

Synthetic Likelihood

(Summary of “Statistical inference for noisy nonlinear ecological dynamic systems”
by Simon N. Wood 2010)

Presented by:

Behrouz Ehsani M. Melanie Stacey

Part I

Characteristics of ecological (biological) dynamic systems

- Chaos:
 - Endogenous dynamic processes
 - Demographic and environmental process noise
- Time sensitivity is because of the joint probability density of the observable data and the process noise.
- Inability to make well-founded statistical inferences.

Problem:

“Inability to make well-founded statistical inferences about biological dynamic models in the chaotic and near-chaotic regimes, ...leaves dynamic theory without the methods of quantitative validation that are essential tools in the rest of biological sciences!” (Wood, 2010)

Solution:

- Simulate the observed data on a system from the dynamic model;
- Summarize the raw data series to phase-insensitive summary statistics;
- Obtain the mean and the covariance matrix of the statistics;
- Create a ‘**synthetic likelihood**’ (/s) that assesses model fit.

This likelihood can be explored using a **Markov chain Monte Carlo** sampler, but one further post-processing step returns pure likelihood-based inference.

Methods

Choosing statistics (S)

- The statistics (s) in the synthetic likelihood \approx the raw data in a conventional likelihood.
- No need for particular statistics to relate to particular parameters.
- It is important to identify a set of statistics, ***sensitive*** to the scientifically important and repeatable features of the data, but ***insensitive*** to replicate-specific details of phase!

Dynamic model and autoregression

A dynamic model predicts the state at the next time step from the state now and possibly from the state at earlier times.

Similarly, autoregression using the observed states can be used to characterize how the observed state, or observed change in state, depends on previous observed states.

Finding appropriate statistics

1. Marginal distribution statistics

Simple moment statistics such as the **mean**, **median** and **standard deviation** . To summarize the 'shape' of the marginal distribution.

How to obtain statistics?

Through polynomial regression of the ordered observed values from the marginal distribution of interest, on some reference quantiles, e.g. uniform on $(0, 1)$. The statistics are the resulting polynomial regression coefficients.

2. Dynamic process statistics

If the structure of a regression model captures a good deal of the dynamics in a data set (fitted model) then the estimates of the **regression coefficients**, β , would be statistics carrying information about dynamic structure.

$$y_t^\alpha = \beta_0 + \beta_1 y_{t-1}^\alpha + \beta_2 y_{t-1}^{2\alpha} + \beta_3 y_{t-1}^{3\alpha} + \varepsilon_t$$

3. Time series statistics

Sensitive to the shape and period of fluctuations, e.g.
the **coefficients of the autocovariance function**
which, truncated at lag 11:

$$N_i = \beta_0 N_{i-12} + \beta_1 N_{i-12}^2 + \beta_2 N_{i-12}^3 + \beta_3 N_{i-2} + \beta_4 N_{i-2}^2 + \varepsilon_i$$

Exploring l_s by MCMC (Metropolis–Hastings method) and further inference

Guess $\theta^{[0]}$, then iterate the following for $k=1, 2, 3, \dots$:

First propose that $\theta^* = \theta^{[k-1]} + \delta^{[k]}$, where $\delta^{[k]}$ is a random vector from a convenient symmetric distribution; then set $\theta^{[k]} = \theta^*$ with probability: $\min [1, \exp\{l_s(\theta^*) - l_s(\theta^{[k-1]})\}]$

and set $\theta^{[k]} = \theta^{[k-1]}$ otherwise.

Theoretical properties of the multivariate normality approximation method

The method uses the approximation: $s \sim N(\mu_\theta, \Sigma_\theta)$

Let, the joint density of s be f_θ . A Taylor expansion of $\log(f_\theta)$ about its mode, μ_θ , gives:

$$\log(f_\theta(s)) \approx \log(f_\theta(\mu_\theta)) + (s - \mu_\theta)^\top \frac{\partial \log(f_\theta)}{\partial s} + \frac{1}{2}(s - \mu_\theta)^\top \frac{\partial^2 \log(f_\theta)}{\partial s \partial s^\top} (s - \mu_\theta)$$

With exponentiating, we will have the approximation:

$$f_\theta(s) \approx k \exp \left[-\frac{1}{2}(s - \mu_\theta)^\top \left(-\frac{\partial^2 \log(f_\theta)}{\partial s \partial s^\top} \right) (s - \mu_\theta) \right]$$

If the right-hand side of this equation is to be a PDF then it is the multivariate normal PDF, with covariance matrix:

$$\Sigma_\theta = (-\partial^2 \log(f_\theta) / \partial s \partial s^\top)^{-1}$$

For a good model with plausible parameter estimates, s should be close to μ_θ , with proximity increasing with increasing raw sample size, n .

Without knowledge of f_θ , μ_θ and Σ_θ are unknown. But, they can be estimated from a sample of s vectors produced by simulation.

Properties of $/s$

Given that $/s$ is a valid likelihood, we can use standard likelihood inference about the model parameters, θ . However, the effective sample size is $\text{dim}(s)$, which will always be small.

In the limit as $n \rightarrow \infty$, $/s$ results in consistent estimators, which

$(\partial^2 l_s / \partial \theta \partial \theta^\top)^{-1}$ are asymptotically unbiased with covariance matrix . The estimators will not be fully asymptotically efficient, and are not guaranteed to be normally distributed.

Robust covariance estimation

The method often produces perfectly reasonable results using the straightforward estimate $\hat{\Sigma}_{\theta}$, however, an estimate which discounts the tails of the distribution of s is better justified theoretically. In addition, if statistics of widely different magnitudes are used, then some care should be taken to ensure numerical robustness.

Campbell's method & numerical preconditioning

The **reweighting** via the w_i down-weights extreme tail observations to ensure statistical robustness. The use of D is standard numerical **preconditioning** to ensure numerical robustness. Operating in terms of the QR decomposition is efficient when computing with the inverse and determinant of the covariance matrix:

1. With S and $\hat{\mu}_\theta$ an $N_s \times Nr$ matrix and
use of preconditioning matrix: where $\bar{d}_i = (\sum_j s_{ij}^2 / N_r)^{1/2}$

$$D = \text{diag}(d)$$

Then form the QR decomposition:

$$\bar{Q}\bar{R} = S^\top D^{-1} / \sqrt{N_r - 1}$$

2. Find the Mahalanobis distance:

$$m_j = (s_j^* - \hat{\mu}_\theta)^\top \bar{\Sigma}_\theta^{-1} (s_j^* - \hat{\mu}_\theta)$$

3. Set $m_0 = \sqrt{N_s} + \sqrt{2}$ and compute the weights:

$$\hat{\mu}_\theta = \sum_j w_j s_j^* / \sum_i w_i$$

4. Redefine and recomputed:

$$w_j = \begin{cases} e^{-(m_j - m_0)^2/2} m_0 / m_j & \text{if } m_j > m_0 \\ 1 & \text{otherwise} \end{cases}$$

and

$$\mathbf{s} = (s_1^* - \hat{\mu}_\theta, s_2^* - \hat{\mu}_\theta, \dots).$$

5. Define $D=\text{diag}(d)$ and $W=\text{diag}(w)$, and then form the QR decomposition:

$$QR = WS^T D^{-1} / \sqrt{\sum_j w_j^2 - 1}$$

6. Compute:

$$\hat{\Sigma}_\theta = DR^T RD \text{ and } \hat{\Sigma}_\theta^{-1} = E^T E, \text{ where } E = (R^T)^{-1} D^{-1}.$$

and $\log |\Sigma_\theta|/2 = \sum_i \log |R_{ii}| + \sum_i \log (d_i)$

which is log likelihood.

Problem

- Observed statistics will be in the tail of the distribution.
- The undermining of the normality approximation for the statistics-not so important if the model doesn't fit anyway.
- MCMC chain may fail to mix properly: because of irregular likelihood based on extreme tails of the statistics distribution, which can cause the chain to become stuck in local maxima.

Solution

A robust *ls* for poorly fitting models

Modify *ls* to satisfy the tail behavior by:

$$g(x, d_0) = \begin{cases} x^2 & \text{if } |x| \leq d_0 \\ k|x|^\gamma + c & \text{if } |x| > d_0 \end{cases}$$

Where

$$k = 2d_0^{2-\gamma}/\gamma \quad c = d_0^2 - kd_0^\gamma$$

Then,

$$\tilde{l}_s = -\frac{1}{2}\log|\hat{\Sigma}_\theta| - \frac{1}{2}g\left(\sqrt{(s - \hat{\mu}_\theta)^\top \hat{\Sigma}_\theta (s - \hat{\mu}_\theta)}, d_0\right)$$

Another approach:

By replacing the Metropolis–Hastings acceptance probability by:

$$\min[1, \exp\{\gamma l_s(\theta^*) - \gamma l_s(\theta^{[k-1]})\}]$$

Unlike the use of robust *ls* this is inefficient when the model actually fits well!

Transformation for normality improvement of s

- The exact multivariate normality of s is not required.
- Nonetheless, the closer s is to multivariate normal, the better the approximation.

Checking the normality assumption and goodness of fit

- (i) Plot the N_r ordered values of $(\mathbf{s}_j^* - \hat{\boldsymbol{\mu}}_\theta)^T \boldsymbol{\Sigma}_\theta^{-1} (\mathbf{s}_j^* - \hat{\boldsymbol{\mu}}_\theta)$ against the quantiles of $\chi_{\dim(\mathbf{s})}^2$.
a departure from a straight-line indicates a departure of the simulated statistics from multivariate normality.
- (ii) Produce normal Q-Q plots for each statistic, \mathbf{s}_i , using N_r replicates. This checks the marginal normality of the statistics, under the model.
- (iii) Produce a normal Q-Q plot for the standardized residuals, for some h near the maximum-likelihood estimate. This checks the normality assumption for the observed statistics.

Question

Why are formal tests of normality not useful here?

&

Why are the normality tests that are applied to the full set of Nr simulated statistics vectors not useful either?

Answer

Formal tests of normality are not useful here, because the dimension of s is usually much too small for formal tests applied to the observed s to have useful power. Conversely, tests applied to the full set of Nr simulated statistics vectors, s^* , will almost always reject normality if we make Nr large enough. This is because they will be sensitive to the far tails of the distribution of s , which are not expected or required to be well approximated by a multivariate normal.

Maximum-likelihood estimation

Consider a problem with two parameters, θ_1 and θ_2 , and denote the output from the converged MCMC chain. A quadratic approximation to \mathcal{L} s in the vicinity of its maximum can then be obtained by quadratic regression, that is, by minimizing:

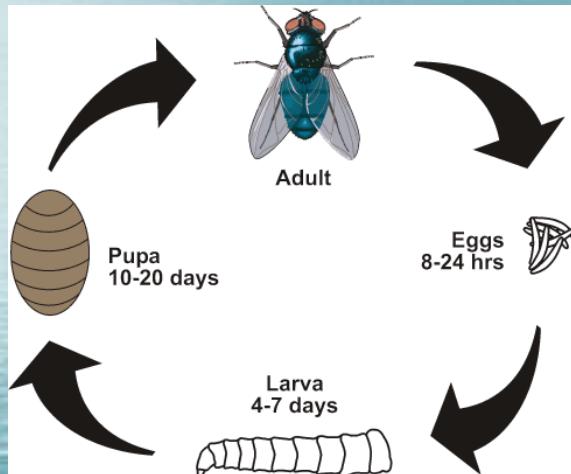
$$\sum_i (l_{si} - \alpha - \beta_1 \theta_{1i} - \beta_2 \theta_{2i} - \beta_3 \theta_{1i} \theta_{2i} - \beta_4 \theta_{1i}^2 - \beta_5 \theta_{2i}^2)^2$$

with respect to and the vector of coefficients β . The resulting quadratic can then be maximized to find $\hat{\theta}$, and the Hessian of \mathcal{L} s can be computed directly from the estimates of β .

Part II

**Example of Synthetic
likelihood in blowflies**

The dynamic nature of the fluctuations in Nicholson's blowfly experiments.



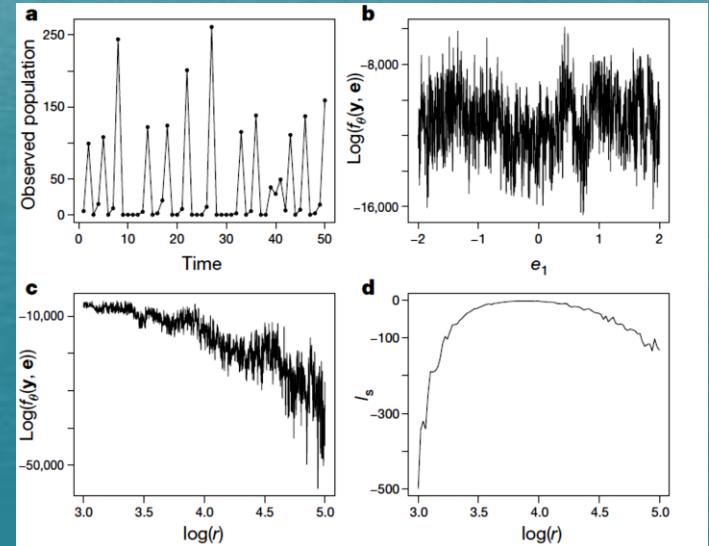
Blowflies are a family of insects in the order Diptera, with 1,100 known species. The name blowfly comes from an older English term for meat that had eggs laid on it. The prototypic ecological model with complex dynamics is the scaled Ricker map:

$$N_{t+1} = rN_t e^{-N_t + e_t}$$

Objective: Can we make statistical inferences about θ ?

$$\theta^\top = (r, \sigma_e^2, \phi)$$

- (a) With $\log(r)=3.8$, $\sigma=0.3$, $\phi=10$, using the Ricker model simulation we get Poisson random deviates.
- (b) The joint probability density function $f_{\theta}(\mathbf{y}, \mathbf{e})$, of data vector \mathbf{y} with noise vector \mathbf{e} from simulation, \mathbf{e} and \mathbf{y} held fixed;
- (c) as (b), but with \mathbf{r} .
- (d) The log synthetic likelihood, I_s , plotted against $\log(r)$ for the Ricker model and the data given in a ($N_r = 500$).



Likelihood-based inference about θ

Usually, we would like to do likelihood-based inference about θ but this requires integration of f_θ over all e , and this is analytically difficult.

Bayesian inference

Requires that we sample replicate vectors e and θ from a density proportional to $f_{\theta e}$, no methods exist to do this since f_θ is so irregular!

Trouble with the conventional methods of statistical inference for dynamic systems

The model must reproduce the exact course of the observed data where the real system itself would not do if repeated!

The trouble is that with conventional or naïve methods, we may be able to make the model reproduce the system exactly, however, in reality if the system were repeated under identical circumstances it would not be identical.

For dynamic models, it is necessary to judge model fit using statistics that reflect what is dynamically important in the data, and to discard the details of local phase.

Steps for synthetic likelihood estimation

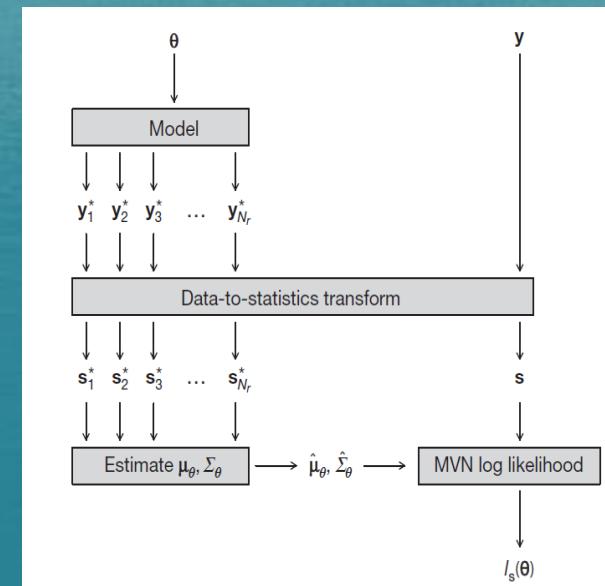
- Reduce the raw observed data, \mathbf{y} , to a vector of summary statistics, \mathbf{s} ;
- Find appropriate statistics, which capture the dynamic structure of the model, such as the coefficients of the autocovariance function and of polynomial autoregressive models;
- Estimate μ_θ , Σ_θ , by simulating model for θ ;

- Based on $\mathbf{s} \sim \mathbf{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$, the synthetic likelihood of any given value of the vector θ can be evaluated:

$$\hat{\boldsymbol{\mu}}_\theta = \sum_i \mathbf{s}_i^*/N_r, \quad \mathbf{S} = (\mathbf{s}_1^* - \hat{\boldsymbol{\mu}}_\theta, \mathbf{s}_2^* - \hat{\boldsymbol{\mu}}_\theta, \dots)$$

$$\hat{\boldsymbol{\Sigma}}_\theta = \mathbf{S}\mathbf{S}^\top / (N_r - 1)$$

$$l_s(\theta) = -\frac{1}{2}(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)^\top \hat{\boldsymbol{\Sigma}}_\theta^{-1} (\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta) - \frac{1}{2} \log |\hat{\boldsymbol{\Sigma}}_\theta|$$



For the Ricker model, Wood (2010) uses the following summary statistics

- autocovariances to lag 5
- the coefficients of the cubic regression of the ordered differences $y_t - y_{t-1}$ on their observed values
- β_1 and β_2 from the autoregression
 $y_{t+1}^{(0.3)} = \beta_1 y_t^{(0.3)} + \beta_2 y_t^{(0.6)} + \varepsilon_t$
- the mean population-dynamic
- number of zeros observed

For the blowfly model, Wood (2010) uses

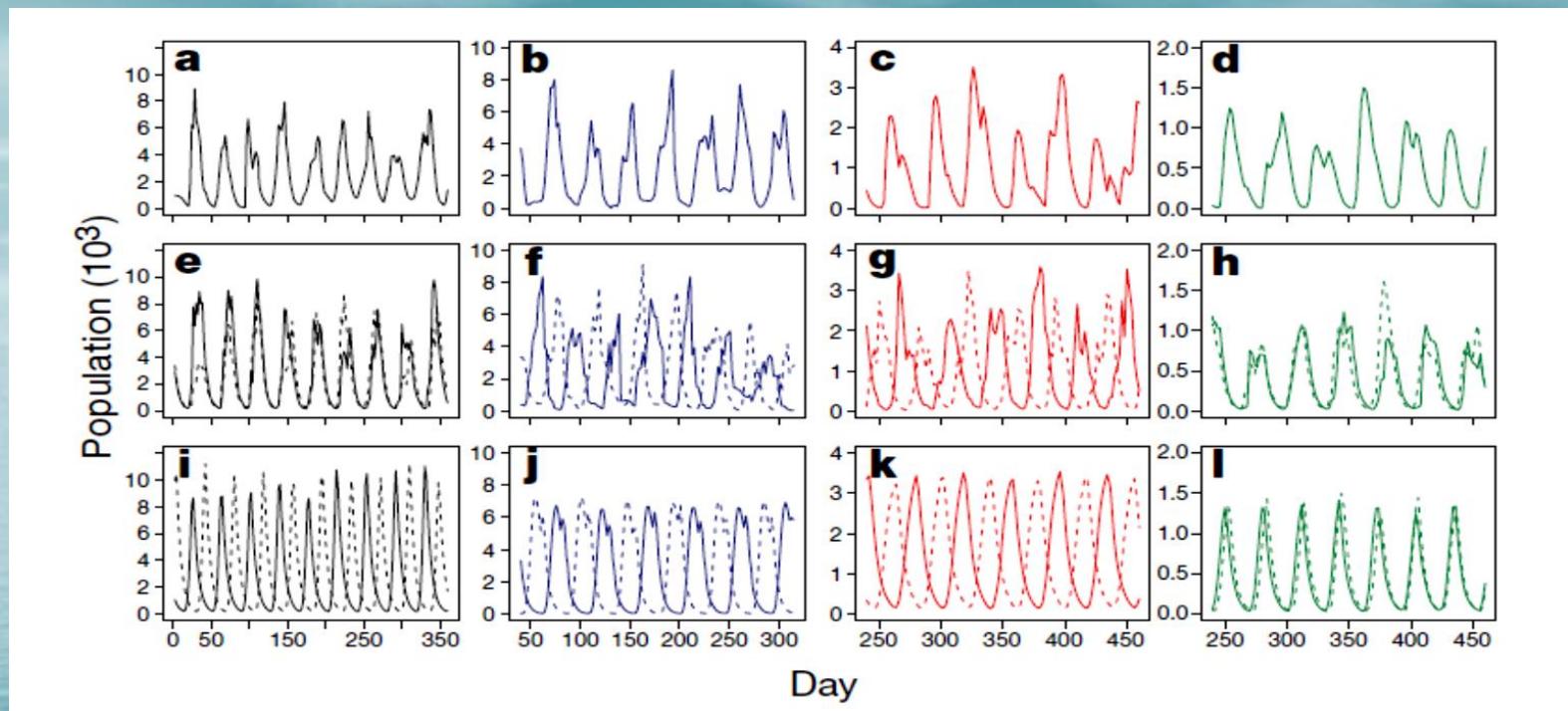
- the autocovariances to lag 11
- the coefficients of the cubic regression of the ordered differences $y_t - y_{t-1}$ on their observed values
- $\text{mean}(N_t) - \text{median}(N_t)$
- number of turning points observed
- coefficients of the autoregression

$$N_i = \beta_0 N_{i-12} + \beta_1 N_{i-12}^2 + \beta_2 N_{i-12}^3 + \beta_3 N_{i-2} + \beta_4 N_{i-2}^2 + \varepsilon_i$$

The model for the adult blowfly populations from Nicholson's experiments was proposed:

$$\frac{dN}{dt} = PN(t - \tau)e^{-N(t - \tau)/N_0} - \delta N(t)$$

Depending on the parameter values, this model has dynamics ranging from stable equilibrium to chaos. Using synthetic likelihood methods, Wood (2010) was able to conclude that the fluctuations are intrinsic to the blowfly population biology rather than a function of "stochastic forcing or excitation of the system".

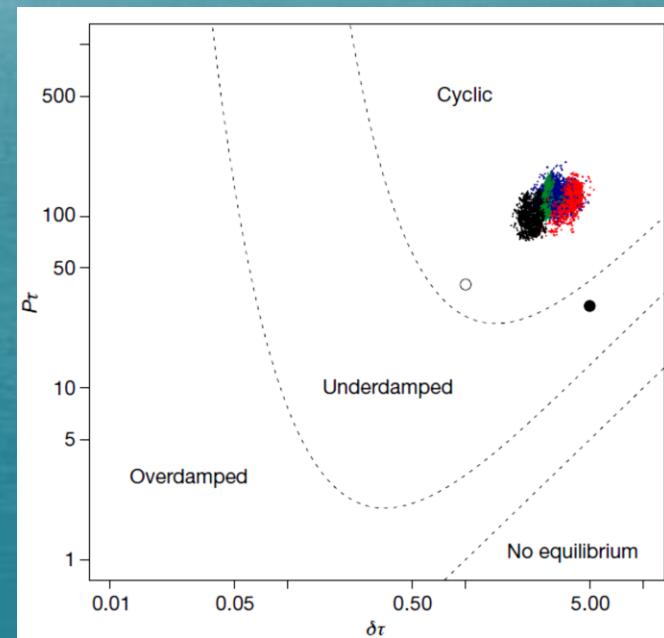


Blow flies data and model runs

a, b, Two laboratory adult populations of blowfly maintained under adult food limitation. **c, d,** as in a and b but maintained under moderate and more severe juvenile food limitation. **e–h,** Two replicates (one solid, one dashed) from the full model ($N_{t+1} = R_t + S_t$) equation fitted separately to the data shown in each of panels a–d, immediately above. **i–l,** As in e–h for the model with demographic stochasticity only. The observations are made every second day. The simulation phase is arbitrary. Notice the qualitatively good match of the dynamics (e–h) of the full model (main equation) to the data, relative to the insufficiently variable dynamics of the model with demographic stochasticity only (i–l).

Model stability

The colored points are samples from the stability-controlling parameter combinations $\delta\tau$ and $P\tau$, plotted (with matching color coding) for each experimental run shown in above. The open and filled circles show stability properties for alternative chain starting conditions: they give indistinguishable results, although the conditions marked by the filled circle lie in the plausible range for external noise-driven dynamics. The dynamics comprise limit cycles perturbed by noise but not driven by noise. The fluctuations are driven by the intrinsic population-dynamic processes, not by random variation exciting a resonance in otherwise stable dynamics.



Conclusion

The fluctuations are an intrinsically population-dynamic processes, not by random variation exciting a resonance in otherwise stable dynamics and would have occurred no matter how constant the experimental conditions, and no matter how large the cultures had been made. The method allowing this conclusion to be reached is widely applicable, and is a general purpose method for well-founded statistical inference about noisy ecological (and other) dynamic models in the chaotic and near-chaotic regimes.

References

- Nicholson, A. J. (1954). An outline of the dynamics of animal populations. *Aust. J. Zool.* 2: 9–65.
- Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466:1102–1104.

