerville, 2016), we think that as always details of implementation are important, thus it is important not to disqualify a method in the early stages of its development.

From an almost sociological viewpoint, we conclude that recommending to abandon established statistical methods like GC, or linear modelling after transformation more generally, in favor of promising yet not fully tested statistical methods may not always help ecological science. More fruitful methodological development might highlight the pros and cons of new developments in statistical inference under a broad range of simulated scenarios.

From a statistical perspective, looking at the various implementations of GC and CCM, it seems that the hardest methodological choice to make is almost invariably not the type of functional form or embedding (linear, nonlinear, etc.) but the conditioning set, i.e. the information that is considered to be known for the prediction (Eichler, 2013). Strategies to better understand how to choose the conditioning set when doing causal inference will be, we believe, a very important feature of ecological interaction inference for the years to come.

References

- Barabás, G., Michalska-Smith, M.J. & Allesina, S. (2017). Self-regulation and the stability of large ecological networks. *Nature ecology & evolution*, 1, 1870.
- Barnett, L., Barrett, A.B. & Seth, A.K. (2009). Granger causality and transfer entropy are equivalent for gaussian variables. *Physical review letters*, 103, 238701.
- Barnett, L. & Seth, A.K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, 223, 50–68.

- Barraquand, F., Louca, S., Abbott, K.C., Cobbold, C.A., Cordoleani, F., DeAngelis, D.L., Elder, B.D., Fox, J.W., Greenwood, P., Hilker, F.M. *et al.* (2017). Moving forward in circles: challenges and opportunities in modelling population cycles. *Ecology Letters*, 20, 1074–1092.
- Barraquand, F., Picoche, C., Maurer, D., Carassou, L. & Auby, I. (2018). Coastal phytoplankton community dynamics and coexistence driven by intragroup density-dependence, light and hydrodynamics. *Oikos*, 127, 1834–1852.
- Berlow, E.L., Neutel, A.M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., Fox, J.W., Jansen, V.A., Iwan Jones, J., Kokkoris, G.D. *et al.* (2004). Interaction strengths in food webs: issues and opportunities. *Journal of animal ecology*, 73, 585–598.
- Brose, U. (2010). Body-mass constraints on foraging behaviour determine population and food-web dynamics. *Functional Ecology*, 24, 28–34.
- Brose, U., Williams, R.J. & Martinez, N.D. (2006). Allometric scaling enhances stability in complex food webs. *Ecology Letters*, 9, 1228–1236.
- Certain, G., Barraquand, F. & Gårdmark, A. (2018). How do MAR(1) models cope with hidden nonlinearities in ecological dynamics? *Methods in Ecology and Evolution*, 9, 1975–1995.
- Charbonnier, C., Chiquet, J. & Ambroise, C. (2010). Weighted-lasso for structured network inference from time course data. *Statistical applications in genetics and molecular biology*, 9.
- Chase, J.M. (2003). Community assembly: when should history matter? *Oecologia*, 136, 489–498.
- Chen, Y., Bressler, S.L. & Ding, M. (2006). Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *Journal of neuroscience 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.
- Cobey, S. & Baskerville, E.B. (2016). Limits to causal inference with state-space reconstruction for infectious disease. *PloS one*, 11, e0169050.
- Coyte, K.Z., Schluter, J. & Foster, K.R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science*, 350, 663–666.
- Detto, M., Molini, A., Katul, G., Stoy, P., Palmroth, S. & Baldocchi, D. (2012). Causality and persistence in ecological systems: a nonparametric spectral granger causality approach. *The American Naturalist*, 179, 524–535.

- Ding, M., Chen, Y. & Bressler, S. (2006). Granger causality: Basic theory and application to neuroscience. *Handbook of time series analysis*, pp. 437–460.
- Eichler, M. (2013). Causal inference with multiple time series: principles and problems. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371, 20110613.
- Fisher, C.K. & Mehta, P. (2014). Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One*, 9, e102451.
- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–438.
- Hampton, S.E., Holmes, E.E., Scheef, L.P., Scheuerell, M.D., Katz, S.L., Pendleton, D.E. & Ward, E.J. (2013). Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology*, 94, 2663–2669.
- Ives, A., Dennis, B., Cottingham, K. & Carpenter, S. (2003). Estimating community stability and ecological interactions from time-series data. *Ecological Monographs*, 73, 301–330.
- Jost, C. & Ellner, S.P. (2000). Testing for predator dependence in predator-prey dynamics: a non-parametric approach. *Proceedings of the Royal Society of London B: Biological Sciences*, 267, 1611–1620.
- Link, J.S. (2002). What does ecosystem-based fisheries management mean. *Fisheries*, 27, 18–21.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.
- Marinazzo, D., Pellicoro, M. & Stramaglia, S. (2008). Kernel-Granger causality and the analysis of dynamical networks. *Physical Review E*, 77, 1–9.
- May, R. (1973). *Stability and complexity in model ecosystems*. Princeton University Press, Princeton, USA.
- Michailidis, G. & d’Alché Buc, F. (2013). Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues. *Mathematical biosciences*, 246, 326–334.
- Mukhopadhyay, N.D. & Chatterjee, S. (2006). Causality and pathway search in microarray time series experiment. *Bioinformatics*, 23, 442–449.
- Murdoch, W., Kendall, B., Nisbet, R., Briggs, C., McCauley, E. & Bolser, R. (2002). Single-species models for many-species food webs. *Nature*, 417, 541–543.

- Mutshinda, C.M., O’ Hara, R.B. & Woiwod, I.P. (2011). A multispecies perspective on ecological impacts of climatic forcing. *Journal of Animal Ecology*, 80, 101–107.
- Mutshinda, C.M., O’Hara, R.B. & Woiwod, I.P. (2009). What drives community dynamics? *Proceedings of the Royal Society B: Biological Sciences*, 276, 2923–2929.
- Paluš, M. (2008). From Nonlinearity to Causality: Statistical testing and inference of physical mechanisms underlying complex dynamics. *Contemporary Physics*, 48, 307–348.
- Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3, 96–146.
- 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.
- Rall, B.C., Guill, C. & Brose, U. (2008). Food-web connectance and predator interference dampen the paradox of enrichment. *Oikos*, 117, 202–213.
- Sims, C. (1980). Macroeconomics and reality. *Econometrica*, 48, 1–48.
- Sugihara, G., May, R., Ye, H., Hsieh, C.h., Deyle, E., Fogarty, M. & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 338, 496–500.
- Suzuki, K., Yoshida, K., Nakanishi, Y. & Fukuda, S. (2017). An equation-free method reveals the ecological interaction networks within complex microbial ecosystems. *Methods in Ecology and Evolution*, 8, 1774–1785.
- Veilleux, B.G. (1979). An analysis of the predatory interaction between paramecium and didinium. *J Anim Ecol*, 48, 787–803.
- Weitz, J. & Levin, S. (2006). Size and scaling of predator-prey dynamics. *Ecol. Lett*, 9, 548–557.
- Wootton, J. & Emmerson, M. (2005). Measurement of interaction strength in nature. *Annual Review of Ecology, Evolution and Systematics*, 36, 419–444.
- Yang, G., Wang, L. & Wang, X. (2017). Reconstruction of complex directional networks with group lasso nonlinear conditional granger causality. *Scientific reports*, 7, 2991.
- 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.

Yodzis, P. (1998). Local trophodynamics and the interaction of marine mammals and fisheries in the benguela ecosystem. *Journal of Animal Ecology*, 67, 635–658.

Yodzis, P. & Innes, S. (1992). Body size and consumer-resource dynamics. *American Naturalist*, 139, 1151.

Appendices and Supplements

Nonlinear Granger causality

Transfer entropy and nonlinear Granger causality

Transfer entropy can be defined as

$$\mathcal{T}_{x \rightarrow y|z} = H(\mathbf{y}^{t_m+1} | \mathbf{y}^{t_m}, \mathbf{z}^{t_m}) - H(\mathbf{y}^{t_m+1} | \mathbf{y}^{t_m}, \mathbf{x}^{t_m}, \mathbf{z}^{t_m})$$

where $\mathbf{y}^{t_m+1} = (y_2, \dots, y_{t_m+1})$ and $\mathbf{y}^{t_m} = (y_1, \dots, y_{t_m})$. The quantity $H(x|y) = H(x, y) - H(y)$ is a conditional entropy, defined with $H(x)$ the Shannon entropy. It has then been shown that the Granger causal measure $\mathcal{G}_{x \rightarrow y|z} = \ln(\frac{\sigma_\eta^2}{\sigma_\epsilon^2})$ where the residuals errors are take from eqs. 2 can be generalized to $\mathcal{T}_{x \rightarrow y|z}$. In the linear case, Barnett *et al.* (2009) proved that $\mathcal{G}_{x \rightarrow y|z} = 2\mathcal{T}_{x \rightarrow y|z}$, so that Granger causality through MAR(1) modelling is a special case of causality defined through transfer entropy.

In general, any method which evaluates whether adding a new time series \mathbf{x} to a dynamical system for variables y_1, \dots, y_n improves prediction of y_i can be defined as a generalised GC method $x \rightarrow y_i | (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$. Quite a number of nonlinear Granger causality inference techniques then fall within this category (e.g., Marinazzo *et al.*, 2008; Paluš, 2008).

Additional results

Example simulation of 2-species stochastic Ricker model

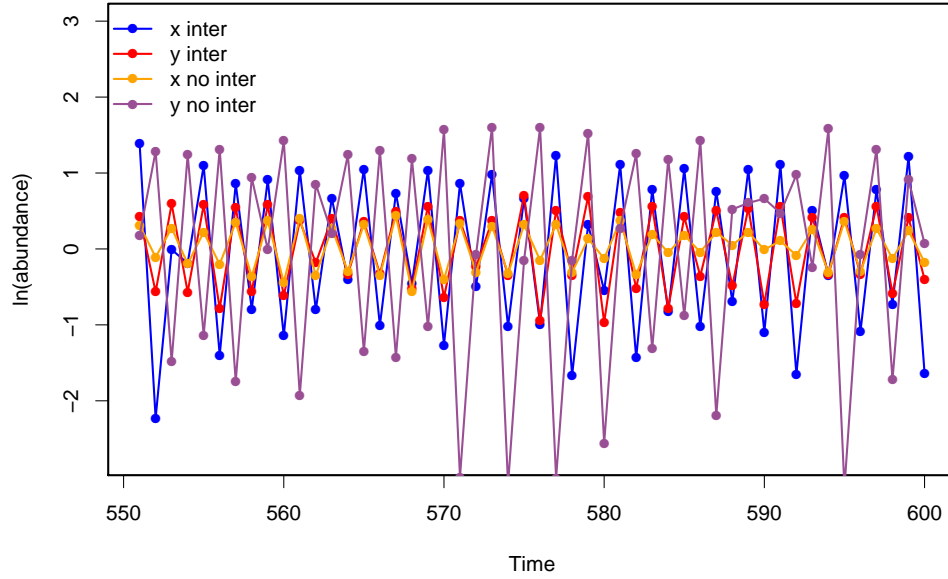


Figure 12: 2-species Ricker model [ref to eqs. of main text] with (blue = species 1, red = species 2) and without (purple = species 1, orange = species 2) competition.

Effect of log-transformation on CCM

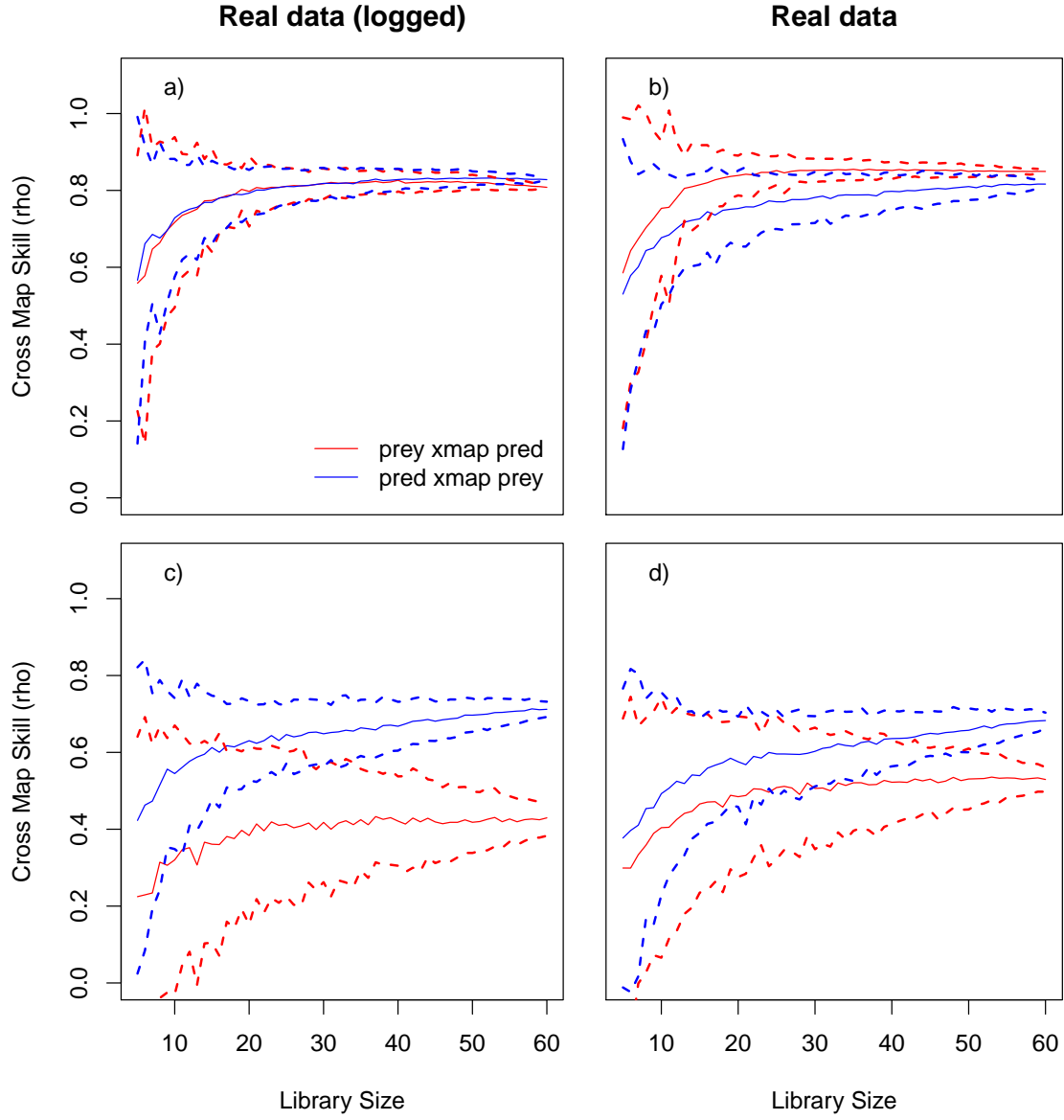


Figure 13: Convergent Cross-Mapping with (left) and without (right) \ln -transform of the data for the Veilleux dataset

Lag order p selection in the $MAR(p)$ framework, for the deterministic competition model

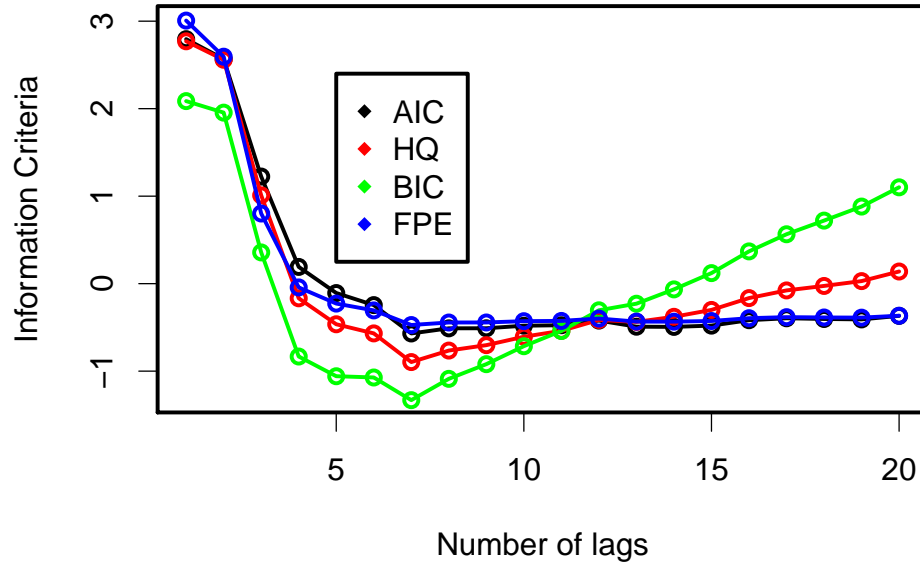


Figure 14: Results of model information criteria vs. lag order for the simulated deterministic competition model of eqs. XX