Supplementary Material for Statistical inference for noisy nonlinear ecological dynamic systems

Simon N. Wood
Mathematical Sciences, University of Bath, Bath BA2 7AY U.K.
s.wood@bath.ac.uk
June 24, 2010

Contents

1 Method Implementation and MCMC output 1
   1.1 Ricker model ................................................................. 2
   1.2 Blowfly models ............................................................ 3
      1.2.1 Demographic stochasticity only .................................. 4
      1.2.2 Alternative blowfly model ....................................... 4
      1.2.3 A better alternative model ...................................... 5

2 Further examples 6
   2.1 Bupalus piniaria in Tentsmuir forest ............................... 6
   2.2 An example with an absorbing state .................................. 8
   2.3 A model cross comparison example ................................... 9

3 Software: the \texttt{sl} package for R 10

1 Method Implementation and MCMC output

This section provides further detail for the examples reported in the paper. The MCMC chains were coded in R 2.10.0 (R Development Core Team, 2009). The major computational cost is in running model replicates, and converting the resulting simulated raw data to statistics: these tasks used compiled C code, called from R. The MCMC simulations used the simplest possible Metropolis Hastings update (see e.g. Davison, 2003). That is...

1. Proposals were of the form $\theta^{[k+1]} = \theta^{[k]} + \delta$ where the $\delta_i$ were usually independent Gaussian deviates, each with fixed standard deviation. For parameter $\tau$, in the blowfly example, $\delta_i$ was -1, 0 or 1 with probability .25, .5 and .25, respectively.

2. Proposals were accepted with probability:

$$\alpha = \begin{cases} 
0 & \text{parameters out of bounds} \\
\min \left[ 1, \exp \left\{ l_\ell (\theta^{[k+1]}) - l_\ell (\theta^{[k]}) \right\} \right] & \text{parameters within bounds}
\end{cases}$$

The ‘bounds’ were simple upper and lower limits on the parameters. The only bounds ever reached were the upper bound for $K$ and the lower bound for $\sigma$, in the Bupalus example in section 2.1. Because almost all parameters were intrinsically positive, the chains were run on log-parameters, with the exception that $\tau$, in the blowfly model, was used untransformed. The results reported here all use the robust covariance estimation approach of the methods section. Doing so makes little quantitative difference to the results, relative to using the simple estimate given in the paper, but it does lead to better mixing of the blowfly chains.
The smoothness of \( l_s \) can generally be improved by always using the same set of uniform random deviates to drive the simulation noise process, for every trial value of \( \theta \), i.e. the same driving deviates are used for each evaluation of \( l_s \) (the deviates of course vary between the \( N_r \) individual replicates generated as part of evaluating \( l_s \)). This trick was used in all the examples presented in this section, except for the demographic stochasticity models for blowflies, where it simply slows computation without improving mixing.

For all the examples presented here, the population models were run for 50-100 ‘burn-in’ steps to ensure that any transient behaviour was lost from the system. This is appropriate for the given data, but is not a universal requirement: some data contain initial transients which should be captured by the model, and in such cases there should be no burn-in period.

### 1.1 Ricker model

The proposal standard deviations for the Ricker model chain were \( .02 \), \( .1 \) and \( .05 \) for the log of \( r \), \( \sigma \) and \( \phi \), respectively. A typical run of this chain is shown here.

![Graphs showing mixing for log r, \( \sigma \), and \( \phi \)]

Note the rapid mixing, despite the severe non-linearity of the dynamics in this case. The acceptance rate was 22%. The confidence interval testing simulation study, reported in the paper, simply ran such chains for 200 replicate datasets, at each of sample sizes of \( n = 50, 100 \) and \( 200 \).

The statistics used for the Ricker model were: (i) The coefficients of the autocovariance function to lag 5. (ii) The coefficients, \( \beta_1 \) and \( \beta_2 \) of the autoregression \( y_{t+1}^3 = \beta_1 y_t^3 + \beta_2 y_t^6 \), where the \( y^3 \) values were centered by subtraction of the mean. This autoregression was suggested by exploratory data analysis, and the structure of the Ricker model. (iii) The coefficients of the cubic regression of the ordered first differences of the simulated or actual observed data on the ordered first differences of the actual observed data. These coefficients summarize the shape of the distribution of the differences in the series. (iv) The mean of the data and (v) the number of zeroes observed in the data. There are 14 statistics in all. The statistics were used in transformed form, as described in the methods section, although this makes little difference to the results.

The figure below shows the normality checking plots described in the methods section: using the numbering from that section, left is plot 1, middle is plot 2 and right is plot 3. Plot 1 shows that \( s \sim N(\mu_\theta, \Sigma_\theta) \) is a good approximation up until the far tails of the distribution. The dashed horizontal line gives the value of the Mahalonobis distance for the observed data: clearly this is well within the region for which the normal approximation is good, and its low value indicates a good fit. Plot 2 shows normal QQ plots for each of the simulated statistics marginally, standardized to have unit standard deviation: again the only departures from normality are in the far (1%) tails. Figure 3 shows a normal QQ plot for the standardized observed statistics: the plot shows no evidence for a problem with the normality approximation.
1.2 Blowfly models

The standard deviation of the proposals for the blowfly chains were .05 for all logged parameters. A chain run for the full model for the first replicate shown in figure 3 of the paper is shown.

The chain appears to converge rapidly, but again does not mix as well as the Ricker example (acceptance rates were 25-35% for the blowfly datasets). $P$ and $N_0$ show fairly strong negative correlation.

The statistics used were: (i) the coefficients of the autocovariance function to lag 11; (ii) the coefficients of the polynomial autoregression given in the paper: this is a crude approximation to the dynamic model itself, so that the regression coefficients, $\beta$, should capture information about the dynamic structure of the model and data series, if the model is any good; (iii) the mean of the data and the difference between the mean and the median; (iv) the ordered first differences regression coefficients as used in the Ricker example; (v) the number of turning points observed in the data. This final statistic was used for the stability plot shown in the paper, and for the replicate full model runs shown in the paper; it was not used for the comparison with the demographic stochasticity only blowfly model, in an attempt to give that model some chance of competing with the full model. There are 23 statistics in total, and they were used untransformed. This was justified by the checking plots below.
Clearly the normality assumption is plausible here, and there is a good fit to the statistics.

### 1.2.1 Demographic stochasticity only

Mixing was slower for the demographic stochasticity version of the blowfly model, and the robust log synthetic likelihood of the methods section was used to drive the MCMC chain, to improve this. The chain for replicate d from the paper is shown: this is the replicate with the lowest population, and hence greatest scope for a good fit to be achieved with only demographic stochasticity.

The acceptance rate was about 6%, and the slow mixing and slow convergence was typical of the fits with this model. The corresponding checking plots are shown here.

Note that the observed value for \((s - \hat{\mu}_\theta) \Sigma_\theta^{-1} (s - \hat{\mu}_\theta)\), lies well above the upper limit of the y axis in the left hand plot, emphasizing the poor fit of the demographic stochasticity only model.

### 1.2.2 Alternative blowfly model

An obvious question is the extent to which the Blowfly stability conclusions (from the full model) are contingent on the particular discretization and noise process used in the model? It would be undesirable for the results to
depend sensitively on such details, and some investigation of the sensitivity is worthwhile. An alternative model using a different (if less natural) daily discretization and an autoregressive noise model is:

\[ N_{t+1} = N_t \exp\{PN_{t-\tau} \exp(-N_{t-\tau}/N_0)/N_t - \delta + \epsilon_t \} \]

where \( \epsilon_t = \epsilon_{t-1} + (1 - \phi)\epsilon_t \) and \( \epsilon_t \sim N(0, \sigma^2) \). All parameters are strictly positive and additionally \( \tau \) is integer and \( \phi < 1 \). This model was also estimated for the 4 blowfly replicates. There appears to be almost no information on the \( \phi \) parameter, and the chains for this model mix more slowly than for the one given in the paper, but the conclusions are very similar. The equivalent of figure 4 in the paper is shown below, and indicates a certain robustness of the conclusions to model details.

1.2.3 A better alternative model

Given the paper’s conclusion that demographic stochasticity is really not important, then the most natural model might be

\[ N_{t+1} = P N_{t-\tau} \exp(-N_{t-\tau}/N_0)\epsilon_t + N_t \exp(-\delta \epsilon_t) \]

(1)

where \( \epsilon_t \) and \( \epsilon_t \) are independent Gamma random deviates with mean 1 and variance \( \sigma^2 \) and \( \sigma^2 \), respectively. All parameters are positive and \( \tau \) is integer. \( \epsilon_t \) and \( \epsilon_t \) both model the combined effect of demographic and environmental stochasticity, in proportions dependent on \( \sigma^2 \) and \( \sigma^2 \). My view is that this is the most sensible model. It results in the following stability figure.
The equivalent of figure 3 in the paper is shown here.

2 Further examples

2.1 Bupalus pineria in Tentsmuir forest

This section provides an additional real data example, illustrating the application of the method to a system with a hidden state variable, and both process and measurement error present. Figure 1a perfectly illustrates the paradox of statistics for ecological dynamics. The 40 year sampling effort required to generate these annual survey data on pupal densities of Bupalus pineria in Tentsmuir forest Fife (see Broekhuizen et al. 1993, 1994, Kendall et al. 2005), results in a timeseries which, in statistical terms, is short and very noisy. One model for these data is that
Figure 1: *Bupalus data and estimated model runs*. 

a. *Bupalus* population observed annually in Tentsmuir forest. 
b-d. Replicate simulations of the population in a, from the estimated *Bupalus* model in the text.

The apparent cycles are driven by a host-parasitoid interaction of the form

\[ H_{t+1} = H_t \exp\{r_0(1 - H_t/K) + \epsilon_t\} \exp(-\alpha P_t^\beta) \]
\[ P_{t+1} = H_t \exp\{r_0(1 - H_t/K) + \epsilon_t\} \{1 - \exp(-\alpha P_t^\beta)\} \]

where the \( \epsilon_t \) are i.i.d \( N(0, \sigma_e^2) \), and the observed pupal densities are described by \( Y_t \sim N(H_t^{25}, \sigma^2) \). Inference about the parameters, \( r_0, K, \alpha, \beta, \sigma_e^2 \) and \( \sigma^2 \) can be conducted using the approach detailed in the paper, and above. The statistics used here were: (i) the coefficients of the autocovariance function to lag 15; (ii) the coefficients of the same sort of cubic regression of ordered differences described for the Ricker model; (iii) the mean of the observed data. There were therefore 20 statistics in total. The transformations described in the methods section were employed.

The standard deviation of the proposal for the Bupalus chain was .05 for all (log) parameters. A typical MCMC run is shown here.

Mixing is not as good as in the Ricker case (although the acceptance rate was 31%), with correlations between \( \beta \) and several other parameters, notably \( K \). In fact it appears that we can only really bound \( K \) from below, which
is the reason that only a one sided interval for $K$ is given below. Essentially the system appears to be under sufficiently tight parasitic control that there is no information on what value the parasitoid-free carrying capacity might take. The measurement error term also mixes somewhat slowly, but does appear to be identifiable.

The normality assumption and goodness of fit plots (see methods section), are as follows.

The model fits well (left), marginally the normality assumption is fine (middle) and the observed statistics are consistent with multivariate normality (right). However, from the left plot, the multivariate normal approximation is slightly out here: it seems that slightly more aggressive downweighting of the tail observations might have been beneficial in the calculation of $\Sigma_\theta$.

Figures 1b-d show typical replicates generated from the fitted model. 95% CIs were $1.7 < r_0 < 2.5, .43 < \alpha < .74, .43 < \beta < .73, .20 < \sigma_e < .66, .05 < \sigma < .20$ and $K > 19$. Clearly we can now make well founded inferences about $Bupalus$ dynamics, whether or not the true state of nature is (near) chaotic.

### 2.2 An example with an absorbing state

The wide applicability of the normality approximation, implied by the methods section, can be well illustrated by considering a version of the Ricker model where the stochasticity is entirely demographic, and the carrying capacity is low, so that extinction of replicate populations is not a rare event. In particular, consider a population with discrete generations and no carry-over of individuals from one generation to the next. Suppose that the number of individuals of generation $t + 1$, produced by an individual of generation $t$, is a Poisson random variable, with mean $r e^{-\alpha N_t}$, where $N_t$ is the size of the population at time $t$. Then we have

$$N_{t+1} \sim \text{Poi}(r N_t e^{-\alpha N_t}).$$

The data shown at top right below were simulated from such a model with $\log r = 3.8$ and $\alpha = 0.18$ (starting from population 10, with a 20 generation ‘burn-in’, prior to time 0).
Approximately 20% of replicates of such simulations become extinct within the period of the data. This tendency to extinction will of course lead to a high number of runs of zeroes in the replicates, \( y^*_j \), simulated as part of the method: At first sight this would appear to present a severe challenge to the approximation that \( s \) should have a multivariate normal distribution. However the complete freedom to choose and transform statistics allows any difficulty to be avoided. In the case of extinction, the simplest device is to qualify each statistic with ‘computed on the data up to the earliest of the time of extinction or the end of the data’. (Many systems with multiple absorbing states can be handled in a similar way, but with large numbers of such states a different tactic would be needed).

The remaining panels, above, show the progress of the MCMC chain for this model. Clearly mixing is good, and the overall acceptance rate was 23%. In the case shown, the statistics were transformed to approximate normality as described in the methods section, although results are very similar if no transformation is used. The statistics used were: the coefficients of the autocovariance function to lag 10; the coefficients of a quadratic autoregression through the origin on the centred data; the mean and the difference between mean and median.

The figure below shows the normality checking plots: From the left plot it is clear that the multivariate normality approximation is plausible except in the far upper tails of the distribution, and the model fit is good. The middle plot shows that the marginal distributions are all well approximated by normals for the middle 98% of the distributions, while the right plot shows no reason to doubt the normality approximation for the statistics of the fitted data itself.

Finally note that the MLEs were \( \log \hat{r} = 3.92 \) and \( \hat{\alpha} = 0.19 \). 95% CIs were \( 3.2 < \log r < 4.8 \) and \( 0.16 < \alpha < 0.21 \), based on large sample likelihood approximation. Based on the quantiles of the converged chains, 95% CIs are the same except that the upper limit on \( \log r \) is 4.6.

### 2.3 A model cross comparison example

This section provides a cross fitting example, as an illustration of the effectiveness of model comparison based on the new method. Consider the two replicate simulated datasets shown below.

The left hand figure was generated by the Ricker model discussed in the paper, but with \( \log r = 3.3 \). The right hand figure is generated by a stochastic version of the discrete logistic model:

\[
N_{t+1} = r_t N_t (1 - N_t), \quad r_t \sim U(\beta - \alpha, \beta), \quad Y_t \sim \text{Poi}(\phi N_t).
\]
where $\alpha < \beta < 4$. In the simulation $\beta = 3.2$, $\alpha = 0.2$ and $\phi = 100$.

Synthetic likelihoods were constructed for both models using the same statistics as in the Ricker example in the paper, except for the count of the number of zeroes, which was omitted (there being no zero counts in the logistic data). Both models were then fitted to both replicates and the maximized $l_s$ was estimated by the quadratic regression method described in the methods section, from the MCMC output. Using the robustified $\tilde{l}_s$ from the methods section to drive the chains, all chains mixed well, with acceptance rates in the range 29%-42%. The normality diagnostics suggested that the normality approximation was good in all cases. Goodness of fit was reasonable for all cases except the fit of the logistic model to the Ricker data, where fit was very poor.

Since both models have 3 parameters, AIC based model selection amounts to choosing the model with the highest maximised likelihood, $\hat{l}_s$, for each dataset. For the Ricker data $\hat{l}_s$ (s.e. in brackets) was -0.7(.1) for the Ricker model and -243(1) for the logistic model. For the logistic data, $\hat{l}_s$ was 4.624(.005) for the Ricker model and 4.963(.004) for the logistic model. So in this case the correct model was selected for both series. Clearly the logistic model could not fit the Ricker data adequately, while the Ricker model did a more reasonable job of fitting the logistic data, albeit not doing quite as well as the logistic itself.

3 Software: the sl package for R

The accompanying R package, sl, provides software implementing all the examples in the paper and supplementary material, as well as functions for fast computation of summary statistics, transformation of statistics and robust computation of the synthetic likelihood. It requires the R language and environment to run, which is available from cran.r-project.org. Package installation instructions are also provided there. Once the package is installed, open its help pages using html help for further information on running the examples.

References