Supplementary Figures: Uncertainty quantification of the effects of biotic interactions on community dynamics from nonlinear time series data

Simone Cenci and Serguei Saavedra

Department of Civil and Environmental Engineering, MIT
77 Massachusetts Av., 02139 Cambridge, MA, USA

Model averaging

0.82	0.43	0.59	NA
NA	0.6	0.67	0.06
0.89	NA	0.4	0.33
0.2	0.15	-0.37	0.73

Cross-validation

0.81	0.42	0.56	NA
NA	0.41	0.43	-0.06
0.88	NA	0.41	0.3
0.12	0.04	-0.39	0.7

Figure S1 Inference of Jacobian coefficients and the importance of the model-averaging process. The figure shows the correlation matrix (Pearson's correlation coefficient) between inferred and analytical coefficients of the Jacobian for the chaotic LV model with 4 species discussed in the main text. The left panel shows the correlation matrix between inferred and analytical Jacobians. The inferred Jacobian is constructed with the model-averaging process (ensemble method) proposed in this work. The right panel shows the correlation matrix when the best Jacobian is selected with cross-validation. NA indicates constant coefficients for which the correlation is not defined. The values shown in both panels are average correlations over the 20 random samples of the time series. Importantly, this figure shows that using model averaging, the overall quality of the inferred Jacobian coefficients is higher that the one generated by the best single model.

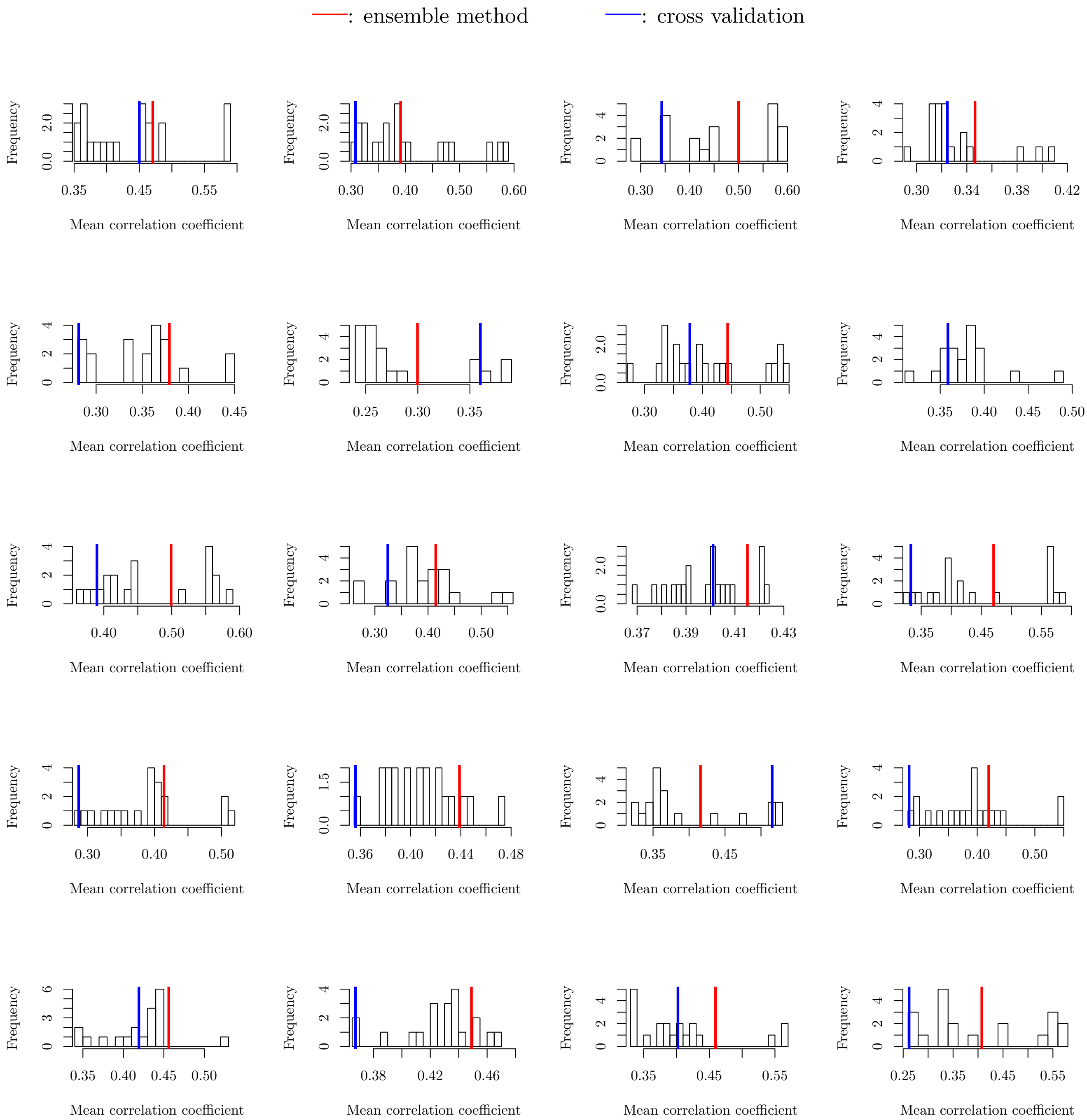

Figure S2 Quality of inference across models The figure shows the distribution of the mean Pearson correlation coefficient between analytical and inferred interactions (elements of the Jacobian matrix). Each panel is one of the 20 samples of the chaotic time series discussed in the main text. The blue and red horizontal lines correspond to the correlations generated by the single best model and the averaged model, respectively. Importantly, this figure shows that the model-averaging procedure tends to move the inference towards models that could not been selected by cross-validation, but exhibit a better correlation with the analytical coefficients.

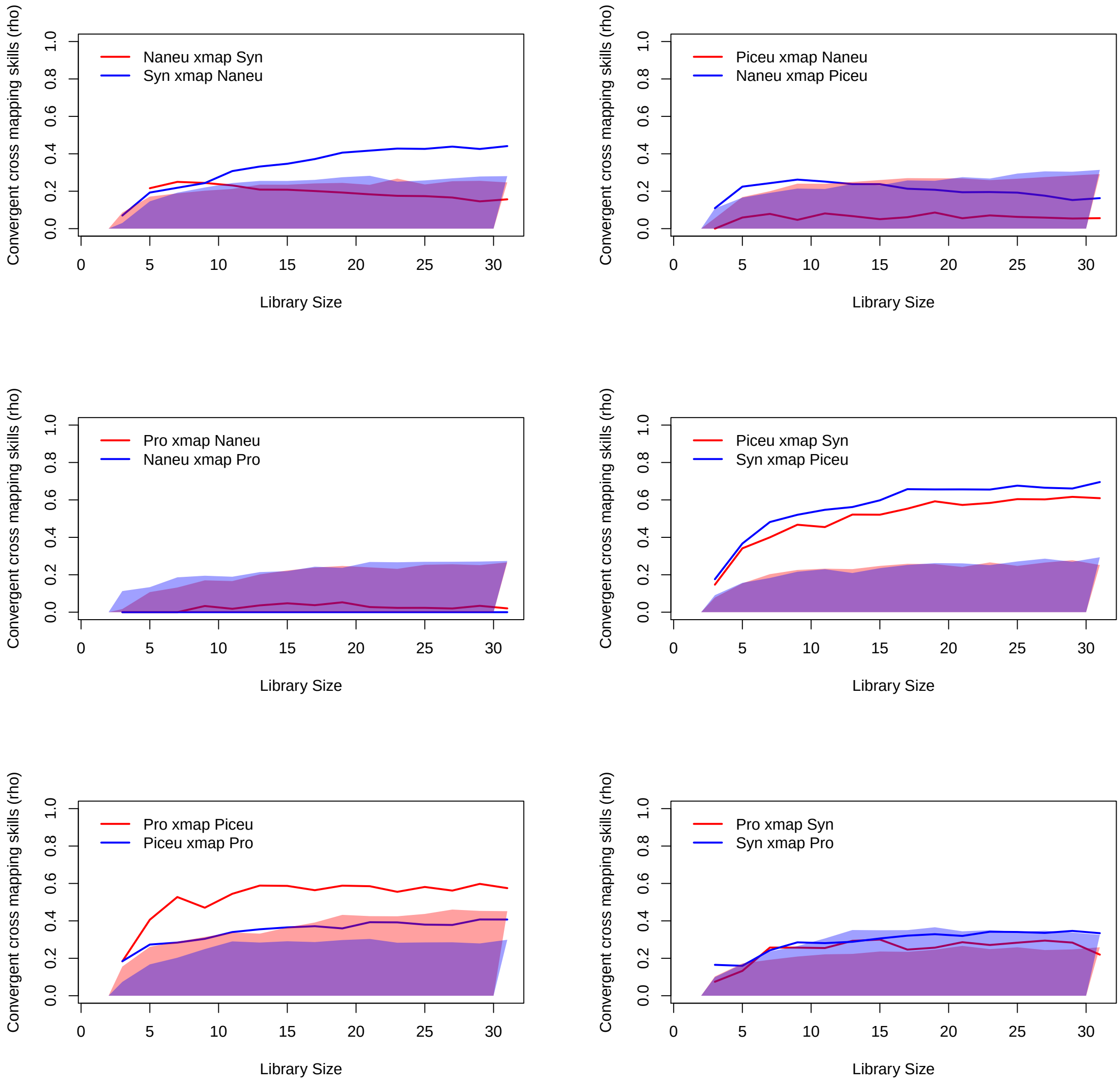

Figure S3 Convergent cross mapping. Results of the convergent cross mapping analysis of the Bermuda Atlantic time series. The shaded area corresponds to the analysis of a surrogate time series. Specifically, to measure the significance of the causal link between variables, we used 50 surrogate time series and the shaded area represents the 5th and 95th percentile of the ensemble. Long-term correlation coefficients larger than the respective shaded area indicate significant causation.

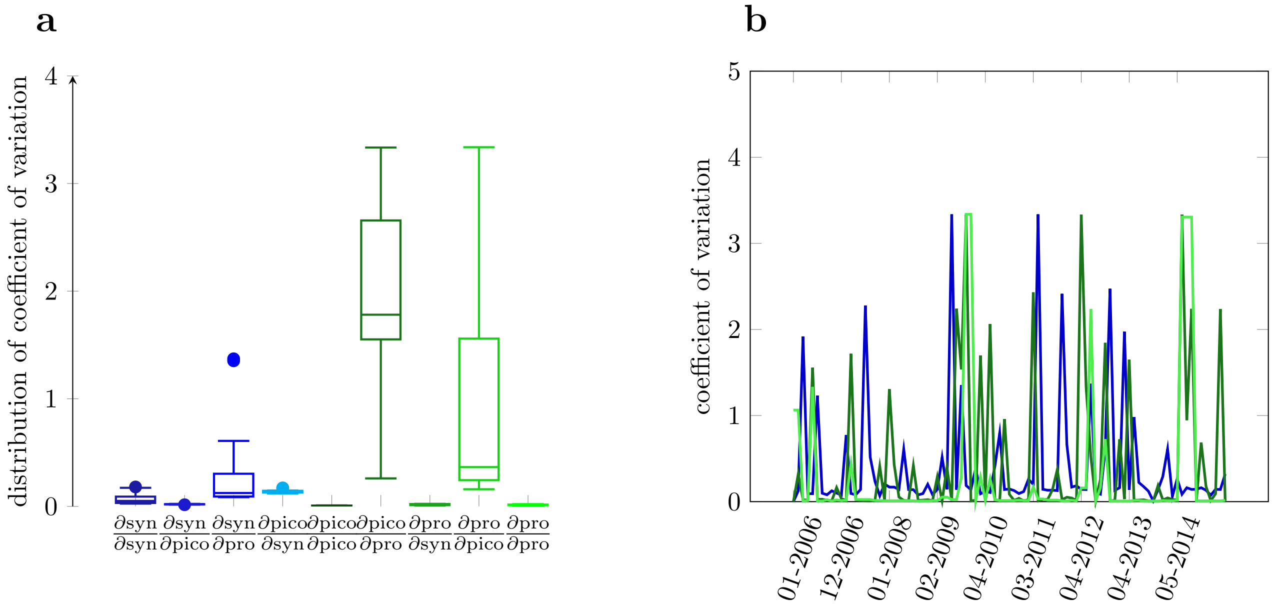

Figure S4 Uncertainty analysis on the Hawaii Ocean time series including the abundance of ***Synenochoccus*****.** The figure shows the pattern of uncertainty on the Jacobian coefficients of the Hawaii Ocean time series when including the abundance of *Synenochoccus* in the regression. Panel **a** shows that the uncertainty on ($\frac{\partial \text{pico}}{\partial \text{pico}}$), ($\frac{\partial \text{pico}}{\partial \text{pro}}$), ($\frac{\partial \text{pro}}{\partial \text{pro}}$) is left qualitatively unchanged while ($\frac{\partial \text{pro}}{\partial \text{pico}}$) shows a different pattern (compare with Figure **??** in the main text). Panel **b** shows the temporal changes of uncertainty on the three least consistently estimated interactions. Note that we did not include the abundance of *Synenochoccus* in the analysis discussed in the main text because the quality of the fit in the training and test sets is significantly reduced. In fact, by including *Synenochoccus* in the analysis, we have $R^2_{\text{training}} = 0.55$ and $\text{RMSE}_{\text{test}} = 0.3$, $\text{RMSE}_{\text{naive}} = 0.65$ (to be compared against the values in the main text).

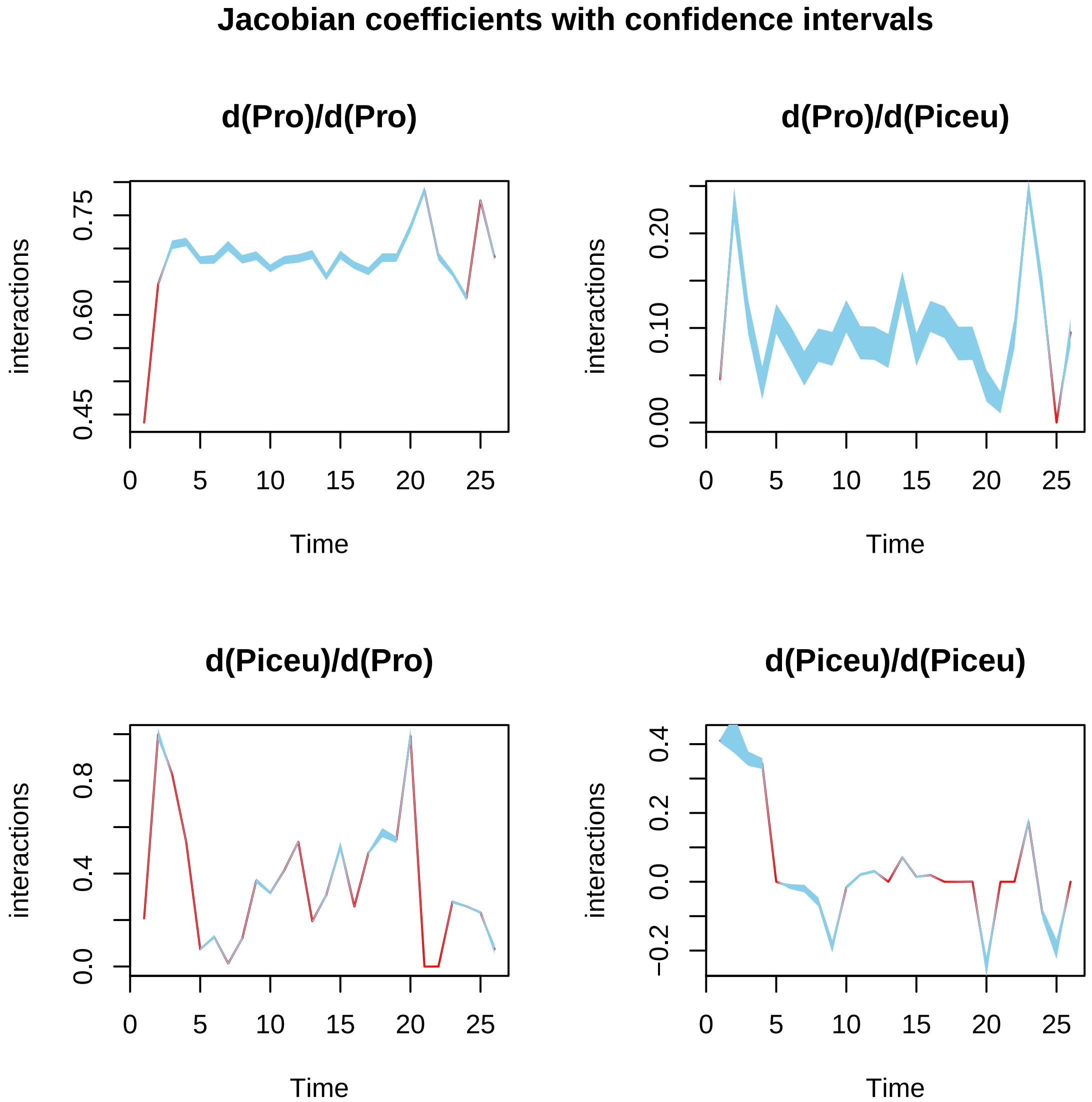

Figure S5 Jacobian coefficients of the Bermuda Atlantic time series with confidence intervals. Shaded area correspond to the 95% confidence intervals computed as discussed in the Method section using the weighted standard error on the coefficient in the optimum ensemble. x-axis are the time points

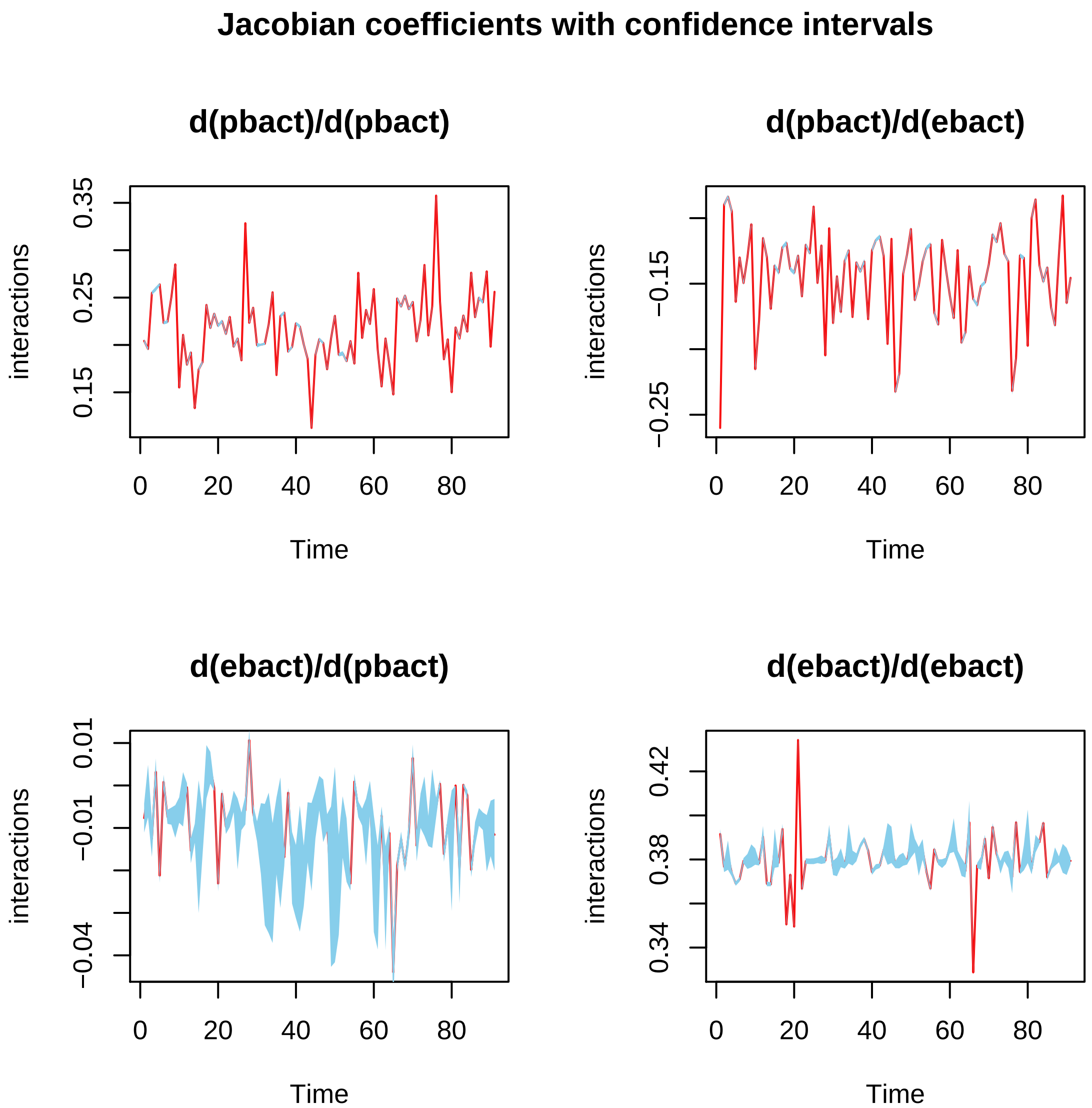

Figure S6 Jacobian coefficients of the Hawaii ocean time series with confidence intervals. Shaded area correspond to the 95% confidence intervals computed as discussed in the Method section using the weighted standard error on the coefficient in the optimum ensemble. x-axis are the time points

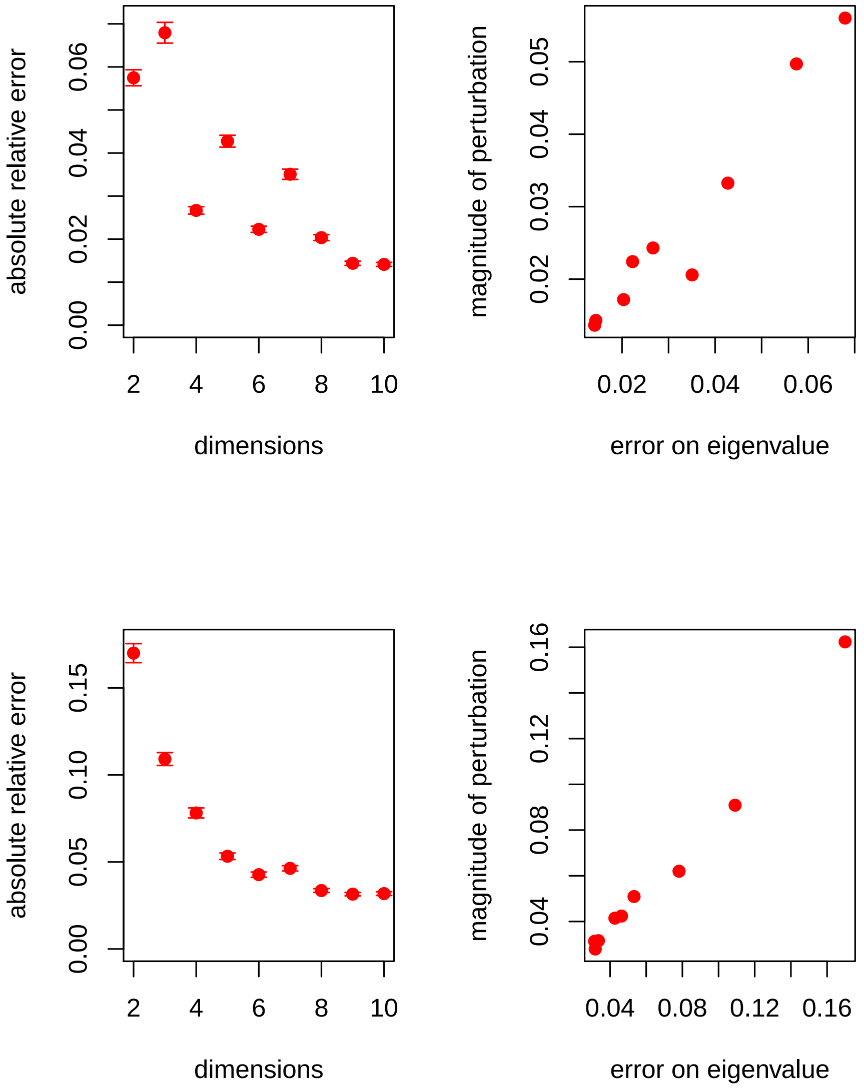

Figure S7 Difference between maximum eigenvalues of approximately similar matrices. The figure illustrates the extent to which small changes in a matrix specification affect the maximum eigenvalue. To produce the figure, we have generated a random matrix with $\mathcal{D}_{ij} \sim \mathcal{N}(0.3, 0.2)$. Then, we perturbed this matrix with a random perturbation generated as $\mathcal{P}_{ij} \sim \mathcal{N}(0, 0.05)$ (top panel) and $\mathcal{P}_{ij} \sim \mathcal{N}(0, 0.1)$ (bottom panel). The absolute relative error was measured as $\epsilon = |\frac{\lambda^D_{\text{max}} - \lambda^P_{\text{max}}}{\lambda^D_{\text{max}}}|$. The magnitude of the perturbation was measured as $\nu = |\frac{\langle\mathcal{D}\rangle - \langle\mathcal{P}\rangle}{\langle\mathcal{D}\rangle}|$. Then, we repeated the experiment for matrices of dimension $d \in [2, 10]$. The statistics are taken over an ensemble of 500 perturbations for each dimension.

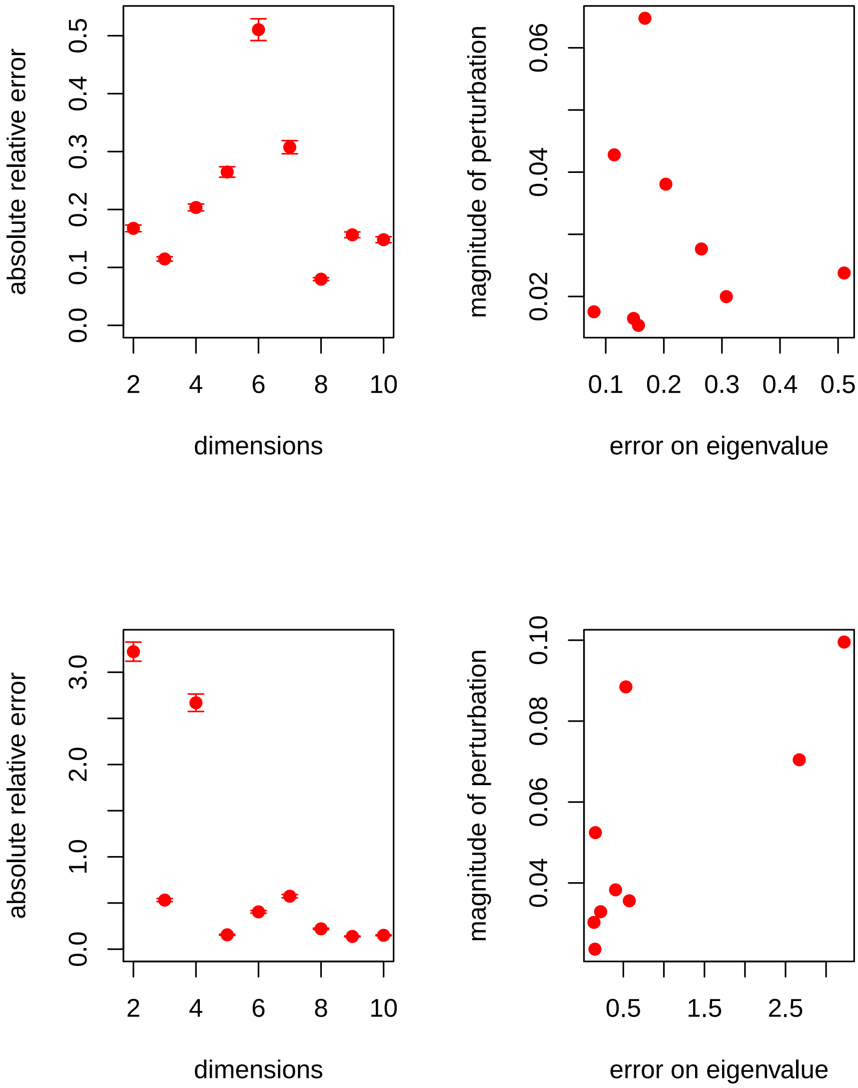

Figure S8 Difference between minimum eigenvalues of approximately similar matrices. Same as Figure **??** but this shows the minimum eigenvalue. The figure clearly illustrates that a small misspecification of a matrix can have large effects on its spectrum.

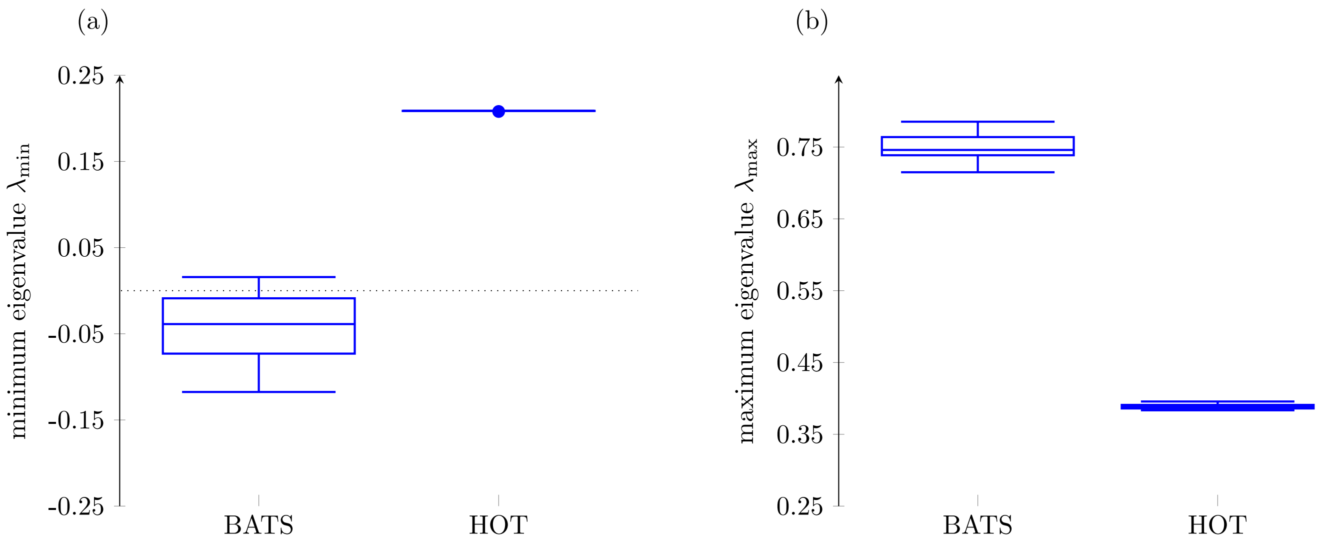

Figure S9 Uncertainty associated with the eigenvalues of the empirical Jacobian coefficients The Figure illustrate the distribution of the minimum (panel (a)) and maximum (panel (b)) eigenvalues of the empirical Jacobian matrices within the ensemble used to constructed the average Jacobian coefficients. As we were expecting based on the result of Figure **??** a large number of eigenvalues (even with different sign) are compatible with the data of the Bermuda Atlantic time series.