

Inferring species interactions using Granger causality and convergent cross mapping

Frédéric Barraquand ^{1,2} , Coralie Picoche ², Matteo Detto ³ & Florian Hartig ⁴

¹CNRS, Institute of Mathematics of Bordeaux

²University of Bordeaux, Institute of Mathematics of Bordeaux

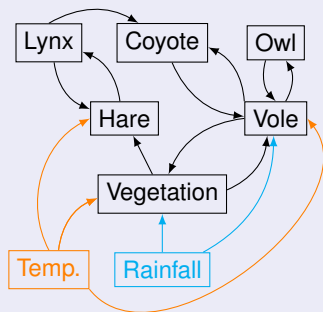
³Princeton University & Smithsonian Tropical Research Institute

⁴University of Regensburg

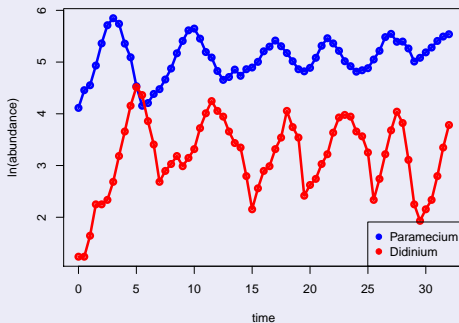
19 January 2022, Maynooth TSE Seminar

Ecological interactions, causality, and time series

Interaction network in a dynamic sense \Leftrightarrow pop. growth rate_i = f(density_j)



Time series data available



Veilleux (1979), predator = red, prey = blue

Granger Causality: back to basics

$\mathbf{x} = (x_t)_{t \in \llbracket 1, T \rrbracket}$ Granger-causes $\mathbf{y} = (y_t)_{t \in \llbracket 1, T \rrbracket} \iff$ including \mathbf{x} in a time series model for \mathbf{y} improves prediction of \mathbf{y} .

Granger Causality: back to basics

$\mathbf{x} = (x_t)_{t \in \llbracket 1, T \rrbracket}$ Granger-causes $\mathbf{y} = (y_t)_{t \in \llbracket 1, T \rrbracket} \iff$ including \mathbf{x} in a time series model for \mathbf{y} improves prediction of \mathbf{y} .

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

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

Granger Causality: back to basics

$\mathbf{x} = (x_t)_{t \in \llbracket 1, T \rrbracket}$ Granger-causes $\mathbf{y} = (y_t)_{t \in \llbracket 1, T \rrbracket} \iff$ including \mathbf{x} in a time series model for \mathbf{y} improves prediction of \mathbf{y} .

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

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

Pairwise Granger causality if $\sigma_{\epsilon}^2 < \sigma_{\eta}^2$.

Granger Causality: back to basics

$\mathbf{x} = (x_t)_{t \in \llbracket 1, T \rrbracket}$ Granger-causes $\mathbf{y} = (y_t)_{t \in \llbracket 1, T \rrbracket} \iff$ including \mathbf{x} in a time series model for \mathbf{y} improves prediction of \mathbf{y} .

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

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

Pairwise Granger causality if $\sigma_\epsilon^2 < \sigma_\eta^2$. $G_{x \rightarrow y} = \ln \left(\frac{\sigma_\eta^2}{\sigma_\epsilon^2} \right)$ Econometrics (Granger, 1969; Sims, 1980; Geweke, 1982, 1984), Neuroscience (Barnett and Seth, 2014), Genetics (Michailidis and d'Alché Buc, 2013), Ecology (Detto et al., 2012).

Granger Causality: back to basics

$\mathbf{x} = (x_t)_{t \in \llbracket 1, T \rrbracket}$ Granger-causes $\mathbf{y} = (y_t)_{t \in \llbracket 1, T \rrbracket} \iff$ including \mathbf{x} in a time series model for \mathbf{y} improves prediction of \mathbf{y} .

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

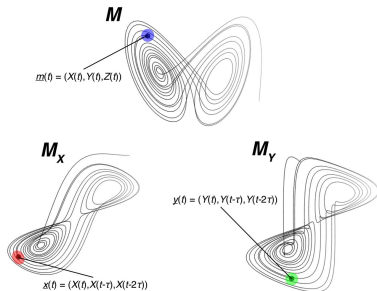
$$y_t = \sum_{i=1}^p a_{1i} x_{t-i} + \sum_{i=1}^p a_{2i} y_{t-i} + \epsilon_t \quad (2)$$

Pairwise Granger causality if $\sigma_{\epsilon}^2 < \sigma_{\eta}^2$. $G_{x \rightarrow y} = \ln \left(\frac{\sigma_{\eta}^2}{\sigma_{\epsilon}^2} \right)$ Econometrics (Granger, 1969; Sims, 1980; Geweke, 1982, 1984), Neuroscience (Barnett and Seth, 2014), Genetics (Michailidis and d'Alché Buc, 2013), Ecology (Detto et al., 2012).

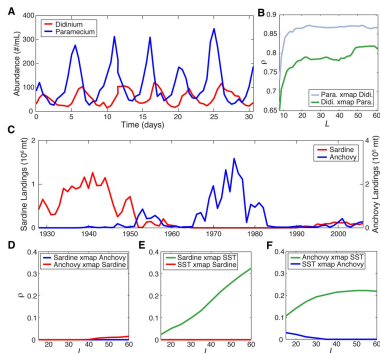
Sugihara et al. (2012). Detecting causality in complex ecosystems. *Science*, 338(6106), 496-500.

- Granger causality has problems with nonlinear processes (“separability” issues), works only for linear stochastic systems.
- New method = **Convergent Cross Mapping** (\exists other methods of state-space reconstruction, but less popular)

CCM: Sugihara et al. (2012)'s Convergent Cross-Mapping



Attractor reconstruction based on Takens theorem: build a 3-variable attractor with one variable 3 times delayed – from Sugihara et al. (2012)



Evaluation of causality with correlation coefficient $\rho(\mathbf{x}, \hat{\mathbf{x}}|M_Y)$ – from Sugihara et al. (2012)

CCM: X causes $Y \Leftrightarrow$ predict X_t values from M_Y . “Causality reversion”.

Granger: X causes $Y \Leftrightarrow Y_{t+h} | (\mathbf{y}_{k \in \text{past}}, \mathbf{x}_{k \in \text{past}})$ better predicts Y_{t+h} than $Y_{t+h} | \mathbf{y}_{k \in \text{past}}$

What is a “linear stochastic system”, really? Multivariate AutoRegressive aka VAR(1) or MAR(1) models

In fact these are linear on the logarithmic abundance scale

$$\mathbf{x}_t = \ln(\mathbf{n}_t), \mathbf{x}_{t+1} = \mathbf{a} + \mathbf{B}\mathbf{x}_t + \mathbf{C}\mathbf{u}_t + \mathbf{e}_t, \mathbf{e}_t \sim \mathcal{N}_d(\mathbf{0}, \Sigma) \quad (3)$$

Ives et al. (2003). Estimating community stability and ecological interactions from time-series data. *Ecol. Monogr.*, 73(2), 301-330.

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

What is a “linear stochastic system”, really? Multivariate AutoRegressive aka VAR(1) or MAR(1) models

In fact these are linear on the logarithmic abundance scale

$$\mathbf{x}_t = \ln(\mathbf{n}_t), \mathbf{x}_{t+1} = \mathbf{a} + \mathbf{B}\mathbf{x}_t + \mathbf{C}\mathbf{u}_t + \mathbf{e}_t, \mathbf{e}_t \sim \mathcal{N}_d(\mathbf{0}, \Sigma) \quad (3)$$

Ives et al. (2003). Estimating community stability and ecological interactions from time-series data. Ecol. Monogr., 73(2), 301-330.

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

Non-zero $b_{ij} \Leftrightarrow$ **Conditional Granger Causality**: $G_{j \rightarrow i | \text{other variables}}$.

What is a “linear stochastic system”, really? Multivariate AutoRegressive aka VAR(1) or MAR(1) models

In fact these are linear on the logarithmic abundance scale

$$\mathbf{x}_t = \ln(\mathbf{n}_t), \mathbf{x}_{t+1} = \mathbf{a} + \mathbf{B}\mathbf{x}_t + \mathbf{C}\mathbf{u}_t + \mathbf{e}_t, \mathbf{e}_t \sim \mathcal{N}_d(\mathbf{0}, \Sigma) \quad (3)$$

Ives et al. (2003). Estimating community stability and ecological interactions from time-series data. Ecol. Monogr., 73(2), 301-330.

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

Non-zero $b_{ij} \Leftrightarrow$ **Conditional Granger Causality**: $G_{j \rightarrow i | \text{other variables}}$.

Why log?

What is a “linear stochastic system”, really? Multivariate AutoRegressive aka VAR(1) or MAR(1) models

In fact these are linear on the logarithmic abundance scale

$$\mathbf{x}_t = \ln(\mathbf{n}_t), \mathbf{x}_{t+1} = \mathbf{a} + \mathbf{B}\mathbf{x}_t + \mathbf{C}\mathbf{u}_t + \mathbf{e}_t, \mathbf{e}_t \sim \mathcal{N}_d(\mathbf{0}, \Sigma) \quad (3)$$

Ives et al. (2003). Estimating community stability and ecological interactions from time-series data. Ecol. Monogr., 73(2), 301-330.

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

Non-zero $b_{ij} \Leftrightarrow$ **Conditional Granger Causality**: $G_{j \rightarrow i | \text{other variables}}$.

Why log? Power-law models with lognormal noise $N_{t+1,1} = e_1^a N_{t,1}^{b_{11}} N_{t,2}^{b_{12}} N_{t,3}^{b_{13}} e^{\epsilon_{1t}}$

Reasons why linear Granger causality might in fact apply to highly nonlinear ecological systems

- Even the simplest MAR/VAR(1) on the log-linear scale are quite robust to nonlinearities (Certain et al. 2018 *Methods in Ecology and Evolution*)
≈ 80 – 95% of signs recovered for 7 two-species models (T=100) + one food web (models with mostly fixed point equilibria).
- In general we use more time lags. MAR(p) models.

$$\mathbf{x}_t = \sum_{q=1}^p \mathbf{B}_q \mathbf{x}_{t-q} + \mathbf{e}_t \quad (4)$$

Nonlinearity can be expressed as higher-order lags. A well-known example with a single time series

- ▶ AR(1): no cycles (noncyclic ACF)
- ▶ AR(2): cycles of any periodicity possible

Reasons why linear Granger causality might in fact apply to highly nonlinear ecological systems

- Even the simplest MAR/VAR(1) on the log-linear scale are quite robust to nonlinearities (Certain et al. 2018 *Methods in Ecology and Evolution*)
≈ 80 – 95% of signs recovered for 7 two-species models (T=100) + one food web (models with mostly fixed point equilibria).
- In general we use more time lags. MAR(p) models.

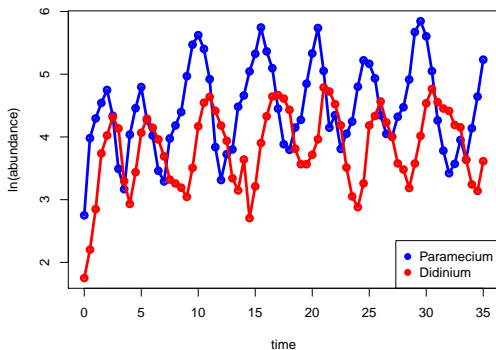
$$\mathbf{x}_t = \sum_{q=1}^p \mathbf{B}_q \mathbf{x}_{t-q} + \mathbf{e}_t \quad (4)$$

Nonlinearity can be expressed as higher-order lags. A well-known example with a single time series

- ▶ AR(1): no cycles (noncyclic ACF)
- ▶ AR(2): cycles of any periodicity possible

We (re)evaluate the relative performances of Granger causality and convergent cross-mapping using a number of time series going from highly nonlinear — sensu *deterministic chaos* — to linear — sensu MAR(1).

Sugihara et al. (2012)'s data and models

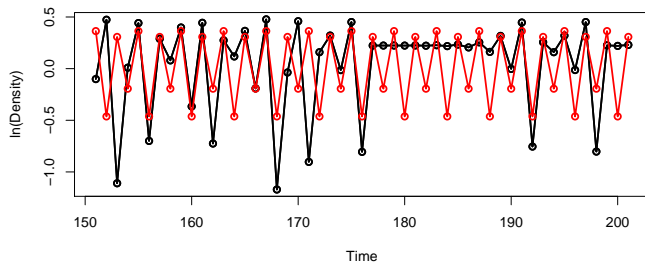


- Veilleux predator-prey data

- Chaotic and deterministic 2-species logistic competition model

$$N_{1,t+1} = N_{1,t}(3.8 - 3.8N_{t,1} - 0.02N_{t,2}) \quad (5)$$

$$N_{2,t+1} = N_{2,t}(3.5 - 3.5N_{t,2} - 0.1N_{t,1}) \quad (6)$$



Additional two-species models - competition with stochasticity

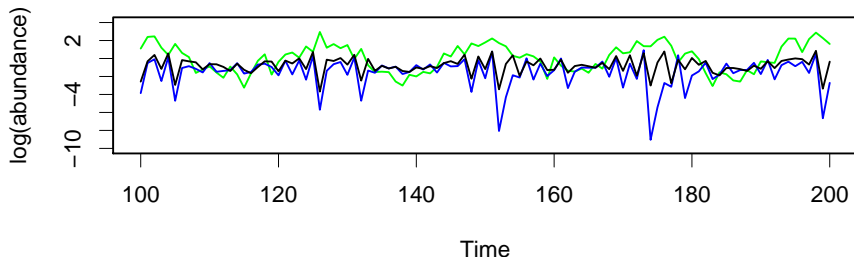
Ricker-based competition models

$$N_{1,t+1} = N_{1,t} \exp(3 - 4N_{1,t} - 2N_{2,t} + \epsilon_{1,t}) \quad (7)$$

$$N_{2,t+1} = N_{2,t} \exp(2.1 - 0.31N_{1,t} - 3.1N_{2,t} + \epsilon_{2,t}). \quad (8)$$

$$N_{1,t+1} = N_{1,t} \exp(3 + 0.5u_t - 4N_{1,t} - 2N_{2,t} + \epsilon_{1,t}) \quad (9)$$

$$N_{2,t+1} = N_{2,t} \exp(2.1 + 0.5u_t - 0.31N_{1,t} - 3.1N_{2,t} + \epsilon_{2,t}). \quad (10)$$



Many-species models

Ricker-based Lotka-Volterra

$$\mathbf{n}_{t+1} = \mathbf{n}_t \circ \exp(\mathbf{r} + \mathbf{A}\mathbf{n}_t + \mathbf{e}_t), \mathbf{e}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \quad (11)$$

$$\mathbf{A} = \begin{pmatrix} -4 & -2 & -0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.31 & -3.1 & -0.93 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.636 & 0.636 & -2.12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.111 & -0.111 & 0.131 & -3.8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & -2 & -2 & -0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.31 & -3.1 & -0.93 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.636 & -0.636 & -2.12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & -2 & -0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.31 & -3.1 & -0.93 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.636 & 0.636 & -2.12 \end{pmatrix} \quad (12)$$

Many-species models

Linear MAR(1) approximation of this Ricker-based Lotka-Volterra model

$$\mathbf{x}_t = \ln(\mathbf{n}_t) \quad (13)$$

$$\mathbf{x}_{t+1} = \mathbf{J}\mathbf{x}_t \quad (14)$$

$$\text{with } J_{ij} = \delta_{ij} + a_{ij}N_j^* \quad (15)$$

Many-species models

Nonlinear (Lotka-Volterra) model with interaction matrix $\mathbf{A} = (a_{ij})$ with 20 species + associated MAR(1) model linearized on the log-density scale

$$a_{ij} = \chi_{ij} [a_{min} + (a_{max} - a_{min})\text{Beta}(2, 2)] \quad (16)$$

$$\chi = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \quad (17)$$

Causality evaluation

2 species data and models (300 timesteps)

- **Pairwise Granger causality tests** with $\alpha = 0.1$ & effect size threshold. Number of time lags p selected by BIC. Wald tests with `grangertest` in the R package `lmtest` (Zeileis and Hothorn, 2002, v0.9-36). For conditional Granger causality testing: `causality` in package `vars` (Pfaff, 2008, v1.5-3), F-tests.
- **Convergent cross-mapping** with $\rho > 0.1$ and $p\text{-value} < \alpha$. P-values are surrogate-based (i.e., permutation-based). `rEDM` (Ye et al., 2018, v0.7.1). Comparison to p-values of Cobey and Baskerville (2016).

For many species models (700 timesteps for 20 species), we did the above pairwise causality analyses with a Benjamini and Hochberg (1995) correction to control the false discovery rate. We also tried **conditional Granger causality**:

- LASSO-penalized MAR(1) models with structured penalties, using the R package `SIMoNe` (Chiquet et al., 2008; Charbonnier et al., 2010, v1.0-3). `SIMoNe` = Statistical Inference for MODular NETworks. *Weighted LASSO approach, combining elements of group and adaptive LASSO.*

Weighted (structured) LASSO

For details see Charbonnier et al. (2010). Latent structure of network \mathcal{P} is structured into \mathcal{Q} classes. We note Z_{iq} the indicator function (a random variable) whose value is 1 if species i belongs in class q (e.g. network module). Follows Ambroise et al. (2009), who use the mixture framework of Daudin et al. (2008).

$$\hat{\mathbf{A}} = \operatorname{argmax} \log \mathcal{L}(\mathbf{Y}, \mathbf{A}; \mathbf{Z}) = \operatorname{argmax} \left\{ \operatorname{Tr}(\mathbf{V}'\mathbf{A}) - \frac{1}{2} \operatorname{Tr}(\mathbf{A}'\mathbf{S}\mathbf{A}) - \|\mathbf{P}^{\mathbf{Z}} \star \mathbf{A}\|_1 \right\} \quad (18)$$

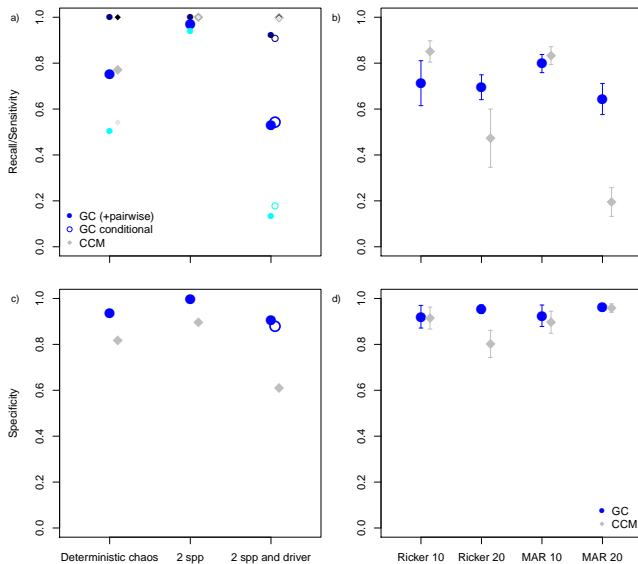
where $\mathbf{P}^{\mathbf{Z}} = (P_{ij}^{\mathbf{Z}}) = \sum_{q,l \in \mathcal{Q}} \frac{Z_{iq}Z_{jl}}{\lambda_{ql}}$ are the penalties encapsulating the network structure. **There is a latent network model which says if two species are connected.**

The particular structure of the model allows to reduce this global LASSO optimization to d LASSO-style problems, which makes it much faster. Tuning of the penalty parameter is then done using BIC (Charbonnier et al., 2010).

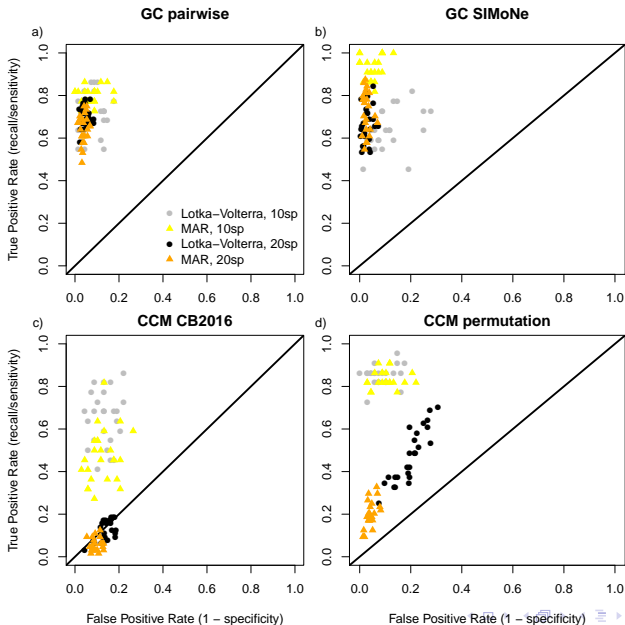
Notational details: $\mathbf{A} = \mathbf{B}'$ with \mathbf{B} from Ives et al. (2003), $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_1, \dots, \mathbf{x}_T]'$ the matrix of observed (log)densities and \mathbf{Y} the one a time step after $\mathbf{Y} = [\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{T+1}]'$.

- $\mathbf{S} = \frac{1}{T} \mathbf{X}'\mathbf{X}$ is the empirical variance-covariance matrix
- $\mathbf{V} = \frac{1}{T} \mathbf{X}'\mathbf{Y}$ is the temporal autocovariance matrix

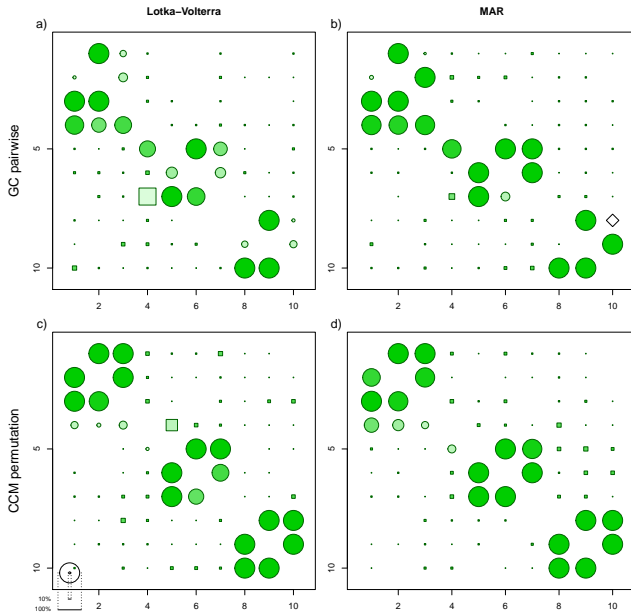
Results: Sensitivity and specificity (two- and many-species models)



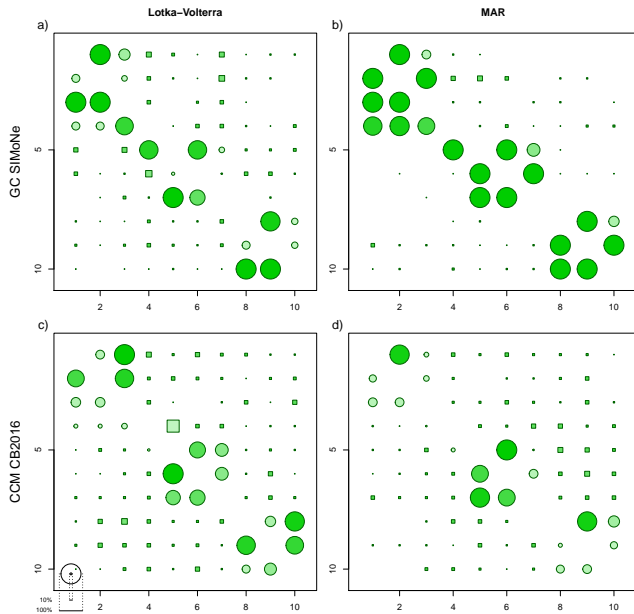
Results: Sensitivity and specificity (two- and many-species models)



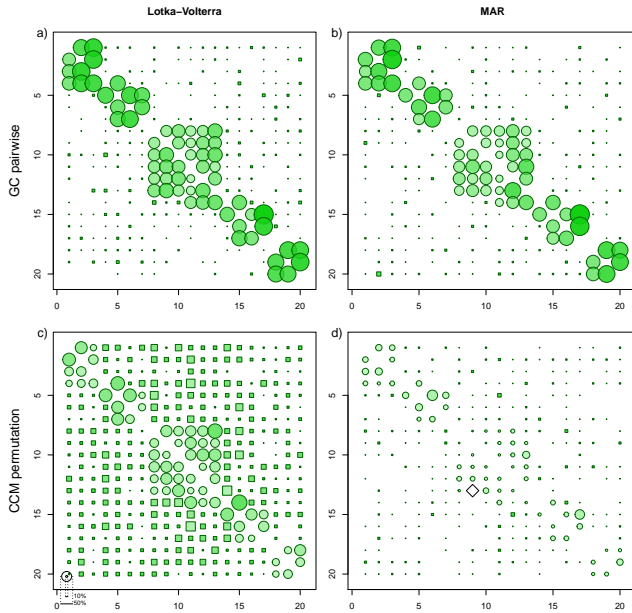
Results: Estimated interaction networks



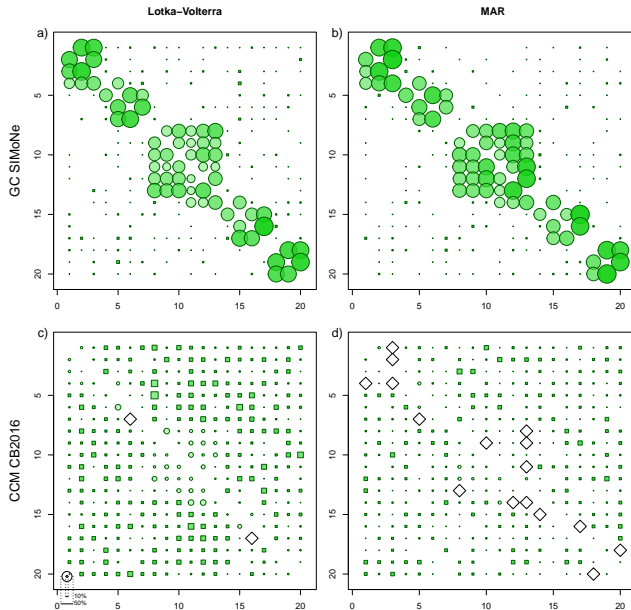
Results: Estimated interaction networks



Results: Estimated interaction networks



Results: Estimated interaction networks



Conclusion: 2-species systems

- **In most 2-species cases of fluctuating community dynamics that we studied both linear Granger causality and convergent cross-mapping work fine.**
- Overlapping data and models with Sugihara et al. (2012): differences likely due to model selection (we used BIC, they used a sequence of Likelihood Ratio Tests).

Conclusion: 2-species systems

- In most 2-species cases of fluctuating community dynamics that we studied both **linear Granger causality** and **convergent cross-mapping** work fine.
- Overlapping data and models with Sugihara et al. (2012): differences likely due to model selection (we used BIC, they used a sequence of Likelihood Ratio Tests).
- Exception: **nonlinear competition with a confounding abiotic driver** (temperature) affecting both species similarly. GC produces false negatives (for the weak interaction) and CCM false positives.

Conclusion: 2-species systems

- In most 2-species cases of fluctuating community dynamics that we studied both **linear Granger causality** and **convergent cross-mapping** work fine.
- Overlapping data and models with Sugihara et al. (2012): differences likely due to model selection (we used BIC, they used a sequence of Likelihood Ratio Tests).
- Exception: **nonlinear competition with a confounding abiotic driver** (temperature) affecting both species similarly. GC produces false negatives (for the weak interaction) and CCM false positives.
- Due to **strong nonlinearity** (high LE, high max growth rate) combined with **an autocorrelated environment**. *Known challenge* even for single-species population dynamics (Lindén et al., 2013).

Conclusion: 2-species systems

- **In most 2-species cases of fluctuating community dynamics that we studied both linear Granger causality and convergent cross-mapping work fine.**
- Overlapping data and models with Sugihara et al. (2012): differences likely due to model selection (we used BIC, they used a sequence of Likelihood Ratio Tests).
- Exception: **nonlinear competition with a confounding abiotic driver** (temperature) affecting both species similarly. GC produces false negatives (for the weak interaction) and CCM false positives.
- Due to **strong nonlinearity** (high LE, high max growth rate) combined with **an autocorrelated environment**. *Known challenge* even for single-species population dynamics (Lindén et al., 2013).
- Bottomline: nothing wrong with using Granger causality for limit cycles and chaotic attractors. Nothing wrong with convergent cross-mapping for stochastic systems. **Both methods define (causal) interactions in the same way.**

Conclusion: Many-species systems (10×10 , 20×20)

- Pairwise GC with Benjamini-Hochberg correction surprisingly efficient. Convergent cross-mapping: false positives & negatives but surrogates tuning might improve performance.

Conclusion: Many-species systems (10×10 , 20×20)

- Pairwise GC with Benjamini-Hochberg correction surprisingly efficient. Convergent cross-mapping: false positives & negatives but surrogates tuning might improve performance.
- Weighted LASSO (modular version of the LASSO) for high-dimensional $MAR(p)$ modelling has comparable performance to pairwise Granger causality tests: why?

Conclusion: Many-species systems (10×10 , 20×20)

- Pairwise GC with Benjamini-Hochberg correction surprisingly efficient. Convergent cross-mapping: false positives & negatives but surrogates tuning might improve performance.
- Weighted LASSO (modular version of the LASSO) for high-dimensional $MAR(p)$ modelling has comparable performance to pairwise Granger causality tests: why?
- We have used a max time lag $p = 1$ with the weighted LASSO because (1) Charbonnier et al. (2010); Chiquet et al. (2008)'s very efficient R package was restricted to $MAR(1)$ modelling and (2) BIC suggested $p = 1$ for 20 species.

Conclusion: Many-species systems (10×10 , 20×20)

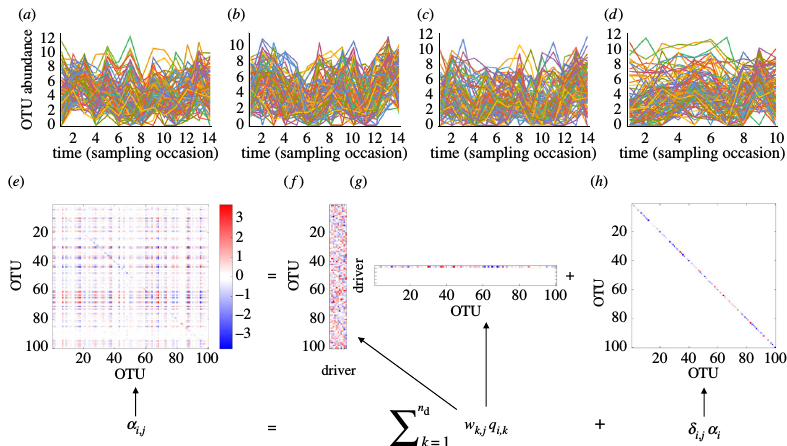
- Pairwise GC with Benjamini-Hochberg correction surprisingly efficient. Convergent cross-mapping: false positives & negatives but surrogates tuning might improve performance.
- Weighted LASSO (modular version of the LASSO) for high-dimensional $MAR(p)$ modelling has comparable performance to pairwise Granger causality tests: why?
- We have used a max time lag $p = 1$ with the weighted LASSO because (1) Charbonnier et al. (2010); Chiquet et al. (2008)'s very efficient R package was restricted to $MAR(1)$ modelling and (2) BIC suggested $p = 1$ for 20 species.
- One could select both p and LASSO penalties simultaneously. Another approach: select both p and matrices using group LASSO (Nicholson et al., 2017a,b; Mainali et al., 2019). R package BigVAR. Recent reviews of high-dimensional VAR/Granger modelling in Basu and Matteson (2021); Shojaie and Fox (2021).

Conclusion: Many-species systems (10×10 , 20×20)

- Pairwise GC with Benjamini-Hochberg correction surprisingly efficient. Convergent cross-mapping: false positives & negatives but surrogates tuning might improve performance.
- Weighted LASSO (modular version of the LASSO) for high-dimensional $MAR(p)$ modelling has comparable performance to pairwise Granger causality tests: why?
- We have used a max time lag $p = 1$ with the weighted LASSO because (1) Charbonnier et al. (2010); Chiquet et al. (2008)'s very efficient R package was restricted to $MAR(1)$ modelling and (2) BIC suggested $p = 1$ for 20 species.
- One could select both p and LASSO penalties simultaneously. Another approach: select both p and matrices using group LASSO (Nicholson et al., 2017a,b; Mainali et al., 2019). R package BigVAR. Recent reviews of high-dimensional VAR/Granger modelling in Basu and Matteson (2021); Shojaie and Fox (2021).
- Note that we have so far avoided modelling strongly autocorrelated environmental forcing + many-species dynamics but this is a common situation!
- Another idea: abandon the sparsity assumption altogether.

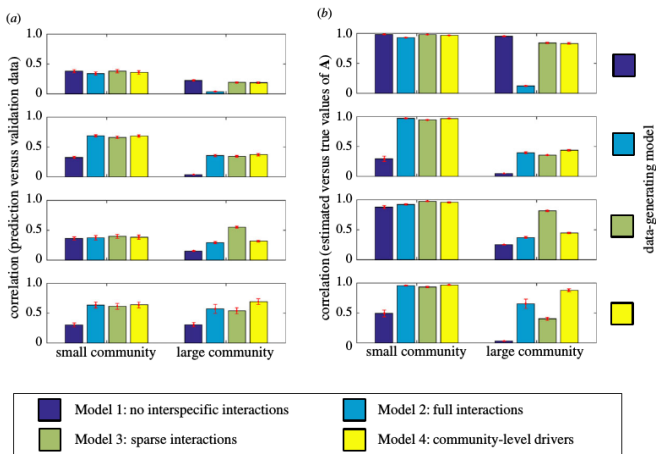
The community-level drivers method – dimension reduction

Ovaskainen et al. (2017) How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings B*, 284: 20170768.



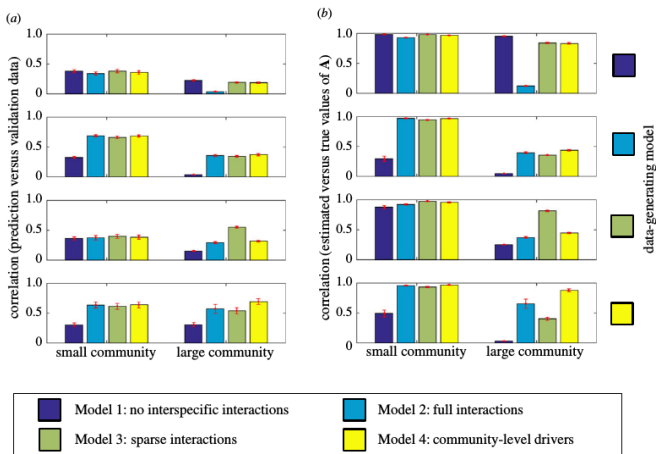
The community-level drivers method – dimension reduction

Ovaskainen et al. (2017) How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings B*, 284: 20170768.



The community-level drivers method – dimension reduction

Ovaskainen et al. (2017) How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings B*, 284: 20170768.



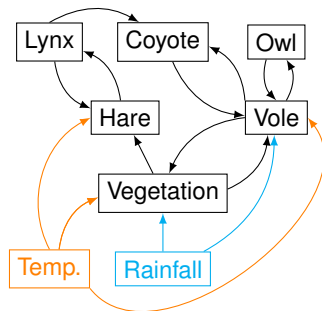
Multiple other low-rank estimation methods in statistics (e.g. Alquier et al., 2020)

Want to know more? Questions?

Now published in *Theoretical Ecology*. Reprint at <https://arxiv.org/abs/1909.00731>

Code at <https://github.com/fbarraquand/GCausality>

Contact frederic.barraquand@u-bordeaux.fr



Follow-up questions:

- how to better interface sparse or low-rank representation (or both, Basu et al. 2019) with modularity and other ecological properties?
- how do links that we infer depend on links that we assume?

References I

- Alquier, P., K. Bertin, P. Doukhan, and R. Garnier, 2020. High-dimensional var with low-rank transition. Statistics and Computing **30**:1139–1153.
- Ambroise, C., J. Chiquet, C. Matias, et al., 2009. Inferring sparse Gaussian graphical models with latent structure. Electronic Journal of Statistics **3**:205–238.
- Barnett, L. and A. K. Seth, 2014. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. Journal of Neuroscience Methods **223**:50–68.
- Basu, S., X. Li, and G. Michailidis, 2019. Low rank and structured modeling of high-dimensional vector autoregressions. IEEE Transactions on Signal Processing **67**:1207–1222.
- Basu, S. and D. S. Matteson, 2021. A survey of estimation methods for sparse high-dimensional time series models. arXiv preprint arXiv:2107.14754 .
- Benjamini, Y. and Y. Hochberg, 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) **57**:289–300.
- Certain, G., F. Barraquand, and A. Gårdmark, 2018. How do MAR(1) models cope with hidden nonlinearities in ecological dynamics? Methods in Ecology and Evolution **9**:1975–1995.
- Charbonnier, C., J. Chiquet, and C. Ambroise, 2010. Weighted-lasso for structured network inference from time course data. Statistical Applications in Genetics and Molecular Biology **9**.
- Chiquet, J., A. Smith, G. Grasseau, C. Matias, and C. Ambroise, 2008. Simone: Statistical inference for modular networks. Bioinformatics **25**:417–418.
- Cobey, S. and E. B. Baskerville, 2016. Limits to causal inference with state-space reconstruction for infectious disease. PloS one **11**:e0169050.

References II

- Daudin, J.-J., F. Picard, and S. Robin, 2008. A mixture model for random graphs. Statistics and computing **18**:173–183.
- Detto, M., A. Molini, G. Katul, P. Stoy, S. Palmroth, and D. Baldocchi, 2012. Causality and persistence in ecological systems: a nonparametric spectral Granger causality approach. The American Naturalist **179**:524–535.
- Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association **77**:304–313.
- Geweke, J. F., 1984. Measures of conditional linear dependence and feedback between time series. Journal of the American Statistical Association **79**:907–915.
- Granger, C., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica **37**:424–438.
- Ives, A., B. Dennis, K. Cottingham, and S. Carpenter, 2003. Estimating community stability and ecological interactions from time-series data. Ecological Monographs **73**:301–330.
- Lindén, A., M. S. Fowler, and N. Jonzén, 2013. Mischaracterising density dependence biases estimated effects of coloured covariates on population dynamics. Population Ecology **55**:183–192.
- Mainali, K., S. Bewick, B. Vecchio-Pagan, D. Karig, and W. F. Fagan, 2019. Detecting interaction networks in the human microbiome with conditional Granger causality. PLoS Computational Biology **15**:e1007037.
- Michailidis, G. and F. d'Alché Buc, 2013. Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues. Mathematical Biosciences **246**:326–334.

References III

- Nicholson, W., D. Matteson, and J. Bien, 2017a. BigVar: Tools for modeling sparse high-dimensional multivariate time series. arXiv preprint arXiv:1702.07094 .
- Nicholson, W. B., D. S. Matteson, and J. Bien, 2017b. Varx-l: Structured regularization for large vector autoregressions with exogenous variables. International Journal of Forecasting **33**:627–651.
- Ovaskainen, O., G. Tikhonov, D. Dunson, V. Grøtan, S. Engen, B.-E. Sæther, and N. Abrego, 2017. How are species interactions structured in species-rich communities? a new method for analysing time-series data. Proceedings of the Royal Society B: Biological Sciences **284**:20170768.
- Pfaff, B., 2008. VAR, SVAR and SVEC Models: implementation within R package vars. Journal of Statistical Software **27**.
- Shojaie, A. and E. B. Fox, 2021. Granger causality: A review and recent advances. arXiv preprint arXiv:2105.02675 .
- Sims, C., 1980. Macroeconomics and reality. Econometrica **48**:1–48.
- Sugihara, G., R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, 2012. Detecting causality in complex ecosystems. Science **338**:496–500.
- Veilleux, B. G., 1979. An analysis of the predatory interaction between *Paramecium* and *Didinium*. Journal of Animal Ecology **48**:787–803.
- Ye, H., A. Clark, E. Deyle, S. Munch, J. Cai, J. Cowles, Y. Daon, A. Edwards, O. Keyes, J. Stagge, M. Ushio, E. White, and G. Sugihara, 2018. rEDM: Applications of Empirical Dynamic Modeling from Time Series.
- Zeileis, A. and T. Hothorn, 2002. Diagnostic checking in regression relationships. R News **2**:7–10.