Model Methods

Aaron M.

27/06/2020

Model description

Discrete time model equations

The age structured viral dynamics model, with all potential transitions is stated in the equations below, with parameters corresponding to table 1, transitions between states occurs in discrete time $(t \rightarrow t+1)$, in each model discrete time setps (t) are set to 0.25 days.

$$S_N(t+1) = S_N(t) + b(t)(S_F(t) + I_F(t)) - (\omega_m + m_j\frac{N}{\kappa} + \beta(I_J(t) + I_N(t) + I_F(t) + I_M(t)))S_N(t) + \omega R_N(t)$$

$$S_J(t+1) = S_J(t) + \omega_m(S_N(t) + Ma) - (\mu + m_j\frac{N}{\kappa} + \beta(I_J(t) + I_N(t) + I_F(t) + I_M(t)))S_J(t) + \omega R_J(t)$$

$$S_M(t+1) = S_M + (t)\mu\frac{S_J}{2} - (m\frac{N}{\kappa} + \beta(I_J(t) + I_N(t) + I_F(t) + I_M(t)))S_M(t)I_M(t) + \omega R_M(t)$$

$$S_F(t+1) = S_F(t) + \mu\frac{S_J}{2} - (m\frac{N}{\kappa} + \beta(I_J(t) + I_N(t) + I_F(t) + I_M(t)))S_F(t) + \omega R_F(t)$$

$$L_N(t+1) = L_N(t) - (\omega_m + m_j\frac{N}{\kappa} + \epsilon)L_N + \rho I_N$$

$$L_J(t+1) = L_J(t) + \omega_m L_N(t) - (\mu + m_j\frac{N}{\kappa} + \epsilon)L_J(t) + \rho L_J(t)$$

$$L_M(t+1) = L_M(t) - (\mu + m_j\frac{N}{\kappa} + \epsilon)L_M(t) + \rho L_M(t)$$

$$L_F(t+1) = L_F(t) - (\mu + m_j\frac{N}{\kappa} + \epsilon)L_F(t) + \rho L_F(t)$$

$$I_N(t+1) = I_J(t) - (\omega_m + m_j\frac{N}{\kappa} + \gamma + \rho)I_J + \beta(I_J(t) + I_N(t) + I_F(t) + I_M(t)))S_N + \epsilon L_N$$

$$I_J(t+1) = I_J(t) - (\mu + m_j\frac{N}{\kappa} + \gamma + \rho)I_J + \beta(I_J(t) + I_N(t) + I_F(t) + I_M(t)))S_J + \epsilon L_J$$

$$I_F(t+1) = I_F(t) - (m\frac{N}{\kappa}\gamma + \rho)I_F(t) + \beta(I_F(t) + I_M(t))S_F(t) + \epsilon L_F(t)$$

$$I_M(t+1) = I_M(t) - (m\frac{N}{\kappa} + \gamma + \rho)I_M(t) + \beta(I_M(t) + I_F(t))S_M(t) + \epsilon L_M(t)$$

$$R_N(t+1) = R_N(t) - (\omega_m + m_j\frac{N}{\kappa} + \omega)R_N(t)$$

$$R_J(t+1) = R_J(t) + \omega_m R_N(t) - (\mu + m_j\frac{N}{\kappa} + \omega)R_J(t)$$

$$R_M(t+1) = R_M(t) + \mu\frac{R_J(t)}{2} - (m\frac{N}{\kappa} + \omega)R_M(t) + \gamma I_M(t)$$

$$R_F(t+1) = R_F(t) + \mu\frac{R_J(t)}{2} - (m\frac{N}{\kappa} + \omega)R_F(t) + \gamma I_F(t)$$

$$M_a(t+1) = M_a(t) + b(R_F(t) + L_F(t)) - (\omega_m + m_j\frac{N}{\kappa})M_a(t)$$

Births occur in seasonal pulses, as previously described (Peel et al. 2014), with timing, amplitude and seasonality derived from three parameters in a modified Gaussian function.

$$b(t) = c \exp^{s\cos^2(\pi t - \phi)} \tag{1}$$

β is derived from a rearrangement of equation (2) in the main text, here N is considered an upper limit to the population, thus N is equal to the carrying capacity κ:

$$\beta = R_0 \left(\frac{(\epsilon + m)(\gamma + m + \rho) - \epsilon\rho}{N(\epsilon + m)} \right) \tag{2}$$

Stochastic transitions between states

Movement between states at each discrete time-step (set as 0.25 days) for individuals is a stochastic process, using a combination of binomial, and multinomial distributions where multiple transition possibilities occur.

e.g. for bats exiting the variable state R, transitions will occur with the following steps:

- The total number of individuals transitioning between states is determined by first summing all of the rates of exiting a current state, e.g. $(R \to exit) = \omega + \frac{N}{\kappa}$.

- We then convert the sum of rates to a probability for each possible transition occurring during a single time-step, e.g. for a rate x, the equivalent probability of transition in one time-step is $p = 1 - e^{-x}$, thus $Pr(R \to exit) = 1 - e^{-(R \to exit)}$

- For each individual, a random draw from a binomial distribution is then performed at each time-step to obtain the number of individuals exiting R e.g. $R_e = bin(R, Pr(R \to exit))$.

- The relative probabilities are then computed for transition to each of the potential states

$$Pr(R \to S) = \frac{\omega}{m\frac{N}{\kappa} + \omega} \tag{3}$$

and

$$Pr(R \to death) = \frac{m\frac{N}{\kappa}}{m\frac{N}{\kappa} + \omega} \tag{4}$$

- A multinomial distribution is used with each probability to determine the destination of each individual $M(R_e, Pr(R \to S), Pr(R \to death))$

Model fitting

The models are considered state-space (hidden Markov) models, with observations k and hidden state variables Z at time t and we seek to identify an unbiased estimate of the likelihood value, (also known as a marginal likelihood) $pr(k|\theta)$, by essentially "marginalising out" (Z). Here we do this using a particle markov chain monte carlo (pMCMC) algorithm. Which combines a particle filter as an unbiased estimator of the marginal likelihood value, and an MCMC acceptence/rejection algorithm (Andrieu, Doucet, and Holenstein 2010).

pMCMC generic example

Particle filters act as an unbiased estimator of marginal likelihood by applying a form of importance re-sampling, to generate an approximate sample from, and make inferences about an unobserved Markov process (Smith 2013). They are becoming popular in disciplines where state-space models are common such as ecology and epidemiology (Kantas et al. 2015; Peters, Hosack, and Hayes 2010; Knape and De Valpine 2012; Fasiolo et al. 2016; Sheinson, Niemi, and Meiring 2014).

In brief, the marginal likelihood $pr(k|\theta)$ for the observed data k, the hidden state Z and the parameter vector θ is considered the joint posterior probability $pr(k|Z,\theta)*p(Z|\theta)$. Using Monte-Carlo approximation for $p(k|\theta)$, a particle filter with j particles which have the possible trajectories/evolution of Z_j, the marginal likelihood can be considered as $pr(k|\theta) \approx \sum_J pr(k|Z_j,\theta) * pr(Z_j|\theta)$.

To run the particle filter algorithm requires the following steps:

1. Initialise the particles with equal weights.

$$
\begin{aligned}
Z_j &\sim pr\left(Z_j|\theta\right) \\
w_j &= \tfrac{1}{J}
\end{aligned}
\tag{5}
$$

2. For each particle j at time t, simulate the initial conditions at the first observed data point.

$$
Z_{jt} \sim pr\left(Z_{jt}|Z_j,\theta\right)
\tag{6}
$$

3. Calculate a probability weighting for each particle based on the results of the simulation and the observed data.

$$
w_{jt} = pr\left(k_t|Z_j,\theta\right)
\tag{7}
$$

4. The marginal likelihood for this data point can be considered the average of each particle weighting.

$$
p\left(k_t|\theta\right) = \frac{1}{J}\sum_J w_{jt}
\tag{8}
$$

5. Normalise the weightings $\frac{w_{jt}}{\left(\sum_J w_{jt}\right)}$, resample with replacement each particle based on their weighting and simulate forward to the next observed data point.

$$
Z_{jt+1} \sim pr\left(Z_{jt+1}|Z_t,\theta\right)
\tag{9}
$$

Repeat steps 3 to 5 for all observed data, an estimate for the total marginal likelihood value can be considered the product of the average particle weights at each observed timepoint.

$$
\mathcal{L}(\theta|k_{1:n}) = \prod_{t=1}^{n}\left(\frac{1}{J}\sum_{j=1}^{J} w_{jt}\right)
\tag{10}
$$

This marginal likelihood value is then used in a traditional MCMC algorithm, in this instance Metropolis-Hastings, where rejection or acceptance of a parameter proposal is based on the current and previous marginal likelihood values.

Model fitting with pMCMC to Boonah bat data

To fit the model with pMCMC to our observed data, we developed a likelihood function which incorporated the three unique longitudinal data-types; Individual serology for Hendra virus antibodies, individual PCR of urine for Hendra virus RNA, and under roost PCR of urine for Hendra virus RNA.

- *Firstly, we consider the observed data:*

  If the sampling time points are $i_1...i_n$, the number of individual samples through time are $N_{i...n}$ of which $k^p_{i...n}$ are PCR positive and $k^s_{i...n}$ are seropositive, the number of under-roost urine samples is $N^u_{i...n}$ of which $k^u_{i...n}$ are PCR positive.

- *Secondly, the hidden state:*

  If the system state of the model at each time point is $Z_{i...n}$, the transition probability density function for Z from the stochastic model, conditioned on the parameters θ, is $pr(Z_i|Z_{i-1}, \theta)$.

Below we define how each data type contributes to the likelihood function in detail:

Individual serological and PCR of urine samples

In the observed data, individual bats are found to be in all four empirical states of serological and PCR positivity:

- PCR positive and seropositive ($k^{P^+S^+}$)
- PCR negative and seropositive ($k^{P^-S^+}$)
- PCR positive and seronegative ($k^{P^+S^-}$)
- PCR negative and seronegative ($k^{P^-S^-}$)

Logically, a relationship between the PCR and serology states is likely and so we consider a joint conditional probability function rather than treating the data as independent. Transitory states and test detection failure are also considered by fitting the coefficients ζ_s and ζ_p which are bounded between 0 and 1. Thus, the simulated implementation of each observed empirical state are as follows:

- $k^{P^+S^+}$: Bats are assumed to be in the I state of the model and are equivalent to the number of I state bats proportionate to the value of the coefficients; $z^{P^+S^+} = I\zeta_p\zeta_s$

- $k^{P^-S^+}$: Bats are assumed to be in either the L or R states of the models, or in the I state of the model; $z^{P^-S^+} = (I + L + R)\zeta_s(1 - \zeta_p)$

- $k^{P^+S^-}$: Bats are considered to be in the I state of the model; $z^{P^+S^-} = I(1 - \zeta_s)\zeta_p$

- $k^{P^-S^-}$: Bats are assumed to be in the S state of the models or in the I state; $z^{P^-S^-} = I(1-\zeta_s)(1-\zeta_p)+S$

If k is a vector of the observed counts of bats in each empirical state, $k = (k^{P^+S^+}, k^{P^-S^+}, k^{P^+S^-}, k^{P^-S^-})$, z is a vector of the simulated prevalences of each state $z = (z^{P^+S^+}, z^{P^-S^+}, z^{P^+S^-}, z^{P^-S^-})$ and N is the total observed bats, we can then propose a joint probability function based on a multinomial distribution, $pr(k^p, k^s|Z, N)$.

$$pr(k^p, k^s|Z) = \frac{N!}{\prod_{j=1}^{4} k_j!} \prod_{j=1}^{4} z_j^{k_j} \tag{11}$$

Pooled under roost samples

If the probability of at least one bat within a pooled urine sample of x individuals being PCR positive for virus RNA can be considered $P_t = 1 - (1-p)^x$, where p is the probability of each individual bat contributing to a pool being positive (Chiang and Reeves 1962). The result of PCR testing on a pooled sample (T_i) will be one of two states, with the following probabilities:

$$T_i = \begin{cases} positive = 1 - (1-p)^x \\ negative = (1-p)^x \end{cases} \tag{12}$$

Therefore, assuming that any one under-roost sample comes from x bats with a prevalence of p, considering the model state $Z(t)$, we aim to identify the likelihood of the contribution of positive under-roost samples at time (t) given (from the predicted GAM values) the number of samples $N^u(t)$, of which $k^u(t)$ were positive, $pr(k^u|N^u, Z)$. As the predicted values from the GAM are for prevalence, the total number of bats ($N^u(t)$) was considered the mean number of samples collected on each sampling occasion.

Due to the nature and challenges of field-sampling, under-roost data is likely to contain a degree of overdispersion, as it is highly exposed to external factors which are difficult to account for in study design. Therefore, the probability is unlikely to follow a standard binomial distribution, where the variance is defined by the mean. To account for this, we used a probability function based around a beta-binomial distribution, which allows for an additional variance parameter to be fitted to account for any overdispersion in the observations. A beta-binomial distribution, is a binomial distribution such $pr(k^u|N^u, Z)$ would be:

$$pr(k^u|N^u, Z) = \binom{N^i}{k^u} p_t^{k^u} (1-p_t)^{N^u-k^u} \tag{13}$$

However, p_t is not constant and is generated from a beta distribution, which takes two shape parameters α and β, $beta(\alpha, \beta)$ thus;

$$pr(k^u|N^u, Z) = \frac{\binom{N^u}{k^u} beta(k^u + O_u p, N^u - k^u + O_u(1-p_t))}{beta(O_u p, O_u(1-p_t))} \tag{14}$$

Here, O_u accounts for overdispersion in the data, and the prevalence of infectious bats p in p_t is derived from the model state $Z(t)$ as $\frac{I(t)\zeta_u}{N(t)}$, where ζ_u is a fitted coefficient accounting for detection failure.

As there is no precise measure of exact bat numbers contributing to a pooled sample, x is a fitted parameter with a prior based on expert knowledge and estimates from the field. For simplicity, only one value of x is derived for all pools on a single sampling occasion, assuming that each pooled sample on average has the same number of bats contributing. We also assume that there is no effect of pooling on the diagnostic tests, and any that does occur should be primarily accounted for by the ζ_u coefficient.

full likelihood function

Considering the above, and the sampling time points (i), we can define the full likelihood function as:

$$\mathcal{L}(\theta|N, N^u, k^{(s,p,u)}) = pr(\theta) * \prod_{i=1}^{16} pr(k_i^s, k_i^p|N_i, Z_i) * pr(k_t^u|N_i^u, Z_i) \tag{15}$$

Initial conditions

Starting parameters are estimated by running a Metropolis-Hastings MCMC algorithm with a deterministic version of the model for 50,000 iterations with a 10,000 iteration burn-in. The median parameter set from each posterior distribution was then used in the pMCMC.

Initial conditions for the bat population are derived from the carrying capacity parameter (κ). A population equal to the carrying capacity is first assumed and this is then split into demographic compartments as follows:

$$N = \kappa \tag{16}$$

$$b = (\omega_m + m_j)\frac{m_j + \mu}{\omega_m} \tag{17}$$

$$S_F = 0.5\left(\frac{N}{\frac{1+m}{\mu} + \frac{b}{\omega_m + m_j}}\right) \tag{18}$$

$$S_M = 0.5\left(\frac{N}{\frac{1+m}{\mu} + \frac{b}{\omega_m + m_j}}\right) \tag{19}$$

$$S_J = (S_F + S_M)\frac{m}{mu} \tag{20}$$

$$S_N = N - S_M - S_F - S_J \tag{21}$$

Infectious and exposed individuals are added as 5% of the population and the model is run for 50 years to reach an equilibrium before fitting to the first data point.

Table 1: Parameter and variable descriptions for supplementary materials

Parameter	Prior	Description
S	-	Number of susceptible bats
I	-	Number of infectious bats
R	-	Number of recovered bats
L	-	Number of latently infected bats
N	-	Total population of bats
θ	-	Parameter vector
Z	-	System state
z	-	Vector of simulated individual bat states
ζ_s	0-1	Individual serology coefficient
ζ_p	0-1	Individual PCR coefficient
ζ_u	0-1	Under roost PCR coefficient
k^s	-	Observed individual seropositive bats
k^p	-	Observed individual PCR positive bats
k^u	-	Predicted no. positive bats contributing to pooled under roost samples
k	-	Vector of observed individual bat states
N^u	-	Number of pooled under-roost samples (mean)
p	-	Probability of bat being positive (infection prevalence)
P_t	-	Probability of pooled sample being positive
O_u	>0	Overdispersion parameter for under roost data
x	0-10	Number of bats contributing to each pooled under roost sample
T_i	-	Pooled under roost sample test state
β	> 0	Infection rate
γ	> 1 day	Recovery rate
ρ	> 1 day	Latency rate
ϵ	> 1 day	Recurrence rate
m	0.186 (0.146-0.225) year^{-1}	Adult mortality rate
m_j	0.500 (0.480-0.520) year^{-1}	Juvenile mortality rate
b	-	Birth rate
c	-	Birth pulse scalar
s	130 (111-150)	Birth pulse synchronicity
ϕ	7.180 (6.787-7.571)	Birth pulse timing
μ	1.37 year^{-1}	Maturation rate
ω_m	0.800 (0.741-0.859) year^{-1}	Maternal antibody waning rate
ω	> 1 day	Antibody waning rate
R_0	> 1	R_0
d	0-25	Pooled sample contributing bats
c_u	1	Env. force scalar
s_u	-	Env. force synchronicity
ϕ_u	-	Env. force timing
κ	4000 (2570-6300)	Environmental carrying capacity

[1] Informative priors are based on normal distributions and are shown with 95% credible intervals in brackets, uninformative priors use a uniform distribution and are shown as a minimum and maximum value
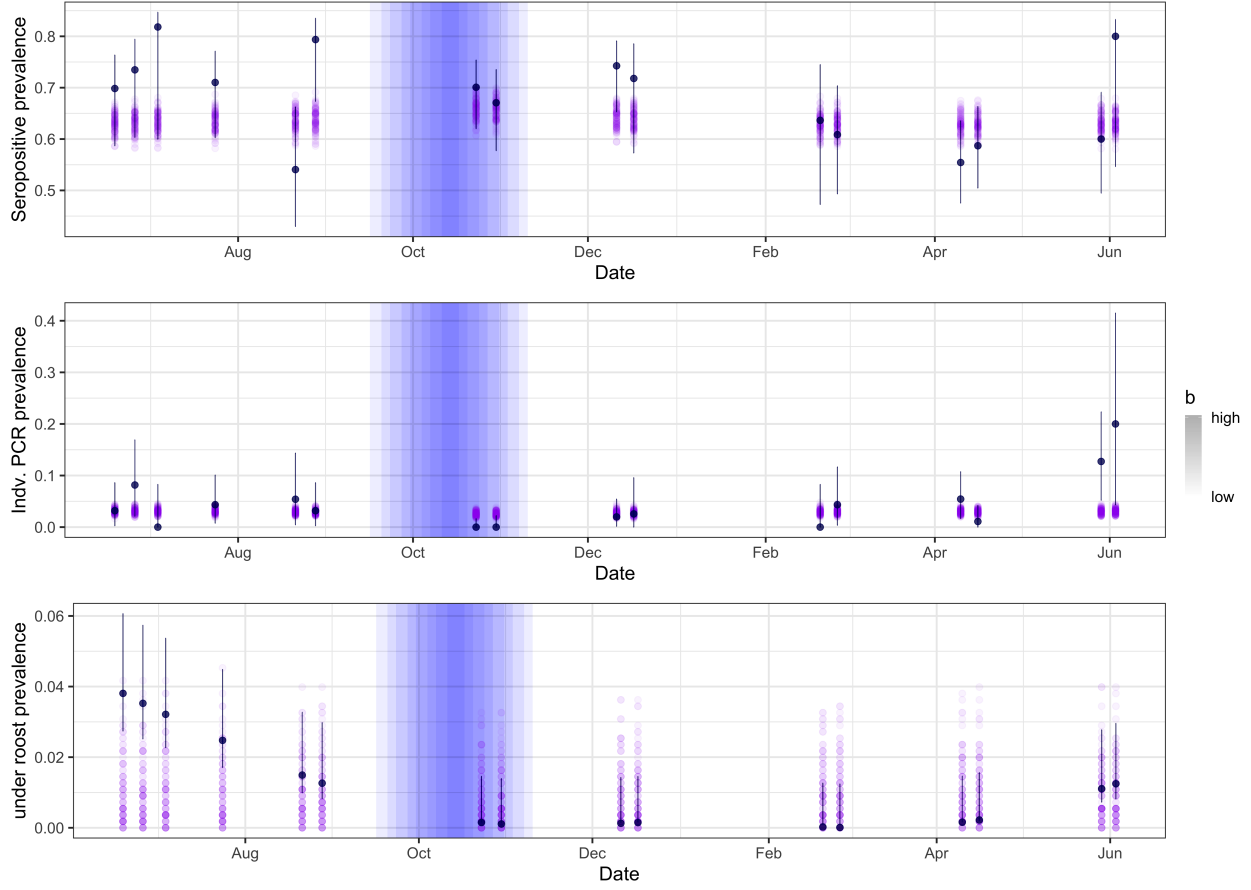
Example output

Figure 1: 100 repeat runs of the maternal immunity SILI model without seasonal forcing (model 2 in table 1), parameterised using the median values from the posterior estimates of each parameter. Dark blue points show the observed data, purple points show the model outputs for each run of the model at the corresponding observed time period. Blue shading shows the approximate timing and amplitude of the birth pulse. The top figure shows the simulated and observed serological prevalence, the middle figure shows the simulated and observed virus RNA prevalence from individual urine samples, and the bottom figure shows the simulated and predicted (using observed data and contributing bats parameter d) virus RNA prevalence from the under-roost samples. Confidence intervals on observed data are calculated as binomial confidence intervals, with a beta prior on the binomial distribution; the shape of the beta prior for individual samples is uninformitive and for under-roost data is derived from the corresponding fitted overdispersion parameters for each data-type.
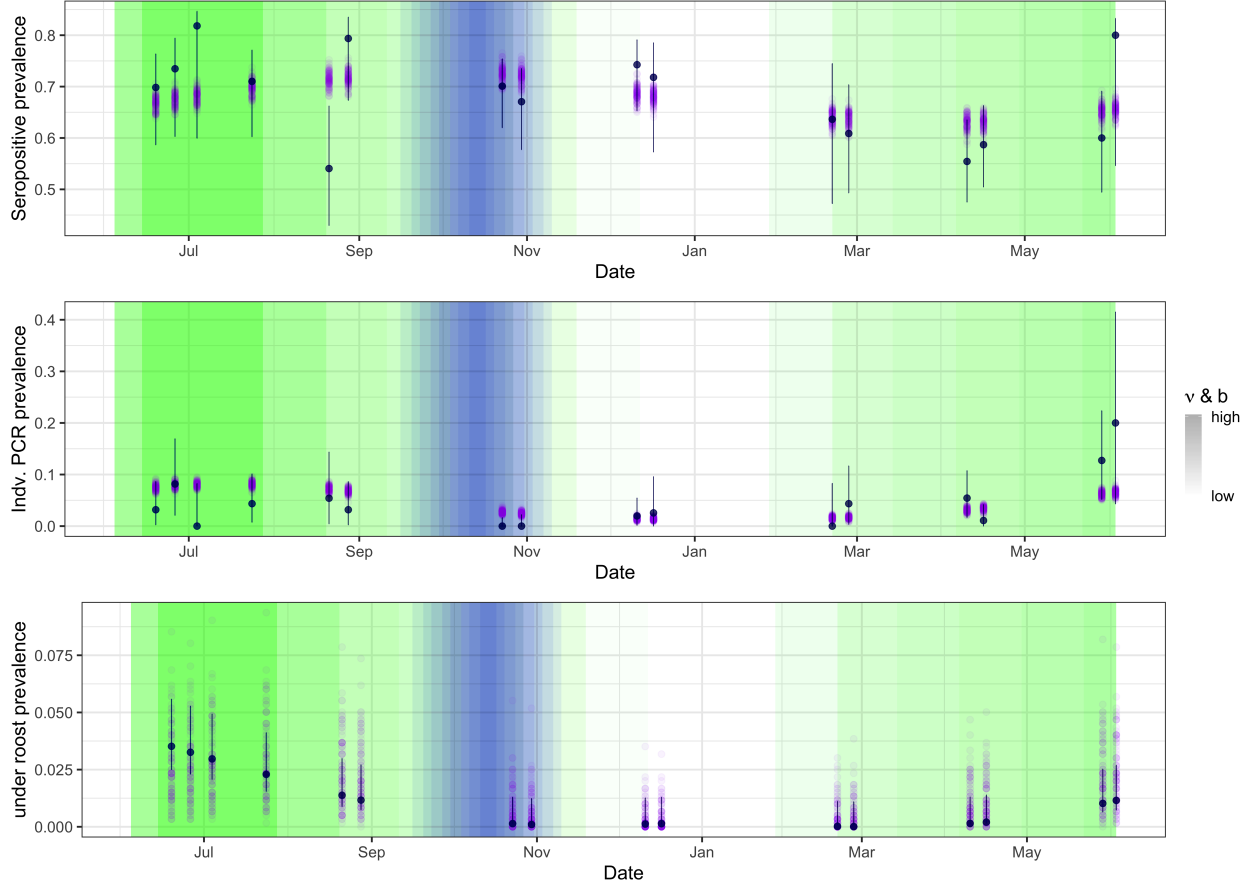
Figure 2: 100 repeat runs of the maternal immunity SILI model with seasonal forcing (model 4 in table 1), parameterised using the median values from the posterior estimates of each parameter. Dark blue points show the observed data, purple points show the model outputs for each run of the model at the corresponding observed time period. Blue and shading shows the approximate timing and amplitude of the birth pulse and seasonal force. The top figure shows the simulated and observed serological prevalence, the middle figure shows the simulated and observed virus RNA prevalence from individual urine samples, and the bottom figure shows the simulated and predicted (using observed data and contributing bats parameter *d*) virus RNA prevalence from the under-roost samples. Confidence intervals on observed data are calculated as binomial confidence intervals, with a beta prior on the binomial distribution; the shape of the beta prior for individual samples is uninformitive and for under-roost data is derived from the corresponding fitted overdispersion parameters for each data-type.

References

Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein. 2010. "Particle Markov Chain Monte Carlo Methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72 (3): 269–342.

Chiang, Chin-long, and William C et al. Reeves. 1962. "Statistical Estimation of Virus Infection Rates in Mosquito Vector Populations." *American Journal of Hygiene* 75 (3).

Fasiolo, Matteo, Natalya Pya, Simon N Wood, and others. 2016. "A Comparison of Inferential Methods for Highly Nonlinear State Space Models in Ecology and Epidemiology." *Statistical Science* 31 (1): 96–118.

Kantas, Nikolas, Arnaud Doucet, Sumeetpal S Singh, Jan Maciejowski, Nicolas Chopin, and others. 2015. "On Particle Methods for Parameter Estimation in State-Space Models." *Statistical Science* 30 (3): 328–51.

Knape, Jonas, and Perry De Valpine. 2012. "Fitting Complex Population Models by Combining Particle Filters with Markov Chain Monte Carlo." *Ecology* 93 (2): 256–63.

Peel, Alison J, JRC Pulliam, AD Luis, RK Plowright, TJ O'Shea, DTS Hayman, JLN Wood, CT Webb, and O Restif. 2014. "The Effect of Seasonal Birth Pulses on Pathogen Persistence in Wild Mammal Populations." *Proceedings of the Royal Society B: Biological Sciences* 281 (1786): 20132962.

Peters, Gareth W, Geoff R Hosack, and Keith R Hayes. 2010. "Ecological Non-Linear State Space Model Selection via Adaptive Particle Markov Chain Monte Carlo (Adpmcmc)." *arXiv Preprint arXiv:1005.2238*.

Sheinson, Daniel M, Jarad Niemi, and Wendy Meiring. 2014. "Comparison of the Performance of Particle Filter Algorithms Applied to Tracking of a Disease Epidemic." *Mathematical Biosciences* 255: 21–32.

Smith, Adrian. 2013. *Sequential Monte Carlo Methods in Practice.* Springer Science & Business Media.