

Using Chapman-Kolmogorov equations in MCMC to analyse S-I-S data

G. T. Innocent, K. O'Reilly, G. Marion & I. J. McKendrick

Chapman-Kolmogorov equations define the rate of change of the probabilities that a system resides in each of several states. In this case we consider the number of infectious animals within a group of N animals as specifying the state in which the system resides. There are therefore $N+1$ possible states ($I \in \{0, 1, 2, \dots, N-1, N\}$), and consequently $N+1$ differential equations used to describe the rate of change of the probability that the group resides in any one of these states. The system is presented in Figure 1, with corresponding ODEs in Equations 1.

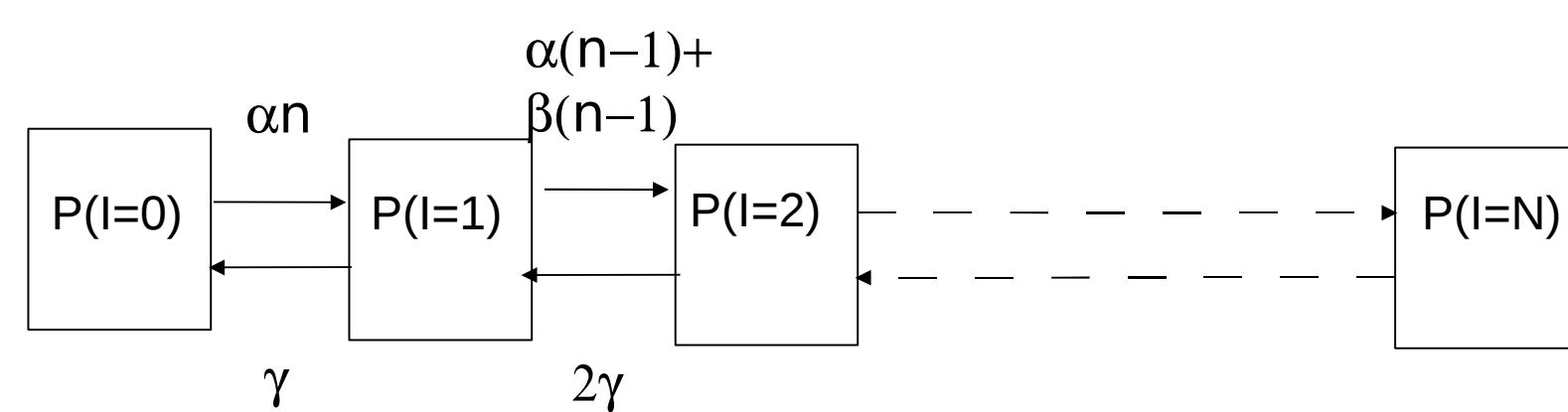


Figure 1 Graphical representation of a system with rates of change. I_0 represents the probability that the system has no infectious individuals, I_1 , that it has 1 infectious individual, etc. The parameters α , β and γ are the rates associated with infection arising from the environment, from an infected individual and recovery, respectively.

$$\frac{dP(I=m)}{dt} = (\alpha(N-m+1) + \beta(m-1)(N-m+1))P(I=m-1) - \gamma m P(I=m) - (\alpha(N-m) + \beta(m)(N-m))P(I=m) + \gamma(m+1)P(I=m+1)$$

Equations 1 The ordinary differential equations that are used to define the rate of change of state of the system over time. The parameters α , β and γ are the rates associated with infection arising from the environment, from an infected individual and recovery, respectively.

The results from solutions to the differential equations at the times at which the group is tested for infection allow us to calculate the likelihood for any parameter set. This approach is used in a standard Metropolis-Hastings method to perform an MCMC analysis of the data. The method is represented in Figure 2.

	P(I=0)	P(I=1)	P(I=2)	P(I=3)	P(I=4)	P(I=5)	
T0 - true	1	0	0	0	0	0	
T1 - dist.	0.72	0.21	0.052	0.0098	0.0012	8.1E-5	
T1 - true	0	0	1	0	0	0	
T2 - dist.	0.26	0.36	0.25	0.091	0.019	0.0019	
T2 - true	0	1	0	0	0	0	

Figure 2 Demonstration of use of Chapman-Kolmogorov equations in determining likelihood for a group with 5 animals, initially susceptible. At time $t=0$ the group is fully susceptible. Therefore the probability that there are no infectious animals is 1, and the probability that there are 1 or more is zero. At time $t=1$ the probabilities of 0, 1, ... 5 infectious animals are as given in the boxes, for an environmental transmission rate of 0.01, infectious transmission rate of 0.1, and recovery rate of 1. If at this time we test the group with a perfect test and there are in reality 2 positive animals, the likelihood of the parameter set used in the Chapman-Kolmogorov equations is 0.052. If we then test again at time $t=2$ and get a single positive animal we have a likelihood of 0.36. Thus the overall likelihood for the two tests is 0.052×0.36

If the true number of infectious animals is reset at each test time to be the correct number then the system can be used to determine the likelihood of the second test result, and so on for a number of tests.

This approach results in a (log) likelihood for all observed tests given a known starting value, for any number of tests, taken at any time points. This quantity can then be used as the basis of a Metropolis-Hastings algorithm to produce a chain of MCMC realisations for the model parameters.

The system was tested by simulating data for a number of scenarios, and comparing the output of the MCMC process to the values used to produce the simulated data. This was repeated for a number of parameter sets. Some typical results are shown in Figure 3 and Table 1

It can be seen from Figure 3 that the result of the MCMC iterations have modes at, or near to the value used for simulation. However, there is a very long tail to the right, leading to large values for the mean and median values in Table 1. The Chapman-Kolmogorov functions also allow us to produce maximum likelihood estimators (MLEs) of the parameters in question, and these are reproduced in Table 1.

The likelihood surface for this model has been examined and it shows a small peak in the region of the MLEs, with a long ridge that extends diagonally in both the transmission-recovery plane and the environment-recovery plane, and hence also in the environment-transmission plane. It should be noted that the MLEs, although optimal estimators of the true value, will have large variances and the assumption of asymptotic normality is unlikely to be valid, given the limited amount of information within the data.

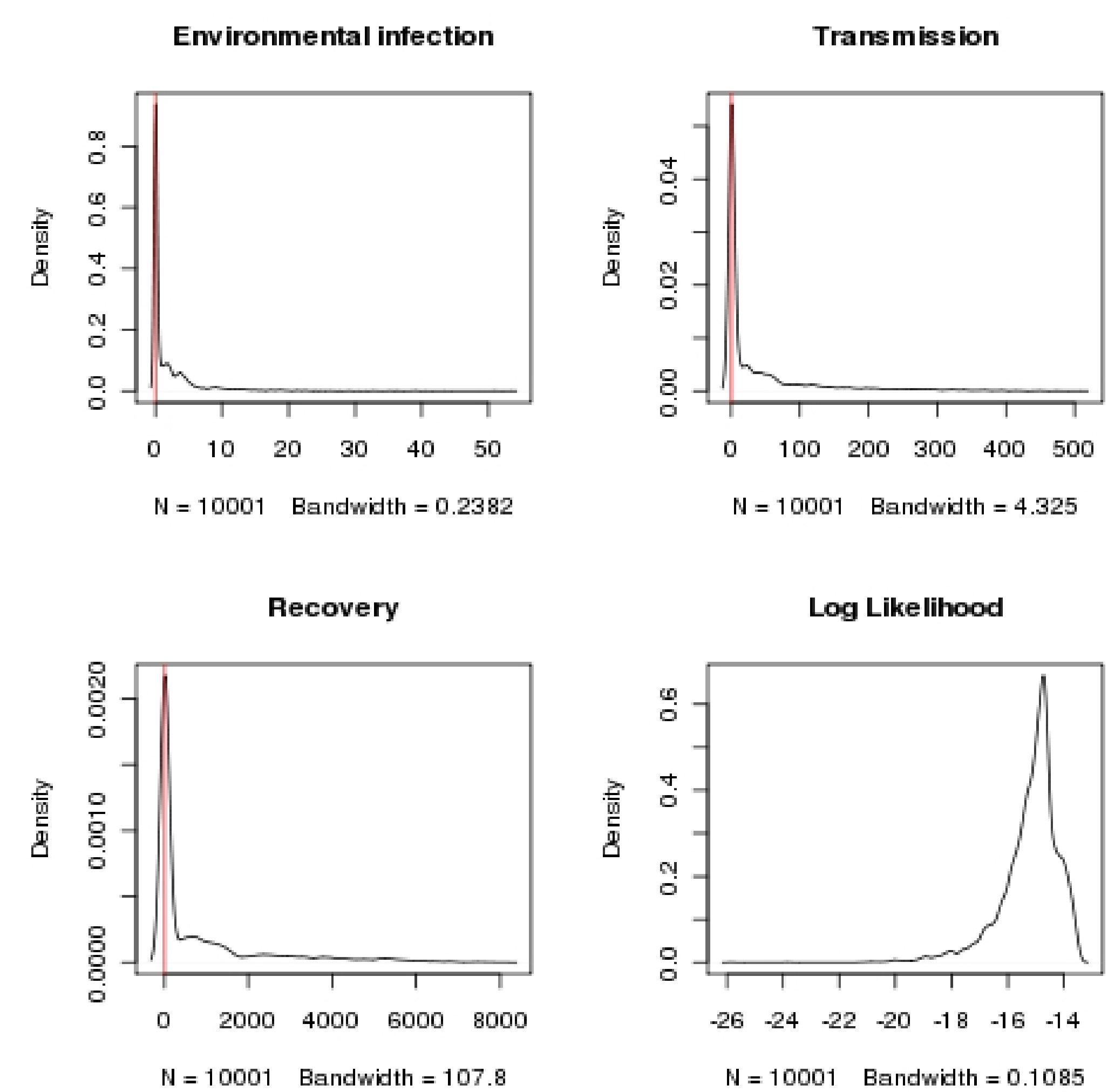


Figure 3 Results from a Metropolis-Hastings algorithm applied to solutions of Chapman-Kolmogorov equations. All parameters have a maximum density (mode) at or near the value used in the simulation, as marked by a vertical red line.

20 animals each tested 30 times

Model	Parameter	True	MLE	MCMC mean	MCMC median
1	α	0.01	0.07	0.08	0.08
	β	0.1	0.06	0.07	0.08
	γ	1	0.7	0.7	0.8
2	α	0.01	DNC	0.02	0.03
	β	0.5	DNC	259	2652
	γ	1	DNC	539	5562
3	α	0.02	0.03	0.01	0.02
	β	0.1	0.1	0.1	0.1
	γ	1	1.1	1.1	1.2
4	α	0.01	0.007	0.005	2
	β	0.1	0.09	0.8	36
	γ	2	2.0	19	821

Table 1 Comparison of values used for simulation and their estimates by the maximum likelihood method (MLE) and MCMC (MCMC mean and MCMC median). Values in blue are close to the values used for simulation; those in red are at least ten-fold away from the values used for simulation; DNC - did not converge.

We are continuing to use simulations to determine the effect of varying the sampling timescale, parameter values, group size and number of tests to determine the effect that these factors have on the estimates for the parameters of the model. It is likely that prior information, or restrictions on the parameter space, reflecting a belief that the timescale upon which tests are performed is comparable to the timescale of disease recovery, will be required to produce robust mean and median estimates for parameters from these models.