

## Supporting Information S3 - 2-species simulation experiment

Supporting Information for *How do MAR(1) models cope with hidden nonlinearities in ecological dynamics?*

Grégoire Certain, Frédéric Barraquand and Anna Gårdmark, *Methods in Ecology and Evolution*.

### GENERAL INFORMATION

The whole set of file required to run our simulation experiment is available at:

<https://doi.org/10.6084/m9.figshare.6127877.v1>

This set of files contains the codes for launching our simulation experiment. They will (i) generate data from theoretical mechanistic models, (ii) fit a MAR(1) model to these datasets and (iii) investigate the accuracy of the fit and of the predictions from the fitted models. The list and description of files is provided below:

- The whole simulation experiment can be run from "1\_MarSimExp\_CommandFile\_DoItAll.R", which will then call all other R files. Amongst other files,
- 10 files called "Modele\_[modelname].R" are dedicated to simulating data from a given underlying community dynamics (that of [modelname]) and fitting MAR(1) models to such dynamics. Among these, 3 files with ending "\_ObsErr" include observation errors.
- 10 files called "MakeGraph\_[modelname].R" are dedicated to extracting the results and plotting the graphs of the corresponding simulation experiment.
- "MakeGraph\_Paper.R" can be run after all the experiments have been done, to prepare the same figures as the one presented in the main document and Supporting Informations S3 and S4.
- "R\_function\_MARanalysis.R" contains utility and plotting functions.

If you want to perform the whole experiment with 1000 repetitions, you should be aware that it takes a couple of days on a 3.5GHz computer to run entirely – the code is *\*not\** optimized for fast run. It will automatically generate a number of folders and subfolders whose architecture is explained below, for a total size of  $\sim 2$  Gb.

We tested the code on several Windows and Linux machines, with R 3.4.3 and some versions below. If you have questions regarding the code, please address them to [Gregoire.Certain@ifremer.fr](mailto:Gregoire.Certain@ifremer.fr) (we do no guarantee assistance, but will do our best). Most of the code only requires the package MARSS, but some of our functions are built upon packages gamlss, gamlss.dist, ade4, mvtnorm, and MASS.

## ARCHITECTURE OF THE SIMULATION EXPERIMENT

When unzipped, all the code files are found in the same folder

`~/MEECertain2017_SI2_SimExp`

While running, the simulation experiment will create other folders. They will be named after each of the 7 data-generating models, for example, results regarding the simulation experiment carried out with the Gompertz data-generating model will be available in:

`~/MEECertain2017_SI2_SimExp/Gompertz_comp`

This folder will contain 3 subfolders, one for each level of process error intensity. Hence, results regarding the Gompertz experiment with process error intensity of 0.1 will be available in:

`~/MEECertain2017_SI2_SimExp/Gompertz_comp/Noise=0.1`

Some simulation experiments (Gompertz, Ricker, and Beverton-Holt competition) were carried without and with observation errors. The results obtained with observation error are to be found one level below, in the folder corresponding to process error intensity of 0.1, as simulation experiments including observation error had the process error fixed at 0.1. Hence, results regarding the Gompertz simulation experiment of process error 0.1 and observation error 0.5 are to be found in

`~/MEECertain2017_SI2_SimExp/Gompertz_comp/Noise=0.1/ObsErr=0.5`

The results of the simulation experiments consist in a list of data storage files (with extension .R) and .pdf figures. A detailed account of the figures is listed below. The data storage files are numerous, so providing a detailed account for them would be too tedious in this document. However, they are all referenced in due place in the code. One result summary table, "Summary\_Stat.txt" will be created in the main directory `~/MEECertain2017_SI2_SimExp` at the end of the experiment. It contains all the summary statistics regarding the evaluation of MAR(1) performances, some of which being displayed in tables S3.2 and S3.3. "Summary\_Stat.txt" is bigger than these tables as it includes more statistics, but it has the same structure.

### NAME AND DESCRIPTION OF FIGURES GENERATED BY OUR EXPERIMENT:

- **Abs(max\_eigenvalue) True Vs Fit:** Plot comparing the maximum eigenvalues of the  $\mathbf{B}$  matrix of the fitted MAR models (x-axis) to the maximum eigenvalues of the Jacobian matrix of the corresponding theoretical model. Red dotted lines corresponds to lines  $x = 0$  and  $y = 0$ , blue lines shows  $y = x$ , and black dotted line shows the linear regression of  $y \sim x$ .
- **Bfit Errors Vs Jtrue:**  $2 \times 2$  plot panel organized as the fitted  $\mathbf{B}$  matrix, that is, the upper-left plot corresponds to  $b_{11}$ . Each dot correspond to a fitted model. The true values of the Jacobian elements are reported on the x-axis, while the error between the value of the fitted  $\mathbf{B}$  matrix and the value of the Jacobian ( $b_{ij} - j_{ij}$ ) is reported on the y-axis. Blue dots correspond to models in which the element of  $\mathbf{J}$  lies within the 95% confidence intervals provided for  $\mathbf{B}$ . Red dots are outside. The red line shows  $y = 0$ , i.e., perfect estimate.

- **Bfit Vs Jtrue\_modelname:**  $2 \times 2$  plot panel organized as the fitted  $\mathbf{B}$  matrix, that is, the upper-left plot corresponds to  $b_{11}$ . Each dot correspond to a fitted model. The true values of the Jacobian elements are reported on the x-axis, while the fitted values of the  $\mathbf{B}$  matrix are reported on the y-axis. Blue dots correspond to models in which the element of  $\mathbf{J}$  lies within the 95% confidence intervals provided for  $\mathbf{B}$ . Red dots are outside CIs. The red dotted lines show  $y = 0$  and  $x = 0$ , the blue solid line shows  $y = x$  (i.e., perfect estimate), and the black dotted line corresponds to the regression  $y \sim x$ .
- **Bfit Vs Jtrue\_sdcol:** Same graph than above, but the color-scale of the dots represents the standard deviation of the time series used to fit the model, i.e. dots near yellow colour represent the most variable time series.
- **Coinertia analysis on errors:** Graphical output of a co-inertia analysis carried out between the table of parameter values drawn for the simulated dynamics and the corresponding errors on the fitted  $\mathbf{B}$  matrices.
- **Coinertia analysis on inference params:** Graphical output of a co-inertia analysis carried out between the table of parameter values drawn for the simulated dynamics and the corresponding estimated values of the fitted  $\mathbf{B}$  matrices.
- **Compare.TrueWithMAR:** These plots show (for the first 100 simulated dynamics, one page per simulation) the time series on which the MAR model has been fitted (left-hand panels) and the corresponding predicted time series produced by the fitted MAR model (right-hand panels). Upper panels show the time series (black line: species 1; red line: species 2; green line: environment), middle panels display phase plots, lower panels show how the environmental covariate a time  $t$  (x-axis) affects species 1 ln-abundance at time  $t + 1$  (y-axis).
- **Enviro TrueVsFit:** Shows how the environmental effect estimated from the fitted MAR model (y-axis) matches the true environmental effect in the dynamics (x-axis) for species 1 (left panel) and species 2 (right panel, for which the true environmental effect is actually 0, i.e. no effect). Blue dots correspond to models in which the true environmental effect lies within the 95% confidence intervals provided by the MAR model. Red dots are outside CIs. The red dotted lines show  $y = 0$  and  $x = 0$ , the blue solid line shows  $y = x$  (i.e., perfect estimate), and the black dotted line corresponds to the regression  $y \sim x$ .
- **Example\_sims\_model\_modelname:** These plots show (for the first 100 simulated dynamics, one page per simulation) the complete time series of the dynamics, over 2000 time steps: 1000 time steps before the PRESS perturbation, and 1000 time steps after the PRESS perturbation. The black line represents species 1; the red line, species 2; and the green line, the environmental driver.
- **Observation Error & Process Error:** For simulation experiments without process error, a graph showing the value (y-axis) of fitted process error on species 1 ( $\Sigma_{11}$ , x-axis) and species 2 ( $\Sigma_{22}$ , x-axis). The true (simulated) process error value is indicated by a blue #. Red dots correspond to simulations in which the 95% *c.i.* of the estimated value do not contain the true

value, blue dots otherwise. In simulation including observation error, fitted observation error  $\nu^2$  is included on the x-axis.

- **Pairplots on underlying and fitted parameters:** These numerous scatterplots, organised as the upper triangle of a correlation matrix, display relationships between (1) the value of the mechanistic parameters of the underlying dynamics and (2) the value of fitted parameters of the MAR model. For readability, scatterplots are represented in color shades, with intense shades corresponding to areas where points accumulate (each point is a simulated dynamics). Blue shades are used in scatterplots involving 2 parameters of the underlying dynamics, red shades are used in scatterplots involving 2 parameters of the fitted MAR model, and purple shades are used in scatterplots involving 1 parameter of the underlying dynamics and 1 parameter of the fitted MAR model.
- **Predictions TrueVsFit:** Comparison of the median predictions of the species' response to a PRESS perturbation obtained from a fitted MAR model (y-axis) with the actual response of the underlying dynamics (x-axis) for species 1 (left panel) and species 2 (right panel). Responses are expressed in term of change in ln-populations densities. The blue, thick line shows the perfect (e.g.  $y = x$ ) match. The thin, dashed line shows the actual regression  $y \sim x$ . Each dot corresponds to a single simulation. Dots colorscale illustrates the distance between the true change and the median prediction in term of inter-quantile range, that is, a value of 0.5 means that the true change lies beyond the bootstrapped distribution of the predictions.
- **Range of Param Values:** Histograms showing the range and frequencies of parameter values that were drawn randomly to simulate the dynamics. Ideally, these should match uniform distributions, but constraints regarding the properties of the simulated dynamics (the existence of a stable equilibrium, and the avoidance of negative, too low or too high values for species equilibrium densities) explain why some distributions depart from uniformity.
- **Residual Plots:** Set of residual diagnostic plots, for each simulation (1000 pages, each corresponding to a simulation).