



Making Predictions of Bacterial Population Dynamics Using a Metropolis-Hastings Algorithm

Mahathi Tallapragada, Susan Rogowski*



Abstract

- We use the logistic growth equation shown below as a toy model for bacterial population and want to estimate the growth rate (r) and carrying capacity (K), our parameters.
- Noisy or sparse data distracts models from an accurate representation of the parameters
- Coding the MCMC- Metropolis-Hastings allows us to approximate realistic parameters for population dynamics from an initial distribution through numerous iterations
- While the code worked, it was computationally expensive and must be improved to be able to apply it to more complex models

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{K}\right)$$

Methods

- We solved for and coded the logistic growth equation using time mesh and Forward Euler techniques on MATLAB software with the ODE45 solver
- The data was taken from a randomly generated set to model the logistic growth equation.
- We applied the MCMC – MH Algorithm to the logistic growth equation on MATLAB.

Algorithm 1 Metropolis-Hastings Algorithm

Inputs: Initial Parameter Guess (θ_0)

for $i = 1, \dots, N$ **do**
Propose new candidate θ^* from $P(\theta)$ (prior distribution)

Calculate

$$\alpha = \frac{P(X|\theta^*)P(\theta^*)}{P(X|\theta_{i-1})P(\theta_{i-1})}$$

Generate r from uniform distribution $\mathcal{U}(0, 1)$

if $r < \min\{1, \alpha\}$ **then**
 $\theta_i = \theta^*$ (accept guess)

else
 $\theta_i = \theta_{i-1}$ (reject guess)

end if
end for

Results

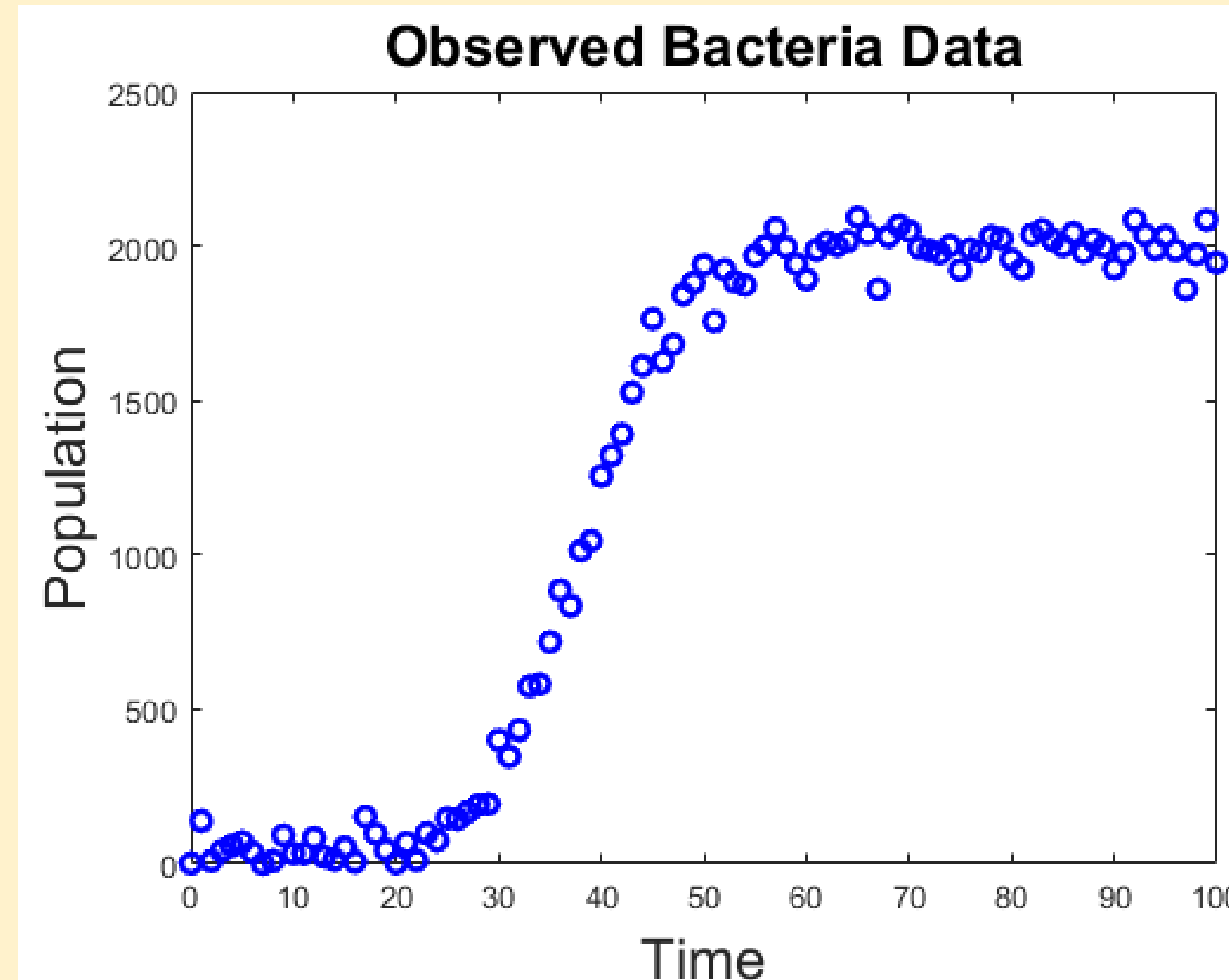


Figure 1: The image above shows the logistic growth model for the generated data.

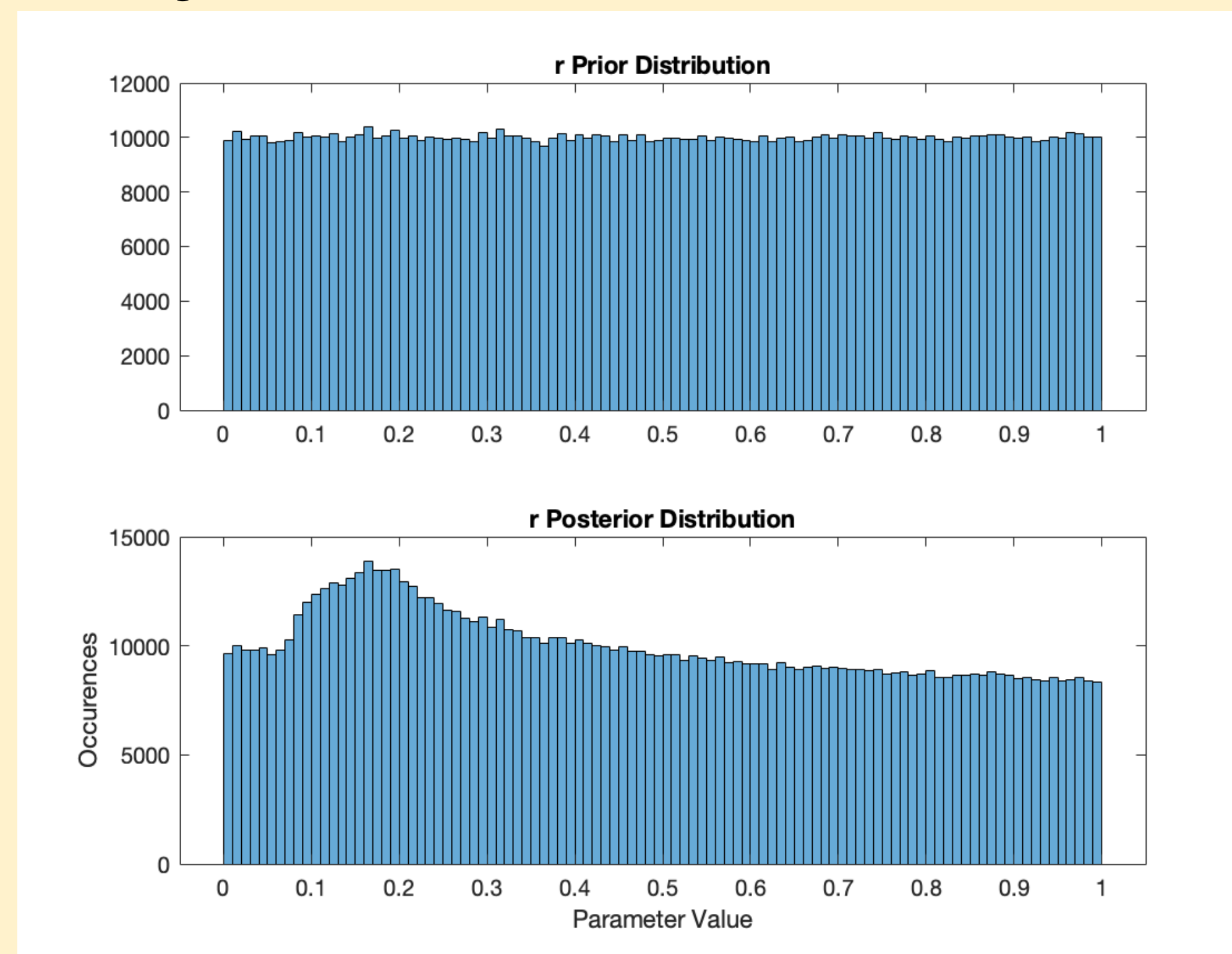


Figure 2: The image shows the r parameter values from the MCMC-MH model, for 1 million runs, with a noticeable peak around 0.2 for the posterior distribution.

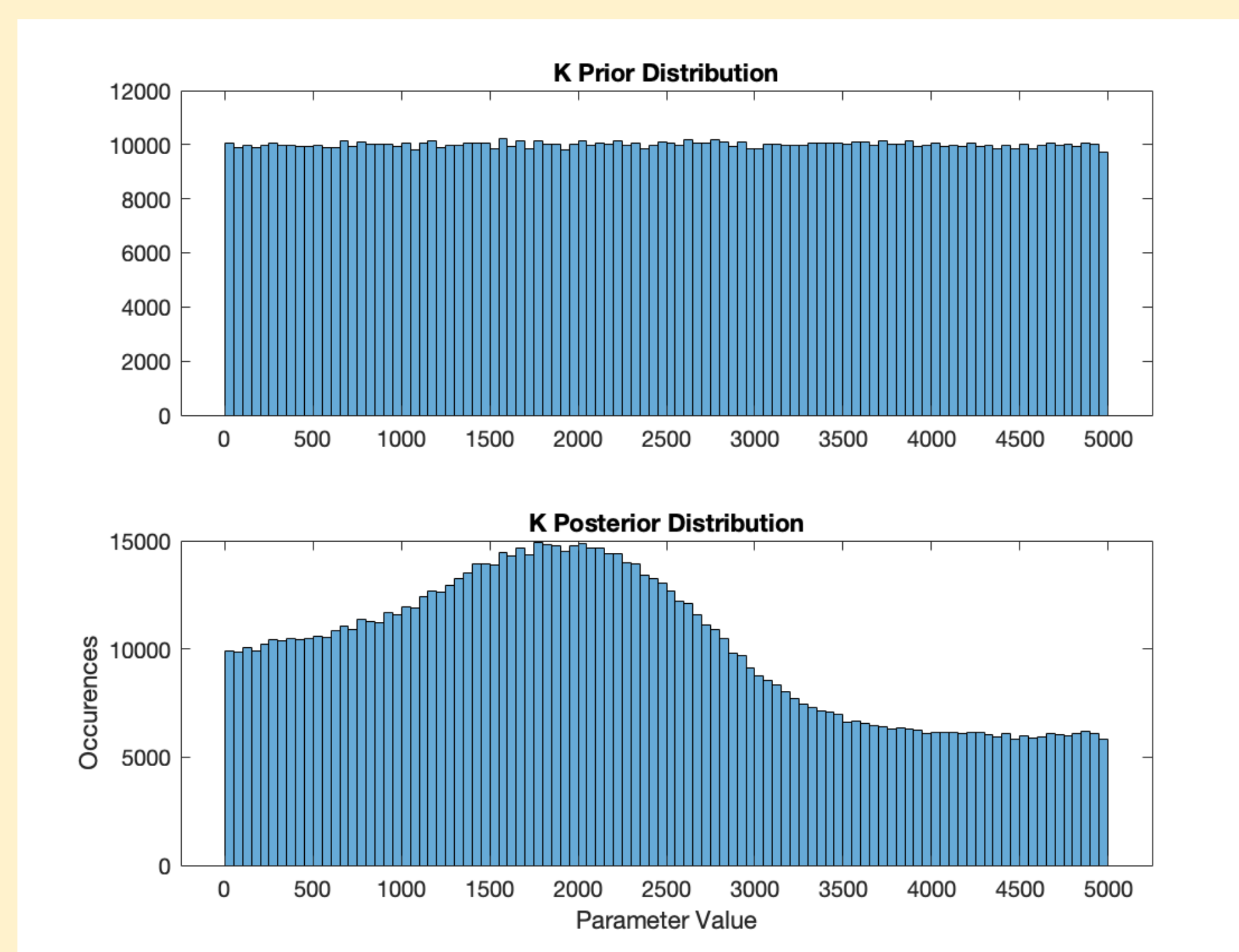


Figure 3: The image shows the K parameter values from the MCMC-MH model, for 1 million runs, with a noticeable peak around 2000 for the posterior distribution.

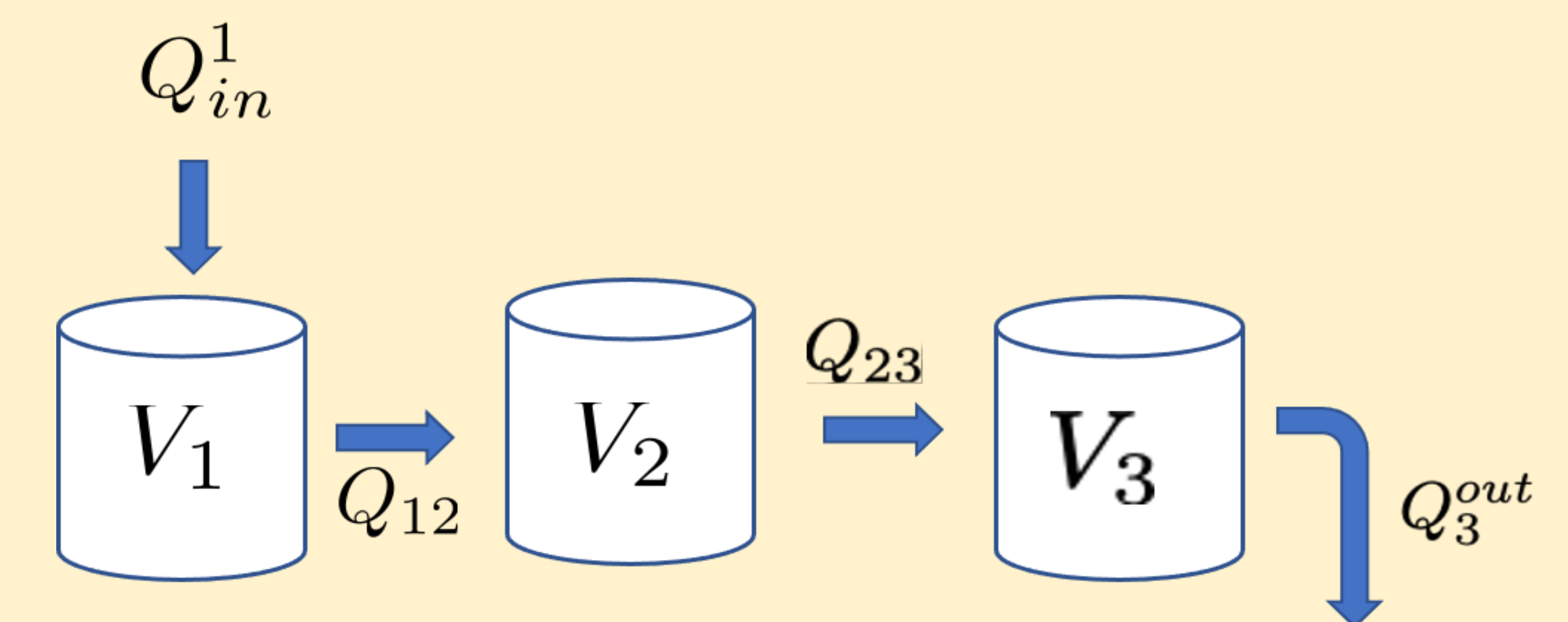
Discussion

- The figure of possible r -values generated by the MCMC-MH algorithm indicates the r -value is most likely in between 0.1 and 0.2
- The figure of possible K -values generated by the algorithm indicates the K -value is around 2000, which is supported by the levelling off of the bacterial population on the logistic growth graph.
- The run time of this code for 1 million runs was around **10.6 hours**; therefore, it must be optimized to become more efficient and less computationally expensive

Future Directions

We will use the logistic growth equation code as a guide to integrate the three-tank problem, where we will work backwards from a known final population to find the initial bacterial population.

Knowing this is useful for being able to further approximate bacterial populations, which can be applied to real-world situations such as wastewater treatment.



References

- Nooranidoost, Mohammad, et al. "Bayesian estimation of Pseudomonas aeruginosa viscoelastic properties based on creep responses of wild type, rugose, and mucoid variant biofilms." *Biofilm* (2023): 100133.
- Edelstein-Keshet, Leah. *Mathematical models in biology*. Society for Industrial and Applied Mathematics, 2005.
- Gelman, Andrew, et al. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.
- Saint-Antoine, Michael. "Markov Chain Monte Carlo (MCMC) for Parameter Estimation (Matlab)." <https://www.youtube.com/watch?v=MbHmgWfXeY&t=959s>

Acknowledgements

I would like to thank my research mentor, Susan Rogowski and the UROP Program at FSU for giving me the opportunity to gain undergraduate research experience.

I would also like to thank Dr. Mohammad Nooranidoost for his much-appreciated insight and guidance.