Vol 466|26 August 2010|doi:10.1038/nature09319

LETTERS

# Statistical inference for noisy nonlinear ecological dynamic systems

Simon N. Wood[1]

**Chaotic ecological dynamic systems defy conventional statistical analysis. Systems with near-chaotic dynamics are little better. Such systems are almost invariably driven by endogenous dynamic processes plus demographic and environmental process noise, and are only observable with error. Their sensitivity to history means that minute changes in the driving noise realization, or the system parameters, will cause drastic changes in the system trajectory[1]. This sensitivity is inherited and amplified by the joint probability density of the observable data and the process noise, rendering it useless as the basis for obtaining measures of statistical fit. Because the joint density is the basis for the fit measures used by all conventional statistical methods[2], this is a major theoretical shortcoming. The inability to make well-founded statistical inferences about biological dynamic models in the chaotic and near-chaotic regimes, other than on an *ad hoc* basis, leaves dynamic theory without the methods of quantitative validation that are essential tools in the rest of biological science. Here I show that this impasse can be resolved in a simple and general manner, using a method that requires only the ability to simulate the observed data on a system from the dynamic model about which inferences are required. The raw data series are reduced to phase-insensitive summary statistics, quantifying local dynamic structure and the distribution of observations. Simulation is used to obtain the mean and the covariance matrix of the statistics, given model parameters, allowing the construction of a 'synthetic likelihood' that assesses model fit. This likelihood can be explored using a straightforward Markov chain Monte Carlo sampler, but one further post-processing step returns pure likelihood-based inference. I apply the method to establish the dynamic nature of the fluctuations in Nicholson's classic blowfly experiments[3–5].**

The prototypic ecological model with complex dynamics is the scaled Ricker map[6], describing the time course of a population $N_t$ by

$$N_{t+1} = rN_t e^{-N_t + e_t} \quad (1)$$

where the $e_t$ are independent N(0, $\sigma_e^2$) 'process noise' terms (assumed to be environmental noise here, for illustrative purposes) and $r$ is an intrinsic growth rate parameter controlling the model dynamics. This model amply illustrates the collapse of standard statistical methods in the face of chaotic or near-chaotic dynamics. Figure 1a shows data from a realization of equation (1) when $\log(r) = 3.8$, and what is observed are Poisson random deviates, $y_t$, with mean $\phi N_t$ (where $\phi$ is a scale parameter), reflecting a reasonably common sampling situation. Suppose that the aim is to make statistical inferences about $\boldsymbol{\theta}^\top = (r, \sigma_e^2, \phi)$ from this data series. Figure 1b, c illustrates the joint probability (density) function, $f_\theta(\mathbf{y}, \mathbf{e})$, of the data vector, $\mathbf{y}$, and noise vector, $\mathbf{e}$, when the (fixed) noise realization and data from the simulation in Fig. 1a are plugged in. Figure 1c shows how $\log(f_\theta)$ varies with $r$, whereas in Fig. 1b $r$ is kept fixed but the first element of the noise realization, $e_1$, is varied. Likelihood-based inference about $\boldsymbol{\theta}$ requires that we integrate $f_\theta$ over all $\mathbf{e}$, something that is analytically intractable and, from Fig. 1b, is numerically intractable as well[2]. Bayesian inference would require that we sample replicate vectors $\mathbf{e}$ and $\boldsymbol{\theta}$ from a density proportional to $f_\theta$: no methods exist to do this in a meaningful way for an $f_\theta$ as irregular as that shown in Fig. 1b, c.

The problem with the conventional approaches is actually philosophical. Naive methods of statistical inference try to make the model reproduce the exact course of the observed data in a way that the real system itself would not do if repeated. Although the dynamic processes driving the system are a repeatable feature about which inferences might be made, the local phase of the data is an entirely noise-driven feature, which should not contribute to any measure of match or mismatch between model and data. Hence, if statistical methods are to correspond with what is scientifically meaningful, 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. In itself, this idea is not new[6–10]. What is new is the ability to assess the consistency of statistics of the model simulations and data without recourse to *ad hoc* measures of that consistency, but in a way that instead gives access to much of the machinery of likelihood-based statistical inference[2].

The first step in the proposed analysis is therefore to reduce the raw observed data, $\mathbf{y}$, to a vector of summary statistics, $\mathbf{s}$, designed to

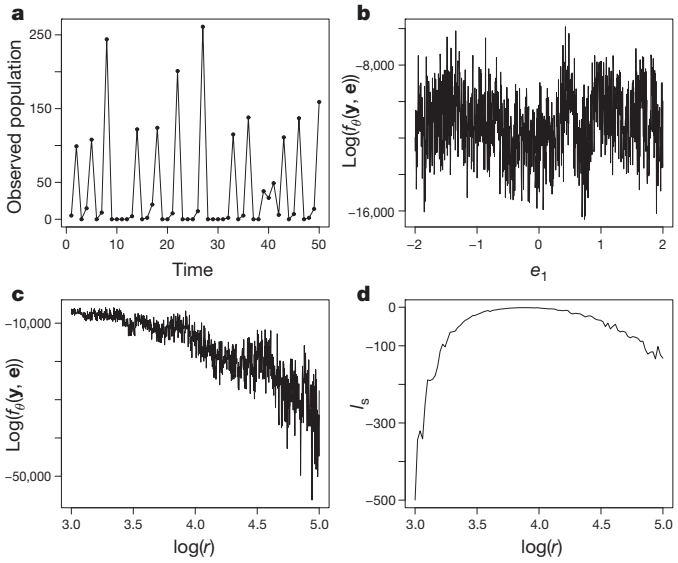


**Figure 1 | Measuring fit of the Ricker model.** **a**, Population data simulated from the Ricker model in the text, observed under Poisson sampling ($\log(r) = 3.8$, $\sigma = 0.3$, $\phi = 10$). **b**, The log joint probability density, $\log(f_\theta(\mathbf{y}, \mathbf{e}))$, of data, $\mathbf{y}$, and random process noise terms, $\mathbf{e}$, plotted against the value of the first process noise deviate, $e_1$, with the rest of $\mathbf{e}$ and $\mathbf{y}$ held fixed. **c**, $\mathrm{Log}(f_\theta(\mathbf{y}, \mathbf{e}))$ plotted against model parameter $r$, again with $\mathbf{e}$ and $\mathbf{y}$ held fixed. **d**, The log synthetic likelihood, $l_s$, plotted against $\log(r)$ for the Ricker model and the data given in **a** ($N_r = 500$).

[1]Mathematical Sciences, University of Bath, Bath BA2 7AY, UK.

capture the dynamic structure of the model (by specifying the components of **s** we define what matters about the dynamics, but not how much it matters). Examples of suitable statistics are the coefficients of the autocovariance function and of polynomial autoregressive models. To ensure appropriate marginal distributions, we can also use the coefficients obtained from polynomial regression of the observed order statistics on appropriate fixed reference quantiles. Using regression coefficients as statistics promotes approximate normality in the distribution of **s**, supporting the key multivariate normality approximation

$$\mathbf{s} \sim \mathrm{N}(\boldsymbol{\mu}_\theta, \Sigma_\theta) \quad (2)$$

The unknown mean vector, $\boldsymbol{\mu}_\theta$, and covariance matrix, $\Sigma_\theta$, are generally intractable functions of the vector of unknown model parameters, $\boldsymbol{\theta}$, but for any $\boldsymbol{\theta}$ they can be estimated by simulating from the model, in which case a sort of synthetic likelihood can be evaluated (Fig. 2).

Based on **s**, the synthetic likelihood of any given value of the vector $\boldsymbol{\theta}$ can be evaluated as follows. First we use the model to simulate $N_r$ replicate data sets, $\mathbf{y}_1^*, \mathbf{y}_2^*, \ldots$, and convert these to replicate statistics vectors, $\mathbf{s}_1^*, \mathbf{s}_2^*, \ldots$, exactly as **y** was converted to **s**. Then we evaluate $\hat{\boldsymbol{\mu}}_\theta = \sum_i \mathbf{s}_i^*/N_r$, $S = (\mathbf{s}_1^* - \hat{\boldsymbol{\mu}}_\theta, \mathbf{s}_2^* - \hat{\boldsymbol{\mu}}_\theta, \ldots)$ and, hence, $\hat{\Sigma}_\theta = SS^\top/(N_r - 1)$ (ref. 11). Dropping irrelevant constants, the log synthetic likelihood is

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

Like any likelihood, $l_s(\boldsymbol{\theta})$ measures the consistency of the parameter values $\boldsymbol{\theta}$ with the observed data, but it is a much smoother function of $\boldsymbol{\theta}$ than $f_\theta$, as Fig. 1d illustrates. Note that a robust estimator[11,12] can be advantageous in calculating $\hat{\Sigma}_\theta$.

The $l_s$ evaluation method is general enough to deal with hidden state variables, complicated observation processes, missing data and multiple data series. Calculated as described, $l_s$ is invariant to reparameterization and is robust to the inclusion of uninformative statistics, so very careful selection of statistics is not necessary. There is complete freedom to transform statistics to improve the approximation in equation (2). Furthermore, $l_s$ behaves like a conventional log likelihood in the $N_r \to \infty$ limit, giving access to much of the machinery of likelihood-based inference.

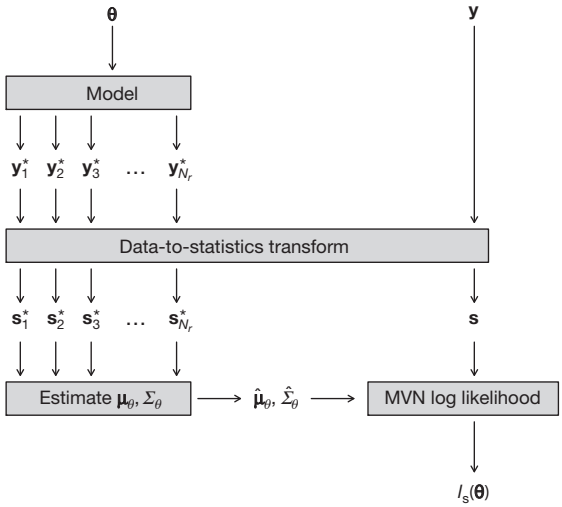


**Figure 2 | Synthetic likelihood evaluation.** Starting at the top, we wish to evaluate the fit of the model with parameter vector $\boldsymbol{\theta}$ to the raw data vector **y**. Replicate data vectors $\mathbf{y}_1^*, \ldots, \mathbf{y}_{N_r}^*$ are simulated from the model, given the value of $\boldsymbol{\theta}$. Each replicate, and the raw data, is converted into a vector of statistics, $\mathbf{s}_i^*$ or **s**, in the same way. The $\mathbf{s}_i^*$ are used to estimate the mean vector, $\hat{\boldsymbol{\mu}}_\theta$, and covariance matrix, $\hat{\Sigma}_\theta$, of **s**, according to the model with parameters $\boldsymbol{\theta}$. We use $\hat{\boldsymbol{\mu}}_\theta$, $\hat{\Sigma}_\theta$ and **s** respectively as the mean vector, the covariance matrix and the argument of the log multivariate normal (MVN) probability density function, to evaluate the log synthetic likelihood, $l_s$.

Finding maximum likelihood estimates, $\hat{\boldsymbol{\theta}}$, by maximizing $l_s$ with respect to $\boldsymbol{\theta}$, cannot usually be achieved by numerical optimizers for smooth functions, as $l_s$ typically has some residual small-scale roughness. Instead, a standard Metropolis–Hastings Markov chain Monte Carlo (MCMC) method[2,13] can be used for this purpose, and to simultaneously explore the range of parameter values consistent with the data. See Methods for implementational details, refinements, model checking and supporting theory.

To illustrate efficacy, I applied the method to the simulated Ricker data of Fig. 1a. Statistics were the autocovariances to lag 5; the coefficients of the cubic regression of the ordered differences $y_t - y_{t-1}$ on their observed values; the coefficients, $\beta_1$ and $\beta_2$, of the autoregression $y_{t+1}^{0.3} = \beta_1 y_t^{0.3} + \beta_2 y_t^{0.6} + \varepsilon_t$ ($\varepsilon_t$ is 'error'); the mean population; and the number of zeroes observed. The resulting 95% confidence intervals were $3.6 < \log(r) < 4.2$, $0.10 < \sigma_e < 0.55$ and $9.1 < \phi < 11.3$: each includes the true value, but a simulation study suggested that such intervals achieve coverage probabilities of 0.84, 0.85 and 0.87 for a sample size of 50, achieving nominal 0.95 coverage for sample sizes of 100–200.

As the main example, Fig. 3a–d shows adult blowfly populations from four runs of Nicholson's classic experiments[4,5]. Three decades after the experiments, the first plausible model for the last three replicates was proposed[3,14]:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = PN(t-\tau)\mathrm{e}^{-N(t-\tau)/N_0} - \delta N(t) \quad (3)$$

Here $N$ is the adult population and $P$, $N_0$, $\delta$ and $\tau$ are parameters. Depending on the parameter values, this model has dynamics ranging from stable equilibrium to chaos. Using careful, but *ad hoc*, parameter estimation methods, the dynamics of the adult food-limited replicate (Fig. 3b) was found to comprise limit cycles[14]: the fluctuations are intrinsic to the blowfly population biology. For the juvenile food-limited replicates (Fig. 3c, d), it was not possible to decide conclusively whether the fluctuations were driven and initiated by random external forcing and/or demographic stochasticity (quasicycles), or were intrinsically driven limit cycles. To progress requires a stochastic version of equation (3) that can produce both types of dynamics, and some means of estimating its parameters from data. Before the development of the method proposed here, estimation was difficult because the dynamics of plausibly parameterized models can be chaotic or near chaotic[14].

I discretized equation (3) with a daily time step. Births and deaths were subject to demographic stochasticity, with rates perturbed by daily environmental noise, yielding

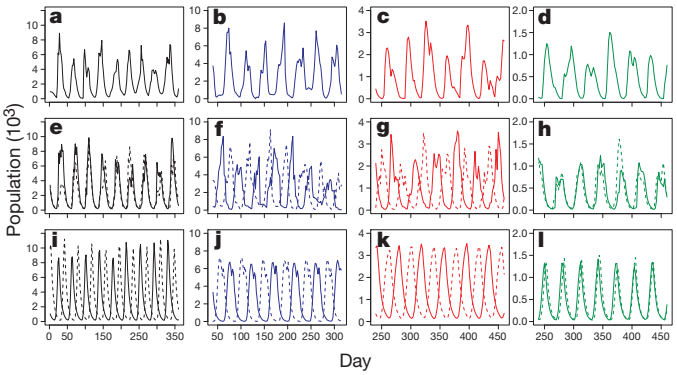


**Figure 3 | Blowfly data and model runs. a**, **b**, Two laboratory adult populations of sheep blowfly maintained under adult food limitation[4,5]. **c**, **d**, As in **a** and **b** but maintained under moderate and more severe juvenile food limitation[4]. **e**–**h**, Two replicates (one solid, one dashed) from the full model (equation (4)) 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 (equation (4)) to the data, relative to the insufficiently variable dynamics of the model with demographic stochasticity only (**i**–**l**).

$$N_{t+1} = R_t + S_t \quad (4)$$

where $R_t \sim \mathrm{Poi}\{PN_{t-\tau}\exp(-N_{t-\tau}/N_0)e_t\}$ denotes recruitment and $S_t \sim \mathrm{binom}\{\exp(-\delta\varepsilon_t), N_t\}$ denotes survival; that is, egg production is an independent Poisson process for each female and each adult has an independent probability of $\exp(-\delta\varepsilon_t)$ of surviving each day. The environmental stochasticity terms, $e_t$ and $\varepsilon_t$, are independent gamma-distributed random deviates with unit means and respective variances $\sigma_p^2$ and $\sigma_d^2$. All parameters are positive and $\tau$ is integer. However, the experiments were conducted in controlled conditions kept as constant as possible, suggesting that a model in which stochasticity is purely demographic is more plausible, a priori. That is, the simplified model with $e_t = \varepsilon_t = 1$ should be compared with equation (4) statistically.

I fitted equation (4) and its simplified version to each experimental replicate, with MCMC chains run for 50,000 iterations. Example replicate simulations from the final chain states are shown in Fig. 3e–h for equation (4) and Fig. 3i–l for the simplified model. The $\chi^2$ fit statistic suggested a good fit ($P > 0.2$) for equation (4) but a very bad fit ($P \ll 0.002$) for the simplified version, in all cases. Differences in the Akaike information criterion were >1,800 in favour of the full model (equation (4)) for all four replicates. Figure 3 suggests that the comprehensive rejection of the simplified model is because demographic stochasticity cannot produce the irregularity of the real cycles.

So equation (4), the stochastic version of equation (3), is not just qualitatively plausible, but actually fits the data quantitatively. Furthermore, uncontrolled variability in the experimental set-up dominates demographic stochasticity, raising the question of whether it drove the fluctuations rather than simply perturbing them. To answer this, Fig. 4 shows the stability diagram of ref. 14 for equation (3), overlaid with 1,500 values of the stability controlling parameters $P\tau$ and $\delta\tau$ randomly sampled from the second half of the MCMC chain for each experimental replicate. The overlaid sets summarize the parameter combinations consistent with the data.

There is extremely strong statistical evidence that the Nicholson blowfly fluctuations are limit cycles perturbed by noise. They are not the result of stochastic forcing or excitation of the system, despite decisive evidence for stochasticity well above demographic levels. The fluctuations are an intrinsically driven feature of blowfly biology 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.

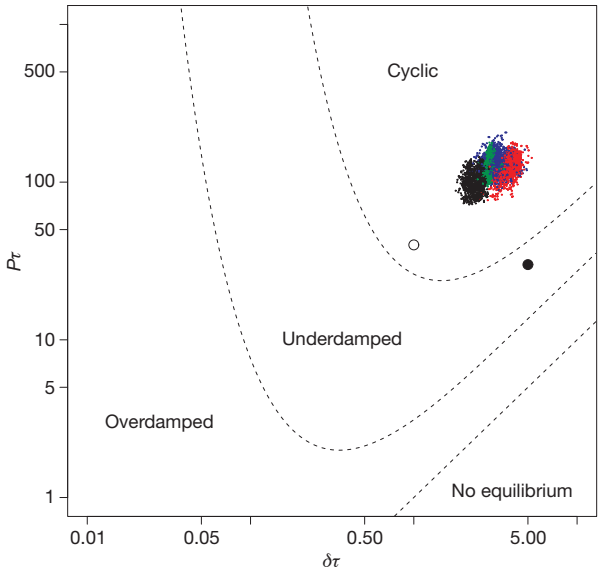


**Figure 4 | Blowfly model stability diagram[3,14].** The coloured points are samples from the stability-controlling parameter combinations $\delta\tau$ and $P\tau$, plotted (with matching colour coding) for each experimental run shown in Fig. 3. The open and filled circles show stability properties 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[14]. 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.

## METHODS SUMMARY

**Exploring $l_s$ by MCMC.** Starting from a parameter guess $\boldsymbol{\theta}^{[0]}$, iterate the following for $k = 1, 2, 3, \ldots$: first propose that $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{[k-1]} + \boldsymbol{\delta}^{[k]}$, where $\boldsymbol{\delta}^{[k]}$ is a random vector from a convenient symmetric distribution; then set $\boldsymbol{\theta}^{[k]} = \boldsymbol{\theta}^*$ with probability $\min[1, \exp\{l_s(\boldsymbol{\theta}^*) - l_s(\boldsymbol{\theta}^{[k-1]})\}]$ and set $\boldsymbol{\theta}^{[k]} = \boldsymbol{\theta}^{[k-1]}$ otherwise.

**Further inference.** For many statistical purposes, the set of $\boldsymbol{\theta}^{[k]}$ values from the converged chain is sufficient. However, in the vicinity of the maximum-likelihood estimate $\hat{\boldsymbol{\theta}}$, $\lim_{N_r \to \infty} l_s$ can be estimated by quadratic regression of the sampled $l_s(\boldsymbol{\theta}^{[k]})$ values on the $\boldsymbol{\theta}^{[k]}$ values, from the converged chain. This allows the use of standard likelihood theory for inference[2,15]. In particular, alternative models can be compared using the Akaike information criterion or generalized likelihood ratio testing[2,8,15]. A useful model-checking diagnostic is that $(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)^\top \hat{\Sigma}_\theta^{-1} (\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta) \sim \chi^2_{\dim(\mathbf{s})}$ if the model fits.

**Blowfly statistics.** The blowfly statistics I used were the autocovariances to lag 11 and the same difference distribution summary as in the Ricker example, along with $\mathrm{mean}\{N_t\}$, $\mathrm{mean}\{N_t\} - \mathrm{median}\{N_t\}$, the number of turning points observed and the estimated coefficients, $\hat{\boldsymbol{\beta}}$, of the autoregression

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

Further details and code are in Supplementary Information and Methods.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.

**Received 4 May; accepted 28 June 2010.**
**Published online 11 August 2010.**

1. May, R. M. Simple mathematical models with very complicated dynamics. *Nature* **261,** 459–467 (1976).
2. Davison, A. C. *Statistical Models* 94–160, 456–458, 605–619 (Cambridge Univ. Press, 2003).
3. Gurney, W. S. C., Blythe, S. P. & Nisbet, R. M. Nicholson's blowflies revisited. *Nature* **287,** 17–21 (1980).
4. Nicholson, A. J. An outline of the dynamics of animal populations. *Aust. J. Zool.* **2,** 9–65 (1954).
5. Nicholson, A. J. The self-adjustment of populations to change. *Cold Spring Harb. Symp. Quant. Biol.* **22,** 153–173 (1957).
6. Turchin, P. *Complex Population Dynamics* 8–11, 47–77 (Princeton Univ. Press, 2003).
7. Kendall, B. E. *et al.* Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. *Ecology* **80,** 1789–1805 (1999).
8. Cox, D. R. *Principles of Statistical Inference* 133–177 (Cambridge Univ. Press, 2006).
9. Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* **162,** 2025–2035 (2002).
10. McFadden, D. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* **57,** 995–1026 (1989).
11. Krzanowski, W. J. *Principles of Multivariate Analysis: A User's Perspective* 60, 211–213, 231–235 (Oxford Univ. Press, 1988).
12. Campbell, N. A. Robust procedures in multivariate analysis. I: Robust covariance estimation. *J. R. Stat. Soc. Ser. C* **29,** 231–237 (1980).
13. Gamerman, D. & Lopes, H. F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* 2nd edn (CRC/Chapman & Hall, 2006).
14. Nisbet, R. M. & Gurney, W. S. C. *Modelling Fluctuating Populations* 285–308 (Wiley, 1982).
15. Silvey, S. D. *Statistical Inference* 68–86 (Chapman & Hall, 1975).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** I am grateful to W. Gurney, R. Nisbet, S. Ellner, P. Turchin and B. Kendall for many discussions of the problem addressed here, and to the participants in the 2009 Statistical Methods for Dynamic Systems Models Workshop in Vancouver for discussion of this particular work. This work is part of the research programme of the EPSRC/NERC-funded UK National Centre for Statistical Ecology.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to the author (s.wood@bath.ac.uk).

## METHODS

This section includes further method details, refinements and theory. See Supplementary Information for MCMC details, example-specific details and further examples.

**Choosing statistics.** The statistics, **s**, have the same role in the synthetic likelihood, $l_s$, that the raw data would in a conventional likelihood. Hence, there is no need for particular statistics to relate to particular parameters any more than there is for particular raw data points to relate to particular parameters in conventional likelihood-based statistics. What is important is to identify a set of statistics that is sensitive to the scientifically important and repeatable features of the data, but insensitive to replicate-specific details of phase.

**Several types of statistic are especially useful.** 1. Marginal distribution statistics. These characterize the marginal distribution of the observations, or first differences of the observations. Here, 'marginal distribution' means the distribution ignoring time ordering. Simple moment statistics such as the mean, median and standard deviation are the most obvious. As useful are statistics summarizing the 'shape' of the marginal distribution. A convenient way of obtaining these is through polynomial regression of the ordered observed values from the marginal distribution of interest, on some reference quantiles. Obvious reference quantiles are uniform on (0, 1), but a better choice is to use the ordered observed values from the raw data as the reference quantiles. The statistics are the resulting polynomial regression coefficients.

2. Dynamic process statistics. A dynamic model predicts the state at the next time step (or the change in state) 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. For example, the structure of the Ricker model suggests that a regression $y_t^\alpha = \beta_0 + \beta_1 y_{t-1}^\alpha + \beta_2 y_{t-1}^{2\alpha} + \beta_3 y_{t-1}^{3\alpha} + \varepsilon_t$ might capture a good deal of the dynamics in the Ricker data ($\alpha$ is a transformation parameter, to be tuned to improve fit). If this is so, then the estimates of the regression coefficients, $\boldsymbol{\beta}$, would be statistics carrying information about dynamic structure. For models with unobserved states, simple autoregressions will often need to be replaced by autoregressions on multiple lagged states (by Taken's theorem[16]).

3. Time series statistics. These are sensitive to the shape and period of fluctuations. Good examples are the coefficients of the autocovariance function for the data series, truncated at some lag.

For example, the statistics that I used for the blowfly models were the autocovariances to lag 11 and the same difference distribution summary as in the Ricker example, along with mean$\{N_t\}$, mean$\{N_t\}$ − median$\{N_t\}$, the number of turning points observed and the estimated coefficients, $\hat{\boldsymbol{\beta}}$, of the autoregression

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

Note that sometimes the selection of statistics will be iterative. That is, after fitting with an initial set of statistics, model checking may identify discrepancies in the fit, which in turn suggest extra statistics to incorporate in a revised synthetic likelihood.

Although the need to find suitable statistics might be viewed as an extra burden, the preparatory work required to do so is no more than the exploratory analysis that constitutes good statistical practice. Similarly, the requirement to explicitly formulate what the model should get right, at the outset, has benefits beyond that of simply permitting estimation.

**Exploring $l_s$ by MCMC and further inference.** The MCMC scheme used in the paper was a simple Metropolis–Hastings method. Starting from a parameter guess $\boldsymbol{\theta}^{[0]}$, iterate the following for $k = 1, 2, 3, \ldots$: first propose that $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{[k-1]} + \boldsymbol{\delta}^{[k]}$, where $\boldsymbol{\delta}^{[k]}$ is a random vector from a convenient symmetric distribution; then set $\boldsymbol{\theta}^{[k]} = \boldsymbol{\theta}^*$ with probability $\min[1, \exp\{l_s(\boldsymbol{\theta}^*) - l_s(\boldsymbol{\theta}^{[k-1]})\}]$ and set $\boldsymbol{\theta}^{[k]} = \boldsymbol{\theta}^{[k-1]}$ otherwise.

For many statistical purposes, the set of $\boldsymbol{\theta}^{[k]}$ values from the converged chain is sufficient. However, in the vicinity of the maximum-likelihood estimate $\hat{\boldsymbol{\theta}}$, $\lim_{N_r \to \infty} l_s$ can be estimated by quadratic regression of the sampled $l_s(\boldsymbol{\theta}^{[k]})$ values on the $\boldsymbol{\theta}^{[k]}$ values, from the converged chain. This allows the use of standard likelihood theory for inference[2,15]. In particular, alternative models can be compared using the Akaike information criterion or generalized likelihood ratio testing[2,8,15]. A useful model-checking diagnostic is that $(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)^\top \hat{\Sigma}_\theta^{-1} (\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta) \sim \chi^2_{\dim(\mathbf{s})}$ if the model fits. See below for further details.

**Theoretical properties of the method.** This section provides some theoretical investigation of the multivariate normality approximation, and of $l_s$, giving theoretical justification for the approach described in the paper.

1. The multivariate normality approximation. The method uses the approximation

$$\mathbf{s} \sim \mathrm{N}(\boldsymbol{\mu}_\theta, \Sigma_\theta) \quad (5)$$

which requires some justification, because even with careful choice of statistics equation (5) is unlikely to be exact (although for many statistics the central limit theorem will imply normality as the raw sample size, $n$, tends to infinity). Therefore, let the true, but unknown, joint density of **s** be $f_\theta(\mathbf{s})$. A Taylor expansion of $\log(f_\theta)$ about its mode, $\boldsymbol{\mu}_\theta$, gives

$$\log(f_\theta(\mathbf{s})) \approx \log(f_\theta(\boldsymbol{\mu}_\theta)) + (\mathbf{s} - \boldsymbol{\mu}_\theta)^\top \frac{\partial \log(f_\theta)}{\partial \mathbf{s}} + \frac{1}{2}(\mathbf{s} - \boldsymbol{\mu}_\theta)^\top \frac{\partial^2 \log(f_\theta)}{\partial \mathbf{s} \partial \mathbf{s}^\top} (\mathbf{s} - \boldsymbol{\mu}_\theta) \quad (6)$$

and as usual the approximation will be more accurate the closer **s** is to $\boldsymbol{\mu}_\theta$. Because $\boldsymbol{\mu}_\theta$ is a mode, $\partial \log(f_\theta)/\partial \mathbf{s} = 0$. So, exponentiating, we have the approximation

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

where $k$ is a constant of proportionality. It is immediately recognizable that if the right-hand side of this equation is to be a probability density function then it is the multivariate normal probability density function, with covariance matrix $\Sigma_\theta = (-\partial^2 \log(f_\theta)/\partial \mathbf{s} \partial \mathbf{s}^\top)^{-1}$. This approximation is familiar in statistics from the Laplace approximation of integrals. For **s** sufficiently close to $\boldsymbol{\mu}_\theta$, we therefore expect $f_\theta$ to be well approximated by the probability density function of $\mathrm{N}(\boldsymbol{\mu}_\theta, \Sigma_\theta)$. For a good model with plausible parameter estimates, **s** should be close to $\boldsymbol{\mu}_\theta$, with proximity increasing with increasing raw sample size, $n$.

Without knowledge of $f_\theta$, $\boldsymbol{\mu}_\theta$ and $\Sigma_\theta$ are unknown. However, they can be estimated from a sample of **s** vectors produced by simulation, as in the main text. Given the Taylor series truncation used to obtain equation (5), we do not necessarily expect it to be a good approximation in the tails of the distribution of **s**. Hence, it may be advantageous to use robust estimators for $\boldsymbol{\mu}_\theta$ and $\Sigma_\theta$, which down-weight **s** vectors from the tails of $f_\theta$. Such a procedure is described below.

2. Properties of the log synthetic likelihood, $l_s$. Given equation (5), consider the properties of $l_s$. First, **s** is observed data, just like the raw data, **y**, and is therefore an equally valid basis for forming a likelihood. Hence, given the approximation in equation (5), $l_s$ is a valid likelihood for $\boldsymbol{\theta}$.

Given that $l_s$ is a valid likelihood, we can use standard likelihood asymptotic results[2,15,17] for inference about the model parameters, $\boldsymbol{\theta}$. However, when using this standard theory the effective sample size is $\dim(\mathbf{s})$, which will always be small, calling into question the accuracy of approximations based on large-sample asymptotics. It is therefore of interest to investigate the properties of $l_s$ as the raw sample size, $n = \dim(\mathbf{y})$, becomes large.

The key to establishing several useful results is to show that $l_s \to \mathrm{E}(l_s)$ as $n \to \infty$ (where E denotes expectation; see, for example, section 4.5 of ref. 15). Let $\boldsymbol{\theta}_0$ denote the true value of $\boldsymbol{\theta}$, such that $\boldsymbol{\mu}_{\theta_0}$ is the corresponding true mean vector of **s**. Let $N_r \to \infty$. Define $\boldsymbol{\varepsilon} = \mathbf{s} - \boldsymbol{\mu}_{\theta_0}$ and make the mild assumption that the elements of **s** are such that $\boldsymbol{\varepsilon} \to \mathbf{0}$ and $|\Sigma_\theta| \to 0$ as $n \to \infty$. Then $l_s$ can be rewritten as

$$\begin{aligned} l_s(\boldsymbol{\theta}) &= -\frac{1}{2}(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta + \boldsymbol{\varepsilon})^\top \Sigma_\theta^{-1} (\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta + \boldsymbol{\varepsilon}) - \frac{1}{2}\log|\Sigma_\theta| \\ &= -\frac{1}{2}(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)^\top \boldsymbol{\Sigma}_\theta^{-1} (\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta) - \boldsymbol{\varepsilon}^\top \Sigma_\theta^{-1} (\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta) - \frac{1}{2}\boldsymbol{\varepsilon}^\top \Sigma_\theta^{-1} \boldsymbol{\varepsilon} - \frac{1}{2}\log|\Sigma_\theta| \\ &= -\frac{1}{2}\mathrm{Tr}\left[ \Sigma_\theta^{-1} \left\{ (\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)^\top + 2(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)\boldsymbol{\varepsilon}^\top + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \right\} \right] - \frac{1}{2}\log|\Sigma_\theta| \end{aligned}$$

For $\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta \neq 0$, as $n \to \infty$ $\boldsymbol{\varepsilon} \to \mathbf{0}$ and the term in curly brackets is dominated by $(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)^\top$, so $l_s \to \mathrm{E}(l_s)$ when $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, both $\mathrm{E}(l_s)$ and $l_s$ tend to $-(1/2)\log|\Sigma_\theta|$.

Let $\hat{\boldsymbol{\theta}}$ be the maximizer of $l_s$. Standard theory (for example section 4.4 of ref. 15) establishes that $\mathrm{E}(l_s)$ is maximized at $\boldsymbol{\theta}_0$. So by the asymptotic convergence of $l_s$ to $\mathrm{E}(l_s)$, just established, $\hat{\boldsymbol{\theta}}$ is a consistent estimator.

Turning to the large-sample distribution of the maximum-synthetic-likelihood estimator, $\hat{\boldsymbol{\theta}}$, the usual Taylor expansion argument (see, for example, section 4.6 of ref. 15) gives

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \approx -\left( \frac{\partial^2 l_s}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right)^{-1} \frac{\partial l_s}{\partial \boldsymbol{\theta}} \quad (7)$$

By standard properties of the expected log likelihood, $\mathrm{E}(\partial l_s/\partial \boldsymbol{\theta}) = 0$ and $\mathrm{cov}(\partial l_s/\partial \boldsymbol{\theta}) = \partial^2 l_s/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$. Because $l_s$ converges to its expected value, in the limit as $n \to \infty$ we find that $\hat{\boldsymbol{\theta}}$ has expectation $\boldsymbol{\theta}_0$ and covariance matrix $(\partial^2 l_s/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top)^{-1}$.

If $\partial l_s/\partial \boldsymbol{\theta}$ had a multivariate normal distribution, then by equation (7) so would $\hat{\boldsymbol{\theta}}$, at least asymptotically. Hence, consider

$$\begin{aligned} \frac{\partial l_s}{\partial \theta_k} &= \frac{\partial \boldsymbol{\mu}_\theta^\top}{\partial \theta_k} \Sigma_\theta^{-1} (\mathbf{s} - \boldsymbol{\mu}_\theta) + \frac{1}{2}(\mathbf{s} - \boldsymbol{\mu}_\theta)^\top \Sigma_\theta^{-1} \frac{\partial \Sigma_\theta}{\partial \theta_k} \Sigma_\theta^{-1} (\mathbf{s} - \boldsymbol{\mu}_\theta) \\ &= \frac{\partial \boldsymbol{\mu}_\theta^\top}{\partial \theta_k} \Sigma_\theta^{-1/2} \mathbf{z} + \frac{1}{2}\mathbf{z}^\top \Sigma_\theta^{-1/2} \frac{\partial \Sigma_\theta}{\partial \theta_k} \Sigma_\theta^{-1/2} \mathbf{z} \end{aligned}$$

where $\mathbf{z} = \Sigma_\theta^{-1/2}(\mathbf{s} - \boldsymbol{\mu}_\theta)$ is a normal random vector, asymptotically distributed as $\mathrm{N}(\mathbf{0}, I)$, where $I$ is the identity matrix. The first term on the right-hand side of

this equation is clearly normally distributed, but the second can only be approximately so. Hence, $\partial l_s/\partial\theta_k$ will be approximately normally distributed if $\Sigma_\theta$ depends only weakly on $\theta_k$ such that $\partial\Sigma_\theta/\partial\theta_k$ is close to the zero matrix and the second term on the right-hand side is negligible, or if $\dim(\mathbf{z}) = \dim(\mathbf{s}) \to \infty$ such that the second term tends to a normally distributed random variable, by the central limit theorem.

We thus have the weak result that $\hat{\boldsymbol{\theta}}$ will have a normal distribution only if $\Sigma_\theta$ depends only weakly on the parameters, or if the number of statistics used is large. This, along with loss of asymptotic efficiency, is the price paid for circumventing the massive irregularity that near-chaotic dynamics give the inferential problem.

In summary, in the limit as $n \to \infty$, $l_s$ results in consistent estimators, which are asymptotically unbiased with covariance matrix $(\partial^2 l_s/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top)^{-1}$. The estimators will not be fully asymptotically efficient, and are not guaranteed to be normally distributed.

**Method refinements.** The remaining sections introduce various useful refinements of the basic method described in the paper. For many applications, using these refinements makes little quantitative difference to the results of applying the method, but they are sometimes necessary, and in any case serve to strengthen confidence in the results.

**Robust covariance estimation.** The method often produces perfectly reasonable results using the straightforward estimate, $\hat{\Sigma}_\theta$, given in the main text. However, the argument of the above section on the multivariate normal approximation suggests that an estimate which discounts the tails of the distribution of $\mathbf{s}$ is better justified theoretically. That is, a statistically robust estimator is somewhat more appropriate. In addition, if statistics of widely different magnitudes are used, then some care should be taken to ensure numerical robustness.

Let $S$ and $\hat{\boldsymbol{\mu}}_\theta$ be as in the paper, with $S$ an $N_s \times N_r$ matrix.

(i) Let $\bar{D} = \mathrm{diag}(\bar{\mathbf{d}})$, where $\bar{d}_i = \left(\sum_j S_{ij}^2/N_r\right)^{1/2}$. Then form the $QR$ decomposition

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

The initial estimate of $\Sigma_\theta$ is $\bar{\Sigma}_\theta = \bar{D}\bar{R}^\top\bar{R}\bar{D}$, and $\bar{\Sigma}_\theta^{-1} = \bar{D}^{-1}\bar{R}^{-1}(\bar{R}^\top)^{-1}\bar{D}^{-1}$. The use of preconditioning matrix $\bar{D}$ ensures that $\bar{R}$ has full numerical rank (and low condition number), so that stable computation of $\bar{R}^{-1}$ is possible.

(ii) Find the Mahalanobis distance, $m_j$, of each $\mathbf{s}_j^*$ from $\hat{\boldsymbol{\mu}}_\theta$. That is, form

$$m_j = (\mathbf{s}_j^* - \hat{\boldsymbol{\mu}}_\theta)^\top \bar{\Sigma}_\theta^{-1}(\mathbf{s}_j^* - \hat{\boldsymbol{\mu}}_\theta)$$

The $m_j$ are used to identify points far into the tails of the distribution of $\mathbf{s}$.

(iii) Set $m_0 = \sqrt{N_s} + \sqrt{2}$ and compute the weights

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

(iv) Redefine and recompute

$$\hat{\boldsymbol{\mu}}_\theta = \sum_j w_j \mathbf{s}_j^* / \sum_i w_i$$

and $S = (\mathbf{s}_1^* - \hat{\boldsymbol{\mu}}_\theta, \mathbf{s}_2^* - \hat{\boldsymbol{\mu}}_\theta, \ldots)$.

(v) Compute $d_i = \left(\sum_j S_{ij}^2/N_r\right)^{1/2}$, define $D = \mathrm{diag}(\mathbf{d})$ and $W = \mathrm{diag}(\mathbf{w})$, and then form the $QR$ decomposition

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

(vi) Now $\hat{\Sigma}_\theta = DR^\top RD$ and $\hat{\Sigma}_\theta^{-1} = E^\top E$, where $E = (R^\top)^{-1}D^{-1}$. The latter is convenient for computation of the log likelihood, as is the fact that $\log|\hat{\Sigma}_\theta|/2 = \sum_i \log|R_{ii}| + \sum_i \log(d_i)$.

The reweighting via the $w_i$ is Campbell's method[12], as described on pp. 231–235 of ref. 11. It down-weights extreme tail observations to ensure statistical robustness. The use of $D$ is standard numerical preconditioning to ensure numerical robustness (see, for example, section 2.9 of ref. 18). Operating in terms of the $QR$ decomposition is efficient when computing with the inverse and determinant of the covariance matrix.

**A robust $l_s$ for poorly fitting models.** When comparing the fits of different models, some models in the comparison may turn out to fit the data poorly. This means that the observed statistics will be in the tail of the distribution of the statistics according to the model, even for the best-fitting parameter values. The consequent undermining of the normality approximation for the statistics is unimportant if the model does not fit anyway, but a more serious problem is that the MCMC chain may fail to mix properly. This failure is because the likelihood based on extreme tails of the statistics distribution can be rather irregular and display local maxima that are pronounced enough that the chain can become stuck in them. A solution is to modify $l_s$ to attenuate the tail behaviour.

For example, first define

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

where $k = 2d_0^{2-\gamma}/\gamma$ and $c = d_0^2 - kd_0^\gamma$. The function $g(x)$ is quadratic in $x$ up to $d_0$, but thereafter grows less quickly than a quadratic if $\gamma < 2$: it is continuous to first derivative. Then define a robust version of $l_s$ as

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

where $d_0$ is the 99th percentile of the $\chi^2_{\dim(\mathbf{s})}$ distribution, for example. The value $\gamma = 0.1$ is quite effective. The robust $\tilde{l}_s$ is exactly $l_s$ for $\mathbf{s}$ not in the far tails of the distribution—that is, if the model fits at all well. Otherwise, $\tilde{l}_s$ decreases as $\boldsymbol{\theta}$ pushes $\mathbf{s}$ into the tails of the distribution of statistics, but the rate of decrease is much less than for $l_s$. Using $\tilde{l}_s$ to determine the acceptance probability, $\alpha$, in the MCMC chain avoids the chain becoming stuck. Of course, $l_s$ is still computed at each step, to be used for inference. This robust approach is used for the demographic stochasticity blowfly model and in the model-comparison example in Supplementary Information.

An alternative approach to getting the chain to move in difficult tail regions is to replace the Metropolis–Hastings acceptance probability by

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

where $\gamma$ is a small positive constant less than one. However, unlike the use of $\tilde{l}_s$, this is inefficient when the model actually fits well.

**Transforming to improve normality of $\mathbf{s}$.** The preceding discussions show that exact multivariate normality of $\mathbf{s}$ is not required, especially if care is taken in estimating $\boldsymbol{\mu}_\theta$ and $\Sigma_\theta$. Nonetheless, the closer $\mathbf{s}$ is to multivariate normal, the better the approximation in equation (5) will be (meaning that it will apply for $\mathbf{s}$ farther from $\boldsymbol{\mu}_\theta$). In particular, the first neglected term in the expansion yielding equation (6) will be even smaller if the statistics have symmetric distributions. The structure of the method allows complete freedom to transform $\mathbf{s}$ to better achieve multivariate normality. See section 7.6 of ref. 11 for a general discussion of such transformation. Here one very simple approach is presented that focuses on improving marginal normality.

(i) Run a pilot MCMC chain, using untransformed statistics, to obtain an estimate $\tilde{\boldsymbol{\theta}}$ close to the maximum-likelihood estimate.

(ii) Simulate a large number, $N$, of replicate statistics vectors using $\tilde{\boldsymbol{\theta}}$.

(iii) For each statistic, plot $N$ quantiles of the standard normal against the ordered replicates of the statistic, and find a piecewise-linear approximation to the resulting plot. These piecewise-linear approximations can be used to transform each statistic to approximate normality.

(iv) Run the full model estimation and inference process with all statistics (observed and simulated) transformed using the piecewise linear transforms from step (iii).

Although this method can substantially improve the multivariate normality assumption, doing so seems to make little practical difference to the results. However, gross violation of the normality assumption would presumably make the transformation step necessary.

**Checking the normality assumption and goodness of fit.** Several checks of the normality assumption, equation (5), are useful.

(i) Plot the $N_r$ ordered values of $(\mathbf{s}_j^* - \hat{\boldsymbol{\mu}}_\theta)^\top \hat{\Sigma}_\theta^{-1}(\mathbf{s}_j^* - \hat{\boldsymbol{\mu}}_\theta)$ against the quantiles of a $\chi^2_{\dim(\mathbf{s})}$ distribution (the log scale is more useful than the raw scale here). Departure from a straight-line relationship with slope one indicates a departure of the simulated statistics from multivariate normality. Note that departures in the upper tail of this plot are expected, and unproblematic, given the preceding theoretical arguments.

(ii) Produce normal quantile–quantile plots for each statistic, $s_i$, using $N_r$ replicates produced for a $\boldsymbol{\theta}$ near the maximum-likelihood estimate. This checks the marginal normality of the statistics, under the model.

(iii) Produce a normal quantile–quantile plot for the standardized residuals $\hat{\Sigma}_\theta^{-1/2}(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)$, for some $\boldsymbol{\theta}$ near the maximum-likelihood estimate. This checks the normality assumption for the observed statistics.

See ref. 11 for further discussion. Note that formal tests of normality are not useful here. The dimension of $\mathbf{s}$ is usually much too small for formal tests applied to the observed $\mathbf{s}$ to have useful power. Conversely, tests applied to the full set of $N_r$ simulated statistics vectors, $\mathbf{s}_j^*$, will almost always reject normality if we make $N_r$ large enough. This is because they will be sensitive to the far tails of the distribution of $\mathbf{s}$, which are not expected or required to be well approximated by a multivariate normal.

It is also helpful to see where the goodness-of-fit statistic $(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)^\top \hat{\Sigma}_\theta^{-1}(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)$ lies on the vertical axis of the plot described in step (i) (the $\mathbf{s}$ in this case is the

observed vector of statistics). Ideally it should lie in the region of the plot where the normality assumption is plausible. Lying too far into the upper tail of the plot would be a cause for concern: it might indicate lack of fit and suggest that the normality approximation may be poor. An observed value above the upper tail of the distribution of the simulated values would certainly indicate lack of fit. A *P* value could be calculated, for a formal fit test, but this provides less information than the plots.

**Maximum-likelihood estimation.** For many inferential purposes, the output of the MCMC chain used to investigate $l_s$ is sufficient, but it is sometimes desirable to find maximum-likelihood estimates and the associated asymptotic covariance matrix. Given the output of the chain, this is straightforward. As an example, consider a problem with two parameters, $\theta_1$ and $\theta_2$, and denote the output from the converged MCMC chain as $\theta_{11}, \theta_{12}, \theta_{13}, \ldots, \theta_{21}, \theta_{22}, \theta_{23}, \ldots$ and $l_{s1}, l_{s2}, l_{s3}, \ldots$.

A quadratic approximation to $l_s$ 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 $\alpha$ and the vector of coefficients $\boldsymbol{\beta}$. The resulting quadratic can then be maximized to find $\hat{\boldsymbol{\theta}}$, and the Hessian of $l_s$ can be computed directly from the estimates of $\boldsymbol{\beta}$. The approach given here generalizes trivially to any dimension for $\boldsymbol{\theta}$.

16. Takens, F. in *Dynamical Systems and Turbulence* (eds Rand, D. A. & Young, L.-S.) 366–381 (Lect. Notes Math. 898, Springer, 1981).
17. Cox, D. R. & Hinkley, D. V. *Theoretical Statistics* 11–58, 250–356 (Chapman & Hall, 1974).
18. Watkins, D. S. *Fundamentals of Matrix Computation* 128–132 (Wiley, 1991).