Efficient likelihood approximation via Gaussian processes: with an application to an existing *Plasmodium vivax* malaria model

The University of Melbourne

Jacob Cumming

May 2024

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Malaria is an infectious disease that poses a significant global health challenge. In 2022, the WHO estimated that malaria had caused over 600,000 deaths that year, most of which were in children under five. The majority of malaria-related deaths are attributable to the *Plasmodium falciparum* species of the parasite. However, the literature is increasingly recognising that the amount of death and severe disease attributable to the *Plasmodium vivax* species is likely traditionally underestimated. The malaria lifecycle is complicated, needing both a vertebrate and mosquito host. In addition, *P. vivax* has a dormant stage that causes relapses.

Various countries with endemic malaria have undertaken significant efforts to reduce the burden of or eradicate malaria. Mathematical disease models are increasingly aiding these efforts by helping to understand the spread of the disease and estimate the effectiveness of possible interventions. Malaria models are complex due to the many staged lifecycle, and in particular, *P. vivax* models need to consider relapses, further complicating the model.

To simulate different scenarios using models, modellers must first calibrate the model parameters to approximate observed disease dynamics, measured by observed data such as case counts or prevalence surveys. Standard techniques to calibrate parameters include maximum likelihood estimation and sampling from a posterior distribution, which both require a likelihood function. As models become increasingly complicated, analytic forms for the likelihood may not exist, or calculating the likelihood may be very computationally burdensome. Some researchers calibrate compartmental models by relying on their deterministic counterparts, which is questionable, as the deterministic model sometimes behaves differently from the stochastic model. Modern likelihood-free techniques, such as approximate Bayesian computation, address this issue but require large model runs to calibrate the parameters.

Drawing on concepts from approximate Bayesian computation, we aimed to improve parameter calibration in malaria models, addressing a recognised need for improvement. We do this by training a Gaussian process to predict how close a model run will be to observed data and extracting a synthetic likelihood.

The thesis follows the following outline: The literature review in Part I discusses fundamental concepts of epidemiological modelling, covering deterministic ordinary differential equation models, stochastic models, and their simulation methods. It then examines malaria and reviews various malaria models. Part I further surveys parameter inference techniques, comparing frequentist and Bayesian approaches, and ends with discussing likelihood-free techniques, particularly approximate Bayesian computation. Approximate Bayesian computation then motivates using Gaussian pro-

cesses to develop a synthetic likelihood, which can be used in place of the unknown true likelihood to use the traditional parameter calibration techniques in complex models. Finally, Part I discusses using Bayesian acquisition functions to train the Gaussian process efficiently.

Part II applies and extends this methodology by calibrating parameters specific to a *P. vivax* model using synthetic observed data. The results and discussion section validates the method and demonstrates that the parameters that produced the observed data are recoverable. Finally, the thesis concludes with a discussion of the findings and outlines avenues for future research.

Chapter 2

Conclusion

Malaria, a significant global health challenge, continues to burden the global health system. Mathematical disease models are increasingly being harnessed to alleviate this burden. This thesis explores ways to maximise the impact of epidemiological models by ensuring they accurately simulate public health scenarios, with a specific focus on the complex issue of *P. vivax* malaria.

The complicated lifecycle of *P. vivax* poses a tough hurdle in disease modelling. This complexity is problematic during the parameter calibration stage, where traditional methods that rely on being about to compute a likelihood are not viable. Likelihood-free methods, while effective, come with a high computational cost due to repeated and inefficient model runs. This has led some researchers to calibrate models using deterministic approximations, failing to capture stochastic models' uncertainty. There is an obvious need for a better calibration methodology.

This research used Gaussian processes to approximate the distribution of the discrepancy function used in approximate Bayesian computation for any set of parameters. The true likelihood function was approximated by manipulating the Gaussian process to create a synthetic likelihood. Empirical evidence demonstrated that this methodology successfully recovered parameters of a *P. vivax* model given simulated data while mitigating the computational overhead. Additionally, this thesis laid out possible further extensions to the methodology.

In conclusion, this thesis has demonstrated the plausibility of a new, more robust method of calibrating malaria models, particularly for *P. vivax.* The use of Gaussian processes and synthetic likelihoods has proven effective in overcoming the infeasibility of traditional calibration methods and the computational cost of likelihood-free calibration methods. Future research should focus on refining these techniques and exploring their application to other infectious diseases. These advancements are crucial for developing more effective interventions and achieving malaria eradication.