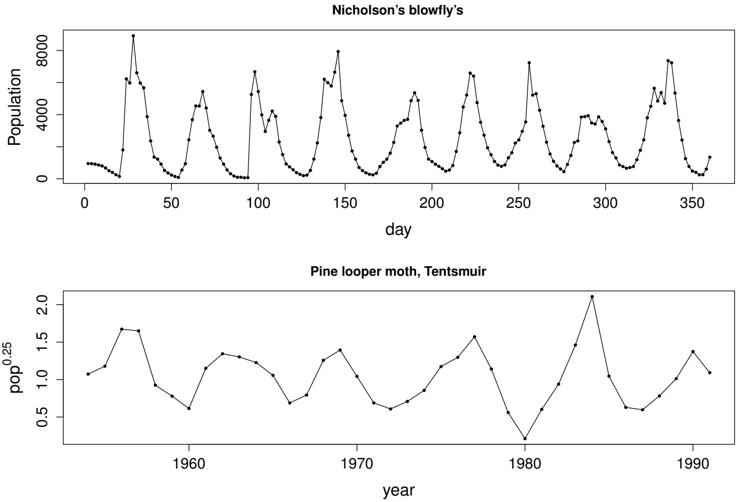
# Simple statistical methods for complex ecological dynamics

**Simon Wood** & **Matteo Fasiolo**
Mathematical Sciences, University of Bath, U.K.
National Centre for Statistical Ecology EPSRC

# Two ecological timeseries

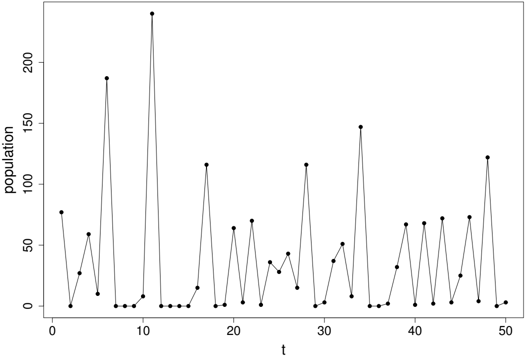


# Two modelling issues

It is appealing to make statistical inferences about mechanism based models of ecological timeseries, but…

- Many scientifically useful ecological dynamic models do not try to get every feature of the data right.
- The population dynamics of small animals (insects, rodents, etc.) can be highly non-linear.

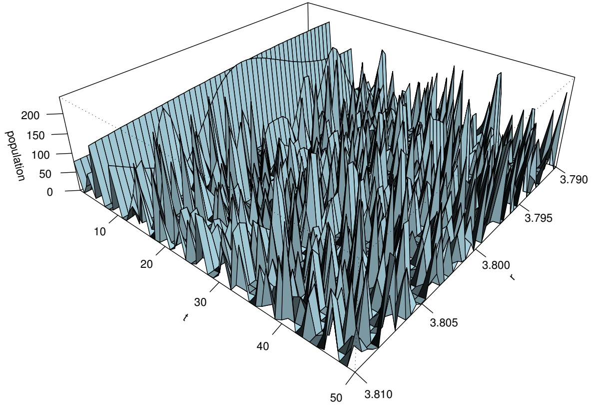
# Simulated Ricker Model Data (toy)



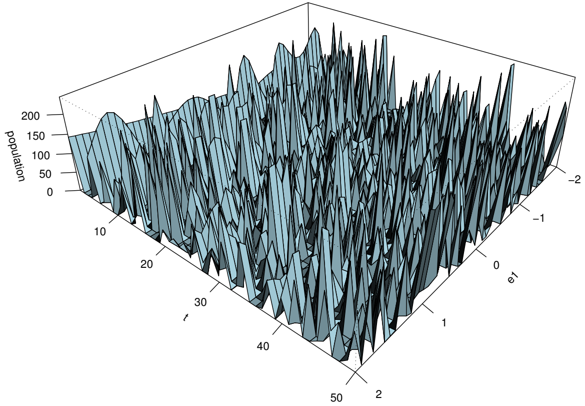

$$N_{t+1} = e^r N_t e^{-N_t + e_t}, \quad e_t \sim N(0, \sigma_e^2), \quad Y_t \sim \text{Poi}(\phi N_t) \ \text{[observed]}$$

...a simple toy model, but like many ecological dynamic systems, it is *very* non-linear...

# Varying $r$ in the Ricker



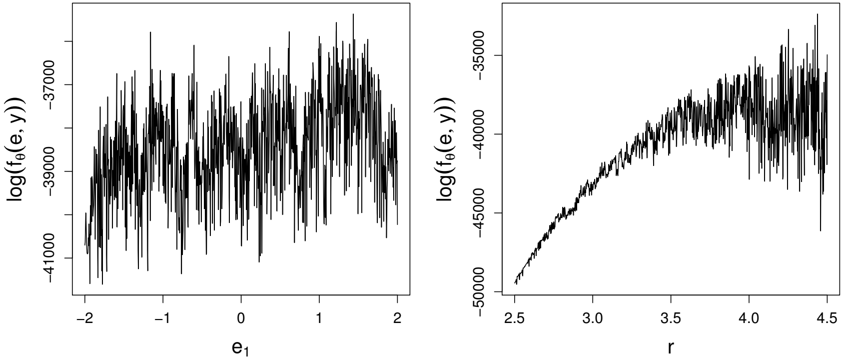
# Varying $e_1$ in the Ricker



# Naive Bayes or Likelihood inference for Ricker

- Suppose we want inferences about $\theta^{\mathrm{T}} = (r, \phi, \sigma_e^2)$, from an observed population data series, $\mathbf{y}$.
- Simple Bayesian or Likelihood approaches both require the joint density $f_\theta(\mathbf{e}, \mathbf{y})$.
  - MLE requires (approximate) evaluation of $\int f_\theta(\mathbf{e}, \mathbf{y}) d\mathbf{e}$ (or an EM approach).
  - Bayesian inference requires samples of $\boldsymbol{\theta}, \mathbf{e}$ from something $\propto f_\theta(\mathbf{e}, \mathbf{y})$.
- We should look at $f_\theta(\mathbf{e}, \mathbf{y})\ldots$

# $\log\{f_\theta(\mathbf{e}, \mathbf{y})\}$ versus $e_1$ and $r$



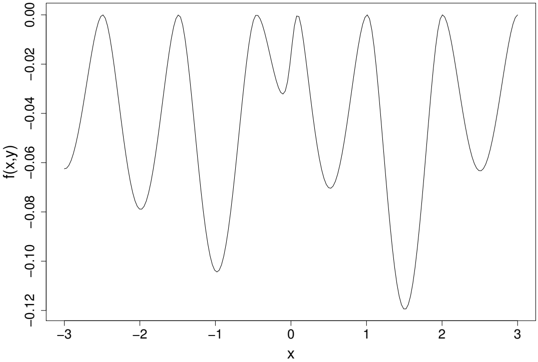

...trying to integrate out the random effects, $\mathbf{e}$, or sample from this density is hopeless.

# Solution 1: change what we be try to model

- We should not try to make the model reproduce features of data that the system would not reproduce itself.
- The exact phase of the series is just a noise feature, which it is not interesting to model.
- We should concentrate on phase insensitive statistics of the series, and on statistics summarizing short term dynamic structure.
- e.g. the ACF, the coefficients of short term autoregressive models, and summaries of the marginal distribution of the observations, or their increments.
- This should be done in a way that is flexible enough to deal with missing data, multiple series and unobserved states.

# But hang on a minute…

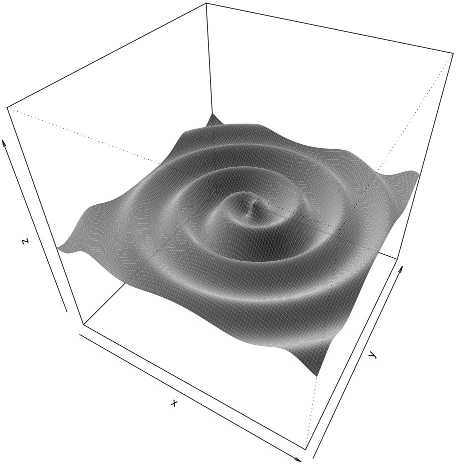
- 1D sections, like that shown through $f_\theta(\mathbf{e}, \mathbf{y})$ are only a partial view.
- For example, here is a section through a function $f(x, z)$…




- Lots of local minimum right?

# $f(x, y)\ldots$

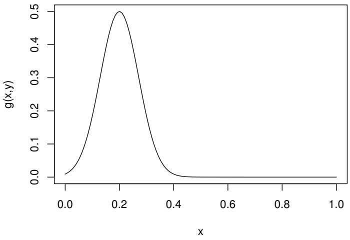
- Wrong!



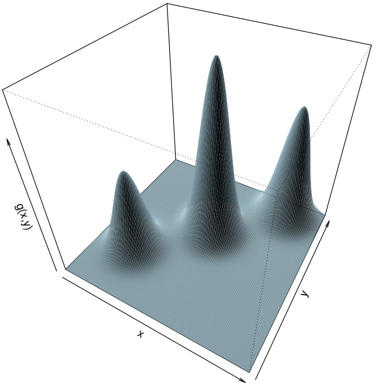

- …it has one global minimum.

# Solution 2: Work with the state

- Recast Ricker problem in terms of state, $n_t$, rather than noise, $e_t$.
- Then plots of $f_\theta(\mathbf{n}, \mathbf{y})$ against any $n_t$ are benignly uni-modal.
- Suggests that problem ought to be tractable using state space methods.
- But 1D views are *still* partial: these transects look lovely



- but are transects through this…




- …need to proceed with caution.

# State space methods

1. Use direct MCMC on states $n_t$, and parameter, $\theta$. Appears to work for Ricker, but
   - Need good starting values for state, which is difficult for a non-linear model with hidden states.
   - Mixing often slow.
   - Hard to prove that it is sampling the target properly.
2. Use filtering, within MCMC or for MC estimation of likelihood.
   - Seems to work, without the initialization and mixing issues.
   - Still hard to prove that the target is being explored properly.

# Filtering in brief

- Idea is to work forward in time, iterating the steps
  1. Prediction:

$$p(n_t|y_{0:t-1}) = \int p(n_t|n_{t-1})p(n_{t-1}|y_{0:t-1})dn_{t-1}$$

  2. Update:

$$p(n_t|y_{0:t}) = p(y_t|n_t)p(n_t|y_{0:t-1})/p(y_t|y_{0:t-1})$$

- In practice replace $p(n_*|y_\cdot)$ by discrete set of $\{n_t^{(i)}\}$ values (particles) with 'importance weights' $\{w_t^{(i)}\}$.
  1. Prediction step then moves the particles, and updates their weights, perhaps by importance sampling.
  2. Update just updates the importance weights.
- The model ought to provide a basis for moving particles.
- In practice periodically resample particles in proportion to their weights to avoid *particle depletion* problems.

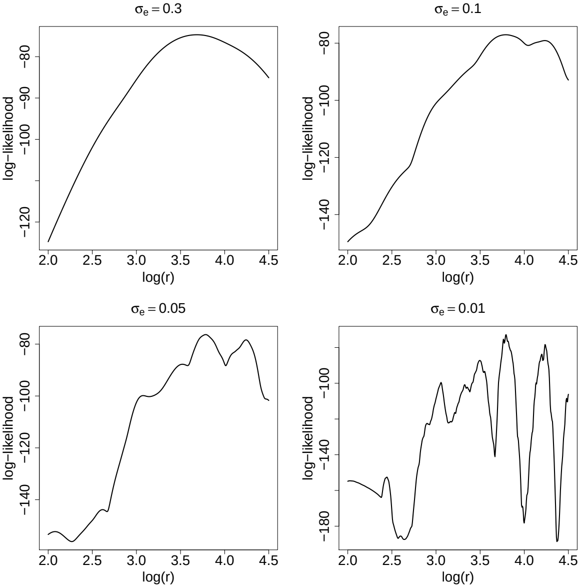
# More particle Filtering

- Likelihood is $p(y_{0:T}) = p(y_0)\prod_t p(y_t|y_{0:t-1})$, where

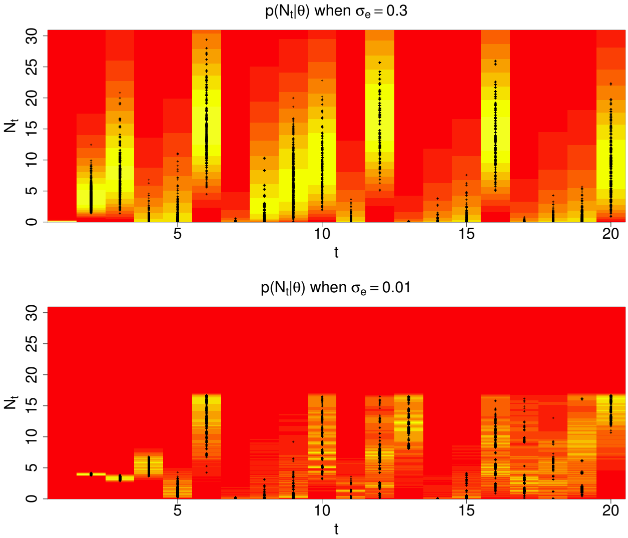
$$p(y_t|y_{0:t-1}) = \int p(y_t|n_t)p(n_t|y_{0:t-1})dn_t \simeq \frac{1}{N}\sum_i p(y_t|n_t^{(i)})w_t^{(i)}$$

- Particle filtering should work, provided the model likelihood is not irregular.
- On toy problems, like the Ricker, it appears to work well, provided there is enough process noise.
- Intuitively it seems that process noise could smooth out the irregularities seen earlier, but generally it's hard to prove anything...
- Let's look at a toy toy model: If we discretise the state space of the Ricker model then the likelihood is exactly and efficiently computable (discrete HMM).

# Exact likelihood for Discrete Ricker as noise declines



# How the particle filter tracks $p(n_t|y_{0:t}, \theta)$




- . . . Density of the state, with attempts to sample by particle filtering. Mixed success.

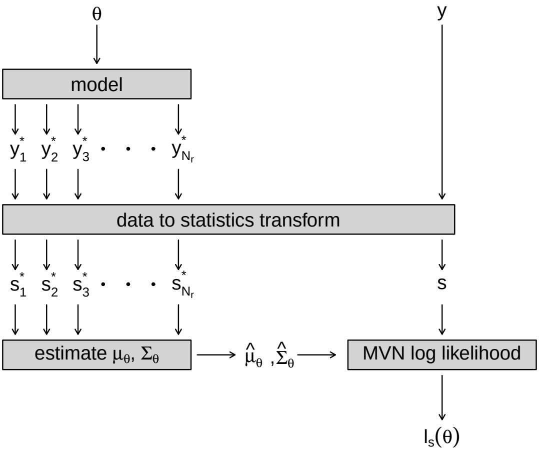
# Statistical dimension reduction. ABC

- The alternative is to try to match carefully chosen *phase insensitive* statistics, $\mathbf{s}$, of the data, in a principled manner.
- Approximate Bayesian Computation, is one approach for stochastic simulation.
  1. For each trial $\theta^*$, data, $\mathbf{y}^*$, are simulated and transformed to statistics, $\mathbf{s}^*$.
  2. The accept/reject decision about $\theta^*$ is then based on replacing the likelihood with $\mathbb{I}(\|\mathbf{s}^* - \mathbf{s}\|_k < \epsilon)$.
- As $\epsilon \to 0$ we simulate from $f(\theta|\mathbf{s})$.
- Snags: need to choose $\|\cdot\|_k$; as $\epsilon \to 0$, the acceptance rate also $\to 0$; works best with few statistics. Tuning needed.

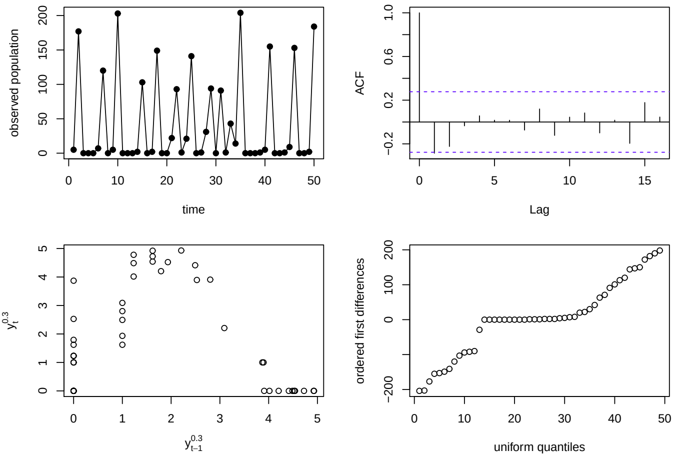
# Less tuning, more assumptions: synthetic likelihood

- Convert the observed data $\mathbf{y}$ into a vector of phase insensitive summary statistics that can be modelled as approx. normal. i.e. $\mathbf{s} \sim N(\boldsymbol{\mu}_\theta, \mathbf{V}_\theta)$, where $\boldsymbol{\theta}$ are the model parameters.
- A 'synthetic' likelihood can be evaluated as follows.
  - Given $\boldsymbol{\theta}$, simulate $N_r$ (=500, here) replicates, $\mathbf{y}_1^*, \mathbf{y}_2^* \ldots$, and process these exactly as $\mathbf{y}$ was processed to obtain replicate statistics, $\mathbf{s}_1^*, \mathbf{s}_2^*, \ldots$.
  - Define $\hat{\boldsymbol{\mu}}_\theta = \sum_i \mathbf{s}_i^* / N_r$, $\mathbf{S} = [\mathbf{s}_1^* - \hat{\boldsymbol{\mu}}_\theta, \mathbf{s}_2^* - \hat{\boldsymbol{\mu}}_\theta, \ldots]$ and hence $\hat{\mathbf{V}}_\theta = \mathbf{S}^{\mathrm{T}}\mathbf{S}/(N_r - 1)$.
  - $l_s(\boldsymbol{\theta}) = -(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)^{\mathrm{T}} \hat{\mathbf{V}}_\theta^{-1} (\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)/2 - \log|\hat{\mathbf{V}}_\theta|/2$ is the log synthetic likelihood.

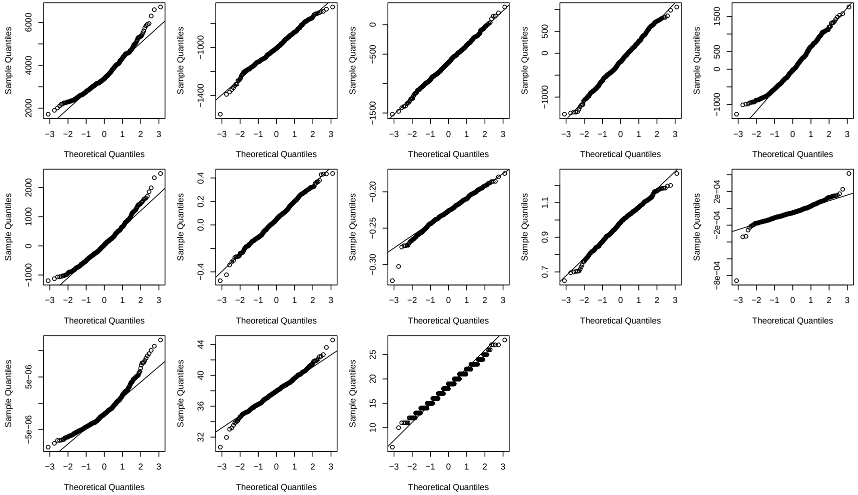
# Synthetic likelihood - picture



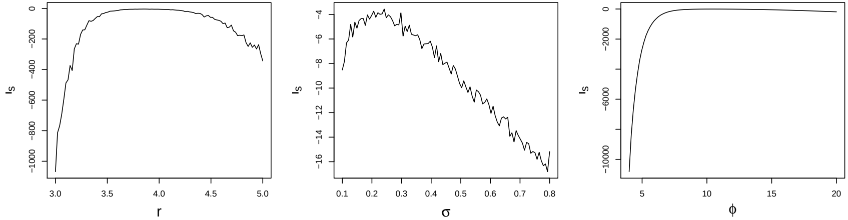
# Ricker example: what statistics?



# Ricker example: statistics $\sim N$?

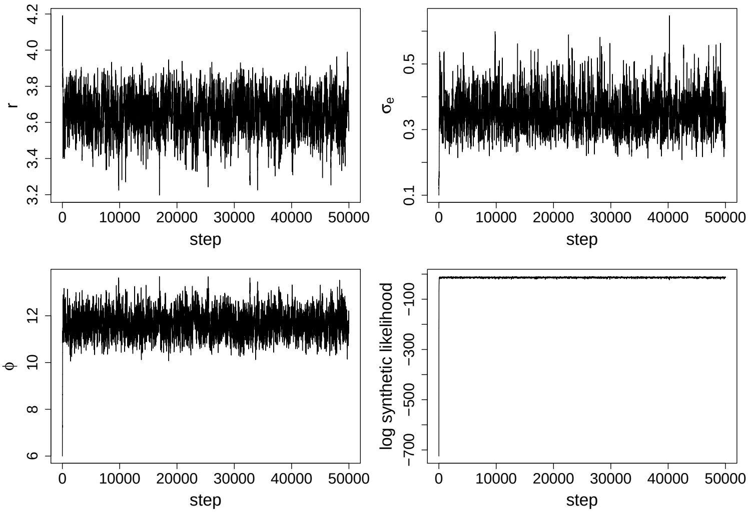


# Ricker example: transects through synthetic likelihood

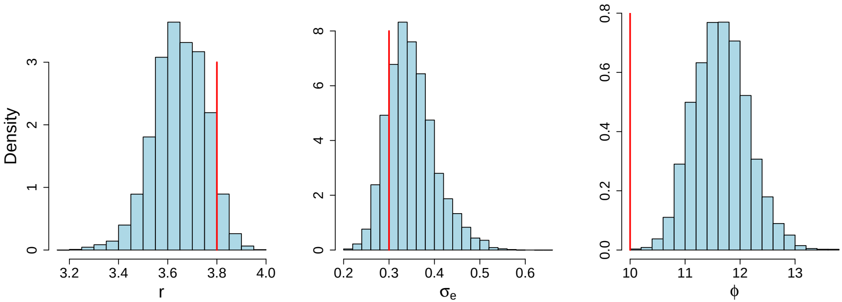



- ...variability is random and small scale.
- Easy to explore $l_s$ by MCMC, or stochastic optimization methods.
- Can estimate $l_s$ in vicinity of maximum by quadratic regression of $l_s$ on parameters.

# Ricker example: MCMC results (25% acceptance)

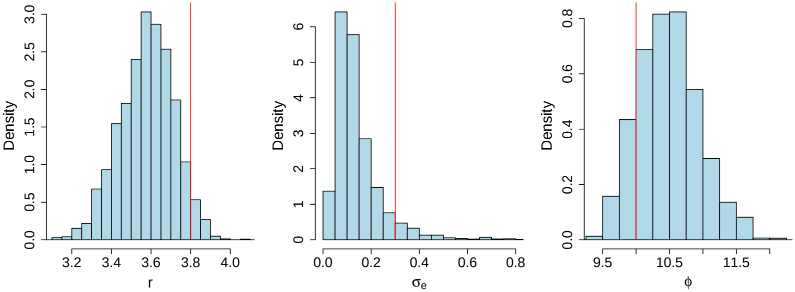


# Ricker example: parameter estimates




- Truth in red, 10000 burn in discarded.

# Ricker direct MCMC on state and parameters



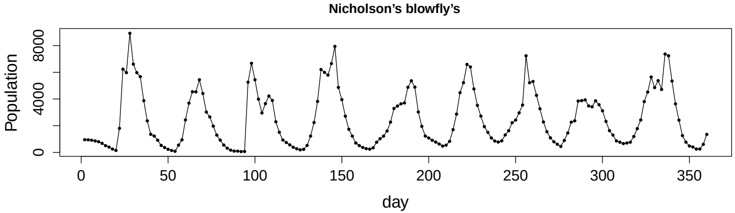

- Truth in red, based on second half of every $100^{\text{th}}$ of 1000000 single term update steps.

# Taking stock

- Nothing is perfect (and in practice nothing is fast)!
  - Direct MCMC is difficult to start and difficult to get to mix.
  - Filtering is problematic at low process noise, and when the parameters are not close to correct.
  - ABC seems to need quite a bit of tuning.
  - Synthetic likelihood has the dodgy normality assumption.
- In simulations with toy models, the direct methods (MCMC and Filtering) have better statistical efficiency, for the correct model, when the process noise is not too low.
- When the model is not aiming to capture everything then only ABC and Synthetic likelihood seem reasonable.
- Let's look at a less toy example. . .

# Nicholson's Blowflies

# Nicholson's Blowfly Data




- Lab population of adult flies known exactly every 2 days, from counts of dead flies and freshly discarded pupal cases.
- Time from egg to adult is around 2 weeks.
- Fecundity is likely to be density dependent.

# Blowfly model

- DDE model (Nisbet and Gurney, 1981) is

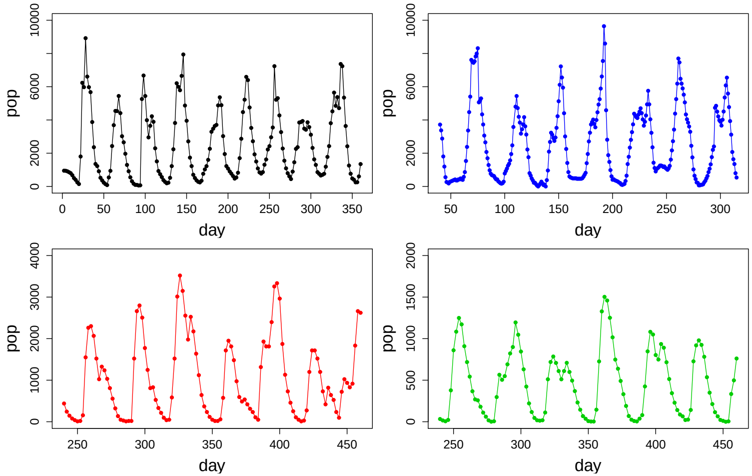
$$\frac{\mathrm{d}N}{\mathrm{d}t} = PN(t-\tau)e^{-N(t-\tau)/N_0} - \delta N(t).$$

- Daily discretisation is

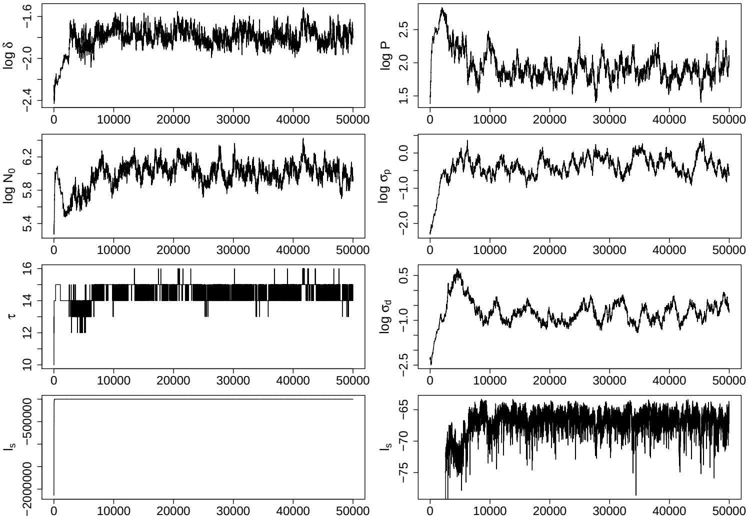
$$N_{t+1} = PN_{t-\tau}\exp(-N_{t-\tau}/N_0)e_t + N_t\exp(-\delta\epsilon_t)$$

  where $e_t$ and $\epsilon_t$ are independent Gamma random deviates with mean 1 and variance $\sigma_p^2$ and $\sigma_d^2$.
- Statistics are ACF coefficients to lag 11, coefficients of the increments regression as for Ricker example, the mean population and coefficients of the auto-regression $y_i = \beta_0 y_{i-6} + \beta_1 y_{i-6}^2 + \beta_2 y_{i-6}^3 + \beta_3 y_{i-1} + \beta_4 y_{i-1}^2 + \epsilon_i$.
- Parameter group $P\tau$ and $\delta\tau$ controls stability.

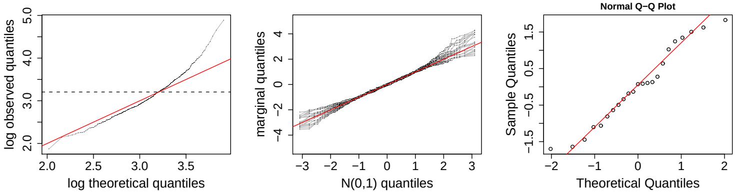
# 4 Experimental replicates



# Blowfly MCMC chain (31% acceptance)

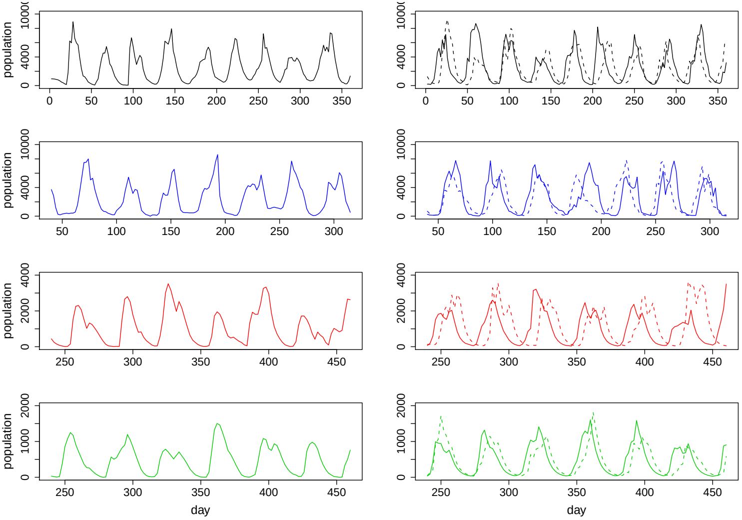


# Blowfly checking plots

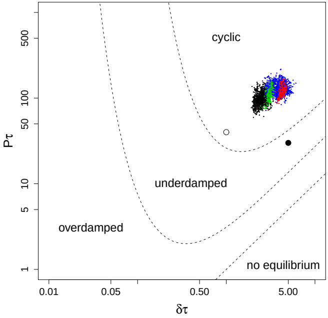



- Left is QQ-plot for simulated $(\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_\theta^{-1} (\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)$ (observed is dashed).
- Middle gives scaled normal QQ plots for each element of simulated $\mathbf{s}$.
- Normal QQ plot for observed $\hat{\boldsymbol{\Sigma}}_\theta^{-1/2} (\mathbf{s} - \hat{\boldsymbol{\mu}}_\theta)$.

# Blowflies: data left, model reps right



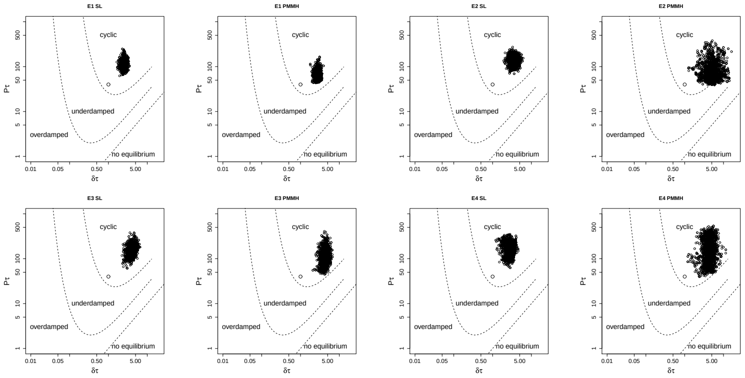
# Blowfly stability




- The dynamics are all limit cycles. The fluctuations are not noise driven.
- This is not obvious a priori: e.g. noise has to be much larger than demographic stochasticity would give, to generate observed variability.

# Blowflies by Filtering

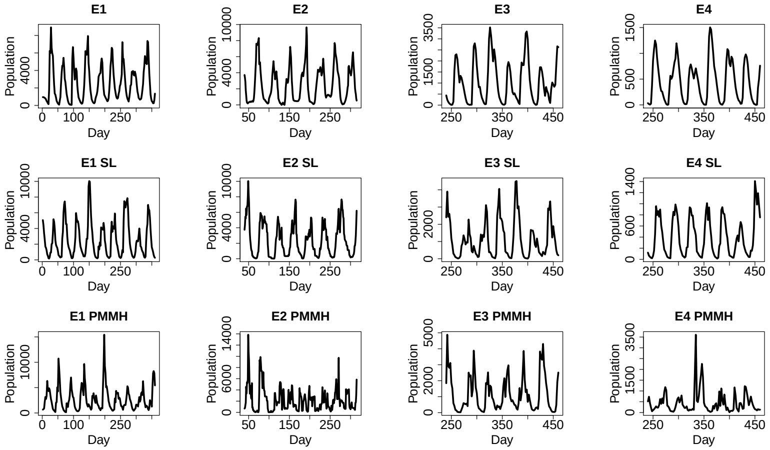
- We tried a similar analysis via filtering.
- Have to add observation noise to model to make this work.



# Which is right?

- The filtering approach puts much more weight on the region of greater dynamic stability.
- For the experimental conditions in E2 there is separate experimental data which is inconsistent with this.
- Further investigation suggests that the filter is suffering severe particle depletion at a couple of times.
- Switching to a heavier tailed error distribution improves matters somewhat.
- But there are at least some clear diagnostic indicators...

# Typical simulations from the fitted models



# Conclusions?

- For a correct enough non-linear dynamic models with enough process noise, filtering is probably best.
- The problem is knowing what's *enough* in the above.
- If you are unsure that the model is right, or it's not supposed to be correct, then information reduction methods such as ABC or Synthetic Likelihood (SL) may be less misleading, and don't seem to be massively worse than filtering in any case.
- SL requires a bit less tuning than ABC, but at the cost of a normality assumption. The latter may be removable via a saddlepoint approximation.

# Some References

Arulampalam, Maskell, Gordon, and Clapp (2002) A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking.

Beaumont, Zhang, and Balding (2002) Approximate bayesian computation in population genetics. Genetics, 162(4):20252035. IEEE Transactions on Signal Processing, 50(2):174188.

Ionides, Bhadra, Atchadé, and King (2011) Iterated filtering. The Annals of Statistics, 39(3):17761802.

Nicholson (1954) An outline of the dynamics of animal populations. Australian Journal of Zoology, 2(1):965.

Wood (2010) Statistical inference for noisy nonlinear ecological dynamic systems. Nature, 466(7310):11021104

Fasiolo, Pya and Wood (nearly complete) Statistical inference for highly nonlinear dynamic ecological models: method comparison.