Statistical inference for noisy nonlinear ecological dynamic systems

2010-01-01186

Simon N. Wood s.wood@bath.ac.uk

Mathematical Sciences, University of Bath, Bath BA2 7AY U.K.

June 24, 2010

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. 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. Since the joint density is the basis for the fit measures used by all conventional statistical methods, 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 it is shown 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 about 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 covariance matrix of the statistics, given model parameters, allowing a synthetic likelihood to be constructed, which assesses model fit. This likelihood can be explored by a straightforward Markov Chain Monte Carlo sampler, but one further post-processing step returns pure likelihood based inference. The method is applied to finally establish the dynamic nature of the fluctuations in Nicholson's classic Blowfly experiments[1,2,3].

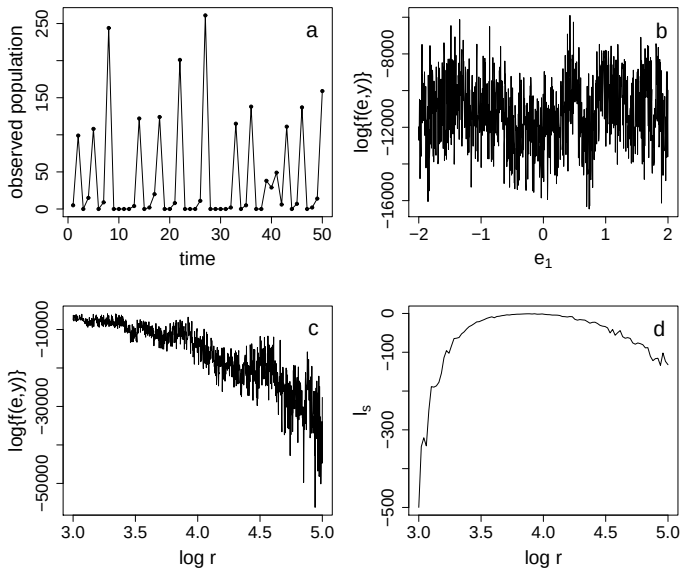

Figure 1: **Measuring fit of the Ricker model a**. Population data simulated from the Ricker model in the text, observed under Poisson sampling ($\log r = 3.8$, $\sigma = .3$, $\phi = 10$). **b**. The log joint probability density, $\log f_\theta(\mathbf{y}, \mathbf{e})$, of data $\mathbf{y}$ and random 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**. $\log f_\theta(\mathbf{y}, \mathbf{e})$ plotted against model parameter r, again with $\mathbf{e}$ and $\mathbf{y}$ held fixed. **d**. The log synthetic likelihood, l_s, against r for the Ricker model and the data given in panel a ($N_r = 500$).

The prototypic ecological model with complex dynamics is the scaled Ricker map[4], describing the time course of a population N_t by

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

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 (1) when $\log r = 3.8$, and what is observed are Poisson random deviates, y_t, with mean ϕN_t, reflecting a reasonably common sampling situation. Suppose that the aim is to make statistical inferences about $\boldsymbol{\theta}^\mathsf{T} = (r, \sigma_e^2, \phi)$ from this data series. Figures 1b and 1c illustrate the joint probability (density) function, $f_\theta(\mathbf{y}, \mathbf{e})$, of data vector, $\mathbf{y}$, and noise vector, $\mathbf{e}$, when the (fixed) noise realization and data from the 1a simulation are plugged in. 1c plots how $\log f_\theta$ varies with r, while 1b keeps r fixed but varies the first element of the noise realization, e_1. Likelihood based inference about $\boldsymbol{\theta}$ requires that we integrate f_θ over all $\mathbf{e}$, something which is analytically intractable, and from figure 1b, is clearly numerically intractable as well[5]. Bayesian inference would require that we sample replicate $\mathbf{e}, \boldsymbol{\theta}$ vectors from a density proportional to f_θ: no methods exist to do this

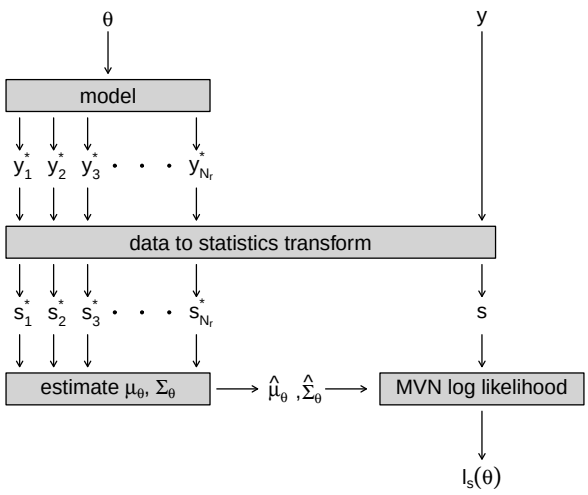

Figure 2: **Synthetic likelihood evaluation**. Starting at the top, we wish to evaluate the fit of the model with parameter vector $\boldsymbol{\theta}$ to the raw data vector $\mathbf{y}$. Replicate data vectors $\mathbf{y}_1^*, \ldots, \mathbf{y}_{N_r}^*$ are simulated from the model, given the value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

in a meaningful way for an f_θ as irregular as that shown in 1b and 1c.

The problem with the conventional approaches is really 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. While 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 which reflect what is dynamically important in the data, while discarding the details of local phase. In itself this idea is not new[4,6,7,8,9]. 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[5].

Defining fit. 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 capture the dynamic structure of the model (by specifying the components $\mathbf{s}$ we define *what* matters about the dynamics, but not *how much* it matters). Example 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 $\mathbf{s}$, supporting the key multivariate normality approximation,

$$\mathbf{s} \sim N(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta). \tag{2}$$

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

Evaluating the synthetic likelihood of $\boldsymbol{\theta}$. For a given value of parameter vector $\boldsymbol{\theta}$...

1. 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 $\mathbf{y}$ was converted to $\mathbf{s}$.

2. Evaluate $\hat{\boldsymbol{\mu}}_\theta = \sum_i \mathbf{s}_i^*/N_r$.

3. Setting $\mathbf{S} = (\mathbf{s}_1^* - \hat{\boldsymbol{\mu}}_\theta, \mathbf{s}_2^* - \hat{\boldsymbol{\mu}}_\theta, \ldots)$, then[10]

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

4. Dropping irrelevant constants, the log synthetic likelihood is

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

Like any likelihood, $l_s(\boldsymbol{\theta})$ measures the consistency of parameter values $\boldsymbol{\theta}$ with the observed data, but it is a *much* smoother function of $\boldsymbol{\theta}$ than f_θ, as 1d illustrates. Note that a robust estimator[10,11] can be advantageous at step 3.

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

Finding maximum likelihood estimates, $\hat{\boldsymbol{\theta}}$, by maximizing l_s with respect to $\boldsymbol{\theta}$, can not usually be achieved by numerical optimizers for smooth functions, as l_s usually displays some residual small scale roughness. Instead, a standard Metropolis Hastings Markov Chain Monte Carlo method[4,12] can be used for this purpose, and for simultaneously exploring the range of parameter values consistent with the data. See the methods section for implementational details, refinements, model checking and supporting theory.

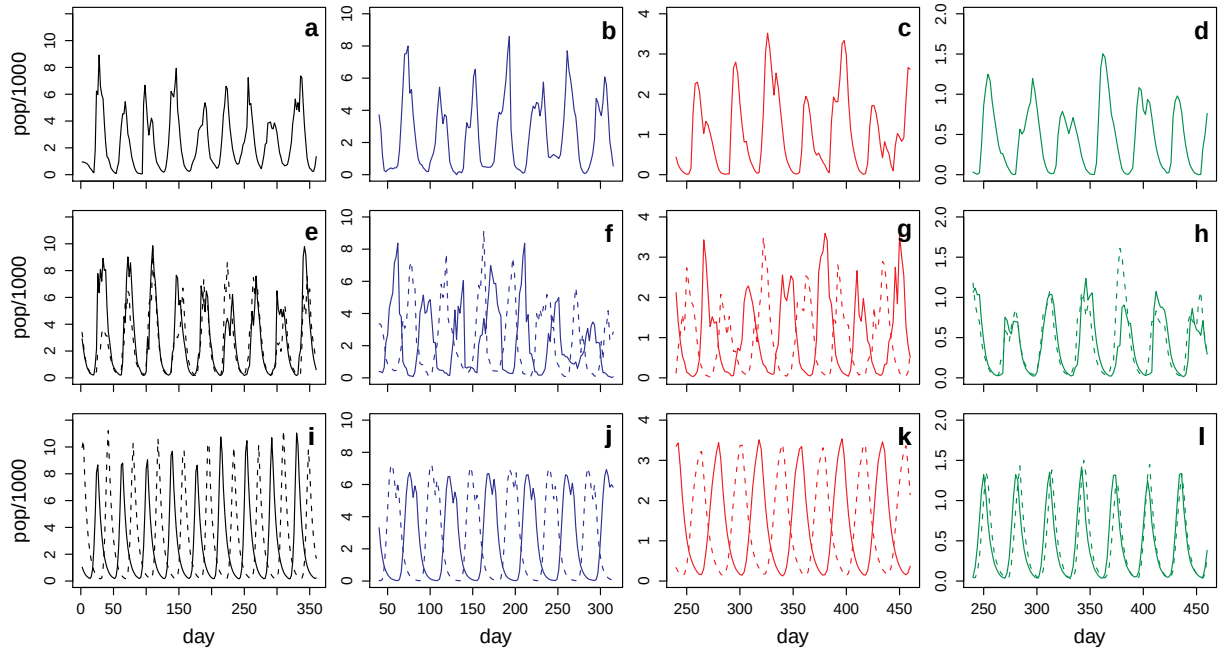

Figure 3: **Blowfly data and model runs**. **a,b**. Two laboratory adult populations of Sheep blowfly maintained under adult food limitation from Nicholson 1957[3] and 1954[2]. **c,d**. As a,b but maintained under moderate and more severe juvenile food limitation[2]. **e-h** each show two replicates from the full model (4) fitted separately to the data shown in each of panels a-d, immediately above. **i-l** as e-h for the demographic stochasticity only model. All observations are every second day. The simulation phase is arbitrary. Notice the qualitatively good match of the dynamics, e-h, of the full model (4) to the data, compared to the insufficiently variable dynamics of the demographic stochasticity only model, i-l.

To illustrate efficacy, the method was applied to the simulated Ricker data of figure 1a. Statistics were (i) the auto-covariances to lag 5, (ii) the coefficients of the cubic regression of the ordered differences, $y_t - y_{t-1}$, on their observed values, (iii) the coefficients of the autoregression $y_{t+1}^3 = \beta_1 y_t^3 + \beta_2 y_t^6 + \epsilon_t$, (iv) the mean population and (v) the number of zeroes observed. Resulting 95% confidence intervals were $3.6 < \log r < 4.2$, $.10 < \sigma < .55$ and $9.1 < \phi < 11.3$: all include the truth, but a simulation study suggested that such intervals achieve coverage probabilities of .84, .85 and .87 for sample size 50, achieving nominal .95 coverage for sample sizes 100-200.

As the main example, figures 3a-d show adult blowfly populations from four runs of Nicholson's classic experiments[2,3]. Three decades after the experiments, Gurney and Nisbet[1,13] provided the first plausible model for the last 3 replicates:

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

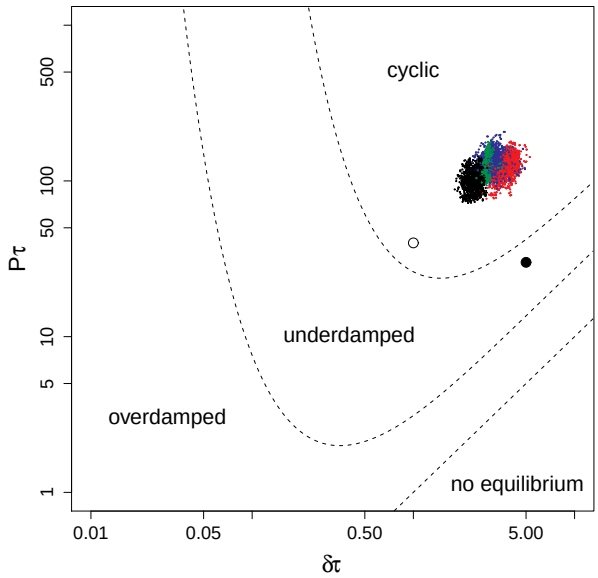

Figure 4: **Gurney and Nisbet's**[1,13] **blowfly model stability diagram**, with samples from the stability controlling parameter combinations, $\delta\tau$ and $P\tau$, plotted for each experimental run shown in figure 3. Colour coding matches figure 3. The open circle and black circle show stability properties for alternative chain starting conditions: they give indistinguishable results, although the black circle conditions lie in Gurney and Nisbet's plausible range for external noise driven dynamics. Clearly, the dynamics are limit cycles perturbed by noise, but are not noise *driven*. The fluctuations are driven by the intrinsic population dynamic processes, not by random variation exciting a resonance in otherwise stable dynamics.

where N is adult population and P, N_0, δ and τ are parameters. Depending on parameter values, this model displays dynamics ranging from stable equilibrium to chaos. Using careful, but ad hoc, parameter estimation methods Nisbet and Gurney concluded that the dynamics of the adult food limited replicate, 3b, are limit cycles: the fluctuations are intrinsic to the blowfly population biology. For the juvenile food limited replicates, 3c and 3d, it was not possible to decide conclusively whether the fluctuations were driven and initiated by random external forcing and/or demographic stochasticity (quasi-cycles), or were intrinsically driven limit cycles. To go further, requires a stochastic version of (3) that can produce both types of dynamics, and some means of estimating its parameters from data. Prior to the method proposed here, estimation was difficult because the dynamics of plausibly parameterized models can be chaotic or near chaotic[13].

(3) was discretized with a daily timestep. Births and death were subject to demographic stochasticity, with rates perturbed by daily environmental noise, yielding

$$N_{t+1} = R_t + S_t, \tag{4}$$

where recruitment is $R_t \sim \mathrm{Poi}\{PN_{t-\tau}\exp(-N_{t-\tau}/N_0)e_t\}$ and survival is $S_t \sim \mathrm{binom}\{\exp(-\delta\epsilon_t), N_t\}$. i.e. egg production is an independent Poisson process for each female and each adult has independent probability $\exp(-\delta\epsilon_t)$ of surviving each day. Environmental stochasticity terms e_t and ϵ_t are independent Gamma random deviates with mean 1 and variance σ_p^2 and σ_d^2, respectively. All parameters are positive and τ 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 $e_t = \epsilon_t = 1$ should be compared to (4) statistically.

(4) and its simplified version were fitted to each experimental replicate, with MCMC chains run for 50000 iterations. Example replicate simulations from the final chain states are shown in figure 3e-h for (4) and 3i-l for the simplified model. The χ^2 fit statistic suggested a good fit ($p > .2$) for (4) but a very bad fit ($p \ll .002$) for the simplified version, in all cases. AIC differences were > 1800 in favour of the full model (4) for all 4 replicates. Figure 3 suggests that the comprehensive rejection of the simplified model is because demographic stochasticity can not produce the irregularity of the real cycles.

So, the stochastic Nisbet and Gurney model (4), is not just qualitatively plausible, but actually fits the data quantitatively. Furthermore, uncontrolled variability in the experimental setup dwarfs demographic stochasticity, begging the question of whether it drove the fluctuations, rather than simply perturbing them? To answer this, figure 4 shows Gurney and Nisbet's stability diagram for (3), overlaid with 1500 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.

Clearly there is extremely strong statistical evidence that the Nicholson's 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 the first general purpose method for well founded statistical inference about noisy ecological (and other) dynamic models in the chaotic and near chaotic regimes.

Methods summary

Exploring l_s by MCMC. Starting from a parameter guess $\boldsymbol{\theta}^{[0]}$, iterate the following for $k = 1, 2, 3 \ldots$

1. Propose $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{[k-1]} + \boldsymbol{\delta}^{[k]}$, where $\boldsymbol{\delta}^{[k]}$ is a random vector from a convenient symmetric distribution.

2. Set $\boldsymbol{\theta}^{[k]} = \boldsymbol{\theta}^*$ with probability $\min[1, \exp\{l_s(\boldsymbol{\theta}^*) - l_s(\boldsymbol{\theta}^{[k-1]})\}]$, otherwise set $\boldsymbol{\theta}^{[k]} = \boldsymbol{\theta}^{[k-1]}$.

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[4,14]. In particular alternative models can be compared using AIC, or generalized likelihood ratio testing[5,7,14]. A useful model checking diagnostic is that $(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)^\mathsf{T} \hat{\boldsymbol{\Sigma}}_\theta^{-1} (\mathbf{s} - \boldsymbol{\mu}_\theta) \sim \chi^2_{\dim(\mathbf{s})}$ if the model fits.

Blowfly statistics. The auto-covariances to lag 11 and the difference distribution summary used in the Ricker example were used, 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 the supplementary material and online methods section.

Acknowledgments

I am grateful to Bill Gurney, Roger Nisbet, Steve Ellner, Peter Turchin and Bruce Kendall for many discussions of the problem addressed here, and to the participants in the 2009 *Statistical Methods for Dynamic Systems Models* in Vancouver, for discussion of this particular work. Also to an anonymous referee for suggesting the demographic stochasticity only blowfly model. The work is part of the research programme of the EPSRC/NERC funded UK National Centre for Statistical Ecology.

References

1. Gurney, W.S.C. and Nisbet, R.M. (1980) Nicholson's blowflies revisited. Nature 287, 17-21.

2. Nicholson, A.J. (1954) An outline of the dynamics of animal populations. Australian Journal of Zoology 2, 9-65.

3. Nicholson, A.J. (1957) The self-adjustment of populations to change. Cold Spring Harbor Symposium on Quantitative Biology 22, 153-173.

4. Turchin, P. (2003) Complex Population Dynamics. Princeton.

5. Davison, A.C. (2003) Statistical Models. Cambridge.

6. Kendall B.E., Briggs, C.J., Murdoch, W.W., Turchin, P., Ellner, S.P., McCauley, E., Nisbet, R.M., Wood, S.N. (1999) Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. Ecology 80, 1789-1805.

7. Cox, D.R. (2006) Principles of Statistical Inference. Cambridge.

8. Beaumont, M.A., Zhang W. and Balding D.J. (2002) Approximate Bayesian Computation in Population Genetics. Genetics, 162, 2025-2035

9. McFadden, D. (1989) A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica. 57, 995-1026.

10. Krzanowski, W.J. (1988) Principles of Multivariate Analysis: A User's Perspective. Oxford.

11. Campbell, N.A. (1980) Robust procedures in multivariate analysis. I: Robust covariance estimation. Journal of the Royal Statistical Society (C) 29, 231-237.

12. Gamerman, D. and Lopes, H.F. (2006) Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference (2nd ed.). CRC/ Chapman and Hall.

13. Nisbet, R.M. and Gurney, W.S.C. (1982) Modelling Fluctuating Populations. Wiley.

14. Silvey, S.D. (1975) Statistical Inference. Chapman & Hall.

Figure legends

1. **Measuring fit of the Ricker model a**. Population data simulated from the Ricker model in the text, observed under Poisson sampling ($\log r = 3.8$, $\sigma = .3$, $\phi = 10$). **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**. $\log f_\theta(\mathbf{y}, \mathbf{e})$ plotted against model parameter r, again with $\mathbf{e}$ and $\mathbf{y}$ held fixed. **d**. The log synthetic likelihood, l_s, against r for the Ricker model and the data given in panel a ($N_r = 500$).

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 $\mathbf{y}$. 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 $\mathbf{s}$, in the same way. The $\mathbf{s}_i^*$ are used to estimate the mean vector, $\hat{\boldsymbol{\mu}}_\theta$, and covariance matrix, $\hat{\boldsymbol{\Sigma}}_\theta$, of $\mathbf{s}$, according to the model with parameters $\boldsymbol{\theta}$. $\hat{\boldsymbol{\mu}}_\theta$, $\hat{\boldsymbol{\Sigma}}_\theta$ and $\mathbf{s}$ are used as respectively the mean vector, covariance matrix and argument of the log multivariate normal probability density function, to evaluate the log synthetic likelihood, l_s.

3. **Blowfly data and model runs**. **a,b**. Two laboratory adult populations of Sheep blowfly maintained under adult food limitation from Nicholson 1957[3] and 1954[2]. **c,d**. As a,b but maintained under moderate and more severe juvenile food limitation[2]. **e-h** each show two replicates from the full model (4) fitted separately to the data shown in each of panels a-d, immediately above. **i-l** as e-h for the demographic stochasticity only model. All observations are every second day. The simulation phase is arbitrary. Notice the qualitatively good match of the dynamics, e-h, of the full model (4) to the data, compared to the insufficiently variable dynamics of the demographic stochasticity only model, i-l.

4. **Gurney and Nisbet's[1,13] blowfly model stability diagram**, with samples from the stability controlling parameter combinations, $\delta\tau$ and $P\tau$, plotted for each experimental run shown in figure 3. Colour coding matches figure 3. The open circle and black circle show stability properties for alternative chain starting conditions: they give indistinguishable results, although the black circle conditions lie in Gurney and Nisbet's plausible range for external noise driven dynamics. Clearly, the dynamics are limit cycles perturbed by noise, but are not noise *driven*. The fluctuations are driven by the intrinsic population dynamic processes, not by random variation exciting a resonance in otherwise stable dynamics.

Methods

This section includes further method details, refinements and theory. See the supplementary material for MCMC details, example specific details and further examples.

Choosing statistics

The statistics, $\mathbf{s}$, play the same role in the synthetic likelihood, l_s, that the raw data would play 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 which 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 via 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 next timestep (or the change in state) from the state now, and possibly the state at earlier times. Similarly, auto-regression 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 $y_t^\alpha = \beta_0 + \beta_1 y_{t-1}^\alpha + \beta_2 y_{t-1}^{2\alpha} + \beta_3 y_{t-1}^{3\alpha} + \epsilon_t$, might capture a good deal of the dynamics in the Ricker data (α 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 auto-regressions will often need to be replaced by auto-regressions on multiple lagged states (by Taken's theorem[15]).

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.

Note that sometimes the selection of statistics will be iterative. i.e. 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.

While 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.

Theoretical properties of the method

This section provides some theoretical investigation of the multivariate normality approximation, and of l_s itself. The section provides theoretical justification for the approach described in the paper.

The multivariate normality approximation

The method employs the approximation

$$\mathbf{s} \sim N(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta) \tag{5}$$

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

$$\log f_\theta(\mathbf{s}) \simeq \log f_\theta(\boldsymbol{\mu}_\theta) + (\mathbf{s} - \boldsymbol{\mu}_\theta)^\mathsf{T} \frac{\partial \log f_\theta}{\partial \mathbf{s}} + \frac{1}{2}(\mathbf{s} - \boldsymbol{\mu}_\theta)^\mathsf{T} \frac{\partial^2 \log f_\theta}{\partial \mathbf{s} \partial \mathbf{s}^\mathsf{T}} (\mathbf{s} - \boldsymbol{\mu}_\theta) \tag{6}$$

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

$$f_\theta(\mathbf{s}) \simeq k \exp\left\{ -\frac{1}{2}(\mathbf{s} - \boldsymbol{\mu}_\theta)^\mathsf{T} \left(-\frac{\partial^2 \log f_\theta}{\partial \mathbf{s} \partial \mathbf{s}^\mathsf{T}} \right) (\mathbf{s} - \boldsymbol{\mu}_\theta) \right\}$$

where k is a constant of proportionality. It is immediately recognizable that if the r.h.s is to be a p.d.f then it is the multivariate normal p.d.f., with covariance matrix $\boldsymbol{\Sigma}_\theta = \left(-\partial^2 \log f_\theta / \partial \mathbf{s} \partial \mathbf{s}^\mathsf{T} \right)^{-1}$. This approximation is familiar in statistics from the Laplace approximation of integrals. So, for $\mathbf{s}$ sufficiently close to $\boldsymbol{\mu}_\theta$ we expect f_θ to be well approximated by the probability density function of $N(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$. For a good model with plausible parameter estimates, $\mathbf{s}$ should be close to $\boldsymbol{\mu}_\theta$, with proximity increasing with increasing raw sample size, n.

Now, without knowledge of f_θ itself, $\boldsymbol{\mu}_\theta$ and $\boldsymbol{\Sigma}_\theta$ are unknown. However they can be estimated from a sample of $\mathbf{s}$ vectors produced by simulation, as in the paper. Given the Taylor series truncation used to obtain (5), we do not

necessarily expect it to be a good approximation in the tails of the distribution of $\mathbf{s}$. Hence it may be advantageous to use robust estimators for $\boldsymbol{\mu}_\theta$ and $\boldsymbol{\Sigma}_\theta$, which down-weight $\mathbf{s}$ vectors from the tails of f_θ. Such a procedure is described in section 'Robust covariance estimation', below.

Properties of the log synthetic likelihood, l_s

Given (5), consider the properties of l_s, itself. Firstly, $\mathbf{s}$ is observed data, just as the raw data, $\mathbf{y}$, is, and is therefore an equally valid basis for forming a likelihood. Hence given the approximation (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[5,14,15] 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 rather 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 $l_s \to \mathbb{E}(l_s)$ as $n \to \infty$ (see e.g. Silvey, 1975, Section 4.5). Let $\boldsymbol{\theta}_0$ denote the true value of $\boldsymbol{\theta}$, so that $\boldsymbol{\mu}_{\theta_0}$ is the corresponding true mean vector of $\mathbf{s}$. Let $N_r \to \infty$. Define $\boldsymbol{\epsilon} = \mathbf{s} - \boldsymbol{\mu}_{\theta_0}$ and make the mild assumption that the elements of $\mathbf{s}$ are such that $\boldsymbol{\epsilon} \to \mathbf{0}$ and $|\boldsymbol{\Sigma}_\theta| \to 0$ as $n \to \infty$. Then l_s can be re-written as

$$\begin{aligned} l_s(\boldsymbol{\theta}) &= -\frac{1}{2}(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta + \boldsymbol{\epsilon})^\mathsf{T}\boldsymbol{\Sigma}_\theta^{-1}(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta + \boldsymbol{\epsilon}) - \frac{1}{2}\log|\boldsymbol{\Sigma}_\theta| \\ &= -\frac{1}{2}(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)^\mathsf{T}\boldsymbol{\Sigma}_\theta^{-1}(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta) - \boldsymbol{\epsilon}^\mathsf{T}\boldsymbol{\Sigma}_\theta^{-1}(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta) - \frac{1}{2}\boldsymbol{\epsilon}^\mathsf{T}\boldsymbol{\Sigma}_\theta^{-1}\boldsymbol{\epsilon} - \frac{1}{2}\log|\boldsymbol{\Sigma}_\theta| \\ &= -\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_\theta^{-1}\left\{(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)^\mathsf{T} + 2(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)\boldsymbol{\epsilon}^\mathsf{T} + \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\mathsf{T}\right\}\right] - \frac{1}{2}\log|\boldsymbol{\Sigma}_\theta| \end{aligned}$$

For $\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta \neq 0$ then as $n \to \infty$, $\boldsymbol{\epsilon} \to \mathbf{0}$ and the term in $\{\cdot\}$ is dominated by $(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)(\boldsymbol{\mu}_{\theta_0} - \boldsymbol{\mu}_\theta)^\mathsf{T}$, so $l_s \to \mathbb{E}(l_s)$ when $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, both $\mathbb{E}(l_s)$ and $l_s \to -\frac{1}{2}\log|\boldsymbol{\Sigma}_\theta|$.

Let $\hat{\boldsymbol{\theta}}$ be the maximizer of l_s. Standard theory (e.g. Silvey[14], section 4.4) establishes that $\mathbb{E}(l_s)$ is maximized at $\boldsymbol{\theta}_0$. So by the asymptotic convergence of l_s to $\mathbb{E}(l_s)$, just established, $\hat{\boldsymbol{\theta}}$ is a consistent estimator.

Turning to the large sample distribution of the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, the usual Taylor expansion argument (e.g. Silvey[14] Section 4.6) gives:

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \simeq -\left(\frac{\partial^2 l_s}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\mathsf{T}}\right)\frac{\partial l_s}{\partial\boldsymbol{\theta}}. \tag{7}$$

By standard properties of the expected log likelihood $\mathbb{E}(\partial l_s/\partial\boldsymbol{\theta}) = \mathbf{0}$ while $\mathrm{cov}(\partial l_s/\partial\boldsymbol{\theta}) = \partial^2 l_s/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\mathsf{T}$. Since l_s converges to its expected value, then in the limit $n \to \infty$, we have that $\hat{\boldsymbol{\theta}}$ has expectation $\boldsymbol{\theta}_0$ and covariance matrix $(\partial^2 l_s/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\mathsf{T})^{-1}$.

Now, if $\partial l_s/\partial\boldsymbol{\theta}$ had a multivariate normal distribution, then, by 7, so would $\hat{\boldsymbol{\theta}}$, at least asymptotically. So,

consider

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

where $\mathbf{z} = \boldsymbol{\Sigma}_\theta^{-1/2}(\mathbf{s}-\boldsymbol{\mu}_\theta)$ is a normal random vector, asymptotically $\sim N(\mathbf{0},\mathbf{I})$. The first term on the r.h.s. 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 (i) $\boldsymbol{\Sigma}_\theta$ depends only weakly on θ_k so that $\partial\boldsymbol{\Sigma}_\theta/\partial\theta_k$ is close to the zero matrix and the second term on the r.h.s., above, is negligible, or (ii) if $\dim(\mathbf{z}) = \dim(\mathbf{s}) \to \infty$ so that the second term, above, tends to a normally distributed random variable, by the central limit theorem.

So we have the rather weak result that $\hat{\boldsymbol{\theta}}$ will only have a normal distribution if $\boldsymbol{\Sigma}_\theta$ 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 $(\partial^2 l_s/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\mathsf{T})^{-1}$. The estimators will not be fully asymptotically efficient, and are not guaranteed to be normally distributed.

Method refinements

This section introduces various useful refinements of the basic method described in the paper.

Robust covariance estimation

The method often produces perfectly reasonable results using the straightforward estimate, $\hat{\boldsymbol{\Sigma}}_\theta$, given in the paper. However, the argument of the above section on 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 $\mathbf{S}$ and $\hat{\boldsymbol{\mu}}_\theta$ be as in the paper, with $\mathbf{S}$ an $N_s \times N_r$ matrix.

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

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

   The initial estimate of $\boldsymbol{\Sigma}_\theta$ is $\bar{\boldsymbol{\Sigma}}_\theta = \bar{\mathbf{D}}\bar{\mathbf{R}}^\mathsf{T}\bar{\mathbf{R}}\bar{\mathbf{D}}$, while $\bar{\boldsymbol{\Sigma}}_\theta^{-1} = \bar{\mathbf{D}}^{-1}\bar{\mathbf{R}}^{-1}\bar{\mathbf{R}}^{-\mathsf{T}}\bar{\mathbf{D}}^{-1}$. The use of pre-conditioning matrix $\bar{\mathbf{D}}$ ensures that $\bar{\mathbf{R}}$ has full numerical rank (and low condition number) so that stable computation of $\bar{\mathbf{R}}^{-1}$ is possible.

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

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

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

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

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

4. Redefine and recompute

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

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

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

$$\mathbf{QR} = \mathbf{W}\mathbf{S}^\mathsf{T}\mathbf{D}^{-1} / \sqrt{\sum w_j^2 - 1}.$$

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

The re-weighting via the w_i is Campbell's method[11], as described on p231-235 of Krzanowski[10]. It down-weights extreme tail observations to ensure statistical robustness. The use of $\mathbf{D}$ is standard numerical pre-conditioning to ensure numerical robustness (see e.g. Watkins[17] section 2.9). 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 fit 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 doesn't 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 which are pronounced enough that the chain can become stuck in them. A solution is to modify l_s in order to attenuate the tail behaviour.

For example, first define

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

where $k = 2d_0^{2-\gamma}/\gamma$ and $c = d_0^2 - kd_0^\gamma$. $g(x)$ 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

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

where d_0 is the 99th percentile of the $\chi^2_{\dim(\mathbf{s})}$ distribution, for example. $\gamma = 0.1$ is quite effective. $\tilde{l}_s$ is exactly l_s for $\mathbf{s}$ not in the far tails of the distribution — i.e. 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 itself. Using $\tilde{l}_s$ to determine the acceptance probability, α, in the MCMC chain avoids the chain becoming stuck. l_s itself is still computed at each step, to be used for inference, of course. This robust approach is used for the demographic stochasticity blowfly model and in the model comparison example in the supplementary material.

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

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

where γ is a small positive constant less than 1. However, unlike the use of $\tilde{l}_s$, this is inefficient when the model actually fits well.

Transforming to improve normality of 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 $\boldsymbol{\Sigma}_\theta$. None the less, the closer $\mathbf{s}$ is to multivariate normal the better approximation (5) will be (meaning that it will apply for $\mathbf{s}$ further from $\boldsymbol{\mu}_\theta$). In particular the first neglected term in the expansion yielding (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 Krzanowski[10], section 7.6, for a general discussion of such transformation. Here one very simple approach is presented which focusses on improving marginal normality.

1. Run a pilot MCMC chain, using untransformed statistics, to obtain an estimate $\tilde{\boldsymbol{\theta}}$ close to the MLE.

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

3. 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.

4. Run the full model estimation and inference process with all statistics (observed and simulated) transformed using the piecewise linear transforms from step 3.

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 *require* the transformation step.

Checking the normality assumption and goodness of fit

Several checks of the normality assumption, (5), are useful.

1. Plot the N_r ordered values of $(\mathbf{s}_j^* - \hat{\boldsymbol{\mu}}_\theta)^\mathsf{T}\hat{\boldsymbol{\Sigma}}_\theta^{-1}(\mathbf{s}_j^* - \hat{\boldsymbol{\mu}}_\theta)$ 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 straightline relationship with slope 1 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.

2. Produce normal QQ-plots for each statistic, s_i, using N_r replicates produced for a $\boldsymbol{\theta}$ near the MLE. This checks the marginal normality of the statistics, under the model.

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

See Krzanowski[10] for further discussion. Note that formal tests of normality are not useful here. The dimension of $\mathbf{s}$ is usually far 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 $(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)^\mathsf{T}\hat{\boldsymbol{\Sigma}}_\theta^{-1}(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)$ lies on the vertical axis of plot 1 (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. 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 of be calculated, for a formal fit test, but this provides less information than the plots.

See the supplementary material for examples of the 3 plots discussed here.

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 θ_1 and θ_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, i.e. 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 $\boldsymbol{\beta}$. The resulting quadratic can then be maximized to find $\hat{\boldsymbol{\theta}}$, while the hessian of l_s can be computed directly from the estimates of $\boldsymbol{\beta}$. Clearly the approach given here generalizes trivially to any dimension for $\boldsymbol{\theta}$.

References

15. Takens, F. (1981). ”Detecting strange attractors in turbulence”. in D. A. Rand and L.-S. Young. Dynamical Systems and Turbulence, Lecture Notes in Mathematics, vol. 898. Springer-Verlag. pp. 366381.

16. Cox, D.R. and Hinkley, D.V. (1974) Theoretical Statistics. Chapman & Hall.

17. Watkins, D.S. (1991). Fundamentals of Matrix Computation. Wiley, New York.