

PHD THESIS

IMPERIAL COLLEGE LONDON

SCHOOL OF PUBLIC HEALTH

Neural network based surrogates for
scalable Bayesian inference on a complex
malaria model

Author:
Giovanni Charles

Supervisors:
Prof. Azra Ghani
Prof. Samir Bhatt
Dr. Seth Flaxman

February 7, 2025

Abstract

Bayesian inference on complex models is challenging, as long simulation times and large observational data sets can make computation infeasible and posterior samplers can struggle with high-dimensional parameter spaces with complex posterior geometries. However, several active areas of research offer methods to address some of these challenges, such as Hamiltonian Monte Carlo (HMC) to use gradient information for robust posterior sampling, Stochastic Variational Inference (SVI) for faster inference which is robust to noisy data, Neural Networks (NNs) and Neural Density Estimators (NDEs) for fast and scalable approximations of deterministic functions and stochastic processes. However, they are yet to find adoption for the more complex, realistic, long running models in the sciences, such as climate, astrological or disease transmission models. Fast and scalable inference for these models would result in timely scientific understanding and policy making which could have a far-reaching impact on our societies. Newer methods of inference are of particular interest in global malaria research, where models have grown in complexity. These often include hundreds of parameters which influence an intricate combination of non-linear dynamics, need to be fit to several decades of global prevalence and incidence data, and simulations for each location take many hours to execute. Researchers have developed these models to meet the growing demands for malaria control and elimination in a widening landscape of potential interventions and confounding dynamics-such as new vaccines, insecticide resistance and a changing climate. Bayesian inference of model parameters affects the predictive power of malaria transmission forecasts when estimating burden or planning intervention strategies, which impacts hundreds of millions of lives globally. New methods are needed to ensure that researchers can perform inference on malaria transmission models as they grow to capture the nuances of modern malaria. This thesis documents my investigations into new methods which could address these needs.

0.1 Extended abstract

Bayesian inference on complex models is challenging, as long simulation times and large observational data sets can make computation infeasible and posterior samplers can struggle with high-dimensional parameter spaces with complex posterior geometries. However, several active areas of research offer methods to address some of these challenges, such as Hamiltonian Monte Carlo (HMC) to use gradient information for robust posterior sampling, Stochastic Variational Inference (SVI) for faster inference which is robust to noisy data, Neural Networks (NNs) and Neural Density Estimators (NDEs) for fast and scalable approximations of deterministic functions and stochastic processes. To date, these methods have been proven on models with a handful of parameters and observations, such as the Susceptible-Infected-Recovered (SIR) and Lotka-Volterra models. However, they are yet to find adoption for the more complex, realistic, long running models in the sciences, such as climate, astrological or disease transmission models. Fast and scalable inference for these models would result in timely scientific understanding and policy making which could have a far-reaching impact on our societies.

Newer methods of inference are of particular interest in global malaria research, where models have grown in complexity. These often include hundreds of parameters which influence an intricate combination of non-linear dynamics, need to be fit to several decades of global prevalence and incidence data, and simulations for each location take many hours to execute. Researchers have developed these models to meet the growing demands for malaria control and elimination in a widening landscape of potential interventions and confounding dynamics - such as new vaccines, insecticide resistance and a changing climate. Bayesian inference of model parameters affects the predictive power of malaria transmission forecasts when estimating burden or planning intervention strategies, which impacts hundreds of millions of lives globally. New methods are needed to ensure that researchers can perform inference on malaria transmission models as they grow to capture the nuances of modern malaria. This thesis documents my investigations into new methods which could address these needs. In the first chapter, I elaborated on the problem of Bayesian inference for parameter estimation, the latest malaria transmission models and provided an overview of related research in HMC, SVI and NDE.

In the second chapter, I investigated the effectiveness of HMC based inference on a simplified malaria model, the “Bi-stable equilibrium” published in (Griffin, 2016). My objectives were to see if (1) it was feasible to implement this malaria model in an auto-differentiable framework, as it had not been done before and required sophisticated programming constructs, (2) HMC was feasible in terms of convergence diagnostics and inference time, and (3) what prior selection would be necessary to improve inference. I evaluated the inference by constructing synthetic data, performing inference and developing prior and posterior predictive checks to assess the inferred immunity, prevalence and incidence relationships. I found that the auto-differentiable model was implementable in JAX (a flexible framework for hardware accelerated and auto-differentiable computing) and I could efficiently produce gradients for a sensitivity analysis (using the less commonly used forward-mode auto-differentiation). When used in conjunction with HMC, the sampler converged within 100 warmup samples requiring over 5 hours to produce 100 samples from the posterior on an 8-core CPU. The diagnostics I developed helped me to tune the prior distribution for faster inference and posteriors which were closer to the true parameters. This workflow illustrates the benefits of modern auto-differentiation and inference frameworks, especially for models which may have previously been cumbersome or infeasible to re-implement in an auto-differentiable way, and may encourage researchers to adopt them for future modelling studies. However, my analyses revealed that many of the model parameters remained un-identifiable despite my best efforts in prior selection, prompting future research into model design for this particular malaria transmission model.

In the third chapter, I investigated the effectiveness of NN-based surrogate models on the same inference problem in order to speed up inference and avoid the re-implementation of existing models into an auto-differentiable framework. My objectives were to see if (1) an NN-based surrogate could approximate the malaria model within a reasonable number of samples, (2) if the surrogate’s outputs could be used for accurate parameter inference, and (3) if this alternative workflow would be quicker than that proposed in the previous chapter. I used a Multi-Layer Perceptron (MLP) as the surrogate for the Bi-stable equilibrium and investigated two parameter sampling strategies (prior and Latin-Hypercube), two output structures (full outputs and fixed outputs - limited to a specific inference problem), and three posterior sampling strategies (NUTS, SVI and annealed SVI). The posteriors which were estimated using each surrogate were compared to the reference

posterior estimated in the previous chapter. I found that NN-based surrogates approximate best with targeted parameter sampling strategies but are unaffected by higher-dimensional outputs. Inference using NUTS was quicker when using a surrogate but would not converge. The only posterior sampling method to have promising results was annealed SVI, which had similar predictive performance to the reference posterior and required 40 minutes to execute (a 7.5x speed improvement). However, the posterior estimated using annealed SVI still had significant discrepancies from the reference posterior. This was the first time annealing had been applied to ameliorate issues in high-dimensional Bayesian inference using SVI. Annealed SVI could be a step towards enabling robust, fast and scalable NN-based inference for high-dimensional inference problems.

In the fourth chapter, I explored the effectiveness of sequence-to-sequence (Seq2Seq) surrogates in enabling Bayesian inference for different observational data with varying time horizons. My objectives were to see if (1) a Seq2Seq surrogate could approximate dynamic model of malaria transmission - the Agent-Based Model (ABM) published in (Griffin, Ferguson, and Ghani, 2014), (2) it could extrapolate temporal dynamics beyond the time horizon in the training set, and (3) it could be used to infer a critical malaria burden parameter (baseline Entomological Inoculation Rate) given a time-series of prevalence data in the region of Kolda, Senegal. I used a Recurrent Neural Network (RNN) to implement the Seq2Seq surrogate, an LHS-based sampling strategy for the input parameters and a NUTS sampler to sample the posterior. I found that the surrogate was able to closely approximate trajectories which were twice as long as its training data for arbitrarily chosen sites without retraining. In this low-dimensional inference setting, the LHS parameter sampling strategy and NUTS posterior sampling strategy were satisfactory - contrary to the high-dimensional inference attempted in the previous chapter. Moreover, the sampler could produce 1000 samples of the posterior in under 12 minutes using the surrogate, despite the underlying Agent-Based model requiring 29 minutes per simulation. These findings suggest that pre-trained surrogates of computationally expensive models could be made available to facilitate cheap, rapid inference for low-dimensional inference problems in a variety of scenarios. This would be particularly impactful in research institutions with limited computing resources, such as those in the Global South which are disproportionately affected by malaria. However, the surrogate did exhibit high approximation errors in a minority of cases without a clear pattern, and so care should be taken if the surrogate is to be used for general purpose modelling.

In the fifth chapter, I investigated the effectiveness of the Seq2Seq surrogate on a high-dimensional inference problem. My main objective was to see if the Bayesian inference workflow which I introduced in chapter three would converge in a realistic, challenging inference scenario - i.e. with a data set of 369 prevalence and incidence studies between 1985 and 2013, 84 parameters to infer, and a likelihood function derived from a stochastic trajectory produced by the ABM I introduced in chapter four. A secondary objective was to see if my posterior estimate entailed differing immunity dynamics to those inferred by most recent published estimates for the same model (*ibid.*). To sample the posterior, I used the annealed SVI method I devised for chapter three. As for the surrogate, I extended the RNN architecture introduced in chapter four to model stochastic outputs. I found that the surrogate approximation of the ABM outputs were at times unsatisfactory, sometimes mistaking seasonal variation in model outputs for noise. However, the inference produced a posterior which was predictive of the observed data, according to surrogate simulations. The diagnostics suggest that more research is required to improve the approximation error of the surrogate in the estimated posterior space and make inference more reliable. My preliminary posterior estimates suggest that there may be a significant update in our understanding of malaria immunity in response to relaxing previous restrictive modelling assumptions and including more recent observational prevalence and incidence data.

These findings show that there is great potential in the application of NN-based surrogates for complex malaria transmission models and scientific models more generally. They can speed up inference and produce estimates which exhibit strong predictive performance. In low-dimensional settings, the surrogates can be trained using sampling strategies which make few assumptions about the inference problem, making them more general, re-usable and appropriate for pre-training. In high-dimensional settings, annealed SVI is a powerful, hitherto unexplored, technique for ameliorating inference when using NN-based surrogates. However, there were important limitations in the approximation quality of NNs and NDEs for estimating the model outputs of the most complex malaria transmission models. Further research is needed to improve it to a standard where they can be used for reliable inference.

0.2 Statement of originality

I, Giovanni Charles, declare that the following thesis is my work under the supervision of Professor Azra Ghani, Professor Samir Bhatt and Dr Seth Flaxman. I have indicated in the text where information, data or images have been derived from other sources. Below I have acknowledged research suggestions and comments, not from my supervisors, which have also directed this work:

Dr Peter Winskill, Dr Robert Verity and Dr Patrick Walker pointed me towards the Bayesian inference problems I tackled in this thesis and gave feedback on my methods for validating the Bayesian models, including synthesising observational data and diagnostic plots.

Dr Peter Winskill and Dr Richard Fitzjohn reviewed my code for the *individual* and *malariasimulation* code bases and Dr Peter Winskill suggested methods for testing and resolving differences between *malariasimulation* and a legacy model.

Dr Jeffrey Imai-Eaton and Dr Ewan Cameron advised me to dig deeper into Bayesian inference challenges instead of my initial plan to explore a wider range of mathematical modelling problems.

0.3 Copyright declaration

The copyright of this thesis rests with the author.

Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

0.4 Acknowledgements

I would like to thank my supervisors for giving me the space for exploratory research and their advice when refining it. I am very grateful for their time and attention. On our first group call, Seth suggested that I look into neural networks to speed up Bayesian Inference. The ensuing rabbit hole engaged me for several years and led to the work you see here. Seth's light-hearted and continued support kept me digging deeper. Sam's rigour and clarity kept me on the straight and narrow. Azra's detailed editing kept my thesis text tight. This research would not have happened without the Wellcome Trust and Medical Research Council funding which Azra and Richard Fitzjohn had won.

I would very much like to thank Peter Winskill. He helped me put together and test the malaria model used in this thesis, and his support, joy and malaria knowledge were an immense help.

Collaborating with Sean Wu has also been a highlight. To have found such a warm, capable, interesting and close collaborator on GitHub fills me with optimism. I have fond memories of our conversations on the design of "Individual" and of other exciting research directions. I hope we meet in person some day.

The malaria group at Imperial College London has been incredibly supportive of my work. My chats and progress updates with the malaria elimination team (Hillary Topazian, Nora Schmit, Matteo Pianella, Katharina Hauck and Anh Pham) have been especially helpful and gave me a much needed routine. Everyone's feedback in the Malaria Group seminars gave me an invaluable impetus and food for thought (in this regard I would particularly like to thank Patrick Walker, Tom Churcher and Joe Challenger).

I am also immensely thankful to my friends and family for their support. Particularly to my parents, my brothers and my closest friends (Chris, Lauren and Josh - The Centipede). They have been the spice of my life and kept me in high spirits. I care for them very deeply.

During my PhD, I suffered a long-term illness. I am grateful to Alex and Scott for nursing me back to some semblance of health. Thanks to them, I made it through the PhD and am once again optimistic about my future.

Contents

0.1	Extended abstract	2
0.2	Statement of originality	4
0.3	Copyright declaration	4
0.4	Acknowledgements	4
0.5	Notation	11
1	Introduction	12
1.1	Bayesian inference for complex models	12
1.1.1	Challenges posed by realistic models	13
1.1.2	Neural networks and Bayesian inference	14
1.2	Malaria and its life cycle	14
1.2.1	The Anopheles vector	15
1.2.2	Clinical malaria	16
1.2.3	Immunity to malaria	16
1.2.4	Malaria burden	17
1.2.5	Malaria control	17
1.2.6	Mathematical models of malaria	18
1.3	Parameter inference for malaria models	20
1.4	Surrogate based parameter inference outside of malaria modelling	21
1.4.1	Gaussian Processes as surrogates	21
1.4.2	Bayes linear emulation for parameter inference	21
1.4.3	Input parameter sampling for surrogate training	22
1.4.4	Neural density estimation	22
1.4.5	Recurrent Neural Networks as surrogates	23
1.5	Contributions	23
2	An auto-differentiable malaria model for efficient parameter inference	25
2.1	Background	25
2.1.1	Observational data	25
2.1.2	Hamiltonian Monte Carlo (HMC)	26
2.1.3	Auto-differentiation	27
2.1.4	Related work on Bayesian inference of malaria model parameters	27
2.1.5	Contributions	28
2.2	Methods	29
2.2.1	Bi-stable equilibrium for a <i>Plasmodium falciparum</i> malaria transmission model	29
2.2.2	Implementation of the equilibrium solution	31
2.2.3	Parameter inference formulation	31
2.3	Bayesian inference workflow	33
2.3.1	Diagnostics on the prior distribution	34
2.3.2	Diagnostics on the posterior samples	34
2.4	Discussion	35
3	Fast black-box Bayesian inference using neural-network based surrogates	43
3.1	Background	43
3.1.1	Surrogates for Bayesian inference	44
3.1.2	Neural networks as surrogates	44
3.1.3	Neural density estimation	45
3.1.4	Stochastic Variational Inference (SVI)	46

3.1.5	Annealing in SVI	47
3.1.6	Contributions	48
3.2	Methods	48
3.2.1	Surrogate formulations	48
3.2.2	Posterior samplers	51
3.2.3	Sample space refinement	52
3.2.4	Surrogate training	53
3.3	Results	54
3.3.1	Surrogate approximation quality	54
3.3.2	Agreement of surrogate posterior estimates with non-surrogate based inference	56
3.3.3	Convergence diagnostics of posterior samplers	57
3.3.4	Speed of inference	59
3.3.5	Surrogate-based inference diagnostics	61
3.4	Discussion	62
3.4.1	Stochastic Variational Inference for black box inference	62
3.4.2	No-U-Turn Sampling for black box inference	63
3.4.3	Curating the input and output space of the surrogate	64
3.4.4	Complex models for black-box inference	64
4	Seq2Seq surrogates for recreating disease transmission dynamics over time	66
4.1	Background	66
4.1.1	Temporal models for malaria transmission	66
4.1.2	Challenges in Agent-Based Model (ABM) development	67
4.1.3	Challenges with parameter inference for ABMs	68
4.1.4	The site calibration problem	68
4.1.5	Related work on parameter inference using ABMs	68
4.1.6	Improving surrogates with inductive biases	70
4.1.7	Contributions	71
4.2	Agent based modelling method	72
4.2.1	Individual: a framework for ABM development	72
4.2.2	Malariasimulation: an Agent-based malaria transmission model	73
4.3	Parameter inference method	81
4.3.1	Surrogate model	81
4.3.2	Sampling strategy	82
4.3.3	Surrogate model architecture	83
4.3.4	Sampling training data	83
4.3.5	Site calibration formulation	84
4.4	Results	84
4.4.1	Surrogate evaluation	84
4.4.2	Parameter inference	85
4.5	Discussion	86
4.5.1	A surrogate for general analysis	86
4.5.2	Generality of the workflow	87
4.5.3	Simplified ABM Development with “Individual” and “malariasimulation”	88
5	Bayesian inference of high-dimensional Agent Based Model parameters	91
5.1	Background	91
5.1.1	Inference using stochastic models	92
5.1.2	Related work	92
5.1.3	Contributions	93
5.2	Methods	93
5.2.1	An Agent Based Model of malaria transmission	93
5.2.2	Site data	95
5.2.3	Surrogate model	98
5.2.4	Surrogate training	98
5.2.5	Parameter Inference	99
5.3	Results	100
5.3.1	Validation of approximation quality	100
5.3.2	Inference of model parameters from synthetic data	101

5.3.3	Inference of model parameters from observational data	101
5.3.4	Execution time for inference	103
5.4	Discussion	103
5.4.1	Surrogate approximation quality	103
5.4.2	Parameter inference	104
5.4.3	Computational performance	104
6	Discussion	113
6.1	Summary of contributions and findings	113
6.2	Limitations and future directions	116
6.3	Conclusions	119

List of Figures

1.1	The life cycle of the malaria parasite, (Warrell and Gilles, 2002)	15
1.2	A global map of malaria vectors, (Sinka et al., 2012)	15
2.1	Selected prior and posterior distribution comparisons	37
2.2	Model based predictive checks	39
2.3	Sensitivity Analysis	40
2.4	Correlation Analysis	42
3.1	A Multi-layer Perceptron	45
3.2	A Normalising Flow	46
3.3	The complete surrogate-based inference workflow	49
3.4	Output activation function	53
3.5	Mean Squared Error of standardised surrogate outputs in the prior distribution	55
3.6	Approximation error on prior distribution disaggregated by the outputs used for likelihood estimation	55
3.7	Approximation error on the surrogate posterior distribution	55
3.8	Approximation error on the surrogate posterior distribution disaggregated by round	56
3.9	Concurrence of surrogate posteriors with underlying posterior at different sample sizes	57
3.10	Concurrence of surrogate posteriors with underlying posterior for each parameter	58
3.11	Log-likelihood of the observational data per round	58
3.12	Log-likelihood of the surrogate posterior predictive distribution	58
3.13	Analysis of surrogate gradients	59
3.14	$\hat{r}$ for surrogate methods using NUTS	60
3.15	Loss for surrogate methods using SVI	60
3.16	Time for training a surrogate	61
3.17	Time for performing inference	61
3.18	A comparison of sensitivity analyses using surrogate models	62
3.19	Posterior predictive checks for surrogate-based inference	63
3.20	Surrogate predictive error plots	65
4.1	An LSTM cell	71
4.2	Individual Simulation Loop	74
4.3	A high-level overview of “malariasimulation”	75
4.4	Human Infection States	76
4.5	Predictive error plot of seq2seq surrogate	84
4.6	Posterior predictive check showing baseline EIR in Kolda, Senegal	85
4.7	Comparison of model fit using seq2seq surrogate and ABM	85
4.8	16 cases with the lowest MSE from the historic benchmark	86
4.9	16 cases with the highest MSE from the historic benchmark	88
5.1	Visualisation of Insecticide Treated Net usage in Africa	97
5.2	Visualisation of global seasonality parameters	97
5.3	Plot of recorded Entomological Inoculation Rate (EIR) against estimated $PfPr_{2-10}$	98
5.4	A Parametric Density Estimating Recurrent Neural Network	99
5.5	Approximation quality of density estimating ABM surrogate	106
5.6	Qualitative analysis of density estimating surrogate trajectories	107
5.7	Posterior estimates for immunity parameters based on synthetic observations	108
5.8	Posterior estimates for site transmission parameters based on synthetic observations	109

5.9	Bayesian p-value plots for surrogate based inference of ABM parameters	110
5.10	The immunity functions for prior and posterior estimates of ABM parameters . . .	111
5.11	Prevalence and incidence trajectories for selected sites sampled from the prior and estimated posterior	112

List of Tables

1	Table of mathematical symbols and their descriptions	11
1.1	Description of variables and parameters in the Ross-Macdonald model	19
2.1	Synthetic site parameters and observations	32
2.2	Prior distribution of the intrinsic parameters	33
2.3	Marginal estimates of the intrinsic parameters	35
2.4	Mean Squared Error (MSE) of posterior estimates	36
2.5	Mean Squared Predictive Error (MSPE) reduction due to prior tuning	36
3.1	The combination of sampling strategies considered for the range and domain of f_{fixed} or f_{eq}	51
3.2	Sampling parameters for the intrinsic parameters	52
3.3	Surrogate output ranges	53
4.3	Default parameter values for “malariasimulation”, set as the mean estimates from Griffin, Ferguson, and Ghani (2014)	80
4.1	Model variables for malariasimulation	82
4.2	Processes for malariasimulation	89
4.4	Hyper-parameters used in the grid search. The selected parameter is emboldened .	90
4.5	The timings per simulation for 10,000 simulations on the training and validation parameter sets	90
5.1	Parameters for Anti-malarial drugs	94
5.2	Representation of countries in sub-Saharan Africa in observational data	96
5.3	Prior distribution of the Agent Based Model parameters	100
5.4	Posterior estimates for ABM model parameters	102

0.5 Notation

Below are definitions for symbols which occur throughout this thesis. For notation which is specific to Imperial College London’s malaria transmission model, please refer to [Table 4.3](#).

Symbol	Description
$\mathcal{X}$	The input parameter space for a model.
$\mathcal{Y}$	The space of valid outputs for a model.
f	Original model to be emulated; $f : \mathcal{X} \rightarrow \mathcal{Y}$
$\tilde{f}_\phi$	Surrogate model parameterised by hyper-parameters ϕ
θ	Parameters of the original model to be inferred
ϕ	Hyper-parameters of the surrogate model
$\hat{\phi}$	Optimal surrogate hyper-parameters
$\mathbf{x}$	Input variables (model parameters), $\mathbf{x} \in \mathcal{X}$
$\mathbf{y}$	Output variables (model outputs), $\mathbf{y} = f(\mathbf{x})$
$\mathcal{L}(\mathbf{x}, \mathbf{y}, \tilde{f}_\phi)$	Loss function measuring the discrepancy between $\mathbf{y}$ and $\tilde{f}_\phi(\mathbf{x})$
$p(\mathbf{x})$	Weighting function over the input space $\mathcal{X}$
N	Batch size for neural network training
Malaria specific symbols:	
Λ	Transmission level parameter, used to characterise a site
η	Demographic parameter, used to characterise a site
pos_M_i	Probability of detection by microscopy for age group i
inc_i	Incidence rate for age group i
b_i	Probability of pre-erythrocytic infection for age group i
φ_i	Probability of clinical infection for age group i
q_i	Probability of detection for age group i
f_{eq}	Equilibrium model function; $f_{\text{eq}} : (\Lambda, \eta, \theta) \mapsto (\text{pos_M}_i, \text{inc}_i, b_i, \varphi_i, q_i)$
f_{fixed}	Fixed surrogate model; a surrogate functional form which tailored to a specific inference problem
Seq2seq specific symbols:	
T	Number of time steps in sequence data
$\mathbf{x}_{1:T}$	Sequence of input variables over time steps 1 to T
$\mathbf{x}_i$	Input variables at time step i
$\mathbf{x}_{<i}$	Sequence of input variables preceding time step i
$\mathbf{y}_i$	Model output at time step i ; $\mathbf{y}_i = f(\mathbf{x})_i$
$\mathbf{T}$	Index set over time steps
Probabilistic surrogate symbols:	
$Y_{\mathbf{x},t}$	Random variable representing ABM outputs at time t given parameters $\mathbf{x}$
$P_{\mathbf{x}}$	Probability distribution of ABM outputs given $\mathbf{x}$
$\mu_{\mathbf{x},t}^\phi$	Mean of surrogate model output at time t given $\mathbf{x}$ and ϕ
$\sigma_{\mathbf{x},t}^{2\phi}$	Variance of surrogate model output at time t given $\mathbf{x}$ and ϕ
Φ	Cumulative distribution function of the standard normal distribution

Table 1: Table of mathematical symbols and their descriptions

Chapter 1

Introduction

The development of complex and realistic models for malaria transmission has been instrumental in the ongoing disease control and prevention efforts. These models help researchers to understand its dynamics (World Health Organization, 2023), assess the burden in data scarce Low and Middle Income Countries (LMICs), evaluate the efficacy of interventions (Bhatt et al., 2015), and plan intricate control measures (Peter Winskill et al., 2017).

However, the utility of these models is often constrained by the computational demands of simulation and the limitations of rudimentary parameter inference methods. Parameter inference in the context of large-scale malaria models have been hampered by long execution times, restrictive modelling assumptions, or their inability to fully capture the dynamics of malaria transmission. These challenges have led to compromises in model complexity and fidelity, which, in turn, impact the reliability and robustness of any insights which researchers make.

I hypothesised that using neural network-based surrogates within the Bayesian inference workflow would help overcome these constraints. By serving as a rapid, high-fidelity approximation of a complex malaria model, a neural network surrogate could drastically reduce simulation times and enable a more thorough exploration of the model parameter space, which has traditionally been computationally prohibitive. The flexibility offered by a fast approximation would allow researchers to relax any restrictive modelling assumptions that have previously restricted the expressive power of malaria models. In doing so, they could construct models which more faithfully represent the intricacies of malaria transmission, its variability across different geographies, and the impact of various intervention strategies.

In this chapter, I will review the research which is related to malaria transmission modelling and surrogate modelling before summarising my contributions to the intersectional field. In [section 1.1](#), I will review why model calibration using Bayesian inference is particularly challenging for complex models. In [section 1.2](#), I will review malaria biology and previously published malaria transmission models. In [section 1.3](#), I will review previously published parameter inference methods for malaria transmission models. Finally, I will summarise my contributions, which culminate in a scalable method for Bayesian parameter inference for a complex malaria transmission model.

1.1 Bayesian inference for complex models

Bayesian inference is an important tool for inferring model parameters in various scientific disciplines. By integrating prior knowledge with observational data, it enables scientists to update their understanding of a system in a clear way. It also allows scientists to quantify the uncertainty in their predictions by outputting a probabilistic posterior distribution. This is particularly important in fields where decision-making is made in critical circumstances or under considerable uncertainty. Bayesian models are especially interpretable since they allow scientists to specify clear modelling assumptions through prior selection and posterior sampling methods, and benefit from a variety of well-established diagnostic methods.

Thus, advances in Bayesian inference methodology would have a wide reaching impact. In infectious disease modelling it has been effective in understanding the control and spread of COVID-19, Ebola and Zika, among other diseases (Flaxman et al., 2020; Adegboye et al., 2021; Charniga et al., 2022). In biology, it has played a role in interpreting biological experimental data and evolutionary trees (Liepe et al., 2014; Cappello, J. Kim, and Palacios, 2023). Macro-economists

have increasingly used it to study market dynamics under uncertainty (Lamperti, Roventini, and Sani, 2018). It is also used in astronomy to analyse data from telescopes, a recent example being inference of the first image of a black hole (Arras et al., 2022).

1.1.1 Challenges posed by realistic models

More formally, Bayesian inference aims to estimate a posterior distribution $p(\theta|y)$ for parameters, θ , and observational data, y . To this end, one defines a joint distribution $p(y, \theta) = p(y|\theta)p(\theta)$ which specifies prior knowledge, $p(\theta)$, as a probability distribution over θ , and a likelihood of observed data $p(y|\theta)$. The alternative methods for Bayesian inference exploit the fact that $p(\theta|y) \propto p(y, \theta)$ to produce samples from the posterior distribution.

In practice, realistic models present a series of significant challenges to Bayesian inference, largely due to the intricate dynamics of these models and the computational demands of simulating and inferring their parameters accurately. These challenges can be split into the computational intensity of realistic models - i.e. difficulties in evaluating $p(y|\theta)$, the complex geometry of posterior distributions - i.e. difficulties in sampling from $p(\theta|y)$, and estimating derivatives for fast inference - i.e. difficulties in evaluating $\frac{\partial}{\partial \theta} p(y, \theta)$.

Many realistic models require extensive computational resources due to their complexity and the granularity of simulation needed. Agent-based models of disease transmission, for example, simulate the interactions of millions of humans (and often households, workplaces or non-human disease vectors) to model the transmission of disease in a population. These models require a great deal of engineering effort to make extensive simulations feasible. Similarly, differential equation-based models face scalability issues as additional realistic details increase the dimensionality of the system, thereby making numerical simulation intractable. Neural networks, proven to be universal function approximators (Hornik, Stinchcombe, and H. White, 1989), have emerged as a solution to speed up simulations across various fields including climate science and fluid dynamics (Kovachki et al., 2022; Sanchez-Gonzalez et al., 2020). These advancements suggest neural networks could significantly reduce the computational burden of traditional simulation methods for other complex systems.

The geometry of posterior distributions introduces another layer of complexity to Bayesian inference. The presence of narrow, skewed, or multi-modal distributions complicates the task for many inference algorithms (Gorinova, Moore, and Matthew D. Hoffman, 2019; Yao, Vehtari, and Gelman, 2021). Sampling methods, particularly Markov Chain Monte Carlo (MCMC) approaches, may exhibit non-converging or incomplete sampling behaviours when faced with such complex distributions. Although derivative-based MCMC methods like the No-U-Turn Sampler (NUTS) (Matthew D. Hoffman and Gelman, 2011) and Hamiltonian Monte Carlo (HMC) (Neal, 2011) offer improvement by efficiently navigating complex geometries, they too can suffer from slow iteration times or remain in localised regions of the posterior space.

Rapid gradient estimation is necessary when implementing the latest MCMC methods, such as HMC and NUTS, which rely on gradient information of model outputs relative to input parameters. For realistic models, particularly those implemented with extensive code-bases, this poses a significant engineering challenge of deriving and maintaining gradient calculation code. Auto-differentiation offers a solution by executing the model within a specialised environment which propagates the gradient of each operation alongside its result, using the chain rule to facilitate the computation of gradients of complicated operations. However, the inclusion of, particularly discrete, stochastic processes in models complicates gradient estimation, as the gradients which are computed for a single simulation may not accurately reflect the expected gradients for a given set of parameters. Traditional approaches like Monte Carlo gradient estimation (Mohamed et al., 2020), although useful, introduce prohibitive computational overheads when combined with parameter inference methods.

Stochastic Variational Inference (SVI) (Matthew D Hoffman, 2013) methods have gained traction for their speed, especially when traditional MCMC methods prove impractically slow. SVI approaches recasts Bayesian inference as an optimisation problem. Instead of sampling from the posterior directly, SVI methods search for a tractable distribution - referred to as the variational model - which closely approximates the posterior. This is done by optimising parameterisations of the variational model to minimise a divergence metric between the variational model and the posterior. The variational model can then be efficiently sampled to produce an approximate posterior. Despite the promising speed improvements, optimisation in SVI remains challenging in realistic

modelling scenarios. SVI has only been proven to converge on smooth, convex objectives (K. Kim et al., 2023). Moreover, the performance of SVI is highly sensitive to choice of variational model, optimiser and objective (Dhaka and Catalina, 2021). In some cases leading to inaccurate and over confident inference (Huang et al., 2018; Mandt et al., 2016).

1.1.2 Neural networks and Bayesian inference

In light of these challenges, Neural Density Estimation (NDE) methods have recently emerged as promising solutions (Cranmer, Brehmer, and Louppe, 2020; Papamakarios, Sterratt, and Murray, 2019; Greenberg, Nonnenmacher, and Jakob H. Macke, 2019; Glockler, Deistler, and Jakob H Macke, 2022). These methods train neural networks to estimate probability densities directly from data, or model simulations in this context. This is particularly well-suited for Bayesian inference frameworks, where deriving a likelihood function from the estimated densities can avoid the difficulties posed by intractable stochastic models. For realistic models, which incorporate a multitude of stochastic processes, the intractability of derived likelihood functions has long been a bottleneck in Bayesian inference. This has traditionally been addressed by methods such as Approximate Bayesian Computation (ABC) and model emulation with Kernel Density Estimators and Gaussian Processes (Lamperti, Roventini, and Sani, 2018; McKinley et al., 2018). However, NDE has proven to be a more scalable, flexible and efficient route to inference than traditional alternatives (Cranmer, Brehmer, and Louppe, 2020). Another advantage of using neural networks in this context is their inherent compatibility with auto-differentiation and stochastic optimisation software.

Despite these promising advances, the application of NDE methods to realistic, computationally demanding models remains relatively unexplored. The majority of research and experimentation with NDE methods to date have focused on smaller, theoretical models to demonstrate the general sampling performance and advantages of NDE for Bayesian inference (Lueckmann et al., 2021). It would fill a gap in research if one were to extend these explorations to more complex and realistic models, where the full spectrum of challenges associated with making realistic inferences can be addressed. Such efforts would validate the efficacy of NDE methods in more demanding scenarios and potentially unlock new avenues for the application of Bayesian inference across a wider range of scientific and engineering domains.

1.2 Malaria and its life cycle

Malaria is a parasitic disease, caused by five known species of the *Plasmodium* genus (*falciparum*, *vivax*, *ovale*, *malariae* and *knowlesi*). However, malaria burden today is predominantly caused by the *Plasmodium falciparum* (*P. falciparum*) and *Plasmodium vivax* (*P. vivax*) parasites (Warrell and Gilles, 2002). Transmission to humans occurs through the bite of an infected *Anopheles* mosquito. Infectious bites introduce parasites, known as sporozoites, into the bloodstream. The sporozoites quickly target and invade liver cells, or hepatocytes, initiating “tissue schizogony”. In tissue schizogony, the sporozoites will transform into merozoites, a form which is capable of rapid asexual reproduction. The build up of merozoites will rupture the hepatocyte, releasing them into the bloodstream and infecting blood cells. The infected red blood cells undergo further cycles of invasion, replication, and rupture. This cycle is called “erythrocytic schizogony”. The release of large numbers of parasites into the bloodstream triggers the body’s immune response, leading to the characteristic symptoms of the disease (*ibid.*). During this blood stage, some parasites differentiate into sexual forms, or gametocytes, ready to be ingested by another *Anopheles* mosquito during a blood meal, thus perpetuating the cycle. The *P. vivax* species is unique in its ability to remain dormant in the liver as hypnozoites. Hypnozoites can re-enter the cell cycle after days or weeks causing relapses of infection.

Upon ingesting gametocytes from an infected human, a female *Anopheles* mosquito facilitates the parasite’s sexual reproduction within her digestive tract. The male gametocytes undergo “exflagellation”, after which they fertilise female gametocytes to produce zygotes. The zygotes then develop into motile ookinetes, which burrow into the mosquito’s midgut wall and form oocysts. Over a period of one to two weeks, each oocyst produces thousands of sporozoites, which, upon the oocyst’s rupture, migrate to the mosquito’s salivary glands. This makes the mosquito capable of transmitting the infection to humans, completing the life cycle of the parasite.

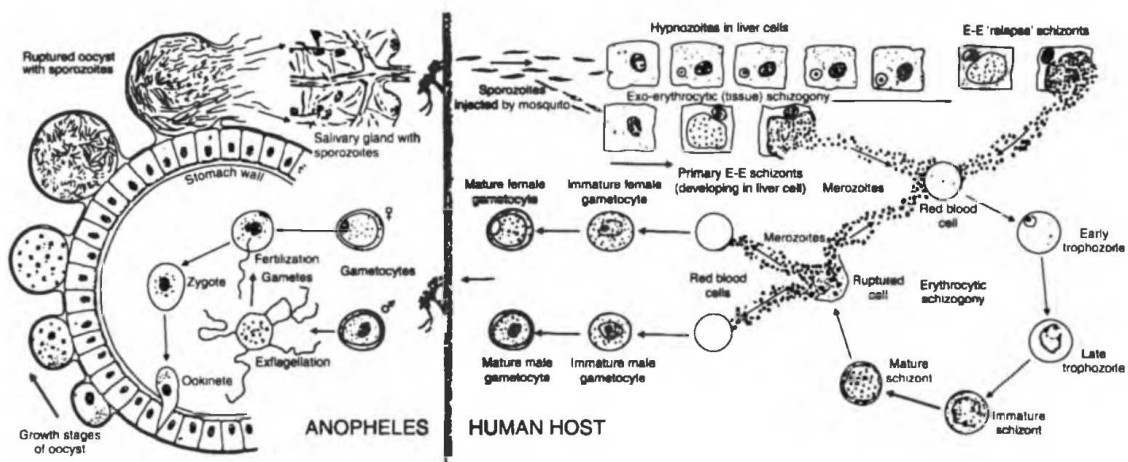


Figure 1.1: The life cycle of the malaria parasite, (Warrell and Gilles, 2002)

1.2.1 The Anopheles vector

The Anopheles mosquitoes, a principal vector of malaria, vary geographically in their density, which in turn influences the dynamics of malaria transmission, see Figure 1.2 for a map of the most abundant malaria vectors. This variation in Anopheles populations impacts malaria transmission through differences in behaviour such as anthropophily (preference for human blood over animal blood) and endophily (the tendency to feed and rest indoors). In sub-Saharan Africa, the most affected area in terms of malaria morbidity and mortality, three types of Anopheles mosquitoes—Anopheles arabiensis, Anopheles gambiae complex, and Anopheles funestus—are predominantly responsible for P. falciparum transmission (Sinka et al., 2012). These species are highly efficient vectors due to their strong preference for feeding on humans.

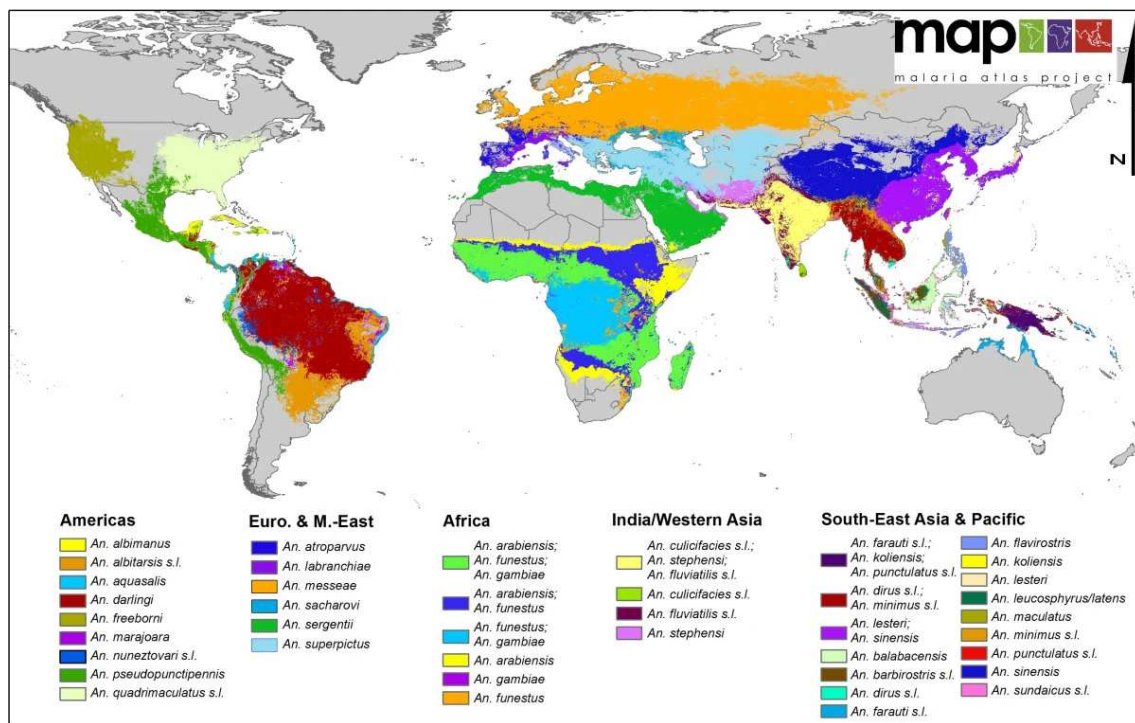


Figure 1.2: A global map of malaria vectors, (Sinka et al., 2012)

The life cycle of Anopheles mosquitoes can be divided into aquatic and adult stages. The aquatic phase includes the development of eggs, larvae, and pupae and can extend up to two weeks, depending on environmental conditions. Once mature, female Anopheles mosquitoes require blood

meals to develop eggs. Following a blood meal, females typically seek environments for resting and egg maturation, a process that takes 2-3 days. After this, mosquitoes will deposit their eggs in standing water, initiating another life cycle. The frequency of oviposition and the development rate of the eggs are both heavily influenced by the ambient temperature (Christiansen-Jucht et al., 2015). For a female mosquito, the interval between ingesting malaria parasites during a blood meal and becoming infectious, known as the extrinsic incubation period, has been estimated to range from 9 to 18 days by (Stopard, Thomas S. Churcher, and Lambert, 2021; Ohm et al., 2018). In this period parasites will develop within the mosquito and they will migrate to the salivary glands, enabling transmission to humans during subsequent feedings.

1.2.2 Clinical malaria

Clinical malaria is triggered by the proliferation of *Plasmodium* parasites within red blood cells, known as parasitaemia. If left untreated, this not only results in the classic symptoms of uncomplicated clinical malaria but has the potential to progress to severe malaria or death. Uncomplicated clinical malaria typically presents with symptoms similar to febrile illnesses, including fever, chills, headaches, and tiredness (Cowman et al., 2016). However, what distinguishes malaria is its periodic recurrence of symptoms, which corresponds with the erythrocytic cycle - as the infected red blood cells rupture in synchrony, releasing new parasites into the bloodstream, patients can experience waves of fever (Warrell and Gilles, 2002). Fortunately, with timely diagnosis and appropriate treatment, uncomplicated clinical malaria is curable.

However, delays in treatment can result from delayed treatment-seeking behaviour by the patient or limited access to healthcare services. Such delays can allow the disease to progress to severe malaria, which is characterised by a range of life-threatening complications, including cerebral malaria, respiratory failure, renal failure, severe anaemia, hypoglycaemia and metabolic acidosis (Cowman et al., 2016). Each of these conditions poses significant risks to the patient and requires urgent and specialised medical treatment to prevent fatality.

Diagnosis is key for the effective treatment of malaria. There are three main methods for diagnosis: Rapid Diagnostic Tests (RDTs), Microscopy and Polymerase Chain Reaction (PCR) - for detailed comparisons see (L. Wu et al., 2015; Madkhali, Ghzwani, and Al-Mekhlafi, 2022). Microscopy is the preferred method of diagnosis as it can be used to identify the species of *Plasmodium* and produce an estimate of parasite density. RDTs can provide a fast and cheap diagnosis, with similar sensitivity, without a clinical laboratory which makes it convenient for rural settings. PCR tests are the most expensive and sensitive of the three making them appropriate for quality control. However, in Low and Middle Income Countries PCR tests are limited.

1.2.3 Immunity to malaria

On first exposure to malaria, individuals are highly susceptible to illness. However, in regions where malaria is endemic, repeated exposure to the parasite results in an accumulation of a natural immunity which significantly reduces the risk of clinical malaria. In endemic settings, the prevalence of parasites is concentrated in children, with a decreasing prevalence with age after young adulthood (Oshagbemi et al., 2023; Makenga et al., 2020). This trend suggests a gradual acquisition of immunity, which not only reduces the likelihood of clinical symptoms but also diminishes the severity of the disease when it does occur. The patterns of mild and severe malaria suggest different immune responses at play across the various stages of the parasite's lifecycle. Although the mechanisms of immunity have been extensively studied, major gaps in our understanding remain (Jong et al., 2021; Barua et al., 2019; Langhorne et al., 2008; Pohl and Cockburn, 2022).

Pre-erythrocytic immunity targets the early stages of malaria infection, before the parasites infect red blood cells. It is characterised by an inflammatory response mediated by cytokines and an adaptive immune response through the recognition of parasites, likely sporozoites, by B and T cells. These immune cells identify and attack the parasite before it is able to establish itself in the liver.

However erythrocytic, or blood stage, immunity is thought to be where most natural immunity against malaria is developed. The immunity, acquired through repeated exposure, primarily targets the merozoites and intraerythrocytic parasites (parasites inside red blood cells). Antibodies produced by the immune system bind to free merozoites and intraerythrocytic parasites, neutralising them, preventing further infection of red blood cells and clinical symptoms in the host.

Anti-gametocyte immune responses target the sexual stages of the parasite within the human host, reducing the number of gametocytes capable of infecting mosquitoes (William E. Collins, Geoffrey M. Jeffery, and J. M. Roberts, 2004). One known mechanism of anti-gametocyte immunity is to attach antibodies to the male gametocytes in the human host in order to disrupt fertilisation after they are ingested by a feeding mosquito (Rener et al., 1980). Anti-gametocyte immunity has been identified as a “bottleneck in the malaria life cycle” (P. E. Duffy, 2021). Since gametocytic parasite densities are found to be relatively low, anti-gametocyte immunity has been of particular interest to researchers seeking a “Transmission-Blocking Vaccine” (TBV).

1.2.4 Malaria burden

The burden of malaria remains a significant global health challenge, with an estimated 249 million cases and 608,000 deaths attributed to the disease in 2022 alone (World Health Organization, 2023). Geographically, malaria burden is concentrated in Africa, accounting for 95% of all malaria cases (World Health Organization, 2020). The youngest have the worst outcomes, with children under five years of age representing 77% of malaria deaths in 2020 (*ibid.*). The predominant cause of these cases and deaths is *Plasmodium falciparum*, responsible for 99.7% of malaria cases in Africa (*ibid.*).

To understand the epidemiology of malaria and evaluate control strategies effectively, three measures of malaria burden are commonly used:

Clinical Malaria Incidence This is the number of clinical malaria episodes within a specific time frame. An episode is typically defined as a new patient with a fever coupled with a positive diagnostic test. Incidence can be determined through active or passive case detection methods. Active case detection involves routinely monitoring a pre-enrolled cohort over the time frame, allowing for an accurate representation of malaria incidence by capturing cases as they occur. In contrast, passive case detection uses existing healthcare systems to identify incoming patients and their contact networks, which, although practical, is known to underestimate the true incidence of malaria due to reliance on symptomatic individuals seeking treatment (Rios-Zertuche et al., 2021).

Parasite Prevalence Representing the proportion of the population infected with malaria at any given time, parasite prevalence is determined through cross-sectional surveys. This measure includes asymptomatic infections, offering a broader perspective on the infection’s spread. However, the prevalence is influenced by the diagnostic methods used, with RDTs and microscopy detecting fewer low density infections compared to PCR testing.

Entomological Inoculation Rate EIR estimates the average number of infectious mosquito bites received by a person per year. This rate is subject to considerable variability, influenced by the methods used for mosquito capture and the estimation of the sporozoite rate, as not all mosquito bites result in transmission. The EIR’s variability is further compounded by environmental heterogeneity and seasonal factors, such as rainfall and temperature, which affect mosquito breeding and survival rates.

All of these metrics are heavily influenced by the vectorial capacity, the rate of contact between infectious mosquitoes and susceptible humans. However, a myriad of confounding factors, such as demographics, seasonality, human habitats, land use, diagnostic methods and wider malaria controls can obscure their precise relationship. Understanding these relationships are essential for predicting and mitigating malaria burden, and so have been the subject of extensive research (Griffin, Ferguson, and Ghani, 2014; Griffin, 2016; Cameron et al., 2015; Reiker, Golumbeanu, et al., 2021).

1.2.5 Malaria control

Many initiatives have been set up to control malaria. Whilst eradication (i.e. complete elimination of the parasite worldwide) is generally considered the ultimate goal, most national programmes strive for local elimination (defined as no ongoing locally-acquired infections whilst allowing for some importation and secondary transmission) (World Health Organization, 2020; Chen et al., 2018; L. Roberts and Enserink, 2007). By 2017, the malaria map had been restricted to the Americas, Asia and the Western Pacific and Africa, where it was most prevalent (Chen et al.,

2018). The World Health Organisation (WHO) established ambitious targets for 2030 (World Health Organization, 2021), including the reduction of incidence of malaria cases and mortality rates by 90% and achieving malaria elimination in at least 35 countries.

However, malaria cases have been on the rise since 2015 (World Health Organization, 2020), due to a range of factors such as COVID-19 (*ibid.*), reduction in funding, environmental changes and insecticide and anti-malarial resistance. This trend was exacerbated in 2020, which saw an alarming increase of 14 million cases and 68,000 deaths compared to the previous year, thanks to many humanitarian crises and the COVID-19 pandemic (*ibid.*). Meeting the WHO targets will require significant improvements in and efficient distribution of current control measures.

The most widespread and effective option so far has been the Insecticide Treated Net (ITN) (World Health Organization, 2023), particularly the Long Lasting Insecticide-treated Net (LLIN). ITNs are hung over beds to protect people while they are sleeping by repelling or killing mosquitoes which attempt to feed. The majority of households in sub-Saharan Africa (70% in 2022) (*ibid.*) have at least one ITN.

Indoor Residual Spraying (IRS) is an alternative form of vector control where an insecticide is applied to the indoor surfaces of a household. Insecticide spray repels or kills mosquitoes providing protection for a large proportion of the day. IRS was estimated to protect less than 6% of the population at risk in 2022 (*ibid.*).

Antimalarials play an important role in the treatment and prophylaxis of malaria, targeting the blood-stage parasites. Previously, Chloroquine and Sulfadoxine-Pyrimethamine (SP) were the frontline treatments, but an increasing parasite resistance to these drugs has necessitated a shift to artemisinin-based combination therapies (ACTs) (Okell et al., 2014; Nguyen et al., 2023). The quality and accessibility of health care services can vary greatly in countries with high malaria transmission, with surveys suggesting that the percentage of patients who received ACTs after seeking care varies from 30% to 93.4% (World Health Organization, 2020).

Additionally, chemoprevention strategies have been adopted in some countries, distributing antimalarials like SP to vulnerable sub-populations during peak transmission seasons. Seasonal Malaria Chemoprevention (SMC) is a strategy which distributes SP to children a few months before the seasonal peak of transmission to maximise the effectiveness of the strategy (World Health Organisation, 2022). Intermittent Preventive Treatment of pregnant women (IPTp) protects women and neonates who are at risk of complications in pregnancy due to malaria or death (*ibid.*). Intermittent Preventive Treatment of infants (IPTi) protects infants who are yet to develop an immunity to malaria and therefore pose an increased risk (*ibid.*).

Vaccines which target the pre-erythrocytic stage of the malaria parasite induce long-term immunity against sporozoites. To date, two pre-erythrocytic vaccines (RTS,S and more recently R21) have been approved as an intervention to protect children (World Health Organization, 2023). Due to a constrained supply of RTS,S, vaccination has only been distributed in Ghana, Kenya and Malawi and Côte d'Ivoire (World Health Organization, 2023; University of Oxford News, 2024). However, the global vaccine coverage is expected to grow as the new R21 vaccine relieves the supply issues.

1.2.6 Mathematical models of malaria

Mathematical models are integral to monitoring the global burden of malaria, providing key global health stakeholders insights into the efficacy of control programs and the potential impact of new strategies. These models simulate the complex interactions between human and mosquito populations, the dynamics of parasite transmission, and the effectiveness of intervention strategies across various settings. The earliest models are referred to as Ross-MacDonald models, which were developed by Ronald Ross, George MacDonald and Alfred Lotka alongside others during the Global Malaria Elimination Programme in 1955 to 1969, see (D. L. Smith, Battle, et al., 2012) for a history of its development. Among the notable models designed for contemporary modelling and intervention planning are those developed by the Swiss Tropical and Public Health Institute (OpenMalaria) (T. Smith, Ross, Maire, Chitnis, et al., 2012), the Institute for Disease Modelling (EMOD DTK) (P. A. Eckhoff, 2011), and Imperial College London (referred to here as the Griffin model) (Griffin, Hollingsworth, et al., 2010).

The Ross-MacDonald family of models were the first to quantitatively facilitate the prediction and control of malaria transmission. They were the first to relate transmission to vectorial capacity and/or a basic reproduction number R_0 , which is now widely used in epidemiology, and to uncover

the dynamics of “super-infection”, where human infections are prolonged by repeated exposure to infectious mosquitoes (D. L. Smith, Battle, et al., 2012). Ross-MacDonald models are characterised as a system of differential equations which describes (1) the force of infection towards humans during blood feeding, (2) the increase of parasite density in the human host, (3) the force of infection towards mosquitoes during blood feeding, and (4) the extrinsic incubation period of the parasite in the mosquito host. I have included an example Ross-MacDonald style model in Equation 1.1 with an explanation of the variables and parameters in Table 1.1. These models often make simplifying assumptions about transmission dynamics such as uniformly distributed blood feeding rates in the human population, a constant mosquito density and human populations are closed to birth and migration. As malaria transmission models have developed, these assumptions have been relaxed and models have become more realistic to meet the need for large scale applicable projections.

$$\begin{aligned}\frac{dx}{dt} &= mabz(1-x) - rx \\ \frac{dy}{dt} &= ax(1-y-z) - ax_{t-v}(1-y_{t-v}-z_{t-v})e^{-gv} - gy \\ \frac{dz}{dt} &= ax_{t-v}(1-y_{t-v}-z_{t-v})e^{-gv} - gz\end{aligned}\tag{1.1}$$

Symbol	Description
x	Proportion of infected humans
y	Proportion of exposed mosquitoes
z	Proportion of infected mosquitoes
m	Mosquito density
a	Biting rate of mosquitoes
b	Probability that an infectious bite infects a human
c	Probability that a mosquito is infected after biting an infectious human
r	Recovery rate of infected humans
g	Death rate of mosquitoes
v	Extrinsic incubation period

Table 1.1: Description of variables and parameters in the Ross-Macdonald model

The Griffin model consists of an agent-based model of human populations coupled with a differential equation-based model for mosquito populations. It allows for detailed parameterisation based on demographics and the heterogeneity of mosquito biting rates. The model simulates the progression of human infections through discrete stages (including sub-patent, asymptomatic and clinical infection stages), incorporating an immunity model which recreates known patterns of incidence and prevalence based on age and exposure (Griffin, Ferguson, and Ghani, 2014). This model can be configured to simulate a variety of interventions, including Insecticide Treated Nets (ITNs), Indoor Residual Spraying (IRS), Seasonal Malaria Chemoprevention (SMC), Intermittent Preventive Treatment in infants (IPTi) and pregnant women (IPTp), larval control, as well as pre-erythrocytic and transmission-blocking vaccines. I used two variants of this model, the Bi-stable equilibrium and the Agent-Based Model, for my experiments and describe them in detail in subsection 2.2.1 and subsection 4.2.2. However, the inference methods which I present could be applied to any dynamic systems model as they make no assumptions which are specific to this use case.

The OpenMalaria model, while offering similar features to the Griffin model in terms of supported interventions, distinguishes itself in its modelling of infection and the use of an ensemble of sub-models to predict outcomes. Instead of simulating discrete infection stages, OpenMalaria simulates a continuous parasite density for each individual in the model. The OpenMalaria model includes several model variants for simulating the probability of clinical infection based on parasite density, including the “base pathogenesis model” (T. Smith, Ross, Maire, Christophe Rogier, et al., 2006) and the “Müller pathogenesis model” (I. Müller et al., 2009). The ensemble approach allows for OpenMalaria to incorporate multiple mathematical formulations for natural immunity levels, transmission heterogeneity, and susceptibility to co-morbidities.

EMOD DTK differs from the Griffin and OpenMalaria models by integrating more sophisticated mosquito ecology and immunity models. The EMOD DTK model allows the researcher to

parameterise extra environmental variables such as temperature and humidity, which influence larval and parasite development within mosquitoes. Similarly to OpenMalaria, EMOD DTK also models parasite densities in human individuals continuously. However unlike OpenMalaria, the strength of the human immune response is modelled to influence infection duration and the recurrence of clinical disease due to the erythrocytic cycle (P. Eckhoff, 2012). The human population model within EMOD DTK is also highly customisable, allowing for the simulation of multiple spatially distributed sub-populations with distinct demographics and transmission rates.

These models allow researchers to quantitatively assess the outcomes of various intervention combinations across different transmission settings, a task that would be impractical, if not impossible, through randomised controlled trials alone. Such models are invaluable for projecting the impacts of funding changes, as illustrated by the US President’s Initiative report, which used the Griffin model to estimate the potential effects of a 44% funding cut on malaria interventions in Africa (Peter Winskill et al., 2017). They also help researchers to interpret and generalise trial data (Thompson et al., 2022; Schmit, Topazian, Natama, et al., 2024; Stuckey et al., 2014; T. A. Smith et al., 2019), support cost-effectiveness analyses (Topazian et al., 2023; Penny et al., 2016) and the prioritisation of global resources for malaria control (Schmit, Topazian, Pianella, et al., 2023; Reiker, Chitnis, and T. Smith, 2019). These models collectively enhance our understanding of malaria dynamics and intervention strategies, providing a robust framework for planning and optimising the global initiatives against this disease.

1.3 Parameter inference for malaria models

Malaria models need to be calibrated to data in order to estimate the many biological and epidemiological parameters which are used for forecasting and planning. The large parameter spaces make searching for the most likely sets challenging. Also, Bayesian methods for calibration, like Markov Chain Monte Carlo (MCMC), for estimating probability densities over the parameter space require large numbers of simulations. For Agent Based Models (ABMs), such as the contemporary malaria transmission models listed above, long execution times are a considerable barrier.

For example, (Griffin, Ferguson, and Ghani, 2014) published Bayesian estimates for the parameters of the Griffin model by calibrating the model to pre-intervention prevalence and incidence data in 31 sites. The published estimates revealed the relationship between prevalence and incidence in different age groups and the proportion of cases which are likely missed by passive case detection. The authors used a bi-stable equilibrium version of the model which made simplifying assumptions about the transmission dynamics in order to reduce the execution time for inference.

Researchers working on the EMOD DTK model avoided the issues of execution time by estimating parameters to sub-components of the model separately. For example, parameters describing the relative types of larval habitat and temperature in the mosquito ecology component of the model were fit to local mosquito abundance data (P. A. Eckhoff, 2011), whereas immunity parameters in the pathogenesis component were fit to malariatherapy¹ data (P. Eckhoff, 2012). Estimating parameters for each sub-component is more convenient as each inference problem operates on a low dimensional parameter space. However, this inference method precludes the possibility of interdependencies between the sub-components.

More recently, surrogate models have been used to infer the parameters of entire models by speeding up their execution time. Surrogate models, or emulators, can quickly provide model outputs for a given parameter set, often approximately and/or for a limited parameter space, in order to speed up the evaluation of a likelihood function. For example, (Cameron et al., 2015) used a kernel regression based surrogate model to emulate the relationship between a parameter set (including treatment probability and an EIR curve) and the resulting prevalence and incidence curves produced by three ABMs (EMOD DTK, OpenMalaria and the Griffin model). The surrogate could then be used in place of the computationally expensive ABMs in conjunction with MCMC to estimate the relationship between incidence and prevalence.

In a more recent example, a surrogate model was used for Bayesian optimisation, in (Reiker, Golumbeanu, et al., 2021). In Bayesian optimisation, the authors optimise the OpenMalaria model’s parameters to minimise an objective function which quantifies the discrepancy between

¹malaria therapy data was collected, before 1963, from studies where participants were infected with malaria to treat other illnesses. The datasets included dynamic observations of parasite counts after infection across the different stages of the parasite’s development cycle (W. E. Collins and G. M. Jeffery, 1999; William E. Collins and Geoffrey M. Jeffery, 2002)

a transmission model and observed data. Since the objective function is computationally expensive to evaluate, the authors employed a Bayesian strategy where they modelled a prior over the objective function and collected evaluations of the objective function which would efficiently update their posterior estimate. The efficiency of each evaluation was dictated by an “acquisition function”. The objective is modelled using a surrogate, typically a Gaussian Process, as they are typically robust for uncertainty quantification. Bayesian optimisation with a carefully designed acquisition function would likely benefit from fewer ABM evaluations than the kernel regression method described above. However, (ibid.) reported a bottleneck in inference which was caused due to the poor scaling properties of Gaussian Processes with respect to the dimensionality of model parameters and the size of the observational data set. Furthermore, it was infeasible to extend their approach to produce probabilistic parameter estimates due to these scalability issues.

1.4 Surrogate based parameter inference outside of malaria modelling

Beyond malaria modelling, the use of surrogate models to accelerate parameter inference has gained significant traction in various domains. The challenge in surrogate modelling is to approximate complex, computationally intensive models, with minimal simulations, striking a balance between accuracy and efficiency. In this section, I review some of the most commonly employed surrogate approaches, their trade-offs and suitability for different types of inference.

1.4.1 Gaussian Processes as surrogates

Gaussian Processes (GPs) have emerged as a versatile and expressive tool for accelerating computationally expensive analyses. Their ability to model non-linear and stochastic functions, coupled with robust uncertainty quantification, makes them particularly appealing for surrogate modelling.

One well known application of GPs as surrogates is in Bayesian optimisation (Gramacy, 2020), where the goal is to identify optimal input parameters for a model based on an objective function. The inherent uncertainty quantification of GPs enables acquisition functions, such as "expected improvement", which balance exploration of the parameter space with exploitation of known optimal regions - which is especially useful in reducing the number of evaluations of the underlying model.

GPs have also been used for uncertainty quantification (UQ) for models which do not natively support it (Kennedy and O’Hagan, 2001; Sawe et al., 2024; Marmin and Filippone, 2022). In such cases, researchers fit GPs to functional input-output pairs of a complex model, where inputs vary according to known uncertainties. Whether the underlying model is deterministic or stochastic, the GP surrogate allows for efficient estimation of uncertainty in the model outputs, reducing the need for repeated runs of the intensive model.

Another important use of GPs is in parameter inference (Z. Li and Tan, 2022). Once a GP is fitted to the outputs of a complex model, it can serve as a surrogate likelihood function for inference methods like Markov Chain Monte Carlo (MCMC) and Stochastic Variational Inference (SVI). This approach has proven effective in many cases; however, challenges such as scalability and computational speed have been highlighted in recent studies, particularly for high-dimensional problems or large datasets.

1.4.2 Bayes linear emulation for parameter inference

Bayes Linear emulation has been widely used to approximate model outputs, particularly for approximate Bayesian computation (I. Vernon et al., 2022; Craig et al., 1997; Goldstein and Huntley, 2017; Bower, Goldstein, and Ian Vernon, 2010). This approach centres around a specific formulation for the surrogate:

$$f_i(x) = \sum_j \beta_{ij} g_{ij}(x) + u_i(x),$$

where $f_i(x)$ represents the surrogate approximation of the model output corresponding to the i -th observation. The first term, $\sum_j \beta_{ij} g_{ij}(x)$, uses a set of deterministic functions ($g_{ij}(x)$), often polynomials, to capture the global structure in the model’s behaviour. Each function are weighted

by coefficients (β_{ij}), which are estimated during the fitting process. The second term, $u_i(x)$, accounts for local variations not captured by the global structure. This term is modelled as a stochastic process, using a covariance function to describe the residual dependencies in the outputs. The surrogate is fitted by determining β_{ij} using either expert opinion or data-driven methods, such as Maximum Likelihood Estimation (MLE) or Maximum A Posteriori (MAP) estimation.

Bayes Linear emulators can be used alongside Bayesian inference methods, including Markov Chain Monte Carlo (MCMC) and Stochastic Variational Inference (SVI), to provide a surrogate likelihood function for posterior estimation (Goldstein and Huntley, 2017). More commonly, however, Bayes Linear emulators are used in *History Matching* (McKinley et al., 2018; Goldstein and Huntley, 2017; I. Vernon et al., 2022), an iterative method for identifying plausible parameter spaces for a model which I describe in the following section.

1.4.3 Input parameter sampling for surrogate training

The choice of input parameters for fitting surrogates could be influential to the accuracy of the resulting approximations. It is particularly of concern in high-dimensional spaces where regions of the parameter space could be insufficiently sampled, resulting in surrogates that fail to capture critical behaviours of the underlying model especially with non-linear dynamics.

Straightforward methods for selecting input parameters include space-filling designs, such as Latin Hypercube Sampling (LHS), or sampling from a prior distribution if one is available (McKinley et al., 2018; Andrianakis et al., 2015; G. Charles, Wolock, et al., 2023). However, these approaches can become inefficient for many real-world inference problems. In such cases, most of the parameter space may correspond to inputs that are irrelevant or implausible for explaining the observed data.

An alternative approach is to use History Matching (HM), which uses a heuristic to identify "plausible" regions of the input space (Goldstein and Huntley, 2017; I. Vernon et al., 2022; McKinley et al., 2018). HM compares model outputs and their associated uncertainties to observational data, systematically excluding regions of the parameter space where the model's outputs are not consistent with observations.

History Matching is conducted in "waves", an iterative process in which the plausible region of the parameter space is refined in successive rounds. In the first wave, an initial set of inputs is chosen using space-filling methods or a prior distribution and corresponding simulations are made. Samples are selected as implausible according to an implausibility metric calculated from the simulations. The plausible input space is then resampled for the next wave until the plausible region is sufficiently small for inference.

1.4.4 Neural density estimation

In recent years, Neural Density Estimation (NDE), in particular Normalising Flows (NFs), have become popular surrogates for flexible and efficient inference (Cranmer, Brehmer, and Louppe, 2020; Papamakarios, 2019; Greenberg, Nonnenmacher, and Jakob H. Macke, 2019; Papamakarios, Sterratt, and Murray, 2019; Glockler, Deistler, and Jakob H Macke, 2022). They are particularly attractive as neural networks have become easier to train with auto-differentiation and have been proven to be expressive enough to approximate any continuous probability distribution (De Cao, Titov, and Aziz, 2019), more formally the universal approximation theorem. There is on going research into how to best integrate these surrogates into a Bayesian workflow (Talts et al., 2020; Cranmer, Brehmer, and Louppe, 2020).

Normalising flows are a class of invertible neural networks that transform a simple, tractable distributions (most often a Gaussian distribution) into a complex target distribution by applying a series of transformations. A benefit of normalising flows, over traditional neural networks, is their ability to compute the probability density of any point in the target distribution through the `**change-of-variables formula**`:

$$p_X(x) = p_Z(z) \left| \det \frac{\partial f^{-1}}{\partial x} \right|,$$

where $z = f^{-1}(x)$ maps the target distribution $p_X(x)$ back to the base distribution $p_Z(z)$. This density estimation makes normalising flows particularly well-suited for Bayesian inference, where the goal is to approximate and sample from posterior distributions.

Normalising flows have been used for approximate Bayesian inference in a variety of ways. One is Neural Likelihood Estimation (NLE) where the NF is trained to approximate an unknown likelihood function by learning the relationship between model inputs and outputs (Papamakarios, Sterratt, and Murray, 2019; Glockler, Deistler, and Jakob H Macke, 2022). Once trained, the NF is used in conjunction with traditional inference methods, such as Markov Chain Monte Carlo (MCMC) or Stochastic Variational Inference (SVI), to sample from the posterior distribution. Neural Posterior Estimation (NPE) takes a more direct approach by training the NF to approximate the posterior distribution itself in an amortised fashion (Greenberg, Nonnenmacher, and Jakob H. Macke, 2019). This means the NF is trained to map any given observation in the model output space to a posterior distribution over the input parameters. Once trained, the NF can provide rapid posterior estimates for new observations without additional training.

Loosely analogous to HM, NDE based approaches have often adopted "sequential sampling" to iteratively refine the surrogate on increasingly important regions of the sample space. In the initial round, the surrogate is trained on the prior distribution. In subsequent rounds, the NF is trained on the posterior estimated from the proceeding round.

1.4.5 Recurrent Neural Networks as surrogates

Recurrent Neural Networks (RNNs) have emerged as a natural choice for surrogate modelling in temporal systems, exploiting the auto-regressive relationships encoded in such models. By introducing temporal inductive biases directly into the surrogate architecture, RNN-based approaches can improve approximation accuracy, reduce parameter size, and potentially stabilise training.

Early RNNs, suffered from the vanishing or exploding gradients problem. When training on long sequences, gradients would either diminish to negligible values (vanishing gradients) or grow exponentially large (exploding gradients), leading to difficulties in optimising the network.

To address this, gated RNN architectures such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) were been developed (Hochreiter and Schmidhuber, 1997; Bahdanau, Cho, and Bengio, 2016). These architectures introduce gating mechanisms to selectively control the flow of information through the network, for example *Forget Gates* to control which parts of the previous state are carried forward; *Input Gates* to determine how much new information should be incorporated into the state; and *Output Gates* to decide which parts of the state influence the next output.

In climate and physical sciences, an alternative RNN variant, the Echo State Network (ESN), has gained popularity for time-series prediction tasks (Jirak et al., 2023; McDermott and Wikle, 2019; Doan, Polifke, and Magri, 2020). ESNs address some of the limitations of traditional RNNs by significantly reducing the number of trainable parameters, while maintaining strong predictive accuracy (Lukoševičius and Jaeger, 2009).

An ESN comprises of a computing reservoir, an encoder network and a decoder network. The reservoir is a sparsely connected RNN with fixed parameters. This reservoir acts as a high-dimensional state space that transforms the input signals into a richer representation. The encoder network is trained to map model inputs into the reservoir's state. The decoder Network is trained to reconstruct the desired model outputs from the reservoir's state. Unlike traditional RNNs, the training is focused solely on the encoder and decoder networks, to reduce the number of trainable parameters and stabilise training.

1.5 Contributions

In this thesis, I introduce methods for performing Bayesian inference on complex models of malaria transmission, leveraging neural network (NN)-based surrogates. NN-based surrogates are designed to facilitate rapid and efficient modelling of malaria dynamics as their parameter sets grow in size and complexity, thereby offering novel avenues for both current and future inference challenges. My work includes sampling strategies for improving convergence during the training of these surrogates, diagnostics for surrogate based inference, as well as updated parameter estimates for a complex malaria model. These estimates could provide deeper insights into malaria transmission dynamics and improve confidence in modelling studies aimed at predicting or controlling malaria transmission.

An auto-differentiable malaria model for efficient parameter inference

In the next chapter, I perform derivative-based inference within the context of a malaria transmission model. I adopt and adapt the bi-stable equilibrium model from (Griffin, Ferguson, and Ghani, 2014), implementing it within an auto-differentiable programming framework for the first time. I redefine the priors over the model parameters to reflect updated knowledge and insights derived from a sensitivity analysis. This approach demonstrates the potential of derivative-based methods in enhancing the accuracy and reliability of inferences drawn from complex epidemiological models.

Fast black-box Bayesian inference using neural-network based surrogates

In chapter three, I introduce a workflow for employing NN-surrogates in derivative-based inference, avoiding the need for re-implementing complex models into auto-differentiable languages. Through this workflow, I explore and compare various sampling strategies, examining their impact on the efficiency and convergence of surrogate training. This comparative analysis not only highlights the versatility of NN-surrogates but also sets the stage for their broader application in complex system modelling.

Seq2Seq surrogates for recreating disease transmission dynamics over time

In the fourth chapter, I apply the NN-surrogate workflow to infer malaria transmission parameters of a computationally demanding agent-based model (ABM). This application showcases the surrogate model's ability to generalise to longer time horizons than those encountered during training and to adapt to arbitrary intervention parameters without the need for retraining. To my knowledge, this is the first time rapid Bayesian inference of model parameters has been achieved for a model of this detail, facilitating timely and informed decision-making in malaria control and prevention efforts.

Bayesian inference of high-dimensional Agent Based Model parameters

In the fifth chapter, I apply the NN-surrogate workflow to estimate the immunity parameters of the same ABM. This chapter illustrates how the workflow effectively addresses a high-dimensional inference challenge in a realistic modelling scenario. The resulting parameter estimates demonstrate a high predictive accuracy for real-world incidence and prevalence data and an alternative modality of immunity dynamics than has been previously published using the Griffin model.

The insights gained from the development and application of NN-surrogates in this thesis extend beyond malaria transmission modelling. I present methodologies which hold promise for addressing similar inference challenges across a variety of domains. By demonstrating the feasibility and effectiveness of NN-surrogates in complex system modelling, this work contributes to the broader field of Bayesian inference, offering avenues for future research and application in diverse scientific and engineering disciplines.

Chapter 2

An auto-differentiable malaria model for efficient parameter inference

2.1 Background

In malaria transmission modelling, parameter inference is essential for understanding of malaria dynamics and producing reliable and consistent predictions. Contemporary malaria models, described in [subsection 1.2.6](#), rely on parameter inference to fit simulations to observable data, namely malaria prevalence and incidence data. To faithfully capture these observable data, contemporary models have introduced complex immunity dynamics which are not directly observable. Developing and fitting using these models is particularly challenging since many biological aspects of malaria immunity remain unknown (Pohl and Cockburn, [2022](#); Deloron and Chougnet, [1992](#); Langhorne et al., [2008](#)). The assumed immunity dynamics have implications for the simulated relationship between incidence, prevalence and underlying malaria transmission (Griffin, Ferguson, and Ghani, [2014](#); Cameron et al., [2015](#); Reiker, Golumbeanu, et al., [2021](#)).

Malaria transmission dynamics are inherently complex and non-linear, which poses further challenges for Bayesian inference. This complexity could lead to situations where parameters exhibit multi-modal posterior distributions or become non-identifiable due to uninformative priors and/or the presence of noisy data. Additionally, the dynamics of the model could result in skewed or narrow geometries, resulting in slow convergence times for commonly used Markov Chain Monte Carlo (MCMC) samplers. The design of the transmission model and prior selection during parameter inference would have an important role in influencing the posterior landscape for successful inference.

Observational data is unfortunately limited and noisy adding another layer of complexity to accurate and reliable inference. Global malaria transmission is highly concentrated in Low to Middle Income Countries (LMICs) in sub-Saharan Africa, regions where surveillance systems are still in the development phase. Health Management Information Systems (HMIS) are being deployed across the continent to systematically track malaria cases and treatments from various healthcare systems in a standardised way. Despite these efforts, the recent reports from the World Health Organization suggest that the effectiveness of routine surveillance is compromised by inadequate record-keeping in the private sector and the prevalence of low treatment-seeking behaviours (World Health Organization, [2021](#)). Apart from routine surveillance, research surveys and limited surveillance initiatives have been carried out with various aims. These surveys, however, can provide disjointed or inconsistent information on malaria burden as they each have their own research goals and methodologies.

2.1.1 Observational data

The most accessible observational data for inferring immunity model parameters is the prevalence of parasites. This data is generally collected through cross-sectional surveys (XSS), wherein a large sample from a population is screened for infection over a brief time window. The methodology of these surveys mostly varies with respect to the diagnostic test used (Rapid Diagnostic Tests (RDT), microscopy, or Polymerase Chain Reaction (PCR) tests), target parasites (usually *Plasmodium falciparum* or *Plasmodium vivax* in sub-Saharan Africa), and the age demographics of the subjects

(often younger populations). The Demographic and Health Surveys programme, which subsumes the Malaria Indicator Surveys (MIS), provide the most up-to-date, regularly collected data from such surveys in LMICs, specifically for children aged between 0 and 59 months (Demographic and Health Surveys, 2022). Additionally, (Snow et al., 2017) has provided a meta analysis on prevalence data published in studies between 1900 to 2017.

The clinical incidence of malaria is tracked through Passive Case Detection (PCD) and Active Case Detection (ACD) strategies. PCD involves testing for malaria among sub-populations that seek medical care. Although it can be seamlessly integrated into existing healthcare services, PCD often yields unreliable incidence estimates due to factors like inconsistent healthcare access, criteria for malaria suspicion, and the likelihood of seeking treatment (Rios-Zertuche et al., 2021). Conversely in ACD studies cohorts of patients in high risk areas are regularly monitored for the study period providing unbiased estimates of incidence. However, the frequency of screenings and cohort size in ACD studies varies, leading to differing incidence estimates. (Battle, Cameron, et al., 2015) published a database of ACD studies of malaria cases (and matched prevalence studies) recorded between 1985 and 2013 providing a valuable resource for understanding prevalence and incidence trends.

2.1.2 Hamiltonian Monte Carlo (HMC)

HMC offers a robust method for efficient inference in models which produce complex posterior geometries, (Neal, 2011). Such posteriors are commonplace in Bayesian inference, where the model complexity and noisy observations would likely lead to challenging posterior landscapes. However, there has been little research on applying Bayesian inference to fit contemporary malaria models and no known cases of derivative-based Bayesian inference at the time of writing. By using gradient information of the target distribution, derivative based approaches can achieve better mixing and faster convergence than less sophisticated counterparts, such as Random Walk Metropolis Hastings (RWMH).

In HMC, each chain is characterised by a position (representing a point in the parameter space) and a momentum (related to the potential energy, or the negative log-likelihood), which together guide the sampling process towards regions of higher posterior probability. The Hamiltonian, $H(q, p)$, is defined as the sum of the kinetic energy, $K(p)$, and the potential energy, $U(q)$, where q represents the position and p the momentum:

$$H(q, p) = K(p) + U(q),$$

The kinetic energy is often modelled using a mass matrix, M , which is typically chosen to be diagonal for computational efficiency, although non-diagonal, positive-definite matrices can be employed to better account for correlations in the posterior space at the cost of increased computation:

$$K(p) = p^T M^{-1} \frac{p}{2}$$

The partial derivative of the Hamiltonian governs the trajectory of p and q over time, t . In practice, the leapfrog sampling method is widely used as a numerical approximation using a step size ε , which incorporates a half step to improve stability compared to Euler’s method for the integration of differential equations. The potential and momentum of each particle under the leapfrog sampling scheme is defined below, see (ibid.) for further implementation details.

$$\begin{aligned} p(t + \varepsilon/2) &= p(t) - (\varepsilon/2) \nabla U(q(t)) \\ q(t + \varepsilon) &= q(t) + \varepsilon M^{-1} p(t + \varepsilon/2) \\ p(t + \varepsilon) &= p(t + \varepsilon/2) - (\varepsilon/2) \nabla U(q(t + \varepsilon)) \end{aligned}$$

The No-U-Turn Sampler (NUTS) extends HMC by adaptively determining the step size and the number of leapfrog steps required in each iteration. This mitigates the need for manual tuning - a process that is both time-consuming and computationally intensive. For complex high-dimensional models, it is unlikely that manual tuning will result in optimal parameterisations. NUTS employs a binary tree expansion mechanism, where leapfrog steps are added either forward or backward in time until a maximum depth is reached or a U-turn is detected. A U-turn occurs when the trajectory of samples turns towards itself, indicating inefficiency in further sampling from regions

beyond the current scope of the binary tree. This is detected with a negative scalar product of the position vectors at the extremes of the binary tree $\delta = (q_{end} - q_{start}) \cdot p_{end}$. After each iteration, NUTS adjusts the step size to steer sampling towards the target acceptance rate.

2.1.3 Auto-differentiation

For the above methods, the computation of gradients is required to perform inference, specifically $\nabla U(q)$. However, in complex models, like those used for malaria transmission modelling, calculating these gradients can be challenging. Automatic differentiation (auto-differentiation) significantly simplifies this step. Auto-differentiation is a computational technique for calculating gradients for computations at execution-time, as opposed to onerous manual or symbolic derivations of derivatives. It achieves this by applying the chain rule to propagate gradients through each operation within a computational process. This series of operations and their interrelationships are tracked using an “execution trace.”

The process of calculating gradients can be executed using “forward-mode” auto-differentiation. In this mode, gradients of the input parameters $x \in R^n$ with respect to themselves are trivially calculated at the beginning of the execution trace. These initial gradients are then propagated forward through the sequence of operations, enabling the computation of gradients of intermediate values, v_i , with respect to the inputs, $\frac{\partial v_i}{\partial x}$, throughout the trace. This procedure continues until gradients for the output values $y \in R^m$ are determined. At the end of an execution one will accumulate a Jacobian matrix, $J \in R^{n \times m}$, which can be expressed as:

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \dots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}. \quad (2.1)$$

Forward-mode auto-differentiation is particularly advantageous when dealing with “tall”, i.e. $n \gg m$, Jacobian matrices and for computing Jacobian-vector products. However, it may not be memory efficient for “wide” matrices, as the intermediate gradients, $\frac{\partial v_i}{\partial x}$, could be large.

An alternative approach is reverse-mode auto-differentiation, where gradients are propagated in the opposite direction, from the output back to the inputs. Gradients are calculated after an execution at the end of the execution trace, starting with trivially computed adjoint values, $\frac{\partial y}{\partial v_i}$, for the final operations and works backward to arrive at J . Reverse-mode auto-differentiation is especially preferred in contexts such as deep learning, where the Jacobian matrix is typically “wide”, i.e. $m \gg n$, involving large parameter models.

2.1.4 Related work on Bayesian inference of malaria model parameters

In malaria transmission modelling, there have been many efforts to apply Bayesian Inference for the estimation of model parameters. However, as of this writing, the use of derivative-based inference methodologies to estimate malaria transmission model parameters remains unexplored. This is potentially an area for improvement in the efficiency and accuracy of parameter inference.

Parameters for the Imperial College London’s model of malaria transmission were estimated using Bayesian inference (Griffin, Hollingsworth, et al., 2010; Griffin, Ferguson, and Ghani, 2014). The authors used prevalence and incidence data from 20 countries in sub-Saharan Africa to estimate values for 28 parameters including immunity and random study effects introduced by incidence data collected through Passive Case Detection (PCD) or weekly Active Case Detection (ACD). A similar method was carried out for a P. Vivax model in (M. T. White et al., 2018), where instead observational data was focused on 11 studies in Papua New Guinea.

In order to simulate incidence and prevalence rates for the likelihood function, a theoretical model of malaria transmission in bi-stable equilibrium was used, which is described in (Griffin, 2016). This model calculates the proportion of individuals in disease-free and endemic states which would persist indefinitely without intervention for a given set of parameters, which can be used to derive stable prevalence and incidence rates for a given age group. The benefit of this model is that simulations can be run extremely quickly compared to agent-based simulations. Unfortunately there is no direct evidence for bi-stable equilibria in practice, except perhaps in elimination scenarios (D. L. Smith, Cohen, et al., 2013), and the widespread interventions that are being deployed today makes malaria unsuited to equilibrium modelling in most scenarios. (Griffin,

Ferguson, and Ghani, 2014), used this likelihood under the assumption that all observational data was collected in pre-intervention scenarios.

The posterior for the model parameters were sampled using a Random Walk Metropolis Hastings (RWMH) kernel for MCMC. In RWMH, candidate samples are drawn from a symmetric proposal distribution, typically a Gaussian distribution centred on the current sample. Samples are accepted according to a parameterised acceptance probability. This technique is easy to implement and effective for small inference problems, however can experience low acceptance rates on high-dimensional, complex inference problems.

Parameters for EMOD’s transmission model (McCarthy et al., 2015) were estimated using an incremental importance sampling method (Raftery and Bao, 2010). Here, a data set of parameters is drawn from a prior distribution to begin inference. Each parameter draw is assigned a weight based on the likelihood for each draw, which is calculated by estimating the likelihood of observed data given EMOD’s simulation outputs. In the incremental step, new parameters are sampled from a multivariate Gaussian distribution centred around the most likely previous draw and a covariance matrix estimated from the weighted distances of the previous draws to the prior distribution. For small unimodal, parameter spaces, this method may explore the posterior distribution quickly with a large number of concurrent samples. However, for more complex inference problems, the majority of the samples will likely fall outside of the accepted posterior distribution requiring a large number of iterations (ibid.).

A study by Cameron et al. (2015) explored the relationship between malaria incidence and prevalence by using three contemporary malaria transmission models: EMOD, OpenMalaria, and the Griffin model. They employed a Random Walk Metropolis-Hastings (RWMH) kernel in conjunction with Parallel Tempering (PT) to sample the posterior space for their Bayesian model.

Parallel Tempering is designed to improve the mixing behaviour of MCMC sampling methods. In this approach, multiple chains are run in parallel, each assigned a distinct temperature, T . These chains sample from an augmented log-likelihood function, $\log(p(x))^{-T}$, where chains assigned higher temperatures have greater degree of freedom to traverse the posterior distribution, potentially overcoming local optima and facilitating a more comprehensive exploration. Samples are swapped between adjacent chains with a carefully selected swap probability. Swapping algorithms for PT have attracted extensive research, please refer to Syed et al. (2022) and N. F. Müller and Bouckaert (2019) for recent advances.

2.1.5 Contributions

I developed the first auto-differentiable implementation of the Griffin et al.’s Bi-stable equilibrium model of malaria transmission. At the time of writing, this is the first auto-differentiable modern malaria model. It includes three distinct models of immunity - pre-erythrocytic immunity, clinical immunity, and immunity to detection - that vary with age and account for heterogeneity in biting rates in the human population.

I conducted Bayesian inference on simulated observational data using a No-U-Turn Sampler (NUTS) sampler, specifically the Numpyro implementation (Phan, Pradhan, and Jankowiak, 2019), to validate a derivative-based MCMC approach. To this end, I proposed a set of diagnostics for use with this model along with carefully tuned priors. These priors were tuned based on our latest understanding of malaria transmission dynamics, a sensitivity analysis, model specific diagnostics, an analysis of posterior artefacts - such as multi-modality and non-identifiability, and commonly used MCMC sampling diagnostics. The diagnostics validated the approach, showing fast convergence and healthy mixing, and produced priors which could be used in future Bayesian inference using this malaria model.

This work builds on the established Bayesian inference workflow to estimate malaria transmission model parameters. This is the first application of derivative-based Bayesian inference to a complex malaria model and, prior to this, no comprehensive analysis including on prior tuning, model and sampling diagnostics, and sensitivity analysis for a complex malaria model has been published. This workflow can produce robust inference and delivers new insights into the dynamics of the Griffin model.

2.2 Methods

Firstly I will describe the bi-stable equilibrium model for malaria transmission, which I will also refer to as the “equilibrium solution”. Then I will define a Bayesian inference problem for inferring the core parameters of the equilibrium solution and implementation details for auto-differentiation and parameter sampling.

2.2.1 Bi-stable equilibrium for a *Plasmodium falciparum* malaria transmission model

The equilibrium solution, originally published in (Griffin, 2016), can be used to calculate the age-stratified steady state prevalence and incidence of the malaria transmission model given a constant number of infectious mosquito bites per day, or Entomological Inoculation Rate, parameterised by Λ .

Human transmission is disaggregated into age and heterogeneity compartments. The average age and heterogeneity of biting rates in each compartment are assumed to be exponentially and log-Normally distributed respectively. I discretised the age compartments into 100 age groups, each group a year wide except for the final group which has no upper bound. Heterogeneity was discretised into 5 groups using Gaussian quadrature. The force of infection Λ_{ij} for each compartment depends on the age group i , heterogeneity group j as well as Λ , defined in Equation 2.2. ξ_j and Ψ_i represent the relative biting rates for a heterogeneity and age group respectively, while b_i is the probability of pre-erythrocytic infection given the average immunity level in the compartment. From Λ_{ij} , I calculated the stable average immunity levels, size of each infection state compartment, and the resulting prevalence and incidence in the population.

$$\Lambda_{ij} = \xi_j \Lambda \Psi_i b_i \quad (2.2)$$

where,

$$b_{ij} = b_0 \left(b_1 + \frac{1 - b_1}{1 + \frac{I_B}{I_{B0}} k_b} \right)$$

The human transmission dynamics for the model can be expressed as a system of differential equations, see Equation 2.3, where susceptible humans, S_{ij} , are infected at rate Λ_{ij} . Depending on the probability of clinical infection ϕ_{ij} , humans progress to clinical infection, D_{ij} , or asymptomatic infection, A_{ij} . Clinical infections will progress to asymptomatic before becoming a sub-patent infection, U_{ij} , and recovering to the S_{ij} state. Λ_{ij} also drives super-infection, where infected individuals are re-exposed, prolonging recovery.

Immunity is accrued on each exposure to malaria, reducing the rate of infection. Pre-erythrocytic immunity, I_B , reduces general force of infection. Acquired clinical immunity, I_{CA} , is immunity to clinical infection. Immunity to detection, I_D , reduces the detection of asymptomatic infection to microscopy. All acquired immunity is boosted by exposure to infectious mosquito bites and decays exponentially. Maternal immunity, I_{CM} , provides protection from clinical disease by transferring a mother’s immunity, I_{C20} , at birth which is not boosted on exposure.

The rates of transition between compartments are defined by r_i , the rate of growth from age group i , $r_{D|A|U}$ the rate of exit from the corresponding infection state, and η the average mortality rate in the population.

$$\begin{aligned} \frac{dS_{ij}}{dt} &= \beta_{S_{ij}} S_{ij} + r_{i-1} S_{i-1} + r_U U_{ij} \\ \frac{dD_{ij}}{dt} &= \beta_{D_{ij}} D_{ij} + r_{i-1} D_{i-1} + \phi_{ij} \Lambda_{ij} (S_{ij} + A_{ij} + U_{ij}) \\ \frac{dA_{ij}}{dt} &= \beta_{A_{ij}} A_{ij} + r_{i-1} A_{i-1} + (1 - \phi_{ij}) \Lambda_{ij} (S_{ij} + U_{ij}) + r_D D_{ij} \\ \frac{dU_{ij}}{dt} &= \beta_{U_{ij}} U_{ij} + r_{i-1} U_{i-1} + r_A A_{ij} \end{aligned} \quad (2.3)$$

where,

$$\begin{aligned}
\beta_{S_{ij}} &= \Lambda_{ij} + r_i + \eta \\
\beta_{D_{ij}} &= r_D + r_i + \eta \\
\beta_{A_{ij}} &= \phi_{ij}\Lambda_{ij} + r_A + r_i + \eta \\
\beta_{U_{ij}} &= \Lambda_{ij} + r_U + r_i + \eta \\
\phi_{ij} &= \phi_0 \left(\phi_1 + \frac{1 - \phi_1}{1 + \frac{I_{CA} + I_{CM}}{I_{C0}} k_c} \right)
\end{aligned}$$

By solving Equation 2.3 at equilibrium I can express formulae for the steady state prevalence and incidence in the system. This calculation first defines the average immunity levels, for each compartment, see Equation 2.4.

$$\begin{aligned}
I_{ij} &= \frac{F_{ij} + (r_i + \eta)I_{i-1}}{\frac{1}{d} + (r_i + \eta)} \\
I_{CM_{ij}} &= I_{C20j} \frac{e^{-\frac{i}{d_m}} - e^{-\frac{i+1}{d_m}}}{r_i d_m}
\end{aligned} \tag{2.4}$$

where F depends on the type of immunity:

$$\begin{aligned}
F_{B_{ij}} &= \frac{\xi_j \Lambda_{ij}}{\xi_j \Lambda_{ij} u_B + 1} \\
F_{C_{ij}} &= \frac{\Lambda_{ij}}{\Lambda_{ij} u_B + 1} \\
F_{D_{ij}} &= \frac{\Lambda_{ij}}{\Lambda_{ij} u_D + 1}
\end{aligned}$$

From average immunity levels, I can derive the size of each infection state compartment, S, D, A, U , see Equation 2.5.

$$\begin{aligned}
D_{ij} &= \frac{r_{i-1}D_{i-1} + \phi_{ij}\Lambda_{ij}Y_{ij}}{\beta_{D_{ij}}} \\
A_{ij} &= \frac{r_{i-1}A_{i-1} + (1 - \phi_{ij})\Lambda_{ij}Y_{ij} + r_D D_{ij}}{\beta_{A_{ij}} + (1 - \phi_{ij})\Lambda_{ij}} \\
U_{ij} &= \frac{r_{i-1}U_{i-1} + r_A A_{ij}}{\beta_{U_{ij}} + (1 - \phi_{ij})\Lambda_{ij}} \\
S_{ij} &= \frac{r_{i-1}A_{i-1} + (1 - \phi_{ij})\Lambda_{ij}Y_{ij} + r_D D_{ij}}{\beta_{A_{ij}} + (1 - \phi_{ij})\Lambda_{ij}}
\end{aligned} \tag{2.5}$$

where,

$$\begin{aligned}
Y_{ij} &= S_{ij} + A_{ij} + U_{ij} = \frac{\pi_{ij} - b_{D_{ij}}}{1 + a_{D_{ij}}} \\
a_{D_{ij}} &= \frac{\phi_{ij}\Lambda_{ij}}{\beta_{D_{ij}}} \\
\beta_{D_{ij}} &= r_D + r_i + \eta \\
\beta_{A_{ij}} &= \phi_{ij}\Lambda_{ij} + r_A + r_i + \eta \\
\beta_{U_{ij}} &= \Lambda_{ij} + r_U + r_i + \eta
\end{aligned}$$

Finally, the prevalence and incidence is estimated from the infection state compartments, see Equation 2.6. In this experiment I define prevalence as *P. falciparum* prevalence as detected by microscopy, a common statistic in malaria surveys, and incidence as the number of clinical cases in the population per day. I modelled the asymptomatic infections detectable by microscopy with probability q_i . Whereas sub-patent infections are undetectable by definition and therefore do not

contribute to prevalence (as detected by microscopy). Clinical incidence is calculated using the rate of clinical infection $\Lambda \cdot \phi$ and the sub-population who are susceptible to (super-)infection.

$$\begin{aligned} \text{inc}_{ij} &= \Lambda_{ij} \phi_{ij} (A_{ij} + U_{ij} + S_{ij}) \\ \text{pos_M}_{ij} &= D_{ij} + q_i A_{ij} \end{aligned} \quad (2.6)$$

where,

$$\begin{aligned} f_D &= 1 - \frac{1 - f_{D0}}{1 + \frac{a}{a_D} \gamma_D} \\ q_{ij} &= d_1 + \frac{1 - d_1}{1 + f_D \frac{I_D}{ID_0} \kappa_D} \end{aligned} \quad (2.7)$$

I used Bayesian inference to fit the free parameters in the above equations to observed prevalence and incidence data. I have listed the free parameters, acceptable ranges and the priors which I selected for those parameters in [Table 2.2](#). These parameters describe the infection dynamics due to pre-erythrocytic immunity and clinical immunity and the probability of detection from a microscopy test due to immunity to detection. The parameters affect the simulated initial and maximum immunity levels as well as their rates of accumulation with exposure, which influence the simulated patterns of prevalence and incidence per age group.

The site parameters, used to characterise each location in the observational data, included Λ and η . These allow for the inference to roughly distinguish between environments with different levels of malaria burden and demographics. In later chapters I will use models which support more detailed site parameters, including age varying mortality, seasonality, vector composition and intervention usage.

2.2.2 Implementation of the equilibrium solution

I implemented the equilibrium solution in a numerical computing framework called JAX, see (Bradbury et al., 2018), which supports forward- and reverse-mode auto-differentiation for all standard arithmetic operations as well as control flows such as while- and for-loops on models defined in the Python programming language. JAX also provides functionality for performance optimisations including Just-In-Time (JIT) compilation, vectorisation, parallelisation and GPU execution. JAX's JIT compilation translates most Python functions into an execution graph to be executed on *XLA*, an optimised execution environment for linear algebra. Domain Specific Language (DSL) primitives such as *vmap* and *pmap* allow users to make their code concurrent by either vectorising a function by transforming the XLA execution graph, or distributing tasks different devices respectively. JAX supports CPU core, GPU and Tensor Processing Units (TPU) devices almost interchangeably allowing users to use all available resources.

The recursive immunity values, in [Equation 2.4](#), and state size variables, in [Equation 2.5](#), are calculated starting with the youngest age group where the previously accumulated immunity, $(r_i + \eta)I_{i-1}$ is zero. The immunity levels for the older age groups can be calculated inductively. This iteration is implemented in JAX using a *scan* primitive. The *scan* primitive is a loop construct which supports both forward- and reverse-mode auto-differentiation and benefits from a JIT compiled loop body, whereas standard Python loops are flattened during JIT compilation.

The calculation of prevalence and incidence for each heterogeneity group is vectorised using JAX's *vmap* primitive. Since immunity and state equations for each heterogeneity group are independent they are calculated concurrently with pre-calculated values of heterogeneity, ξ_j .

I published my auto-differentiable implementation of the equilibrium solution as a Python package called *dmeq*, see (G. Charles, 2024b). At the time of writing, this is the first JAX (or hardware accelerated and auto-differentiable) implementation of a malaria model. This implementation is the first step towards more computationally efficient malaria models, which can support more ambitious inference, forecasts and control optimisation research.

2.2.3 Parameter inference formulation

Similarly to previous parameter inference efforts in Griffin, Ferguson, and Ghani (2014) and Cameron et al. (2015), I defined a parameter inference problem in which I estimate model parameters based on observed parasite prevalence and clinical incidence statistics for many sites in sub-Saharan Africa. However, for this exercise I simulated observational data from the model, in

order to focus on inference diagnostics and suitable priors to include in the workflow, instead of data analysis and cleaning.

Let θ be a candidate parameter set, with my prior knowledge of these parameter values listed in Table 2.2. The simulated observations were made at six transmission sites characterised by the EIR, $\Lambda \in \mathbb{R}^{+6}$ and demography (defined as average age, $\eta_{\text{obs}} \in \mathbb{R}^{+6}$). Let $\text{pos_M}_a(\theta, \Lambda, \eta)$ and $\text{inc}_a(\theta, \Lambda, \eta)$ be functions which map model parameters onto detectable parasite prevalence and clinical incidence respectively, where a is an age range for the statistic. In practice, pos_M_a and inc_a are implemented by simulating the equilibrium solution and aggregating the detectable prevalence and incidence at age group a . I selected five age ranges, which loosely correspond to those found in the malaria surveys of detectable parasite prevalence and clinical incidence used in Griffin, Ferguson, and Ghani (2014).

Λ	η	$\text{pos_M}_{[2,10]}$	$\text{pos_M}_{>10}$	$\text{inc}_{<5}$	$\text{inc}_{[5,15]}$	$\text{inc}_{>15}$
3.9	0.000050	232	232	660	899	668
15.0	0.000028	388	329	1817	1840	1076
20.0	0.000045	414	330	2088	1950	1248
100.0	0.000035	461	335	4737	3888	2967
150.0	0.000051	432	373	5503	4734	4051
418.0	0.000031	437	425	8439	8142	8236

Table 2.1: Synthetic site parameters and observations drawn for 6 sites of increasing malaria transmission. Malaria transmission in each site is characterised by Λ . Age in each site is exponentially distributed with rate η . $\text{pos_M}_{[2,10]}$ and $\text{pos_M}_{>10}$ represent positive microscopy counts for synthetic XSS, each with a population size of 10^5 , for the 2-10 and over 10 age groups respectively. $\text{inc}_{<5}$, $\text{inc}_{[5,15]}$ and $\text{inc}_{>15}$ represent positive case counts for synthetic ACD studies, each tracking 1,000 participants for 1 year, for the under 5, 5-15 and over 15 age groups respectively. The prevalence and incidence results were simulated by executing the equilibrium solution with latent parameters θ^* and the simulated Λ and η for each site.

To generate synthetic observational data, I first simulated latent model parameters and subsequently generated prevalence and incidence data by executing the equilibrium solution with those latent parameters. Let θ^* be the latent model parameters, which I sampled from their respective prior distributions. I selected the site parameters, listed in Table 2.1, to be consistent with the range of Λ found in the previous parameter inference study published in (ibid.). η is sampled from a uniform distribution in $[100^{-1}, 40^{-1}]$, assuming an average age between 40 and 100. The latent prevalence and incidence for each age group, a , was calculated using $\text{pos_M}_a(\theta^*, \Lambda, \eta)$ and $\text{inc}_a(\theta^*, \Lambda, \eta)$ respectively. The observed prevalence and incidence data, $\mathcal{Y}$, was sampled from synthetic XSS and ACD studies respectively. The XSS studies were assumed to have a sample size of $N = 10^5$ and detect malaria according to the Binomial distribution. The ACD studies were assumed to track 1,000 participants for 1 year, i.e. $\text{PYO} = 10^3 \times 365$, and find new malaria cases according to the Poisson distribution. This is more formally written in Equation 2.8

$$\begin{aligned}
\mathcal{Y} = \{ & \text{n_pos_M}_{2-10}^*, \text{n_pos_M}_{>10}^*, \text{n_inc}_{<5}^*, \text{n_inc}_{[5,15]}^*, \text{n_inc}_{>15}^* \} \\
& \text{n_pos_M}_{2-10}^* \sim \text{Binom}(N, \text{pos_M}_{2-10}(\theta^*, \Lambda, \eta)) \\
& \text{n_pos_M}_{>10}^* \sim \text{Binom}(N, \text{pos_M}_{>10}(\theta^*, \Lambda, \eta)) \\
& \text{n_inc}_{<5}^* \sim \text{Pois}(\text{inc}_{<5}(\theta^*, \Lambda, \eta)) \\
& \text{n_inc}_{5-15}^* \sim \text{Pois}(\text{PYO} \times \text{inc}_{5-15}(\theta^*, \Lambda, \eta)) \\
& \text{n_inc}_{>15}^* \sim \text{Pois}(\text{PYO} \times \text{inc}_{>15}(\theta^*, \Lambda, \eta))
\end{aligned} \tag{2.8}$$

The likelihood function for inference is defined as the joint likelihood of Binomial and Poisson likelihoods, for prevalence and incidence respectively, calculating the probability of a number of positive microscopy and clinical incidence cases given the simulated rates for a candidate parameter set θ , see Equation 2.9.

Parameter	Support	NI Prior	Tuned Prior
Pre-erythrocytic immunity:			
κ_b	$\mathbb{R}^+$	Uniform(0, 10)	LogNormal(0, 0.25)
u_b	$\mathbb{R}^+$	Uniform(0, 10)	LogNormal(0, 1)
b_0	[0, 1]	Beta(1, 1)	Beta(1, 1)
IB_0	$\mathbb{R}^+$	Uniform(0, 1000)	Normal(50, 20) ($25 < IB_0 < 75$)
Clinical immunity:			
κ_c	$\mathbb{R}^+$	Uniform(0, 10)	LogNormal(0, 0.25)
u_c	$\mathbb{R}^+$	Uniform(0, 10)	LogNormal(0, 1)
ϕ_0	[0, 1]	Beta(1, 1)	Beta(2, 1)
ϕ_1	[0, 1]	Beta(1, 1)	Beta(1, 5)
IC_0	$\mathbb{R}^+$	Uniform(0, 1000)	Normal(25, 10) ($5 < IC_0 < 50$)
PM	[0, 1]	Beta(1, 1)	Beta(1, 1)
dm	$\mathbb{R}^+$	Uniform(0, 1000)	Normal(50, 20) ($5 < dm < 100$)
Immunity to detection:			
κ_d	$\mathbb{R}^+$	Uniform(0, 10)	LogNormal(0, .25)
u_d	$\mathbb{R}^+$	Uniform(0, 10)	LogNormal(0, 1)
d_1	[0, 1]	Beta(1, 1)	Beta(1, 1)
ID_0	$\mathbb{R}^+$	Uniform(0, 1000)	Normal(25, 1) ($5 < ID_0 < 50$)
fd_0	[0, 1]	Beta(1, 1)	Beta(1, 1)
γ_d	$\mathbb{R}^+$	Uniform(0, 10)	LogNormal(0, 2)
ad_0	$\mathbb{R}^+$	Uniform(40, 100)	Normal(70, 1) ($40 < ad_0 < 70$)
rU	$\mathbb{R}^+$	Uniform(0, 10)	LogNormal(0, 1)

Table 2.2: Intrinsic parameters to be inferred. The “Support” column lists the valid sets of values for each parameter. The “NI Prior” column lists the non-informative prior distributions which were initially selected to investigate the likelihood space. The “Tuned Prior” column lists the priors which were selected to avoid spurious immunity dynamics.

$$\begin{aligned}
\mathcal{L}(\mathcal{Y}|\theta) = & Pr(n_pos_M_{2-10}^*; N, pos_M_{2-10}(\theta, \Lambda, \eta)) \times \\
& Pr(n_pos_M_{>10}^*; N, pos_M_{>10}(\theta, \Lambda, \eta)) \times \\
& Pr(n_inc_{<5}^*; inc_{<5}(\theta, \Lambda, \eta)) \times \\
& Pr(n_inc_{5-15}^*; inc_{5-15}(\theta, \Lambda, \eta)) \times \\
& Pr(n_inc_{>15}^*; inc_{>15}(\theta, \Lambda, \eta))
\end{aligned} \tag{2.9}$$

I implemented this model in a probabilistic programming language called Numpyro, (Phan, Pradhan, and Jankowiak, 2019). I chose it so that I could easily integrate it with *dmeq* and for its range of MCMC kernels. I used the No-U-Turn-Sampler (NUTS) variant of the HMC kernel with an initial step size of 1 and acceptance probability of 0.8.

2.3 Bayesian inference workflow

In this section, I describe the diagnostics I used to select the prior distribution for each model parameter. These analyses build on well known general Bayesian workflows for checking prior and posterior distributions, particularly constructing synthetic data to validate the Bayesian joint distribution, and apply them to the domain specific setting of Bayesian inference for fitting parameters of malaria transmission models, see Gelman, Vehtari, et al. (2020) and Talts et al. (2020) for details. I began with non-informative priors within each parameter’s support, referred to as the “NI Priors”. Here, I use “non-informative” to describe a distribution which does not incorporate prior knowledge into the model. However, it is worth noting that some “NI Priors” reduced the valid support for their parameters and, moreover, in a non-linear system even uniform priors can influence the posterior distribution. After reviewing my analyses, including sensitivity analyses, immunity curve analysis and prior and posterior predictive checks, I selected prior distributions which were more informative and consistent with malaria dynamics described in the literature. I refer to the distributions selected after these analyses as the “Tuned priors”.

I then present a comparative analyses on the posterior estimates for the model parameters given the two prior distribution sets. I include traditional convergence diagnostics as well as model specific posterior predictive checks for immunity dynamics. I also analyse the posterior parameter space to show correlations and compare the estimates to the ground truth, latent parameter values.

2.3.1 Diagnostics on the prior distribution

My prior predictive checks involved visualising the model’s immunity probability, incidence and prevalence values for each age group and across all transmission intensities. Immunity probability outputs were considered anomalous if they were “flat”, or output a similar immunity probability for all age groups. Flat probability curves for medium to high burden environments would conflict with the general consensus that immunity is acquired with age, making the older population less susceptible to infection. Prevalence/incidence curves were also considered anomalous if they were flat, as this would conflict with the consensus that malaria predominantly affects younger age groups. The final prior predictive checks are shown in [Figure 2.2](#) for 1000 draws from the NI and tuned priors.

The most effective mitigation for flat immune response was to select log-Normal prior distributions in place of the uniform distributions for the the hill coefficients κ_b , κ_c and κ_d . When these coefficients are set to lower values, they allow for a larger accumulation of immunity over time. The NI priors were biased towards flatter immunity dynamics, especially for immunity to detection, see [Figure 2.2a](#). However, these changes had little effect on the prevalence probabilities per age, likely due to inherent sensitivity issues in the model.

I conducted a sensitivity analysis using derivative-based local sensitivity measures. I first calculated the derivative of the likelihood function, using automatic differentiation, with respect to the model parameters, i.e. $\frac{d\mathcal{L}(\mathcal{Y}|\theta)}{d\theta}$, at 1000 points drawn from the NI and the tuned priors. I also disaggregated the likelihood function into its independent parts, i.e. the prevalence and incidence likelihoods for each site, in order to visualise the influence of each component. The gradients are shown in [Figure 2.3](#). It is important to note that the sensitivity values are based on synthetic data created from one parameter draw and may not reflect the model sensitivity to real observational data. I used forward-mode differentiation to compute the Jacobian of the above likelihood function, which mitigated numerical and out-of-memory errors which I experienced for the reverse-mode alternative.

The sensitivity analysis in [Figure 2.3a](#) showed that the likelihood is not sensitive to many parameters. In particular it showed that the rate of recovery from sub-patent infection, rU , would not be adequately explored using the NI priors. In response, I set the rU prior to a log Normal distribution, since the likelihood was more sensitive to lower rU values, or a longer duration of sub-patent infection. Tuning the remaining parameters had smaller effects on their sensitivity, with parameters for initial immunity levels remaining the most sensitive, i.e $b0$, $\phi i0$, $\phi i1$ and $d1$; and some parameters remaining particularly insensitive to tuning, i.e. $ad0$ and dm .

The sensitivity analysis disaggregated by model output showed that parameter sensitivity is dependent on the available data. Incidence likelihoods are insensitive to immunity to detection parameters, i.e. $ID0$, $ad0$, $d1$, $fd0$ and γ_d , whereas prevalence likelihoods are extremely sensitive. Some parameters are more informed by age based likelihoods. For example, PM and dm are most informed by prevalence and incidence at lower age groups.

2.3.2 Diagnostics on the posterior samples

The posterior estimates, see [Table 2.3](#), show varying success in terms of Bayesian updates towards the observed data. Some posterior estimates displayed a strong update towards the true parameter, e.g. $b0$, ϕ_0 , ϕ_1 and d_1 . The majority of parameters had little discernible update, and were still dominated by their prior distributions, which was consistent with the sensitivity analysis. Of the weakly sensitive parameters, IB_0 , IC_0 , ID_0 , ad_0 and d_m benefited the most from prior tuning, as the posterior mean for the NI priors remained far from their latent values. The error in the posterior estimates are listed in [Table 2.4](#) in terms of mean squared error and visualised in [Figure 2.1](#) terms of prior and posterior distribution comparisons.

Despite this, estimates inferred with both the NI and the Tuned priors fit well to the observed data, as is shown in [Figure 2.2f](#). The absolute change in the Mean Squared Prediction Error (MSPE), listed in [Table 2.5](#), suggests that prior tuning has a limited (if any) effect on model

prediction quality. The limited effect of prior tuning on predictive error despite the clear positive effect on parameter inference further points towards identifiability issues in the model.

Correlation in the posterior samples was investigated using pair plots, see [Figure 2.4](#). The pair plots revealed stronger relationships between parameters in the posterior after prior tuning. For example, my tuned priors for ϕ_0 and ϕ_1 had less overlapping density leading to steeper immunity accumulation over time. Initial and slope parameters, like d_1 and κ_d , were more accepted at low values of the slope parameter.

As a result, the NUTS sampler required $1.7\times$ longer to draw samples from the posterior using NI priors (8 hours, 42 minutes) than Tuned priors (5 hours, 11 minutes). Both sets of samples ($n=400$) were drawn from 4 parallel chains, after 100 warm-up samples on each chain, on an 8-core Intel i7 CPU. I measured forward simulation times for the equilibrium solution for each site to be $436 \text{ ms} \pm 4.06 \text{ ms}$. Thus, the cause of the long parameter inference times is likely due to large numbers of forward simulations in low acceptance regions. The prior tuning steered the sampler towards regions of the parameter space with high acceptance probability, which caused a noticeable speed up in sampling in the case of this complex model. However, contemporary malaria models require in the order of hours for forward simulations making this method of parameter inference infeasible, an issue which I address in the next chapter.

	NI								Tuned								latent_value
	mean	std	median	5.0%	95.0%	n_eff	r_hat		mean	std	median	5.0%	95.0%	n_eff	r_hat		
kb	5.066	2.874	4.920	0.724	9.485	552	1.003		1.730	1.929	1.118	0.057	3.855	644	1.005		0.297
ub	3.484	2.997	2.362	0.001	8.437	405	1.010		1.474	1.744	0.957	0.046	3.275	867	1.005		2.171
b0	0.712	0.070	0.706	0.592	0.815	624	1.006		0.746	0.066	0.737	0.637	0.845	383	1.019		0.745
IB0	493.852	298.085	496.748	26.575	904.282	259	1.025		51.181	9.352	51.263	37.330	67.093	1280	0.998		58.737
kc	5.465	2.529	5.412	1.474	9.273	770	1.002		1.868	1.491	1.579	0.530	2.912	290	1.029		2.499
uc	2.865	2.410	2.231	0.003	6.708	807	1.000		0.907	0.771	0.687	0.034	1.826	1059	0.997		0.391
phi0	0.753	0.115	0.744	0.590	0.969	521	1.007		0.904	0.068	0.914	0.807	1.000	515	1.007		0.962
phi1	0.227	0.057	0.228	0.130	0.317	272	1.032		0.169	0.055	0.173	0.087	0.261	227	1.038		0.172
IC0	121.102	51.632	110.364	48.270	196.237	317	1.018		102.861	15.650	100.864	79.286	127.450	756	1.005		103.927
PM	0.407	0.255	0.343	0.021	0.817	585	1.004		0.427	0.244	0.396	0.051	0.825	802	1.004		0.730
dm	418.639	273.329	367.565	6.180	850.305	598	1.001		199.912	49.074	199.191	147.878	246.929	598	1.014		182.862
kd	6.527	2.188	6.679	3.412	10.000	721	1.015		6.152	5.716	4.624	1.036	11.719	827	1.004		7.250
ud	4.367	2.906	3.973	0.016	8.667	1174	0.999		1.204	1.259	0.790	0.051	2.735	1073	1.003		2.499
d1	0.607	0.016	0.606	0.581	0.631	269	1.028		0.615	0.018	0.616	0.587	0.644	512	1.006		0.619
ID0	20.604	4.707	20.344	13.514	28.448	700	1.006		24.692	1.845	24.841	21.771	27.147	1143	0.997		25.626
fd0	0.575	0.274	0.607	0.165	0.980	985	1.000		0.656	0.263	0.727	0.243	0.996	895	1.007		0.958
gd	4.986	2.901	5.069	0.292	9.074	1778	1.001		1.359	1.716	0.813	0.035	2.854	822	1.003		6.689
ad0	18446.075	10299.738	18433.989	2475.442	34610.334	1164	1.004		10909.862	735.362	10920.983	9838.957	12032.443	690	1.006		10555.968
rU	5.013	2.978	4.937	0.511	9.555	1525	0.996		1.672	2.288	0.963	0.040	3.954	659	1.004		6.905

Table 2.3: Marginal estimates of the intrinsic parameters in the model along with diagnostic metrics. n_eff is the Effective Sample Size (ESS) (Stan Development Team, 2018), a measure of information content in the posterior sample. The ESS can be compared with the total number of samples in this experiment, 1000. The ESSs are generally high, indicating uncorrelated samples within each chain. However in the minority of cases, ESSs exceeded the number of samples indicating negative autocorrelation in certain chains, a sign that further investigation is necessary. r_hat , or the Gelman-Rubin value, is a metric for convergence (Stan Development Team, 2018), numbers close to 1 indicate that posterior draws are consistent within and between chains. 10 chains were used for sampling. $latent_value$ is the ground truth, latent parameter value which was used to generate the observational data

In the interest of reproducibility, I have published the code I wrote to perform these analyses and visualisations. You can access the code using details listed under (G. Charles, 2024d).

2.4 Discussion

The differentiable implementation of the bi-stable equilibrium model successfully facilitated a derivative-based Bayesian inference workflow. The diagnostics demonstrated a high predictive accuracy for both prevalence and incidence rates across all age groups, despite the coarse age bands used in the observational data. The immunity dynamics suggested by the posterior parameter estimates agreed with the ground truth dynamics, despite many parameters being non-identifiable. The MCMC diagnostics revealed high Effective Sample Size (ESS) and strong convergence statistics, even with non-informative priors, suggesting that the NUTS sampler is robust and forgiving of less precise prior information. This is useful in the context of complex epidemic models, where exact priors are often difficult to ascertain due to the inherent uncertainties and incomplete data.

The gradient-based sensitivity checks was a quick and effective method for analysing the informativeness of the prior distributions within complex models. Traditional approaches to sensitivity

Parameter	NI MSE	tuned MSE	MSE reduction
kb	24.8	0.202	24.6
ub	13.6	5.47	8.08
b0	0.0447	0.0143	0.0304
IB0	1.23e+05	204	1.23e+05
kc	1.09	0.012	1.08
uc	9.9	0.561	9.34
phi0	0.00832	0.0227	-0.0144
phi1	0.00136	0.000227	0.00113
IC0	85.2	78.9	6.36
PM	0.218	0.0667	0.151
dm	1.06e+05	686	1.06e+05
kd	26	0.139	25.8
ud	9.68	3.28	6.4
d1	0.00233	0.000557	0.00177
ID0	22.6	63.9	-41.3
fd0	0.199	0.145	0.0538
gd	1.6e+03	2.65e+03	-1.05e+03
ad0	1.53e+08	1.96e+05	1.53e+08
rU	12.5	32.4	-19.9

Table 2.4: Mean Squared Error (MSE) of samples drawn from the posterior (n=400) under the non-informative (NI) and tuned priors. MSE values are relative to the true latent parameters listed in [Table 2.3](#)

	MSPE reduction				
output →	inc_0_5	inc_15+	inc_5_15	prev_10+	prev_2_10
EIR ↓					
3.9	414	426	-159	-94.1	36.6
15	179	499	129	-139	-34.8
20	-534	-928	-2.46e+03	-61.8	239
100	-762	-1.08e+03	-360	262	-47.2
150	2.72e+03	872	78.9	-37.6	150
418	-2.78e+03	-307	1.55e+03	569	78.7

Table 2.5: Mean Squared Predictive Error (MSPE) reduction due to prior tuning. The difference between the MSPE of posterior predictive draws (n=400) with respect to the latent prevalence and incidence rates for each age group.

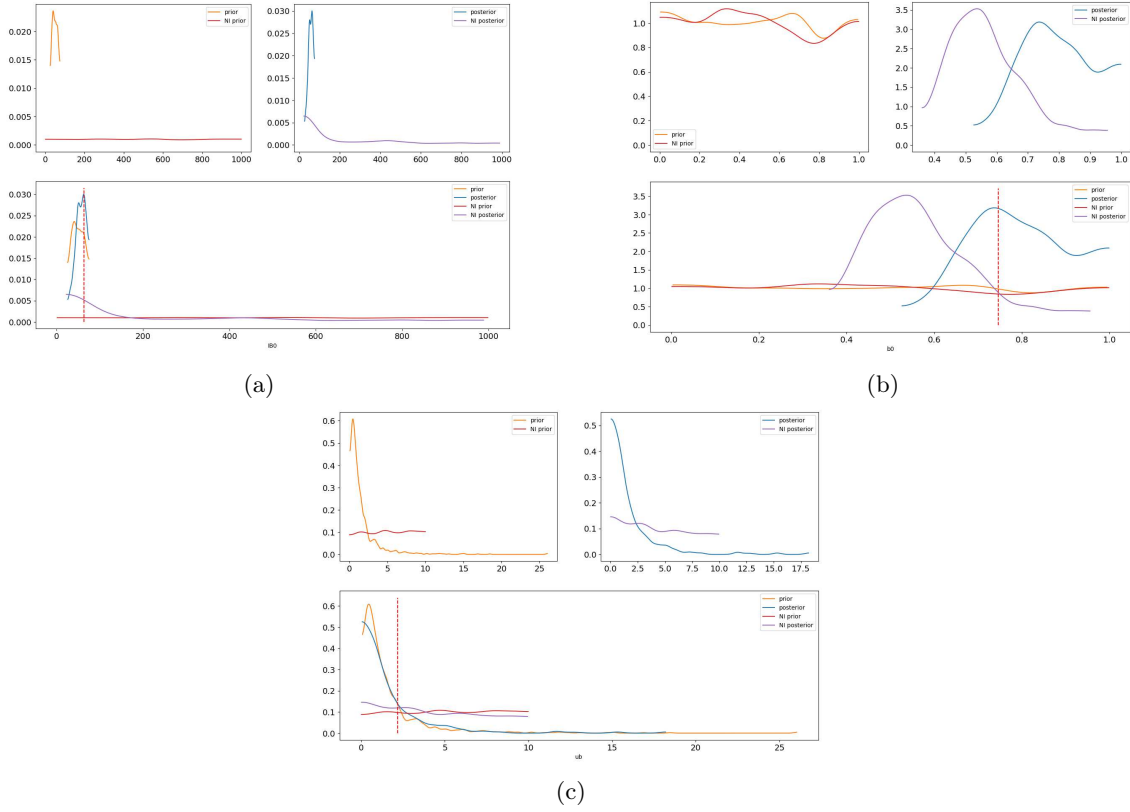


Figure 2.1: Selected prior and posterior distribution comparisons

Above are kernel density estimates of the prior ($n=1000$) and posterior ($n=400$) distributions of three pre-erythrocytic immunity parameters for both the non-informative (NI) and tuned priors. The latent parameter value is plotted as a red horizontal dashed line. [Figure 2.1a](#) shows how the more informative tuned prior can lead to posteriors which more reliably infer the true value. [Figure 2.1b](#) demonstrates how the tuned priors can lead to improved posterior updates for the b_0 parameter despite its particular prior remaining uninformative. However, in some cases, such as the estimates for u_b depicted in [Figure 2.1c](#), where the parameter is unidentifiable and the estimated posterior does not update from the prior.

analysis typically require multiple executions of the model, varying parameter sets each time to evaluate their impact on the model’s outputs. However, by using the gradient information associated with each parameter draw, the method is more computationally efficient and further encourages experimentation with prior tuning. Sensitivity analyses are crucial for understanding Bayesian inference, yet their publication is infrequent for more complex models due to the technical and computational hurdles.

For the malaria model, the sensitivity checks correctly identified the influential model parameters and observational data. The influential model parameters clearly displayed the largest average gradients and were subsequently validated through MCMC analyses. Moreover, by disaggregating the likelihood function during the sensitivity analysis, it became possible to discern how future data collection efforts could be improved to refine parameter estimates. There was no clear pattern in sensitivity of model parameters to the transmission intensity in the observed site. Variations in sensitivity were likely due to distributional artefacts in the likelihood function, i.e. larger variance with higher Poisson rates and smaller variance with extreme Binomial success probabilities. Further investigation would be required to identify and correct any unwanted biases. Parameters pertaining to immunity to detection were more sensitive to prevalence data, whereas maternal immunity parameters showed a pronounced influence from data pertaining to younger age groups. Such analyses could play a role in survey design to increase the predictive power of our epidemic models and shed light on under-researched biological mechanisms of malaria.

One could extend this work to perform a sensitivity analysis on the posterior estimates. In this study, the goal of sensitivity analysis was to present a quantitative estimate of the influence

of perturbations in the prior space on the likelihood function in order to assess identifiability and informativeness. Other sensitivity analyses assess the influence of prior parameters and likelihood formulation on the posterior estimates, as outlined in Gelman, Carlin, et al. (2021). Such analyses are important for quantifying uncertainty in posterior estimates and validating modelling assumptions. However, given the non-differentiability and long execution time of traditional inference methods, extensive quantitative sensitivity analyses of the posterior are computationally expensive and an active area of research, see the latest advances in (Paananen et al., 2019; Y. Kim, Berliner, and D. H. Kim, 2015).

In light of the above diagnostics, parameter tuning was important for ensuring reasonable assumptions in the underlying dynamics of malaria transmission and improving the estimation of the latent parameters. The diagnostics and sensitivity analysis conducted revealed a notable challenge inherent to complex epidemiological models: the non-identifiability of many parameters. Non-identifiability is where multiple parameter values yield similar model outputs, complicates the inference process and can obscure our understanding of disease dynamics. Moreover, it became apparent that certain regions within the model’s parameter space generated implausible immunity dynamics. Such outcomes not only detract from the model’s utility but suggest a careful reconsideration of the parameter values and assumptions embedded within the model. To mitigate these issues, I refined the prior parameter space to ensure that the tuned parameters were biologically plausible and improved the model’s performance during inference. The parameter tuning improved on Griffin, Ferguson, and Ghani (2014) priors with a more detailed justification for each parameter informed by the most current understanding of malaria immunity dynamics. Furthermore, parameter inference need not be limited to observational data in the form of prevalence and incidence. More detailed data on immune responses to infection, as is used in (P. A. Eckhoff, 2012), could also mitigate these identifiability issues.

The adoption of JAX to implement the malaria model yielded performance optimisations for forward simulation. The Just-In-Time (JIT) compilation of the model resulted in considerable speed improvements, while executing on a GPU led to moderate performance gains. It’s important to acknowledge that not all operations benefit equally from GPU acceleration and many of the computational tasks involved in malaria transmission modelling do not exhibit the high degree of parallelism found in, for instance, neural network operations. Nevertheless, these advancements collectively contribute to a more efficient analytical workflow. A large performance improvement was achieved by vectorising the model computation across the heterogeneous biting rates. This optimisation is device independent and can be applied to any epidemiological model which handles sub-populations independently before aggregating their statistics. Since forward simulation is used for most epidemiological analysis tasks, it could allow for speed improvements across other tasks such as intervention planning, optimisation or counterfactual analysis.

Despite these improvements, the practicality of re-implementing more complex models in JAX remains a challenge from a software engineering standpoint. The existing code bases for such models can be extensive and intricate, potentially incorporating features not easily supported by JAX, such as resizable populations or auto-differentiable discrete processes (Quera-Bofarull et al., 2023). In light of these considerations, in the next chapter I explored alternative methods that aim to create a differentiable surrogate model without re-implementation.

There are several alternatives to using the JAX ecosystem for hardware accelerated auto-differentiation based parameter inference. Stan (Stan Development Team, 2018), Julia (Turing.jl) (Ge, Xu, and Ghahramani, 2018), Pyro (Bingham et al., 2018) and Tensorflow Probability (Abadi et al., 2016) all provide compilation, auto-differentiation and high-level constructs for performing Bayesian inference. I had no reason to suspect any platform to be more appropriate for this particular inference task.

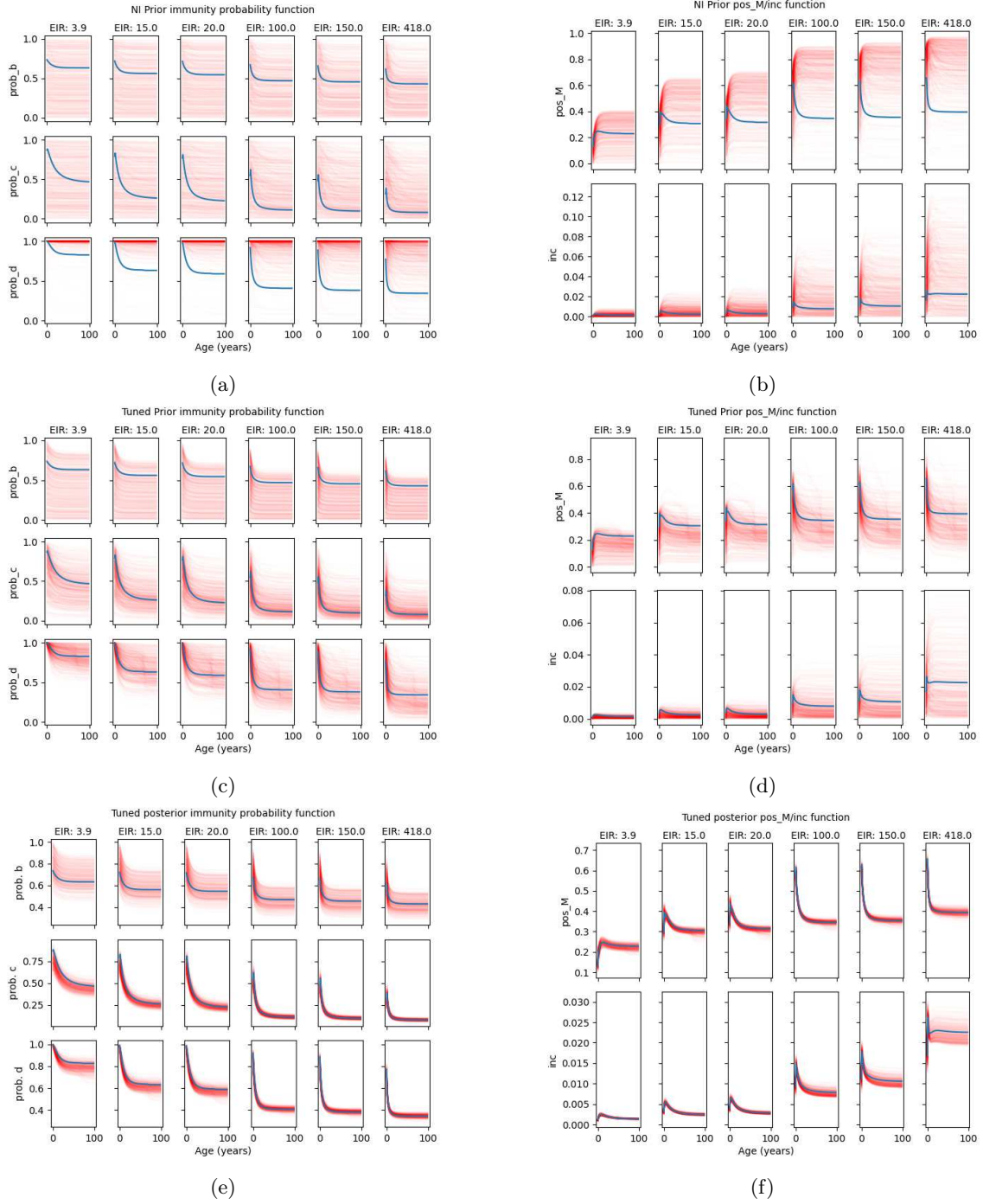
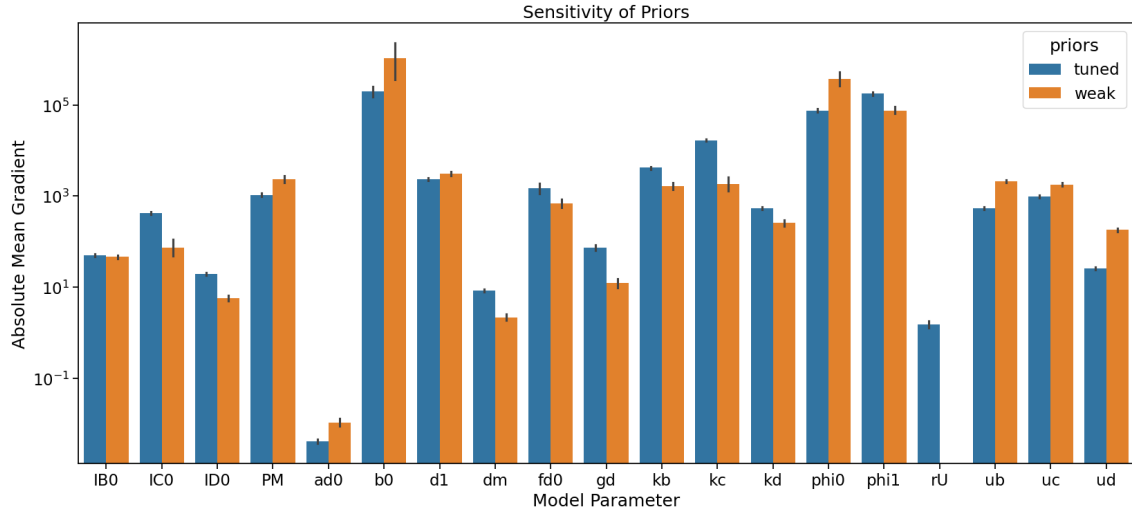


Figure 2.2: Model based predictive checks

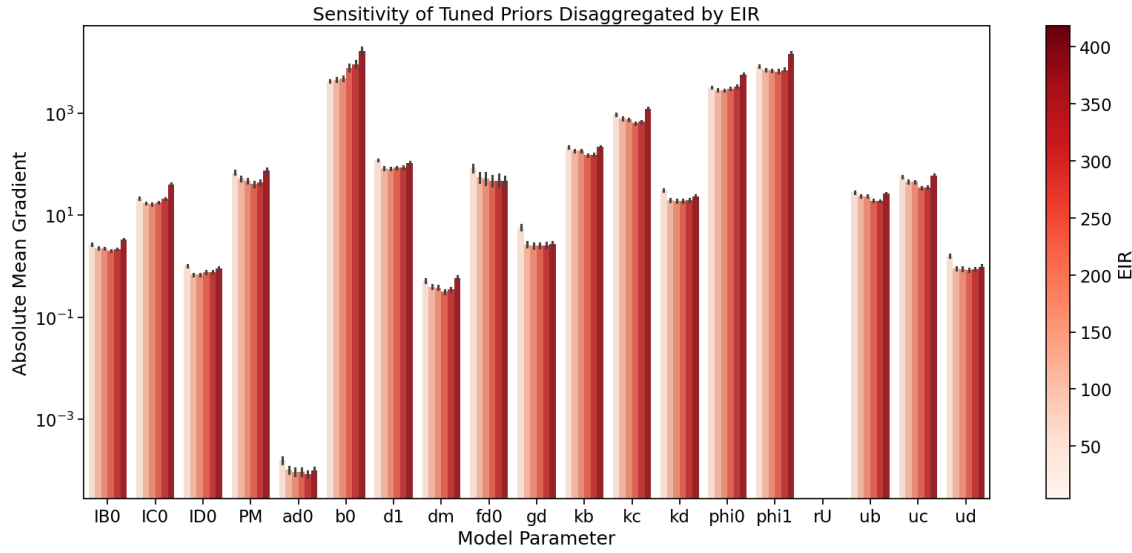
a, **c** and **e** show three functions of immunity probability per age group for the each site in the experiment, labelled with each site's EIR. *prob. b* shows the probability of a pre-erythrocytic infection; *prob. c*, the probability of clinical infection; and *prob. d*, the probability of non-detection by microscopy.

b, **d** and **f** show prevalence and incidence functions for each site in the experiment. *pos_M* shows the probability of detection by microscopy per age group; *inc* shows the rate of clinical infection by age group.

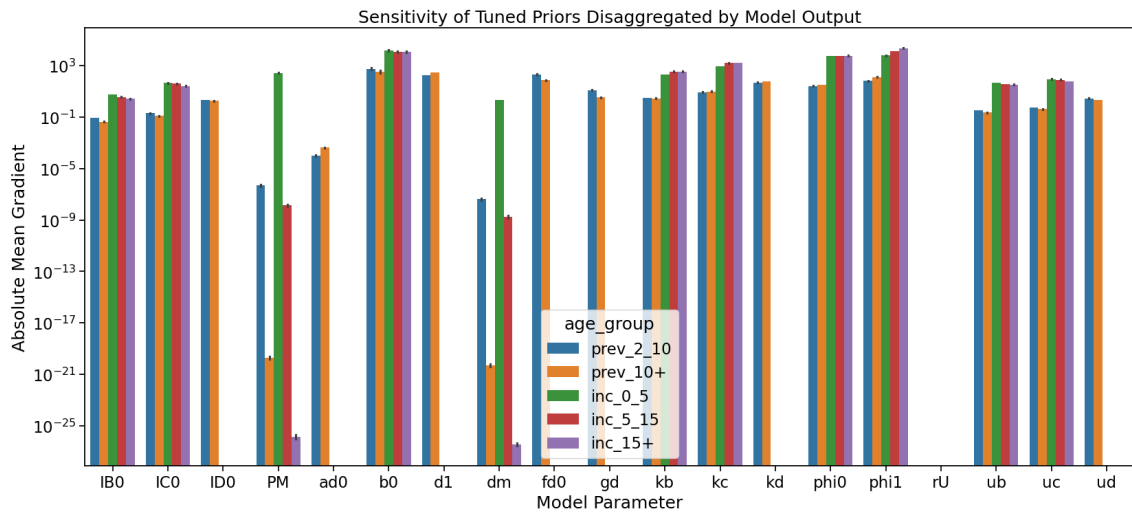
a, **c**, **b**, and **d** are plots of 1000 draws from prior distributions whereas **e** and **f** are plots of posterior estimates using the tuned priors.



(a)



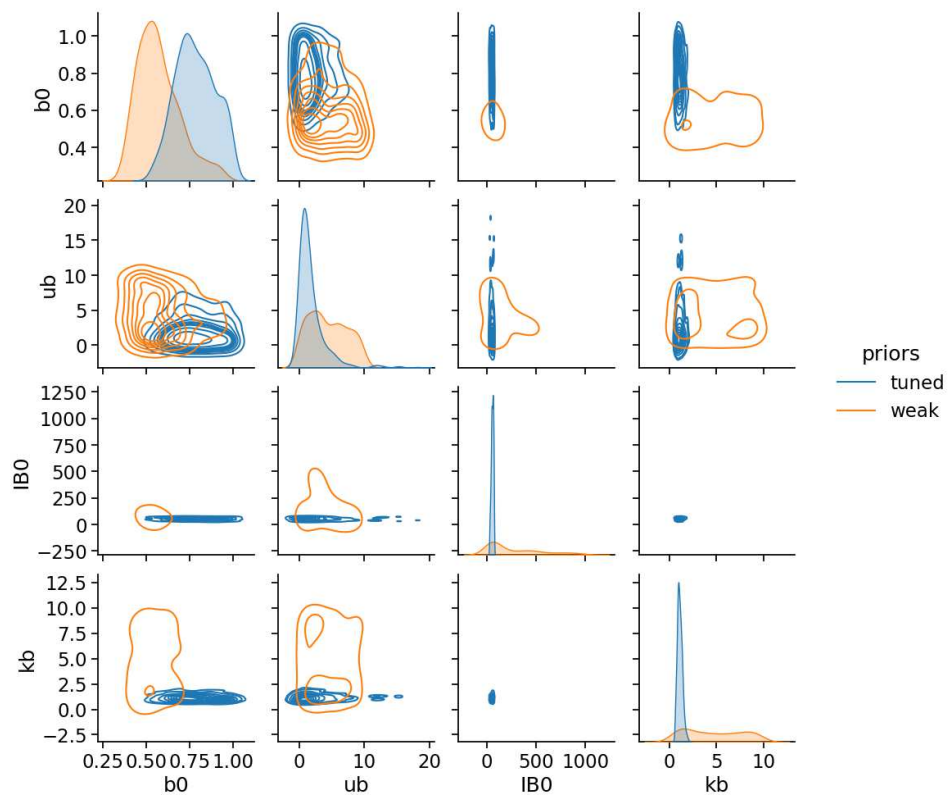
(b)



(c)

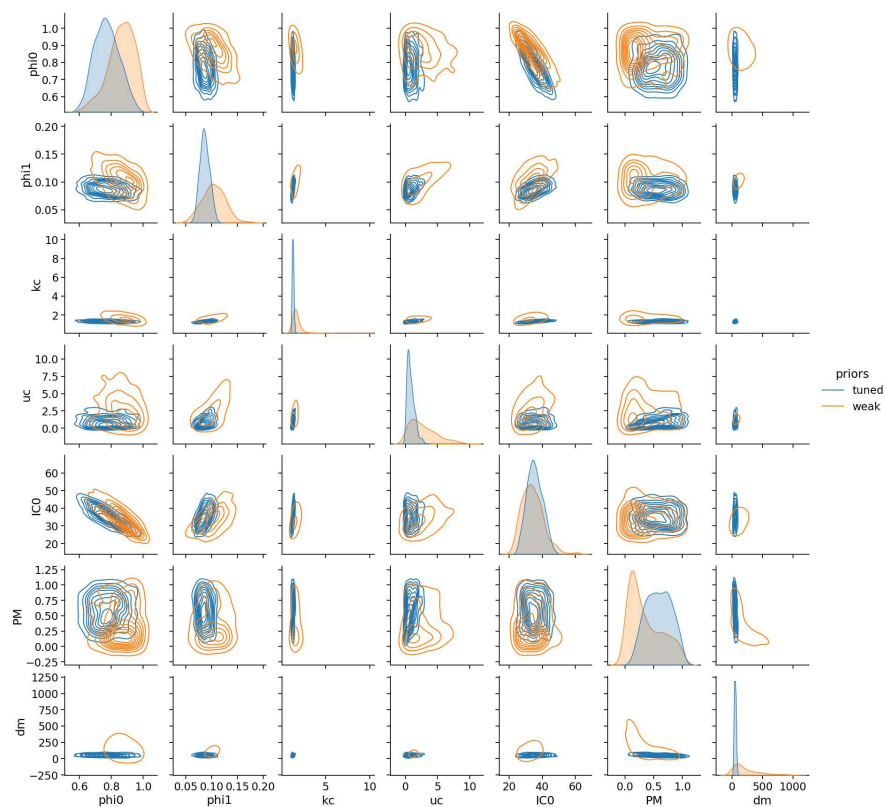
Figure 2.3: All bar plots show the Absolute Mean Gradient of the posterior measure with 95% confidence intervals shown as error bars. Figure 2.3a shows the difference in gradients between the NI and tuned priors for the overall likelihood. Figure 2.3b and Figure 2.3c show the gradients for the tuned priors only, disaggregating the likelihood by the site EIR and model output respectively.

Correlation in the posterior estimates of pre-erythrocytic immunity parameters



(a)

Correlation in the posterior estimates of clinical immunity parameters



(b)

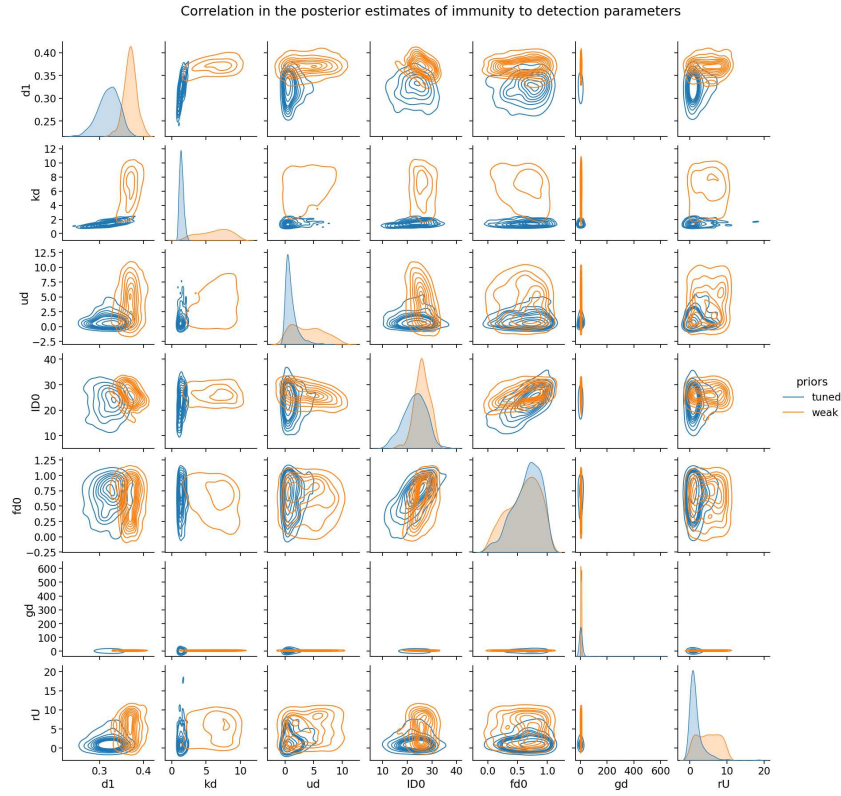


Figure 2.4: Pair plots of the posterior estimates of immunity parameters for pre-erythrocytic immunity [Figure 2.4a](#), clinical immunity [Figure 2.4b](#) and immunity to detection, [Figure 2.4c](#). Kernel density estimates were made for 1000 posterior samples from each set of priors (NI and tuned) and are visualised using contour plots.

Chapter 3

Fast black-box Bayesian inference using neural-network based surrogates

3.1 Background

In the previous chapter, I presented a Bayesian inference workflow for inferring the parameters of a Bi-stable equilibrium solution for a malaria transmission model. I introduced the No-U-Turn Sampler (NUTS) which used gradient information to reliably sample the posterior, prior and posterior diagnostics to validate and interpret the results and tuned priors which encode knowledge about malaria immunity and improve posterior sampling. However, the equilibrium solution is a simplified version of Griffin et al’s malaria transmission model. On a CPU, forward simulations required less than a second to execute and NUTS required over 5 hours to produce 100 samples of the posterior.

For more complex models, high computational requirements remain a limiting factor. Contemporary agent-based malaria models can take, in the order of hours, to run forward simulations, since they often need to simulate large populations for decades in order to produce outputs which can be fit to epidemiological data. In a study which calibrated the OpenMalaria model (Reiker, Golumbeanu, et al., 2021), the parameter inference workflow took between 7 and 14 days. Additionally, the need to run for several days on computing clusters increase carbon emissions and could act as a barrier to entry and restrict access to researchers with limited computational resources. This is particularly problematic for malaria models, as researchers in the regions most affected by malaria — in the Global South — may not have access to the necessary computing infrastructure. In practice, researchers are often forced to adopt simpler models to avoid long inference times, which limits the detail and applicability of our models to realistic scenarios.

Another drawback of the derivative-based workflow discussed in the previous chapter is the need to (re-)implement a model within an auto-differentiable framework. This can pose a considerable challenge, particularly for complex epidemiological models which demand a significant software engineering effort to implement. Although demanding, it is not insurmountable. For instance, (Chopra et al., 2023) implemented a differentiable agent-based model (ABM) for COVID-19 transmission, demonstrating that ABMs can be adapted to support gradient-based inference methods. However, as model complexity increases, researchers face challenges in propagating gradients through discrete randomness (Arya et al., 2023) and avoiding vanishing/exploding gradients (Quera-Bofarull et al., 2023; Andelfinger, 2021).

Current research is rising to these challenges. Discrete randomness can be differentiated using approximations, such as Gumbel softmax (Jang, Gu, and Poole, 2017), or alternative auto-differentiation frameworks (Krieken, Tomczak, and Teije, 2021). This area of research holds promise for making gradient-based methods more accessible and effective across a broader range of epidemiological models. However the issue of long execution time will remain.

As an alternative, I propose the use of surrogate models. Surrogates have the potential to accelerate forward simulations by orders of magnitude while being differentiable, eliminating the need for the re-implementation of models in an auto-differentiable way and providing a route to

derivative-based inference. Surrogate models closely emulate original models but offer a speed advantage by approximating the model outputs given a set of model parameters. Surrogate modelling is a mainstay in Bayesian Optimisation where it is used to emulate a costly objective function. It is also commonly used in climate modelling and fluid simulation in order to speed up large scale simulations (Lam et al., 2023; Sanchez-Gonzalez et al., 2020) and even perform parameter inference (Kovachki et al., 2022).

Neural network-based surrogates are particularly attractive, as they are inherently differentiable, seamlessly integrate with ecosystems which support gradient-based optimisation methods, such as PyTorch, Tensorflow and JaX (Paszke et al., 2019; Abadi et al., 2016; Bradbury et al., 2018), and more conveniently scale to large data sets. This makes them valuable tools for inference, providing a practical and efficient alternative to direct simulation of complex models.

3.1.1 Surrogates for Bayesian inference

The first surrogate applied to a modern malaria model was published in Cameron et al. (2015). The authors' approach used a functional regression method to develop a fast surrogate capable of producing estimates for prevalence and incidence from the OpenMalaria, EMOD and Imperial College London's Agent-Based Models (ABMs) (Thüring et al., 2016; P. A. Eckhoff, 2011; Griffin, Ferguson, and Ghani, 2014). Their surrogate was trained using samples from noisy simulations derived from the ABMs simulating a less computationally demanding, low population size making Bayesian inference more feasible and efficient. Despite its benefits, this approach had several limitations including its sensitivity to the kernel parameters used in the functional regression. For the high-dimensional and non-linear model dynamics we see in malaria models, it remains unclear whether the kernel used could accurately capture the model outputs across all relevant areas of the parameter space. In the same study, the authors noted that careful kernel selection and model sampling strategies were needed to ensure that the surrogate could reliably replicate the outputs of high-dimensional epidemiological models.

The challenges of surrogate modelling in a large parameter space were also acknowledged in McKinley et al. (2018). Their study posited that in Bayesian inference, the posterior densities are often concentrated in a small region of the overall parameter space. They argued that a more strategic sample selection, which would concentrate on more of the yet unknown posterior space, would improve the accuracy of a surrogate during inference. This intuition led to the authors devising an "implausibility" metric as a heuristic to identify regions of the parameter space which are useful for surrogate training. As the surrogate model, they proposed a Gaussian Process (GP), which they trained iteratively: starting with a Latin Hypercube Sampling (LHS) strategy to initially cover the parameter space, followed by the rejection of implausible regions and resampling around the plausible samples. This method, known as History Matching, was applied in Andrianakis et al. (2015) to calibrate an HIV model using prevalence data from Uganda, demonstrating its utility in practical epidemiological studies.

A more sophisticated surrogate was introduced in Reiker, Golumbeanu, et al. (2021) to emulate the OpenMalaria model. They proposed a GP Stacked Generalisation (GPSG) surrogate, which ensembles predictions from a neural network, regression splines, and a random forest under a Gaussian Process meta-learner. The GPSG was used in a Bayesian Optimisation workflow, designed to iteratively refine the surrogate by optimising the fit of 11 OpenMalaria outputs. However, the scalability of GPs was cited as a limiting factor, as the computational requirements scale cubically with the size of the input parameter space due to the need for inverting large covariance matrices in order to perform inference. This scaling issue presents a significant challenge, especially when dealing with high-dimensional data, making it impractical for large-scale applications. Consequently, although the GPSG method provided valuable insights, it did not facilitate a fully Bayesian estimation of the posterior distribution.

3.1.2 Neural networks as surrogates

Given that neural networks (NNs) are not commonly used as surrogates in infectious disease modelling, I will include a primer on their application and functionality in this context. Neural networks are models inspired by biological neural networks. They consist of a network of "nodes" which act as basic computational units connecting model inputs to outputs. Each node performs a simple linear transformation on its input data, $y = wx + b$, where w , the weight, b , the bias, are NN parameters and x is the input and y the output.

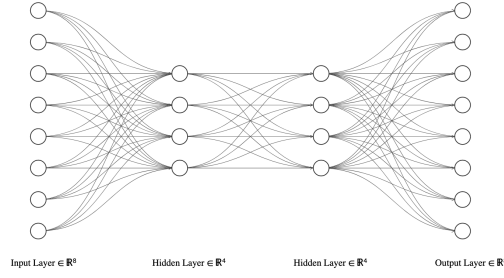


Figure 3.1: A $8 \times 4 \times 4 \times 8$ multi-layer perceptron - each circle represents a "node" which has bias and weight hyper-parameters in order to process its inputs with a linear transformation followed by a non-linear function. The arrows show how outputs from previous layers are used as inputs to subsequent layers. This image was generated using [NN SVG](#)

A multi-layer perceptron (MLP) is a typical neural network architecture. In an MLP, nodes are organised into layers, each layer, indexed by i , processing a vector of inputs x_i to produce a vector of outputs x_{i+1} . The output of one layer becomes the input for the next layer, except for the final layer, which produces the network's overall output. Between these layers, activation functions are used to introduce non-linearities into the model, enhancing its ability to capture complex patterns in data that linear models cannot. This sequential processing allows for the modelling of complex relationships between high-dimensional inputs and outputs.

In the context of surrogate modelling, the aim is to optimise the neural network's weights and biases to minimise the discrepancy between the surrogate and the emulated model. To this end, one specifies a loss function to quantify this discrepancy using sampled inputs and outputs from the emulated model and optimise the weights and biases, in a process known as training. The most widespread method for optimising neural networks is Stochastic Gradient Descent (SGD).

Given that neural networks tend to be overparameterised—having more parameters than necessary to simply fit the training data—regularisation techniques are used to prevent overfitting. Overfitting occurs when a model learns the details and noise in the training data to an extent that it negatively impacts the performance of the model on new data. Common regularisation methods include dropout, where outputs from randomly selected nodes are set to zero during training, and penalty-based methods, which augment the loss function to penalise large weights and biases, thereby encouraging simpler models which generalise better to unseen data. These techniques help ensure that the surrogate model remains robust and performs well across various input parameters.

3.1.3 Neural density estimation

Neural Density Estimators (NDEs) are designed to produce probabilistic estimates using neural networks. There are several methods by which NDEs can achieve this. One simple approach is the parametric method, where neural networks parameterise a pre-defined tractable probability distribution, effectively learning the inputs which best fit the parametric distribution to the data.

A more flexible approach, which does not prescribe a parametric distribution, is a normalising flow (NF), which uses neural networks to estimate the density in a wider range of settings. NFs train neural networks to learn a differentiable, invertible transformation (T and its inverse T^{-1}) from a source of noise sampled from a tractable random process ($u \sim p(u)$) into the desired distribution ($q(x)$ for model outputs x). With an NF, one can sample the model by transforming noise $T(u)$ and estimate the density of any output using the inverse transform and its Jacobian, defined in [Equation 2.1](#), $q(T^{-1}(x))|det(J_{T^{-1}}(x))|$. The possible distribution is only limited by the expressiveness of T . In practice, NFs layer a sequence of k transformations $T_{1..k}$, as is done with conventional neural networks, to learn more complex distributions, this is shown in [Figure 3.2](#).

Neural Autoregressive Flows (NAF) represent a particularly efficient and expressive variant of NFs. NAFs reformulate the target distribution $q(x)$ as a product of univariate densities, with each density conditioned on all preceding dimensions:

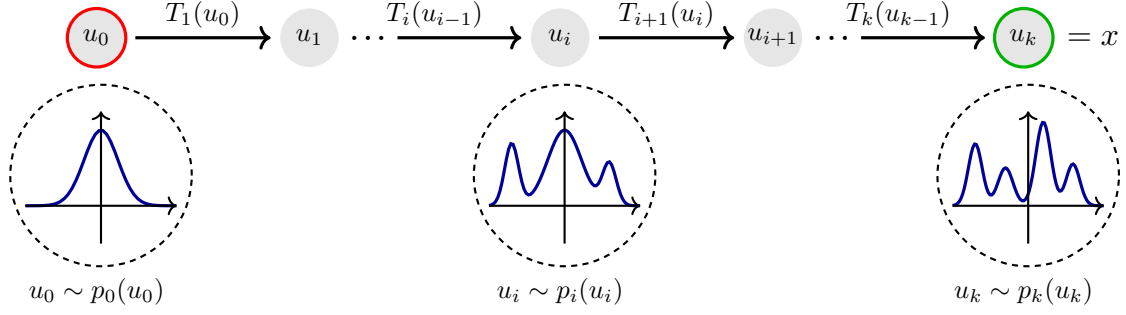


Figure 3.2: A Normalising Flow. A tractable source of randomness u_0 , depicted with a normal distribution, is transformed by a sequence of differentiable, invertible transformations $T_{1..k}$, which are learned by neural networks. Until the distribution of the outputs u_k are close to the desired distribution. Image adapted from (Riebesell and Bringuier, 2020)

$$q(x) = q(x_1) \prod_{i=2}^n q(x_i | x_{<i})$$

This formulation, where each variable x_i depends only on the variables before it ($x_{<i}$), is the basis for the term "autoregressive". Each univariate density $q(x_i | x_{<i})$ in the sequence is modelled using a modified NF, which includes an invertible and differentiable transformation T_i and a conditioning function c_i . This decomposition ensures that the entire transformation's Jacobian matrix is lower triangular, since each transformation T_i is dependent only on preceding variables, $x_{<i}$. This guarantees that the Jacobian's determinant can be computed efficiently, making density evaluations in high-dimensional spaces particularly efficient.

Many variants of NAFs, including Masked Autoencoder for Distribution Estimation (MADE) (Papamakarios, Pavlakou, and Murray, 2018) and Block Neural Autoregressive Flow (B-NAF) (De Cao, Titov, and Aziz, 2019) are, theoretically, universal density approximators. This asserts that autoregressive flows can approximate any real, continuous distribution, provided the transformations T_i and the conditioners c_i are learned with universal function approximators, given wide or deep enough architectures. In practice, these assertions are contingent on finding the correct hyper-parameterisations for the neural networks in the first place.

3.1.4 Stochastic Variational Inference (SVI)

NDEs are most often trained using SVI in the context of simulation based inference (Papamakarios, 2019; Glockler, Deistler, and Jakob H Macke, 2022; Greenberg, Nonnenmacher, and Jakob H. Macke, 2019), where they are used for fitting probabilistic models which include latent variables, denoted as z , to a dataset of observations, denoted x . These observations are assumed to be sampled from a distribution $p(x)$. The goal in SVI is to optimise a variational distribution $q_\phi(z)$, where ϕ are parameters for the variational distribution, to closely approximate the posterior distribution $p(z|x)$. In order to do this, one minimises a discrepancy measure between these two distributions. A commonly used measure of discrepancy is the forward Kullback-Leibler (D_{KL}) divergence, leading to the following objective:

$$\hat{\phi} = \arg \min_{\phi} D_{KL}(q_\phi(z) || p(z|x)) \quad (3.1)$$

where D_{KL} can be expressed as:

$$D_{KL}(q_\phi(z) || p(z|x)) = \int q(z) \log \frac{p(z|x)}{q_\phi(z)} dz \quad (3.2)$$

$$D_{KL}(q_\phi(z) || p(z|x)) = \mathbb{E}_{q_\phi}[\log p(x, z) - \log q_\phi(z)] + \log p(x) \quad (3.3)$$

Since the term $\log p(x)$ is often intractable or not directly available, one instead optimises a lower bound to this loss, known as the Evidence Lower Bound (ELBO) or the variational bound. The ELBO is expressed as

$$\text{ELBO}(q_\phi(z)||p(z|x)) = \mathbb{E}_{q_\phi}[\log p(x, z) - \log q_\phi(z)]$$

Across large datasets, this expectation can be approximated by drawing N samples from the variational model $q_\phi(z)$ to allow for optimisation:

$$\text{ELBO}(q_\phi(z)||p(z|x)) \approx \frac{1}{N} \sum_{i=0}^N (\log p(x, z_i) - \log q_\phi(z_i)) \quad \text{where } z_i \sim q_\phi(z)$$

In practice, optimisation involves sampling a mini-batch of data from $q(z)$, computing the ELBO as the loss, and performing stochastic gradient descent (SGD) on the model parameters to improve the fit of the NDE to the data.

However, integrating NDEs into Bayesian inference workflows can be approached in multiple ways, each with its own set of advantages and challenges. For instance, Neural Likelihood Estimation (NLE) involves training an NDE to estimate an unknown likelihood function. This allows the use of traditional posterior samplers, such as Markov Chain Monte Carlo (MCMC), offering flexibility since the NLE can be adapted for use with different sampling techniques (Papamakarios, Sterratt, and Murray, 2019). Conversely, Neural Posterior Estimation (NPE) involves training an NF to directly estimate the posterior distribution itself (Greenberg, Nonnenmacher, and Jakob H. Macke, 2019). The requirement of invertibility does impose limits on the choice of neural network architectures, and these methods have struggled with complex posterior geometries. Significant research efforts are currently directed towards improving the performance of invertible NDEs (Papamakarios, Pavlakou, and Murray, 2018; De Cao, Titov, and Aziz, 2019). Often these inference methods are applied sequentially, meaning that parameter estimation occurs in rounds, with posterior estimates produced from each round being used to refine the training of the NDE for the subsequent round. This iterative process helps to address the challenge of training surrogates to approximate a small, yet unknown, posterior space in a large parameter space, as identified by McKinley in (McKinley et al., 2018).

Most studies which present these workflows have evaluated their methods on simple, theoretical models, such as M/G/1 queues, Gaussian mixtures, SIR, Lotka-Volterra, Brock and Hommes, and dual moon models (Papamakarios, Sterratt, and Murray, 2019; Greenberg, Nonnenmacher, and Jakob H. Macke, 2019; Lueckmann et al., 2021; Lamperti, Roventini, and Sani, 2018). These models have a handful of parameters with simple, uncorrelated relationships to the outputs. At the time of writing, neural density estimation workflows have not yet been applied to models as complex as modern malaria models.

3.1.5 Annealing in SVI

Annealing is a parameter search and optimisation technique that has seen success in a variety of contexts, including parameter inference in complex models. The technique is especially useful for global optimisation problems, where conventional algorithms tend to converge to local optima. Drawing inspiration from physical systems, annealing introduces the concept of "internal energy" to parameter search and optimisation algorithms: in a "high energy" or "hot" state, the algorithm is allowed to explore a wide region of the parameter space. This state is conducive to escaping local optima since the algorithm can make larger, more exploratory moves. In a "low energy" or "cold" state, the algorithm focuses its search on a narrower region, optimising towards local optima as it cools down and settles.

This method is famously a key part of simulated annealing (Kirkpatrick, Gelatt, and Vecchi, 1983; Tsallis and Stariolo, 1996), a technique designed to find global optima of complex objectives within large parameter spaces. In applications involving continuous search spaces, the algorithm simulates the movement of a particle which navigates the parameter space towards the nearest optimum, using gradient information, with its momentum influenced by the temperature of the system—higher temperatures allowing for broader exploration.

Annealing has also been adapted for use with Markov Chain Monte Carlo (MCMC) in an approach called Parallel Tempering or Exchange Replica MCMC (Syed et al., 2022; Sambridge, 2014). This involves running several MCMC chains in parallel at different energy levels. The energy level of a chain is used to calculate an exponential coefficient to augment the likelihood function and thereby influence whether it explores the posterior distribution or the wider parameter space. The coldest chain uses the original likelihood function and samples from the posterior.

Meanwhile, hotter chains have coefficients less than one, which flattens the target distribution thereby allowing a broader exploration. States are allowed to swap between chains according to specific swapping criteria, facilitating exploration across different regions of the posterior, which is particularly advantageous when dealing with multi-modal distributions.

Annealing principles have also been applied to Stochastic Variational Inference (SVI) (Mandt et al., 2016; Huang et al., 2018). Given SVI’s tendency to produce overconfident estimates, researchers have investigated modifications to the optimisation process to steer it towards exploring a broader distribution space. These adaptations often involve tweaking the loss function to adjust the balance between the weight of the prior probability and the likelihood of the observed data. Such adjustments have led to improvements in predictive likelihoods and more stable optimisation.

3.1.6 Contributions

My key contribution is the development and evaluation of a neural network-based surrogate workflow to facilitate inference on the malaria model introduced in the previous chapter. This workflow does not require a re-implementation of the model into an auto-differentiable framework, as a neural-network based surrogate provides a differentiable likelihood function in its stead. This makes the proposed method more applicable to complex models which are not trivially differentiable such as Agent-Based Models. I have illustrated the workflow in [item 3.3](#).

I evaluated twelve variations on the workflow: two proposal distributions for surrogate training (Prior vs LHS), two surrogate formulations (Fixed vs Full) and three posterior sampling strategies (NUTS vs SVI vs Annealed SVI). My analyses focused on the effects of these strategies on each surrogate’s approximation error, the accuracy of inference estimates, as well as how these metrics are affected by sample size. Additionally, I assessed the implications on inference speed offered by each workflow, considering both training and inference times. This analysis includes insights into the trade-offs between approximation accuracy and computational efficiency, proposing feasible workflows which can be adopted depending on specific research needs and resource availability.

Furthermore, I have replicated several diagnostic tests from the previous chapter using the evaluated surrogate models to demonstrate that the gains in speed do not undermine the interpretability of the results.

This is the first evaluation at the time of writing to apply Neural Density Estimation for Bayesian inference on a high-dimensional, complex model. Likelihood-free studies have evaluated on low-dimensional toy models and previous inference studies on modern malaria models have used surrogate models which do not well scale to high-dimensional settings. This study reveals that in high-dimensional settings Sequential Stochastic Variational Inference coupled with an annealed likelihood can prove effective for parameter inference. It also shows that MCMC based samplers and SVI alone may not converge to useful posteriors.

3.2 Methods

I begin by detailing the two surrogate formulations I considered to accelerate the parameter inference problem. The primary goal of these surrogates is to emulate the underlying model in an auto-differentiable way to facilitate quicker parameter inference. Therefore, I construct surrogates to speed up the calculation of prevalence and incidence statistics so that the likelihood function, defined in [Equation 2.9](#), and its gradients can be estimated more quickly. It is worth noting that I do not propose Neural Density Estimation for estimating the likelihood, as is the case in similar Simulation Based Inference methods in the literature (Papamakarios, Sterratt, and Murray, 2019; Glockler, Deistler, and Jakob H Macke, 2022; Lueckmann et al., 2021), since likelihood function in this case is well established.

The alternative formulations I propose have trade-offs in terms of flexibility and diagnostic interpretability against the specificity to the observed dataset. By exploring these trade-offs, this study aims to identify the most effective strategies for improving inference speed without sacrificing the accuracy and reliability of the results.

3.2.1 Surrogate formulations

The following is an adaptation of a formulation I have previously published in (G. Charles, Wolock, et al., 2023).

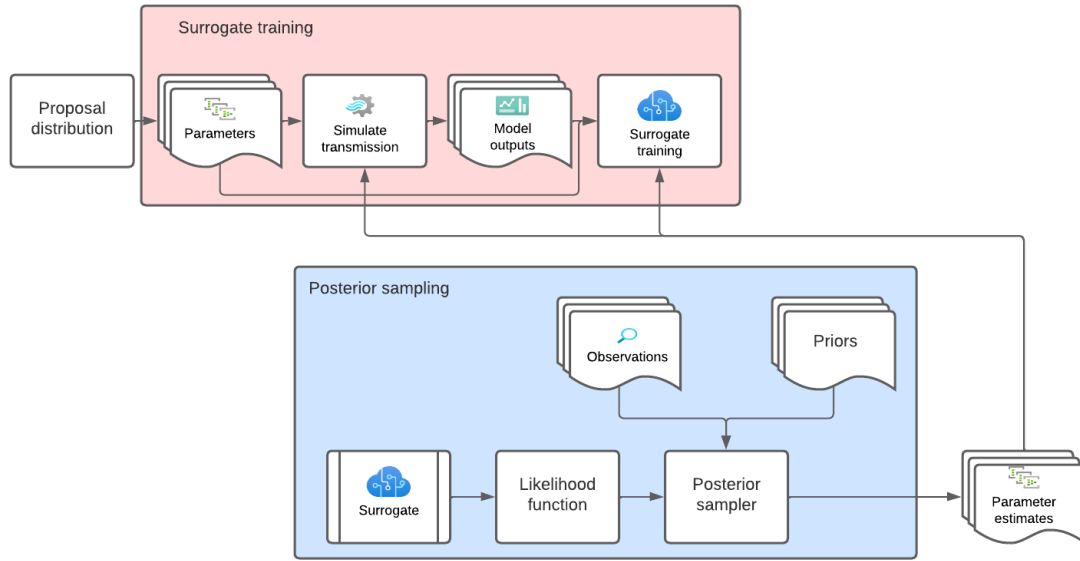


Figure 3.3: The complete surrogate-based inference workflow used in this study. The flow begins in the top left and results in parameter estimates produced in the bottom right.

1. Model parameters are sampled from the parameter space according to a proposal distribution.
2. These parameters are simulated, preferably in parallel, to calculate corresponding outputs.
3. A surrogate is trained to produce similar outputs conditioned on model parameters.
4. The posterior is estimated using either Stochastic Variational Inference (SVI) or No-U-Turn Sampling (NUTS), which in turn uses a likelihood function which is calculated using model outputs provided by the trained surrogate.
5. Posterior estimates are added to the surrogate training data, the surrogate is retrained, and the posterior is re-estimated. This is referred to as sequential sampling.

Let $f : \mathcal{X} \rightarrow \mathcal{Y}$ be the model to be emulated. I refer to $\tilde{f}_\phi$ as a surrogate model. Let $\mathcal{L}(\mathbf{x}, \mathbf{y}, \tilde{f}_\phi)$ be the loss of the surrogate model with respect to $\mathbf{y}$ (where $\mathbf{y} = f(\mathbf{x})$).

The goal of surrogate modelling is to find a parameter set that replicates the original model accurately over the entirety of $\mathcal{X}$. That is, we want to find $\hat{\phi}$:

$$\hat{\phi} = \arg \min_{\phi} \int_{\mathcal{X}} p(\mathbf{x}) \mathcal{L}(\mathbf{x}, \mathbf{y}, \tilde{f}_\phi) d\mathbf{x}$$

Here, $p(x)$, allows me to ascribe a relative importance for accurately recreating certain areas of the parameter space. Note that if $p(x)$ is uniform, this surrogate modelling procedure is akin to empirical risk minimisation, described in (Vapnik, 1991). I outline two alternatives for $p(x)$ later in this section.

In practice, it is impossible to integrate over the entirety of $\mathcal{X}$, as the parameter space is a high-dimensional infinitely large space (see Table 2.2), so I instead optimised over samples $x_i \sim p(\mathbf{x})$, where $i \in 0 \dots N$. I then investigated the sample size, N , to effectively approximate the underlying model.

$$\hat{\phi} = \arg \min_{\phi} \sum_{i=1}^N \mathcal{L}(\mathbf{x}_i, \mathbf{y}_i, \tilde{f}_\phi)$$

I assumed the mean squared error (MSE) as my loss function making my objective function:

$$\hat{\phi} = \arg \min_{\phi} \sum_{i=1}^N \left\| \mathbf{y}_i - \tilde{f}_\phi(\mathbf{x}_i) \right\|^2$$

Full & Fixed surrogate formulations

It is helpful to view the bi-stable equilibrium model, introduced in the previous chapter, as a function, see Equation 3.4. As input it takes the model parameters to infer, θ , and parameters to characterise each site, Λ and η , representing the transmission level and demographics respectively. The outputs include the probability of detection by microscopy pos_M_i , the incidence rate inc_i , and the probabilities of pre-erythrocytic infection, clinical infection and detection b_i, ϕ_i, q_i for each age group i .

The likelihood function, see Equation 2.9, uses aggregated microscopy and incidence outputs and calculates a likelihood of the observational data considering the study type and sample size.

$$f_{\text{eq}} : (\Lambda \times \eta \times \theta) \rightarrow (\text{pos_M}_i \times \text{inc}_i \times b_i \times \phi_i \times q_i) \ (i \in [0 \dots 100]) \quad (3.4)$$

I will refer to a *full* surrogate as one which attempts to emulate f_{eq} in its entirety. The surrogate would act as a drop-in replacement for the equilibrium model. It would similarly need to be executed one time for each site and its outputs aggregated to calculate a likelihood value. Most of its outputs would be redundant for likelihood calculations, however they provide contextual outputs which could regularise training.

An alternative formulation, which I will refer to as *fixed*, trains a surrogate with a functional form which is specific to my inference problem, see Equation 3.5. Given a parameter set, the surrogate is trained to produce only the aggregated prevalence and incidence rates which are relevant for the current problem $\text{pos_M}_{2-10}^j, \text{pos_M}_{>10}^j, \text{inc}_{<5}^j, \text{inc}_{5-15}^j, \text{inc}_{>15}^j$. In this scenario, the surrogate would have to learn a lower dimensional input and output space, but would have to be re-trained if one would like to introduce observational data with new age ranges or sites.

$$f_{\text{fixed}} : (\Lambda \times \eta \times \theta) \rightarrow (\text{pos_M}_{2-10} \times \text{pos_M}_{>10} \times \text{inc}_{<5} \times \text{inc}_{5-15} \times \text{inc}_{>15}) \quad (3.5)$$

Both of the formulations described decompose the likelihood function into independent calculations for each site. This decomposition reduces the dimensionality of the outputs required of the surrogates. Such a reduction could simplify and accelerate the training of the neural networks, as the models now have fewer interactions and complexities to learn.

Such a formulation is not limited to this specific model but can be broadly applied to any parameter inference problem where the likelihood is constructed from model outputs across multiple sites, a common scenario in epidemiological studies.

Strategy name	Intrinsic parameters	Site parameters	Outputs
<i>prior_fixed</i>	prior	fixed	fixed
<i>lhs_fixed</i>	lhs	fixed	fixed
<i>prior_full</i>	prior	lhs	full
<i>lhs_full</i>	lhs	lhs	full

Table 3.1: The combination of sampling strategies considered for the range and domain of f_{fixed} or f_{eq}

Prior & Space filling sampling strategies

For each formulation, I sampled model parameters from the input space, $\mathcal{X}$, calculated the corresponding outputs in $\mathcal{Y}$ using the bi-stable equilibrium model, and stored the input-output pairs for training. All of the model parameters represented in $\mathcal{X}$ are positive, continuous and some were bounded in $[0, 1]$. I considered two strategies for sampling values, *prior* and *lhs*.

The *prior* strategy samples training points from the prior distribution, as is proposed for surrogate training in (Jabot et al., 2014), with distributions which were selected in the previous chapter. If an informative prior distribution over the parameter space is known, it could be used to improve the coverage of the training set on a sub-space in $\mathcal{X}$, leading to better approximations during inference. This relies on high-posterior densities lying in that same sub-space.

The *lhs* strategy instead samples the parameters from a Latin Hyper-Cube, as suggested in (G. Charles, Wolock, et al., 2023; Andrianakis et al., 2015), with margins of the hypercube selected to include a sensible range of values for each parameter. With computationally intensive models, it is unlikely that an informative prior is available, and a range of valid values requires much less subjective decision making than the distributional approach used in *prior*. Using this strategy would allow researchers to perform prior tuning and predictive checks using the surrogate.

I combined each surrogate formulation with each sampling strategy to experiment with four surrogates, listed in Table 3.1. For the *full* formulations, site parameters were sampled according to the *lhs* strategy, whereas for the *fixed* formulations, the site parameters from the observational data, see Table 2.1, were used for training set construction. The hyper-parameters for each sampling strategy are listed in Table 3.2. For each sampling strategy, the data set is standardised by removing the mean and scaling to unit variance.

3.2.2 Posterior samplers

With a surrogate model that produces model outputs for prevalence and incidence, I can look towards sampling from the posterior distribution based on the evidence calculated from observations. I explored three alternative posterior sampling methods.

No-U-Turn Sampler (NUTS)

The first method applied the same NUTS kernel as described in the previous chapter. However, given the convergence diagnostics I initially observed - with $\hat{r}$ values far greater than one and effective sample sizes between 1 and 20 despite 500 posterior samples - I increased the number of warm-up samples from 100 to 500. On further investigation, it became apparent that increased warm-up samples would not improve convergence. These findings are discussed in subsection 3.3.3.

Stochastic Variational Inference (SVI) with B-NAF

The second method was Stochastic Variational Inference (SVI) using a Block Neural Autoregressive Flow (B-NAF) variational model. B-NAF is a Neural Autoregressive Flow described in detail in (De Cao, Titov, and Aziz, 2019). I configured the B-NAF model with five flows, each containing 8×8 blocks. The model was trained over 50,000 training steps, with each training step using $N = 128$ samples from the likelihood to estimate the Evidence Lower Bound (ELBO) loss. After training, I sampled 500 posterior samples from the variational model. This training procedure equates to optimising the following objective:

$$\hat{\phi} = \arg \min_{\phi} \frac{1}{N} \sum_{i=0}^N \log p(x, z_i) - \log q_{\phi}(z_i) \quad \text{where} \quad z_i \sim q_{\phi}(z)$$

Parameter	Support	Prior	LHS Range
Pre-erythrocytic immunity:			
κ_b	$\mathbb{R}^+$	LogNormal(0, 0.25)	[0, 10]
u_b	$\mathbb{R}^+$	LogNormal(0, 1)	[0, 10]
b_0	$\mathbb{R}^+$	Beta(1, 1)	[0, 1]
IB_0	$\mathbb{R}^+$	Normal(50, 20) ($25 < IB_0 < 75$)	[25, 75]
Clinical immunity:			
κ_c	$\mathbb{R}^+$	LogNormal(0, 0.25)	[0, 10]
u_c	$\mathbb{R}^+$	LogNormal(0, 1)	[0, 10]
ϕ_0	[0, 1]	Beta(5, 1)	[0, 1]
ϕ_1	[0, 1]	Beta(1, 2)	[0, 1]
IC_0	$\mathbb{R}^+$	Normal(25, 10) ($5 < IC_0 < 50$)	[5, 50]
PM	[0, 1]	Beta(1, 1)	[0, 1]
dm	$\mathbb{R}^+$	Normal(50, 20) ($5 < dm < 100$)	[5, 100]
Immunity to detection:			
κ_d	$\mathbb{R}^+$	LogNormal(0, 0.25)	[0, 10]
u_d	$\mathbb{R}^+$	LogNormal(0, 1)	[0, 10]
d_1	[0, 1]	Beta(1, 1)	[0, 1]
ID_0	$\mathbb{R}^+$	Normal(25, 10) ($5 < ID_0 < 50$)	[5, 50]
fd_0	[0, 1]	Beta(1, 1)	[0, 1]
γ_d	$\mathbb{R}^+$	LogNormal(0, 2)	[0, 10]
ad_0	$\mathbb{R}^+$	Normal(70, 1) ($40 < ad_0 < 100$)	[40, 100]
rU	$\mathbb{R}^+$	LogNormal(0, 1)	[0, 10]
Site parameters:			
Λ	$\mathbb{R}^+$	-	[0, 500]
η	$\mathbb{R}^+$	-	[20, 40]

Table 3.2: Intrinsic parameters to be inferred. The "Support" column shows valid the valid sets of values for each parameter. The "Prior" column lists the prior distribution used for inference and the Prior sampling strategies. The "LHS Range" column lists my restrictions on the range of values which can be sampled from the Latin-Hypercube

Annealed Stochastic Variational Inference (Annealed SVI)

In the third method, which I refer to as “annealed SVI”, I adapted the ELBO objective to balance the importance of the evidence relative to the prior likelihood according to a temperature parameter T . In my case, the target conditional distribution in the ELBO can be expressed in terms of the prior distribution $p(z)$ and the evidence $p(x|z)$:

$$\text{ELBO}(q_\phi(z)||p(z|x)) = \mathbb{E}_{q_\phi}(\log p(z) + \log p(x|z) - \log q_\phi(z))$$

I annealed the evidence using a temperature T , leading to the following objective:

$$\hat{\phi} = \arg \min_{\phi} \frac{1}{N} \sum_i \log p(z_i) + \log p(x|z_i)^{-T} - \log q_\phi(z_i)$$

As an annealing schedule, I used a geometric temperature schedule where each round, $i \in 0 \dots 9$, optimised at a constant temperature, $T = r^{-j}$, where $j = 9 - i$ and $r = 0.45$. Note that at each round the temperature would decrease until $T = 1$ for the tenth, and final, round when the unchanged ELBO was optimised.

3.2.3 Sample space refinement

Here I address the issue of non-informative training sets. The initial training sets sampled with the above strategies may not be adequate for training surrogates capable of accurately sampling the posterior space during inference, as the parameter space is very large. To overcome this, I performed sequential posterior sampling, aimed at refining the surrogate training set towards more informative regions for posterior sampling.

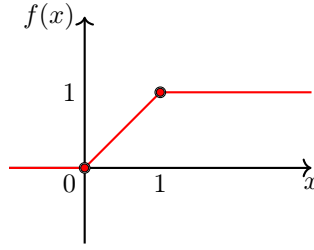


Figure 3.4: Output activation function. The function I used to constrain the surrogate to produce valid outputs for inference

Output	Range
Fixed surrogates:	
pos_M	$[0, 1]$
inc	$\mathbb{R}^+$
Full surrogates:	
pos_M	$[0, 1]$
inc	$[0, 1]$
b	$[0, 1]$
ϕ	$[0, 1]$
q	$[0, 1]$

Table 3.3: Surrogate output ranges

As in previously published sequential sampling methods, (Papamakarios, Sterratt, and Murray, 2019; Greenberg, Nonnenmacher, and Jakob H. Macke, 2019), I carried out inference in rounds, with each round’s estimated posterior being added to the training set for the next round. The premise is that the estimated posterior will progressively converge towards the true posterior, provided that the initial samples are capable of training a surrogate that generates sufficiently informative posterior estimates.

In the first round I drew n samples from the proposal distributions outlined in [section 3.2.1](#). In each subsequent round, I drew n model samples from the posterior using the samplers outlined in [subsection 3.2.2](#). Each sample resulted in $n \times 6$ model runs, since there were six different sites in this study, which was sub-sampled to create a new training set of size n . Each surrogate was then retrained from scratch on a concatenation of the initial training samples and posterior training samples from all preceding rounds. This sequential sampling process was run for 10 rounds. I explored values of n ranging between 10^2 and 10^4 , resulting in total sample sizes ranging between 10^3 , 10^5 , in order to assess convergence behaviours.

As with previous studies, I attempted to preserve the posterior sampler’s state between rounds. In the case of the NUTS sampler, I tried to preserve the state by setting the initial state of the chains to the final posterior sample from the previous round. However, this lead to unstable inference since the sampler would often struggle to find a suitable starting point for inference. To complete inference, I sampled the initial state for each round from the prior distribution. In the case of the SVI samplers, I preserved the state by re-using the optimised hyper-parameters from the previous round at the start of optimisation for each round.

3.2.4 Surrogate training

I a multi-layer perceptron as the surrogate architecture, with 3 layers of 256 linear units and rectified linear units as the activation function. The outputs of the surrogate were clipped to restrict them to valid ranges, see [Figure 3.4](#) for an example output which was restricted to $[0, 1]$. I imposed this constraint so that the surrogate outputs would always be valid for the likelihood function and not interfere with inference. The range for each output is listed in [Table 3.3](#).

The gradients in the network were optimised using Adam after each mini-batch of 100 samples to minimise a mean squared error loss function. For regularisation, I set a dropout rate of 20% in

the hidden layers. The training set was randomly shuffled after each epoch. The surrogate and optimiser were implemented using Flax (Heek et al., 2023) and Optax (Babuschkin et al., 2020) respectively.

Two test data sets were constructed to evaluate the surrogate at different levels of generality. The *surrogate posterior* test sets were sampled using each of the posterior samplers outlined in subsection 3.2.2. The *prior* test set consisted of 600 out-of-sample input/output pairs generated using the *prior_fixed* and sampling strategy.

I have made the code publicly accessible to reproduce these experiments in (G. Charles, 2024e).

3.3 Results

After executing the above alternatives, I evaluated the approximation quality of surrogates, how it was influenced by the sample size and strategy, and how it affected the parameter inference. I compared the surrogate outputs to outputs computed from the auto-differentiable model, posterior estimates to a reference posterior which I computed for the previous chapter and the speed of inference. These results revealed the convergence behaviour and speed of inference of the alternative workflows for estimating malaria transmission parameters.

3.3.1 Surrogate approximation quality

The approximation quality was evaluated against validation inputs and outputs from the prior distribution and the estimated posterior distributions. The approximation quality was measured using the mean squared error (MSE) for both standardised - centred at zero and scaled to unit variance, see Figure 3.5, and the surrogate outputs in their original scale, Figure 3.6. The standardised MSE was used to assess the accuracy of the surrogate training process, while the MSE on the unscaled outputs provided insights into the errors on the original scale, which more directly affects downstream tasks such as parameter inference. It is important to note that the posterior distributions estimated using the surrogates may vary from the reference posterior.

As expected, surrogates that were trained on prior data demonstrated the best performance on unseen prior samples. Interestingly, the surrogate trained on the *prior_full* data showed comparable performance to *prior_fixed* despite its higher dimensional output. This can be explained by the universal approximation theorem, which states that neural networks can approximate any measurable function given that it has enough hidden units (Hornik, Stinchcombe, and H. White, 1989). In this example, the NN architecture was sufficiently wide and deep to approximate both the *prior_fixed* and the *prior_full* functions. Since *prior_fixed* is in fact a functional composition of *prior_full* and an aggregation, one could surmise that the higher dimensional output provided by the *prior_full* function is redundant and that both surrogates learn similar, low-dimensional representations in their hidden layers leading to similar performance. However, a functional analysis of the trained surrogates would need to be carried out to confirm such claims. For training set sizes larger than 6,000 samples, there were diminishing returns with respect to improvements in the approximation quality for the *prior_full* and *prior_fixed* proposal distributions. The LHS proposal distributions showed less reliable improvement with increased sample size, especially in the *lhs_full* case, pointing to the importance of curating the distribution of the training set.

Without standardisation, I observed that incidence was better approximated than prevalence outputs. This discrepancy is likely due to the larger scale of plausible prevalence values compared to incidence. It is standard practice to train neural networks on standardised data and independently standardising each type of data can improve training performance. However, this approach may lead to uneven approximation performance when recovering the original data distribution. Further exploration is needed to develop methods that ensure equal approximation quality across different types of data. This also suggests that the model's overall predictions, after inference, could be more sensitive to discrepancies in prevalence outputs.

However, when assessing the quality of the estimated posteriors, all proposal distributions which used the annealed SVI sampler consistently approximated the underlying model best, see Figure 3.7 and Figure 3.8. The annealed SVI sampler produced sound approximations even at the minimum sample size, 100 samples per round. More surprising was the finding that increasing the number of rounds in sequential sampling did not improve the approximation quality for the other posterior samplers. This suggests that the assumption underlying sequential sampling - that the

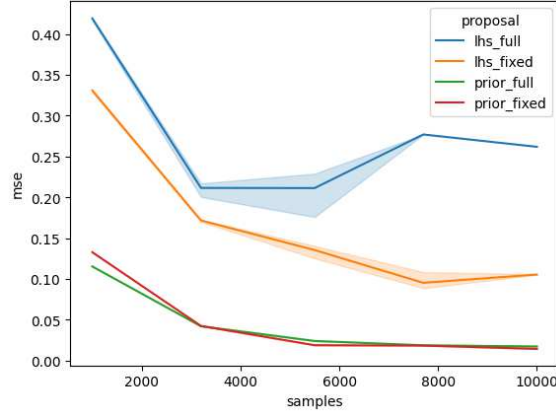


Figure 3.5: Mean Squared Error of standardised surrogate outputs on out-of-sample parameter draws from the prior distribution. Values are aggregated across the different posterior samplers, with the mean shown as a solid line and the 95% confidence interval shown by the shaded area. The samples on the X-axis represent the number of samples which were drawn for training each surrogate. With the LHS proposal functions, the distributional shift of the validation set led to a non-monotonic decrease in the approximation error of the surrogate with respect to training set size - which is not ideal if you would like to assign a simulation budget.

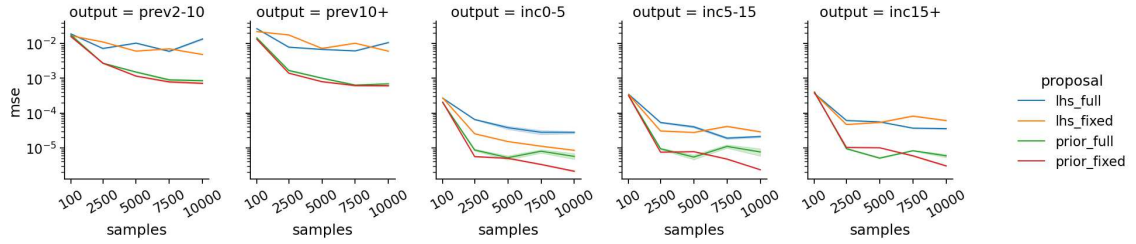


Figure 3.6: Mean Squared Error of surrogate outputs on parameter draws from the prior distribution in their original scale, disaggregated by the type of output. This was evaluated on the "prior" validation set, using the surrogates which were trained on the initial training set, i.e. no posterior sampling and surrogate re-training had occurred. The surrogate trained on the lhs_full proposal distribution produced 7 MSEs greater than 10^{10} which were removed as outliers. The 95% confidence interval is shown in the shaded area. The interval is not visible in most cases as the MSE is plotted on a log scale to accommodate the large variation in errors.

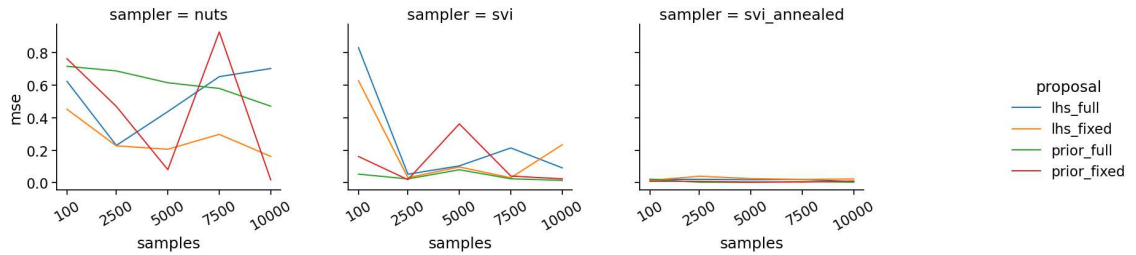


Figure 3.7: Mean Squared Error of surrogate outputs on parameter draws from the posterior distribution, estimated using each surrogate. This was evaluated on each surrogate after 10 rounds of sequential posterior sampling with their respective posterior samplers. The X axis labelled "samples" refers to the number of training samples introduced before each round of posterior sampling. Each point represents a single execution of inference, since inference was computationally expensive even in the surrogate setting, therefore no confidence intervals are shown.

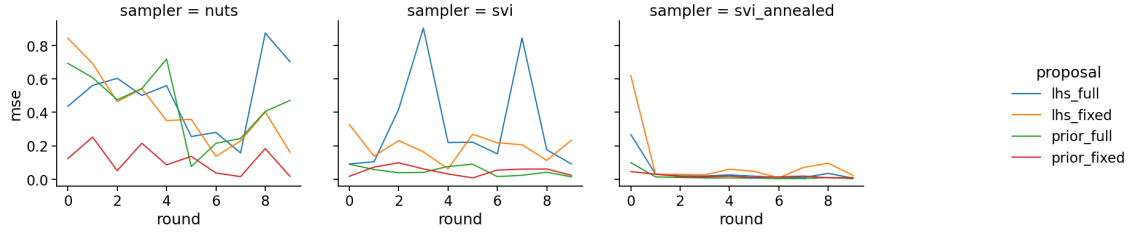


Figure 3.8: Mean Squared Error of surrogate outputs on parameter draws from the posterior distribution, estimated using a surrogate after each round. This was evaluated on each surrogate using a sample size of 10,000 for each round. Each point represents a single execution of inference, since inference was computationally expensive even in the surrogate setting, therefore no confidence intervals are shown.

estimated posterior in early rounds adequately covers the true posterior - may not hold in those contexts despite even the largest sample size of 10,000 samples per round.

3.3.2 Agreement of surrogate posterior estimates with non-surrogate based inference

I then evaluated how the posterior estimates using the surrogates with different sample sizes agreed with non-surrogate based posterior estimates. This analysis focused on the concurrence between posterior estimates obtained using each surrogate model and those generated by the underlying model, which I refer to as the reference posterior. For the reference posterior, estimates were made using the NUTS sampler as outlined in the previous chapter. Posterior estimates for the surrogates were sampled using each posterior sampler listed in subsection 3.2.2 with the likelihood function calculated from surrogate outputs.

To quantify the agreement between the surrogate-derived posteriors and the reference posterior, I used a two-sample Kolmogorov-Smirnov (KS) test on each model parameter (see Figure 3.9a and Figure 3.10). The KS test poses the null hypothesis that two one-dimensional empirical distributions are the same. It uses the metric, $D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|$ where $D_{n,m}$ is the KS statistic, $F_{1,n}(x)$ and $F_{2,m}(x)$ are the empirical distributions of the two samples, n and m are the sample sizes, and $\sup_x$ represents the supremum function over all possible values of x . A large value of $D_{n,m}$ suggests a significant difference between the two distributions, leading to the rejection of the null hypothesis. It is worth noting that as a one-dimensional test, the KS test does not account for correlations between parameters.

I also considered the C2ST metric for quantifying the agreement between posteriors (see Figure 3.9b). This metric has become popular since it was revisited by (Lopez-Paz and Oquab, 2018) and used as the preferred method for benchmarking Simulation Based Inference methods by (Lueckmann et al., 2021; Glockler, Deistler, and Jakob H Macke, 2022). In essence, the C2ST metric is calculated by training a classifier to discriminate between the two posteriors and analysing the test accuracy. If the test accuracy is close to $\frac{1}{2}$, or "chance-level", then the posteriors are likely the same. Conversely, if the test accuracy is higher the posteriors are likely different. Further details on hypothesis testing using C2ST is detailed in (Lopez-Paz and Oquab, 2018). It is an easily interpretable method for comparing probability distributions over high dimensional input spaces, since classifiers which are robust to high dimensional data can be used. I followed the evaluation proposed by (Lueckmann et al., 2021) by training a 2 layer multilayer perceptron with $10 \times n$ nodes, where n was the dimensionality of my posterior space, over 10,000 epochs and evaluated the test accuracy over 5 fold cross-validation.

I observed that posterior estimates using the annealed SVI generally showed the closest agreement with the reference posterior, with KS statistics averaging at around 0.2 compared to 0.4 for NUTS and 0.5 for SVI. My inference diagnostics, which will be detailed in the next section, revealed that the NUTS sampler suffered from convergence issues during inference.

I also found that the C2ST metric was less helpful than the KS statistic for diagnosing inference issues. The KS statistic provided parameter-level information on the divergence in the posterior estimates and exhibited much more sensitivity to the different surrogate methods and sample sizes which I investigated.

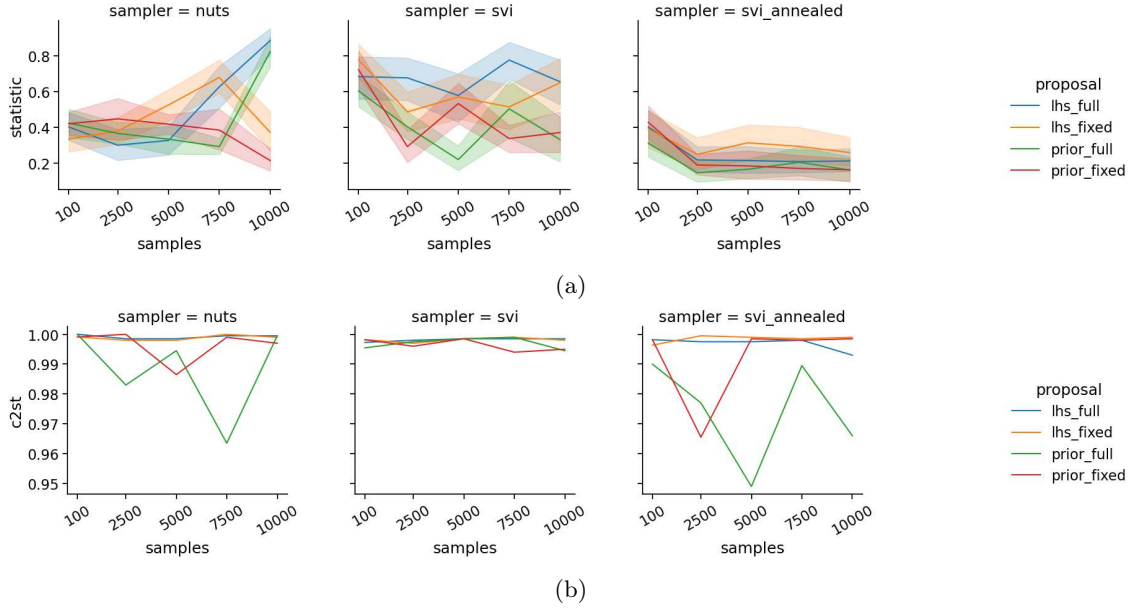


Figure 3.9: Concurrency of surrogate posteriors with underlying posterior at different sample sizes. In [Figure 3.9a](#) I estimated concurrence by taking the mean of the KS statistics for each parameter, solid line, with the 95% confidence interval, shaded region, between the reference posterior and a posterior estimated with each posterior sampler using surrogate after 10 rounds of sequential sampling. [Figure 3.9b](#) instead uses the C2ST metric for comparing posterior estimates. The X axis, "samples", refers to the number of samples introduced in each round.

Additionally, I evaluated the model fit likelihood (calculated using the underlying model) of the observed data given the posteriors estimated by the surrogates. This evaluated both the approximation quality and the suitability of the posterior samples for explaining the observed data. See [Figure 3.11](#) for a summary of these likelihoods for each posterior sampling method at increasing sample sizes. The annealed SVI posterior sampler was the only method to provide comparable model fits to the reference posterior. Other methods produced more variable, and thus less reliable, model fits.

Lastly, I evaluated the likelihood of the posterior predictive distributions estimated using the surrogates, given the latent model variables θ^* . This measure helped compare the approximated posterior predictive distribution against the latent data-generating process. See [Figure 3.12](#) for a plot of the likelihoods for each surrogate. Again, the surrogate trained using the annealed SVI posterior sampler generated a posterior predictive which was closest to the reference posterior predictive. Again, the annealed SVI sampler generated posterior predictive distributions which were most viable. The variability in the predictive performance of the NUTS sampler suggested that the sequential sampling was divergent.

3.3.3 Convergence diagnostics of posterior samplers

The convergence diagnostics of NUTS revealed poor quality samples, with $\hat{r}$ values much greater than 1 and effective sample sizes mostly below 10 for 500 sampling steps, despite increasing the warm-up period from 100 (which was sufficient to generate the reference posterior) to 500 steps. This suggested that the surrogate models approximate the model in such a way that makes posterior geometries that are more challenging to sample. On closer inspection, the surrogates produced far noisier and discontinuous gradients compared to the underlying model, as illustrated in [Figure 3.13](#), which could be a contributing factor. However, the SVI samplers showed very minor improvements in ELBO loss after 50 samples, see [Figure 3.15](#) for the losses which I recorded during training. Both scenarios suggest a rapid convergence of SVI on the estimated posterior and room for further performance improvements with meta-optimisation techniques such as early stopping and learning rate scheduling.

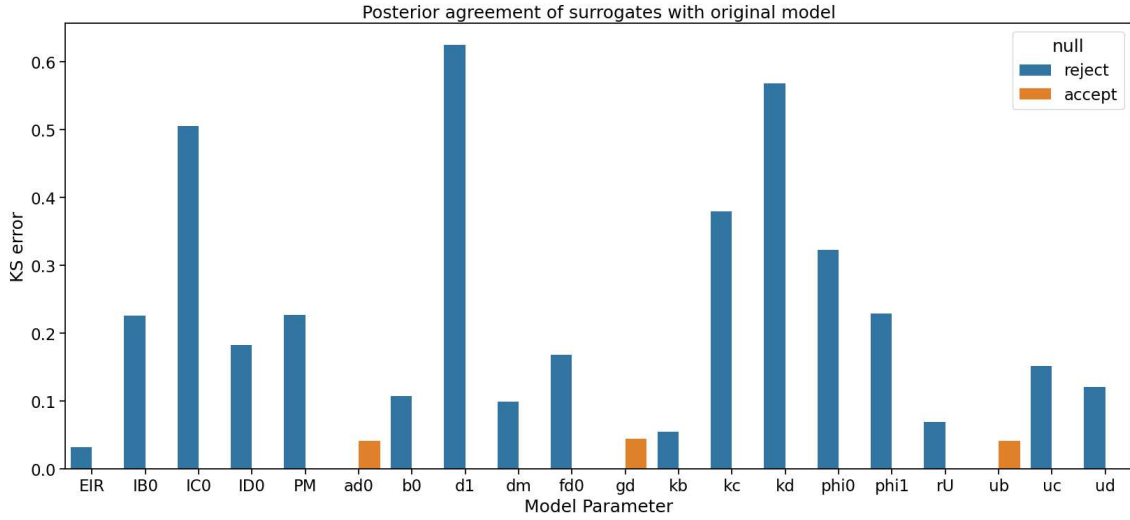


Figure 3.10: Concurrence of posterior, estimated by the surrogate trained with the "prior_fixed" proposal and "svi_annealed" posterior sampler after 10 rounds, against the underlying posterior estimated using "nuts" on the underlying model. The colour of each bar represents whether the null hypothesis, that the values of each posterior were drawn from the same distribution, can be accepted at the 95% confidence interval.

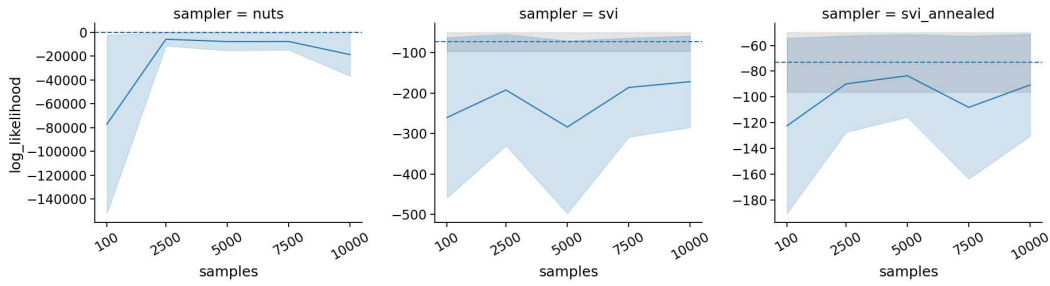


Figure 3.11: The log-likelihood of observational data evaluated on the underlying model using posteriors estimated from each posterior sampler after the final round of sequential posterior sampling. Each surrogate was trained using the "prior_full" proposal distribution. The mean likelihood is shown by the solid line and the 95% confidence interval is shown by the shaded region. The log-likelihood of the observational data using the reference posterior is represented by the dashed line and grey region representing the mean and 95% confidence interval respectively.

bb

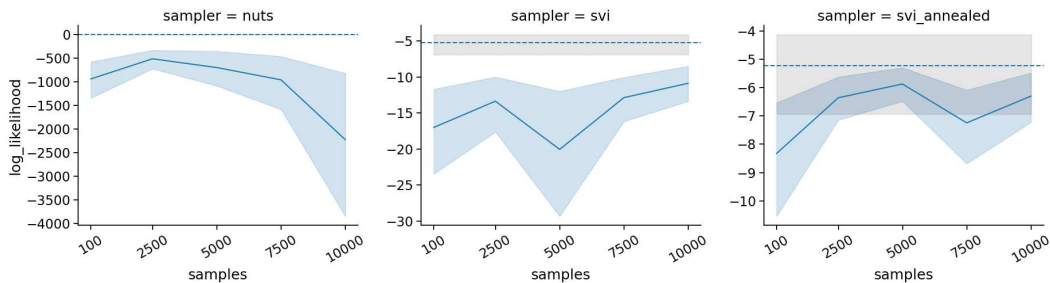


Figure 3.12: The log-likelihood of the posterior predictive distribution, estimated by each posterior sampler after each round of posterior sampling, estimated using the underlying model. Each surrogate was trained using the "prior_full" proposal distribution. The mean likelihood is the solid line and the 95% confidence interval is the shaded region. The dotted line and the grey area shows the mean and 95% interval respectively of the log-likelihoods of the reference posterior predictive.

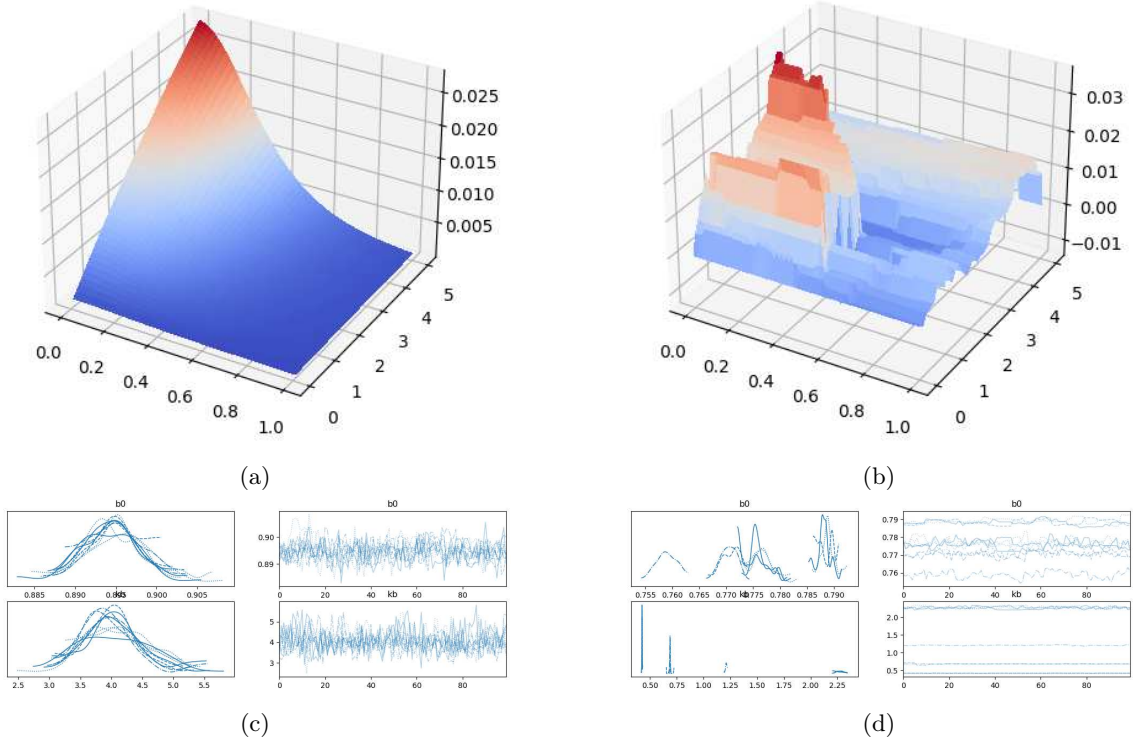


Figure 3.13: An illustration of the effects of non-smooth gradients on No-U-Turn-Sampler (NUTS) based MCMC. I plotted the gradient of the likelihood function, listed in Equation 2.9, as estimated by the underlying model - shown in Figure 3.13a - and estimated by a surrogate trained on 10,000 samples from the "prior_fixed" proposal distribution - shown in Figure 3.13b. The gradients were plotted with respect to two model parameters: b_0 and κ_b across a 50×50 grid of valid values. Traces from their respective posterior samples using NUTS were plotted in Figure 3.13c and Figure 3.13d. The surrogate which produces non-smooth outputs is not suitable for NUTS sampling as chains easily become lost in local optima.

3.3.4 Speed of inference

To assess the potential speed improvements offered by surrogate-based workflows, I timed the various components within the workflow. This was in order to estimate the speed improvements that could be expected under different scenarios to facilitate faster inference and experimentation. The timing data were aggregated from inference runs conducted during the above convergence experiments, listed in Figure 3.16 and Figure 3.17. All experiments were executed using JAX in a vectorised manner on a single Nvidia A100 GPU.

The sampling phase of the underlying model is the most predictable part of the workflow, as it scales linearly with the number of samples. This scaling is also dependent on the available computational resources. In this case, solving the equilibrium solution for a single sample took approximately 3.53×10^{-4} seconds, and batch sampling remained constant at around 2 seconds for each sample size due to parallelism on the GPU. However, for complex malaria agent-based models (ABMs) such as EMOD, OpenMalaria, and the Griffin model, sampling a single instance could take hours, making it the most computationally intensive component of the workflow.

Despite differences in output dimensions, neural network training times remained relatively consistent across the different surrogates, requiring under 50 seconds to complete 1,000 epochs for up to 10,000 samples. Training time variability can also depend on optimisations within JAX's execution environment and its just-in-time (JIT) compilation.

Inference times using the NUTS sampler varied significantly. Running NUTS with 500 warmup samples and 500 posterior samples took between 3 to 14 minutes, with an average of 11 minutes. This was a substantial improvement over the 5 hours required to sample the reference posterior on the CPU. Conversely, sampling using Stochastic Variational Inference (SVI) sampler, which involved training over 50,000 samples from the surrogate before generating 500 posterior samples,

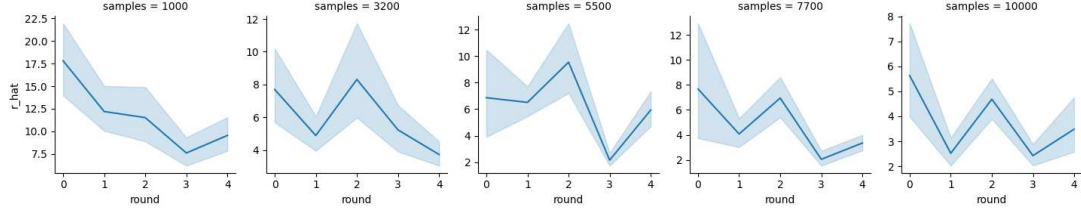


Figure 3.14: $\hat{r}$ for surrogate methods using NUTS. The mean $\hat{r}$ measure for all estimated parameters is shown as the solid line and 95% confidence interval is shaded

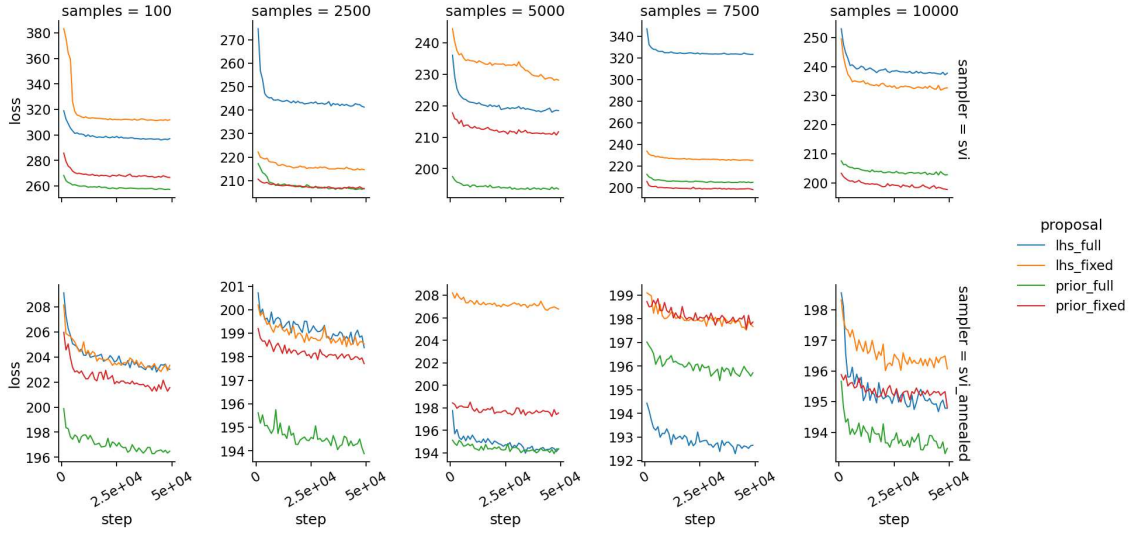


Figure 3.15: Loss, negative log likelihood as estimated by the surrogate, during SVI training for each proposal method and sample size. These losses were recorded on the final round of posterior sampling.

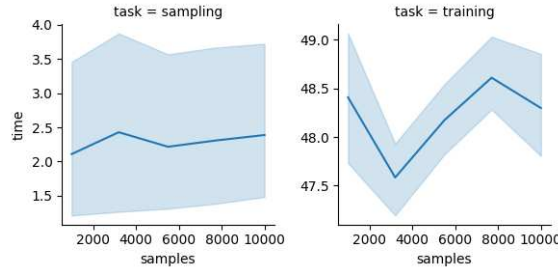


Figure 3.16: Time, in seconds, for sampling from the underlying model and training each surrogate. The time was evaluated on the initial round of sequential posterior sampling.

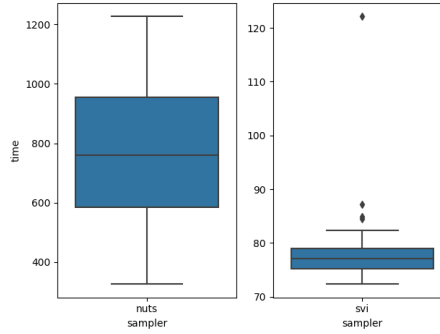


Figure 3.17: Time, in seconds, for sampling the posterior distribution for the NUTS and SVI posterior samplers. Timing values were aggregated over all sample sizes and each round as they were similar. The annealed SVI sampler had similar inference times to the SVI sampler.

took an average of 3 minutes and 30 seconds with much less variability. Although these inference times provide a general idea of execution times in a surrogate-based workflow, they could increase with the addition of more observational data, which would necessitate more frequent surrogate executions. The NUTS execution time, in particular, may be more sensitive to changes in the surrogate’s approximation error.

In practical terms, from start to finish, the NUTS and SVI-based inference workflows, with 10,000 samples and without sequential posterior sampling, took on average 11 minutes and 57 seconds and 4 minutes and 27 seconds, respectively. Sequential posterior sampling would have a multiplicative effect on the runtime which depends on the number of rounds. After 10 rounds of SVI-based inference, this workflow would have taken 40 minutes, which was an 8 times speed improvement on the 5 hour workflow I presented in the previous chapter. These results show the potential of these methodologies as rapid alternatives for inference and experimentation.

3.3.5 Surrogate-based inference diagnostics

I recreated the inference diagnostics which I introduced in the previous chapter to assess if these faster surrogate-based workflows could reliably replicate the diagnostic methods and to highlight any potential issues with the surrogates. As demonstrated earlier, the *full* surrogates, which produce the complete model outputs, are particularly advantageous as they can be used for these diagnostics. Coupled with the annealed SVI posterior sampler, the *prior_full* surrogate does not compromise on approximation or inference quality. For these diagnostics I used the *prior_full* surrogate trained with 10,000 samples, as it performed best in the above experiments.

I firstly re-ran a sensitivity analysis to determine if the gradients produced by the surrogates aligned on average with those from the underlying model. The results indicated that the sensitivity of each model parameter was broadly consistent across different surrogates. However, the sensitivity of some the parameters I identified as correlated and non-identifiable, such as ad_0 and r_u , was overestimated. The general variability of gradients was also higher than those produced using the underlying model. A comparison of the sensitivity is listed in [Figure 3.18](#). Despite the discrepancies, a surrogate based sensitivity plot could still provide accurate insights into the importance of parameters in a probabilistic model.

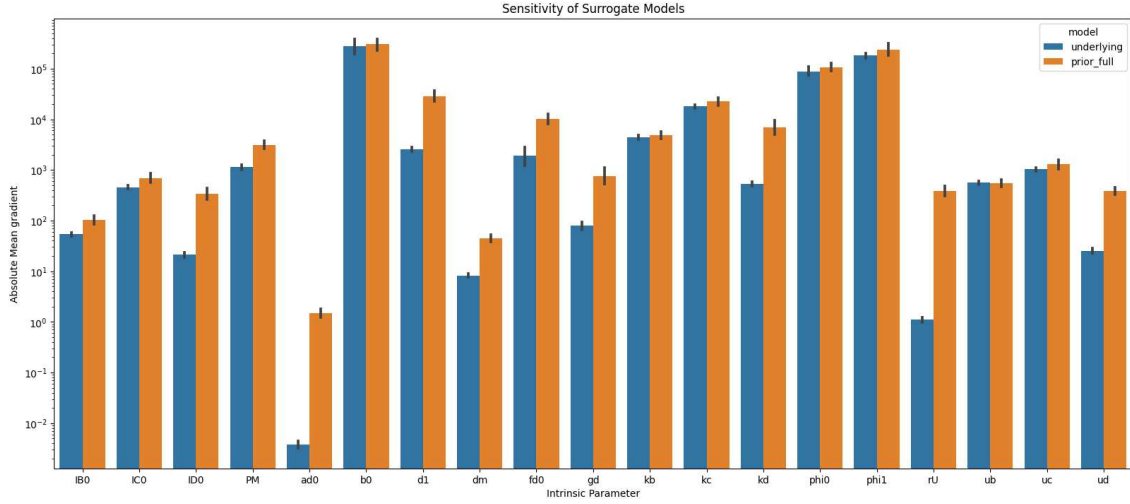


Figure 3.18: A comparison of sensitivity analyses using surrogate models. The local sensitivity, was estimated for each surrogate by calculating the absolute gradient of the likelihood function, see Equation 4.14, over 600 samples from the prior distribution. The error bars show the 95% confidence interval of the estimate.

I then performed posterior predictive checks to analyse the immunity by exposure and the prevalence/incidence by age. In both the prior and posterior space these, generally showed close agreement with the true values Figure 3.19. Bayesian p-value plots also showed a strong predictive performance of the posterior sampler, see Figure 5.9. These plots show consistency with those produced with the underlying model, making them suitable for prior tuning and sanity checking posterior draws.

Lastly, I used predictive error plots to reveal areas where the surrogate struggled in approximating the model in both the prior and the posterior, see Figure 3.20. These plots indicated that the surrogate generally did not exhibit systematic approximation issues when the validation data was drawn from the prior distribution. However in the posterior, outputs tended towards the mean. Predictive error plots, especially disaggregated by output, can help diagnose surrogate training issues. This is especially useful when training opaque surrogates such as neural networks.

3.4 Discussion

3.4.1 Stochastic Variational Inference for black box inference

Stochastic Variational Inference (SVI) with annealing showed a lot of promise for black-box inference for high-dimensional, complex inference problems. Both of the SVI samplers converged rapidly to an estimate and were more computationally efficient than NUTS. However these estimates were not necessarily close to the true posterior distribution. With the baseline SVI sampler, this caused a scenario where inaccuracies in the estimated posteriors, lead to unstable and divergent posterior samples across rounds.

Annealing the evidence term in the SVI objective appeared to address this instability by gradually transforming the surrogate training set from the prior distribution to the posterior distribution. At the start of sequential inference, the hot temperatures produced a likelihood function that was closer to the and easily sampled prior. In later rounds, the incrementally colder temperatures shifted the likelihood towards the more challenging, and yet unknown, posterior likelihood. The gradual transition helped to stabilise the sequential sampling process and ensured more reliable convergence. The empirical results showed that annealing achieved a good fit to the observed data, exhibited strong predictive performance, and demonstrated a high degree of agreement with the reference posterior.

Moreover, the annealing approach is not limited to the specific parameter inference problem addressed in this study. It could be broadly applied to any simulation-based inference problem, even in the commonly studied likelihood-free context, to stabilise the convergence of sequential posterior sampling.

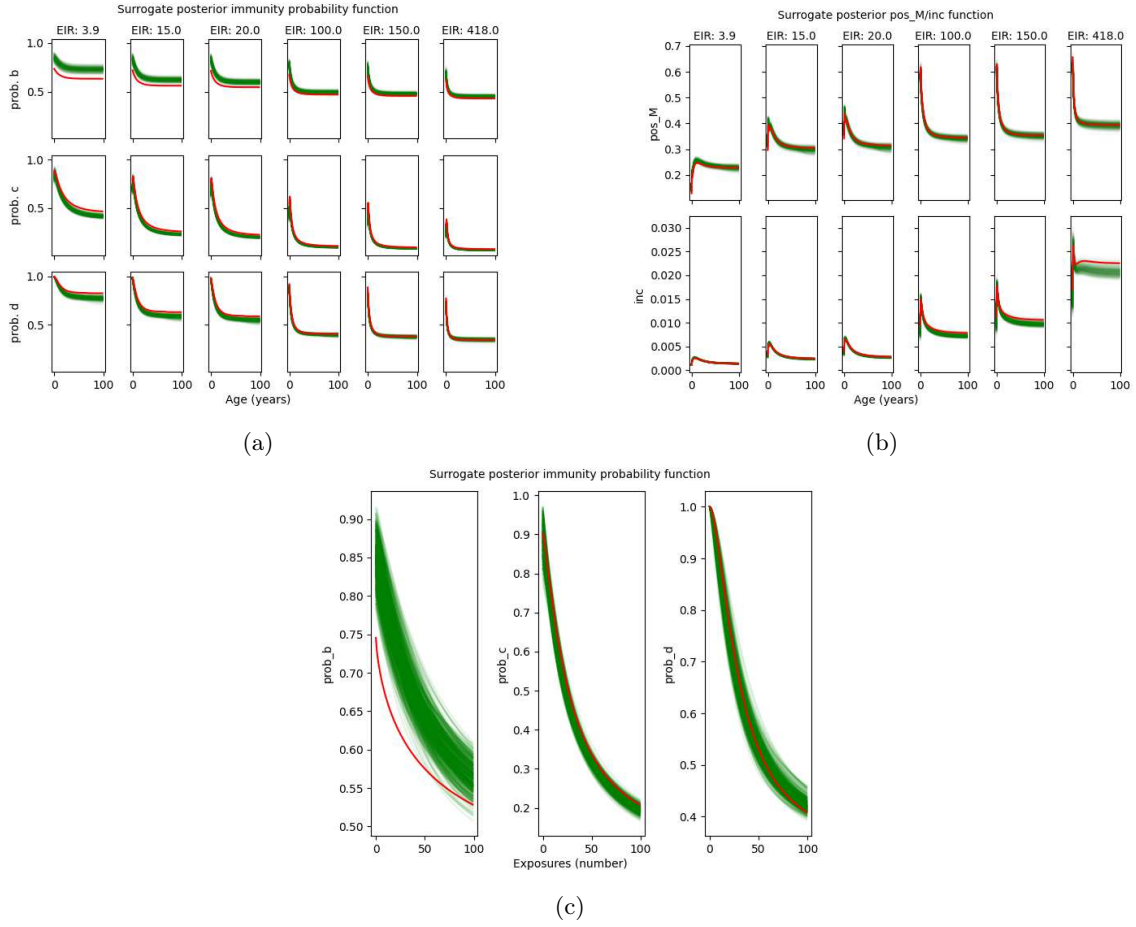


Figure 3.19: Posterior predictive checks for surrogate-based inference. I recreated the posterior predictive checks on the immunity, prevalence and incidence outputs of the model, introduced in [Figure 2.2](#), to qualitatively compare the posterior estimates (in green) to model outputs which would be generated by the true latent values (in red). The surrogate was trained using the "prior_full" proposal distribution with the annealed SVI posterior sampler and a sample size of 10,000 per round.

However, further research is needed to refine the design of annealing schedules and explore adaptive annealing strategies. In this study, the hyper-parameters for annealing - such as the geometric rate of decay, the maximum temperature, and the temperature during SVI optimisation - were determined largely through trial and error and convenience. This approach, while practical, is sub-optimal. Hyper-parameter optimisation techniques could search for more effective parameters, or adaptive annealing as described in (Huang et al., 2018; Mandt et al., 2016) could further improve the performance and stability of annealed SVI for simulation based inference. Additionally, the use of a global annealing coefficient for all observations may not be ideal, as the scale of evidence can vary significantly between different observations. Tailoring annealing coefficients to individual observations could provide more nuanced control and improve overall inference accuracy. While there has been some research on the reliable application of annealing to SVI, this area remains under-explored, especially as a method to mitigate surrogate approximation error. The current body of work does not address how annealing can be adapted to optimally improve the performance of surrogate models within Simulation Based Inference workflows.

3.4.2 No-U-Turn Sampling for black box inference

I found the No-U-Turn Sampler (NUTS) to be the least effective, primarily because of its poor convergence behaviour when the likelihood was approximated using Neural Network (NN) surrogates, likely due to the noisier gradients it produced. While there has been some research into training neural networks to produce smooth functions (Anil, Lucas, and Grosse, 2019), this field

is still developing. Current methods often compromise the expressivity of the networks, making them less suitable for complex inference tasks where approximation quality is a key factor.

There has also been research on enhancing MCMC methods to be robust against non-smooth potential functions (Chaari et al., 2015). These efforts aim to develop MCMC algorithms which are robust to the irregularities and discontinuities that arise from noisy gradients. However, these techniques still require significant computational resources and inference time, making them less appealing than the SVI alternatives.

3.4.3 Curating the input and output space of the surrogate

Selecting a proposal distribution which is close to the posterior is important for effectively training surrogates in complex models. In contexts where the parameter space is vast and intricate, an informed proposal distribution helps in concentrating the surrogate training on regions that are most relevant for accurate posterior estimation. Conversely, using uninformative priors, such as those generated by a Latin Hypercube Sampling (LHS) distribution, tends to result in surrogates with a higher approximation error, even as the number of training samples increases.

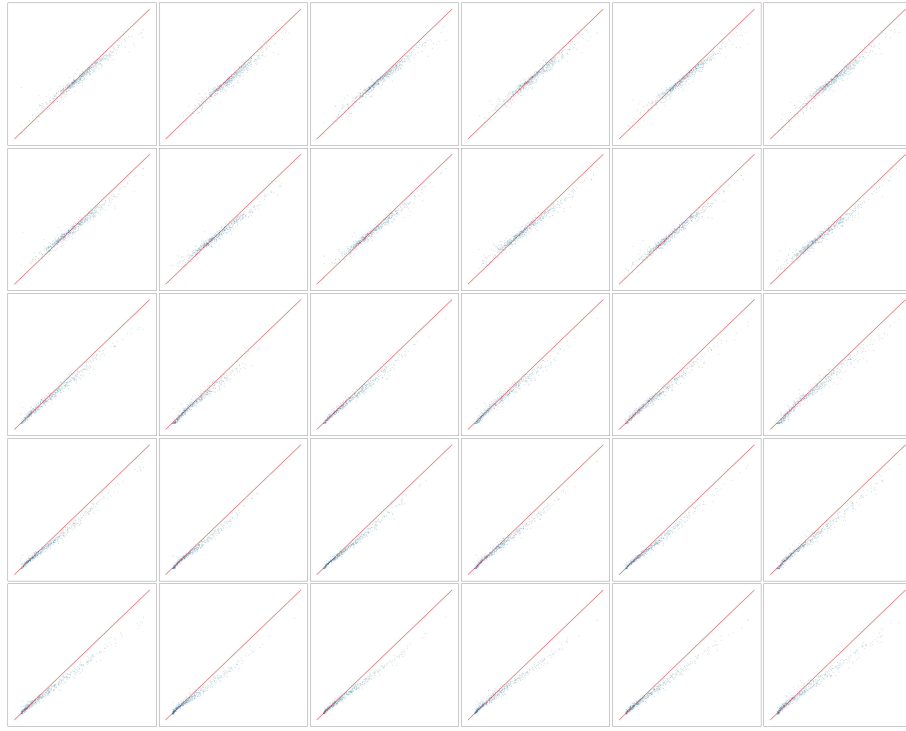
Neural networks (NNs) offer the significant advantage of being able to model high-dimensional outputs, which can include contextual outputs valuable for diagnostics. However, when using an uninformative proposal distribution, such as Latin Hypercube Sampling (LHS), the training of surrogates which produce high-dimensional *full* outputs can become unstable. In these cases, the validation loss can sometimes increase with the sample size, indicating that the model is struggling to generalise and learn effectively from the training data. This instability is likely due to the surrogate being exposed to a wide range of parameter values, many of which are irrelevant or less informative for the posterior estimation.

In contrast, with the *prior* proposal distribution, the *full* outputs provide comprehensive diagnostic information without increasing the approximation error. The flexibility and fidelity of the *prior_full* surrogate could enable researchers to perform a variety of modelling tasks, outside of parameter inference, much more efficiently. By simulating counterfactual or future scenarios, optimising parameters, and conducting other analyses using the surrogate, researchers can produce research in a fraction of the time that they typically require for simulation.

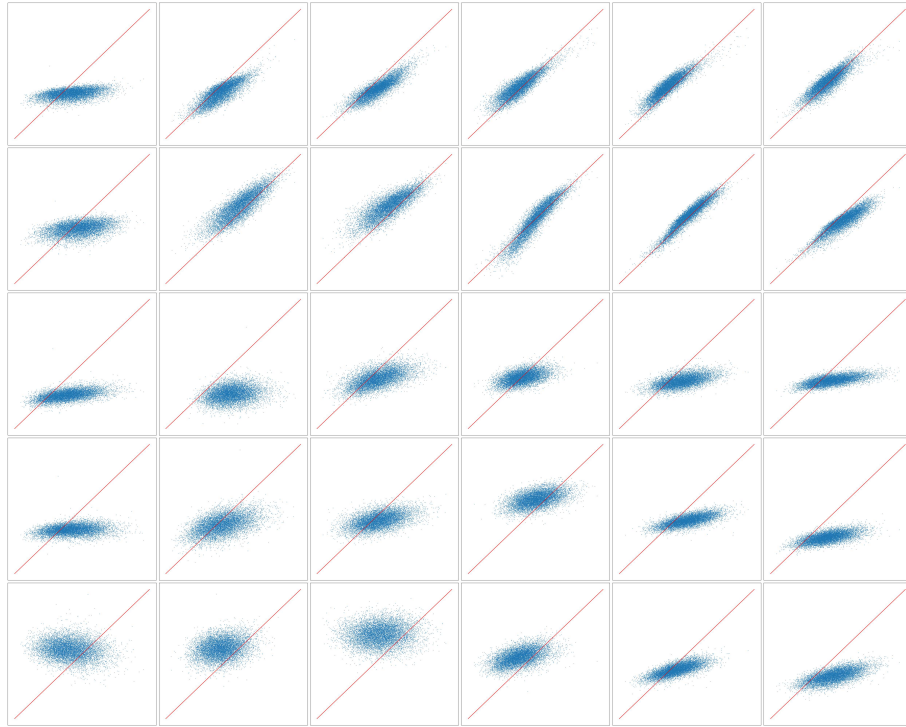
3.4.4 Complex models for black-box inference

There is a pressing need to study surrogate-based inference methods for more complex models. In applying the latest simulation-based inference methods, SNLE and SNVI (Papamakarios, Sterratt, and Murray, 2019; Glockler, Deistler, and Jakob H Macke, 2022), to this malaria model I have uncovered several shortcomings which were not previously apparent. These issues, such as instability in training with uninformative priors and the challenges in accurately approximating high-dimensional posteriors, highlight the limitations of current methodologies when scaled beyond simple, toy models with only a few parameters.

It would be beneficial for the wider scientific community to develop benchmarks which evaluate surrogate-based inference methods on more complex and realistic models. Such benchmarks would provide a more rigorous assessment of each method’s capabilities and limitations, guiding future research and development efforts. They would help identify specific areas where improvements are needed and drive innovation in creating more robust and reliable inference techniques. Creating these benchmarks would require a significant engineering effort. Developing complex models suitable for benchmarking involves careful decision-making to establish representative ground truths and observational data. However, it would push the boundaries of current inference methodologies, promoting the development of more advanced techniques capable of handling the complexities of real-world models.



(a)



(b)

Figure 3.20: Predictive error plots for model outputs at each site. I validated the predictive error against out-of-sample prior draws (Figure 3.20a) and posterior draws (Figure 3.20b). Each point represents a point in the validation data, with its x and y position corresponding to the surrogate prediction and underlying model output value respectively. The red lines represent $x = y$, i.e. no approximation error. Each plot varies model outputs ($pos_M_{2-10}, pos_M_{>10}, inc_{<5}, inc_{5-15}, inc_{>15}$) on the vertical axis and site (listed in Table 2.1) on the horizontal axis.

Chapter 4

Seq2Seq surrogates for recreating disease transmission dynamics over time

Please note that this research was previously published in (G. Charles, Wolock, et al., 2023) (with the code published in (G. Charles, 2024a)) and (G. D. Charles and S. L. Wu, 2021). All of the writing, programming and experimental work in (G. Charles, Wolock, et al., 2023) was my own with supervision and editing from the co-authors. I have adapted much of this text to contextualise it within the rest of this thesis. Since this paper was published, “malariasimulation” an agent-based malaria transmission model, see (G. Charles, Peter Winskill, et al., 2022), received contributions from a wider community of researchers to add or modify interventions, resolve bugs and optimise performance. At the time of writing I am the largest contributor to the original malariasimulation repository. The design and most of the programming of the “individual” library was done by me, with contributions from the community to add more expressive primitives to support a wider range of Agent-Based Models and for code optimisations. At the time of writing I am the largest contributor to the individual repository. The content for (G. D. Charles and S. L. Wu, 2021) was written by my collaborator, Dr. Sean Wu, and was not used in this text.

4.1 Background

In the previous chapter, I established that neural networks could serve as effective surrogate models for high-dimensional complex inference problems associated with malaria transmission. However, a limitation of the model I discussed is that it assumes a stable equilibrium in transmission. Equilibrium transmission does not represent most real-world conditions since malaria transmission is influenced by various interventions and environmental factors. With this in mind, I shift the focus in this chapter towards developing surrogate models which are designed for recreating temporal outputs, i.e. a sequence of outputs over time. These outputs are also referred to as trajectories.

4.1.1 Temporal models for malaria transmission

Two primary methodologies have been used to model infectious disease dynamics: differential equation-based (DE) models and Agent Based Models (ABMs). Both have been applied to malaria transmission modelling, but offer advantages and challenges in the simulation of disease spread and mitigation through interventions. DE models are typically quicker to simulate, even using numerical methods, for a low-dimensional system of equations. However, ABMs may be preferred when one is studying more granular interactions or a large number of interventions.

The Ross-Macdonald model, a DE model, was widely used during the first Global Malaria Eradication Programme (GMEP) from 1955 to 1969, see (D. L. Smith, Battle, et al., 2012) for a detailed evolution of the model. Despite its broad usage, the model’s simplistic assumptions, such as (a lack of) human immunity and a constant mosquito density, has limited its applicability to modern malaria modelling, which requires the simulation complex intervention scenarios and our latest understanding of the disease dynamics.

As malaria pathology research and malaria control has progressed, more sophisticated DE models have been published to study the implications on public health. For instance, (O. Collins and K. Duffy, 2022) incorporated bed nets and treatment and drug resistance to forecast their effects on transmission in Nigeria; (Adegbite et al., 2023) incorporated a travelling human population to study the effects of human movement on transmission; and Imperial College London have published a complex DE model of malaria transmission which incorporated a range of interventions, including insecticide-treated nets, indoor residual spraying, Artemisinin Combination Therapies, and seasonal malaria chemoprevention (Hellewell et al., 2023). Additionally, it modelled the detailed immunity dynamics and current research on vector behaviours such as growth, infection, and oviposition. These models can simulate more realistic disease transmission trajectories in theoretical and controlled environments, which is useful for informing current policy-making. However, as these models incorporate increasingly complex scenarios, such as multiple simultaneous interventions, they can become computationally infeasible due to the numerical challenges of solving high-dimensional systems of differential equations.

In contrast, ABMs simulate interactions within a population of agents, each representing a discrete entity whose behaviours are designed to reflect the known dynamics of disease transmission. ABMs are favoured in scenarios requiring detailed simulation of complex interventions due to their predictable computational complexity, which generally scales linearly with the size of the simulated population. This scalability makes ABMs particularly suited for handling increased complexity without exponential increases in computation time experienced when increasing the dimensionality of DE models.

Among the prominent ABMs for malaria are EMOD, OpenMalaria, and the Imperial College London model (P. A. Eckhoff, 2011; Thüring et al., 2016; Griffin, Ferguson, and Ghani, 2014). Their detail and flexibility are unmatched, as they provide functionality for understanding and exploring strategies against malaria transmission in a wide array of hypothetical scenarios. However, they require substantial computational resources, often running for hours to complete a single simulation.

4.1.2 Challenges in Agent-Based Model (ABM) development

Modern malaria transmission ABMs are most commonly implemented in systems programming languages like C/C++, see [subsection 1.2.6](#), as their execution speed is more easily optimised when developers have explicit memory management and low-level algorithmic control. However, systems programming requires advanced software engineering skills and optimised codebases can become difficult to parse, maintain and extend. Moreover, scientific researchers with less programming expertise may find it challenging to engage with such models and explore new hypotheses, inhibiting research in the field.

Many scientific researchers are more familiar with high-level scripting languages like R and Python. These languages offer several advantages for academic and research settings: they are easier to learn and use, particularly for those without a formal background in computer science; as interpreted, dynamically typed languages, researchers need not concern themselves with issues in object oriented programming, type systems, memory management or compilation; they also enable quick, iterative development, which is convenient for testing new hypotheses and making frequent adjustments to models; finally, both R and Python have rich ecosystems of libraries and frameworks which support data analysis, visualisation, and computational statistics.

Several ABM development frameworks have emerged to take advantage of the benefits provided by high-level programming languages, often through domain-specific languages (DSLs) to balance the trade-off between optimisation with flexibility. Some examples include: NetLogo, "A multi-agent programmable modelling environment" ([NetLogo Home Page 2024](#)); Mesa (Kazil, Masad, and Crooks, 2020), an open-source Python library for agent-based modelling. It allows users to create, analyse, and visualise ABMs with an interface aligned with Python's programming principles; and Agents.jl (Datseris, Vahdati, and DuBois, 2022), built in the Julia programming language, combining the ease of use found in high-level languages with the performance closer to that of low-level languages. These tools make ABM development more accessible to researchers who may not be expert programmers but want to create detailed simulations. By providing user-friendly interfaces and abstracting away some of the lower-level computational details, these frameworks help lower the barrier to entry for running sophisticated simulations.

4.1.3 Challenges with parameter inference for ABMs

Using Agent Based Models (ABMs) presents new obstacles to parameter inference. Their lengthy run times make Bayesian inference methods such as MCMC infeasible. The stochasticity in ABM outputs complicates matters further, as the probability of any output given the model parameters is intractable for all but trivial ABMs, which makes it challenging to define a likelihood function for inference. Finally, gradient estimation for ABMs is challenging due to the simulation of discrete randomness and the potential for vanishing or exploding gradients in long simulations, leaving modellers with less efficient, gradient-free methods, which are known to scale poorly in high-dimensional settings.

Building on the findings presented in the previous chapter, where neural network surrogates were used to estimate the equilibrium outputs of a malaria model, I extend the surrogates to estimate temporal outputs. ABMs do not produce a single equilibrium output but a sequence of outputs over time. This increases the dimensionality and complexity of the data, requiring a new formulation of surrogates which can effectively model such sequences as inputs and reproduce sequential outputs.

4.1.4 The site calibration problem

One of the parameter inference problems researchers face when using ABMs of malaria transmission is the “site calibration” problem. The purpose of site calibration is to estimate the baseline Entomological Inoculation Rate (EIR), or the number of infectious bites per person per unit time, for a given location, where “baseline” refers to the counterfactual EIR in conditions prior to any interventions being applied. Site calibration involves aligning model predictions of malaria prevalence with a time-series of observed historical prevalence. The model predictions are conditioned on “site parameters” which reflect the impact of various environmental factors and interventions in a particular location and could vary with time. These site parameters could include factors such as rainfall, vector composition, and records of historical interventions. Once a baseline EIR has been estimated, a modeller could simulate any counterfactual trajectory or predict future trajectories given current or alternative intervention strategies.

The prevalence statistics used for site calibration are usually gathered through cross-sectional studies (XSS). In these studies, researchers conduct routine surveys using Rapid Diagnostic Tests (RDT) or microscopy to test a portion of the population. These studies are relatively simple since they do not require a longitudinal tracking of individuals, allowing for consistent and frequent data collection over time. For instance, the Demographic and Health Surveys (DHS), see (Demographic and Health Surveys, 2022) regularly carry out XSS surveys across Sub-Saharan Africa, focusing on children aged between 5 and 59 months. Outside of the DHS, (Snow et al., 2017) have compiled a sizeable database of prevalence studies dating back to 1900, providing a valuable historical perspective on malaria prevalence.

Site calibration has been used in several important studies and global health reports, serving as a useful tool for decision-making regarding malaria interventions. A study for The President’s Malaria Initiative (PMI), a US governmental initiative, used site calibration to fit the Imperial College ABM to prevalence data for 1,020 administrative units across Sub-Saharan Africa (SSA) (Peter Winskill et al., 2017). This calibration allowed them to estimate the potential impacts of funding adjustments, informing the PMI’s decisions around future funding. Simulations for The Global Technical Strategy (GTS) for Malaria, a report published by the World Health Organisation, featured results from a modelling study which used site calibration to align Griffin model simulations with observed malaria prevalence data across SSA in order to estimate the burden of incidence of malaria - a less readily available metric (World Health Organization, 2021). The Malaria Intervention Tool (MINT) uses site calibration to align the Griffin model to a policy maker’s location of interest so that they can evaluate various scenarios of bed net distribution in their region (Sherrard-Smith et al., 2022).

4.1.5 Related work on parameter inference using ABMs

When calibrating computationally intensive infectious disease models, researchers frequently resort to optimisation techniques, see (Luo et al., 2018; Sauboin et al., 2015; Knight et al., 2015). The objective in these scenarios is to minimise discrepancies between observed time-series of disease prevalence and the outputs generated by the ABM. Common metrics for measuring these

discrepancies include absolute errors or likelihood functions. Since gradient estimates for ABMs are challenging, these optimisation efforts rely on gradient-free methods such as grid-search and Nelder–Mead search, for low-dimensional convex optimisation problems, despite being limited in their efficiency and scalability. Another drawback of using optimisation is that they yield point estimates without providing insights into uncertainty. This limitation gives no indication of the robustness of the results or the range of potential outcomes under different scenarios, which is critical in infectious disease modelling given the widespread impact of public decision making.

With more computational resources, researchers have applied probabilistic inference techniques for ABM calibration. These methods include Incremental Mixture Importance Sampling (IMIS) (Raftery and Bao, 2010), Approximate Bayesian Computation (ABC) (McCreesh et al., 2017), and surrogate modelling for Markov Chain Monte Carlo (MCMC) (Cameron et al., 2015). Due to the large computational requirements, these methods have not been used for reoccurring inference problems like site calibration, which is typically performed before modelling malaria transmission in any specific location.

In IMIS, modellers set up a Bayesian framework with a prior distribution over model parameters and specify a likelihood with respect to observed data with the goal to sample from a posterior distribution. This process is iterative: simulations are run drawing model parameters from the prior distribution, after which the likelihood of these samples given the observed data is calculated. Then an acceptable region of parameters is determined, and resampling occurs from this accepted region. The process continues until the posterior distribution is sufficiently approximated.

Instead in ABC, the likelihood function is assumed to be intractable. This is common with Agent Based Models (ABMs), where the stochastic nature of the outputs and the complexity of the models often mean that there is no closed form for the probability of outputs given model parameters. ABC avoids the need for a direct likelihood calculation by using a similarity measure between observed data and simulated data, rejecting simulations which do not meet a predefined threshold of similarity.

Surrogate modelling allows for fully Bayesian inference (Lamperti, Roventini, and Sani, 2018; Cameron et al., 2015; G. D. Charles and S. L. Wu, 2021). A tractable surrogate model is trained using a dataset of reference simulations and is then used in place of the ABM to rapidly calculate a likelihood for MCMC. This has been applied to parameter inference in many high-dimensional ABMs of infectious disease.

IMIS, ABC and surrogate models for MCMC require a considerable amount of computational resources for inference. These methods require running large numbers of computationally intensive simulations, many of which result in low likelihood outputs that are not informative for sampling the posterior. This makes these approaches inefficient and time-consuming. The reliance on extensive simulation runs to approximate the posterior distribution or to filter through acceptable simulations in ABC places can be impractical for routine use in policy-modelling or rapid response scenarios.

Advances in surrogate modelling across various scientific domains, such as physics and climate science, have introduced surrogate models which can generalise to different scenarios by incorporating inductive biases. Inductive biases steer surrogates towards learning more generalised model dynamics, enabling surrogates to perform robustly across a range of scenarios without the need for retraining. In one example, Sanchez-Gonzalez et al. (2020) proposed a surrogate which learned fluid simulations in from single time-step simulations of thousands of particles. Their research demonstrated that once the surrogate had learned inter-particle interactions, it could generalise to simulate more complicated physical conditions, with more particles on longer time horizons. In another compelling example, Kovachki et al. (2022) developed surrogates for climate models. These surrogates were trained to produce “super-resolution” simulation outputs, generating high-resolution outputs from lower-resolution data. The key to their approach was in training the surrogate to learn a discretisation-invariant mapping from model parameters to outputs. So that the surrogate could be trained on quickly generated low-resolution training data and produce outputs at alternative resolutions in a fraction of the time required by traditional models.

By adopting similar inductive biases, I hypothesised that one could develop surrogates capable of learning from a short, cheap to simulate, simulations and generalise to long simulations. This flexibility would allow a single, well-trained surrogate to adapt to various geographical settings or intervention strategies without the need for re-training. The potential to train such models once and apply them across multiple inference problems could result in substantial computational savings, and speed up response times in policy-making scenarios where rapid decision-making is needed.

4.1.6 Improving surrogates with inductive biases

To exploit the sequential nature of outputs from ABMs, I propose sequence-to-sequence (seq2seq) surrogate models to recreate ABM outputs to facilitate Bayesian Inference. These models are designed to take sequences of model inputs over time inputs and produce sequences of prevalence or incidence values. This sequential approach allows me to introduce an inductive bias towards learning autoregressive relationships, where each output depends on present and past inputs. If the seq2seq surrogates learn the autoregressive relationship which underlies the temporal dynamics of disease transmission, it would theoretically be able to predict model outputs over longer time horizons than those present in the training data. The benefits of this approach include training seq2seq surrogates on shorter sequences than those used in inference, reducing computational demands of training set generation, and applying the same surrogate to inference problems varying sequence lengths.

Recurrent Neural Networks (RNNs) are a class of neural networks which are suited to processing sequential data, making them ideal for tasks that involve sequential input. A distinguishing feature of RNNs is that they model a hidden state, denoted as h_i , which represents information learned from prior elements in the sequence for subsequent predictions. This ability to “remember” previous inputs via the hidden state is how RNNs could recreate temporal dynamics.

The fundamental computational unit of an RNN is known as a cell, which updates the hidden state for each time step. The typical computation within an RNN cell can be described by the formula for the hidden state:

$$h_i = \tanh(Ux_i + Wh_{i-1} + b)$$

Here, U and W represent the weight matrices for the current input x_i and the previous hidden state h_{i-1} respectively, while b is a bias term. The activation function $\tanh$ (hyperbolic tangent) introduces non-linearity into the model. Non-linear activation functions are an important ingredient in neural networks for learning complex patterns in data, including non-linear autoregressive functions.

RNN architectures often layer cells in order to learn more complex sequences. In a multi-layered RNN, the hidden state output from one layer becomes the input to the next layer. This stacking of layers enables the network to learn more complex representations of the autoregressive dynamics in the hidden states of the deeper layers. RNNs can be used to generate sequences by outputting the sequence of hidden states produced by the final layer.

One of the key advantages of RNNs, compared to other autoregressive models, is their operational efficiency in processing sequences. Since each element of the sequence is processed exactly once and the results are accumulated through the hidden states, RNNs are particularly efficient at relating current inputs to the history of previous inputs. This efficiency is crucial in applications where real-time processing of incoming data streams is required, such as in speech recognition or real-time anomaly detection in network systems.

While RNNs are highly effective for sequential data processing, they are not without their challenges. One issue is the problem of vanishing and exploding gradients, which commonly occurs when training RNNs on very long sequences. This problem stems from the nature of back-propagation in RNNs, where gradients of the loss function are propagated back through each time step of the input sequence to update the parameters. During this backward pass, the gradients can either shrink exponentially towards zero (vanish) or grow exponentially towards infinity (explode) as they pass through the recurrent cell many times. If the gradients vanish, the update to the weights becomes infinitesimally small, which slows down the learning process or halts it altogether. Conversely, if the gradients explode, it leads to large updates that cause learning to diverge, often manifesting as numerical overflow errors. Both scenarios result in unstable training dynamics.

To address these issues, cell architectures such as Long Short-Term Memory (LSTM) units and Gated Recurrent Units (GRUs) have been developed. These architectures incorporate mechanisms such as “forget gates” to help regulate the flow of information through the network and maintain stable gradients when inferring long sequences.

The LSTM cell, in particular, includes three gates which manage the memory and information: a “Forget gate” which decides what information is discarded from the cell state; “Input gate” which updates the cell state with new information from the current input; and an “Output gate” which determines what part of the current cell state is used to compute the output. The complete of operations is depicted in [Figure 4.1](#).

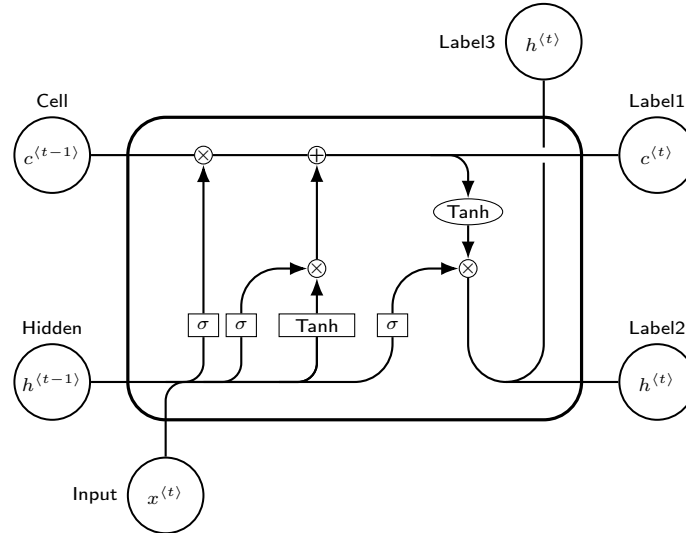


Figure 4.1: An LSTM cell

An LSTM cell. As with an RNN state, the hidden state h_t is output at every time step and re-used to calculate outputs for subsequent time steps. However the LSTM cell introduces a cell state c_t to regulate the flow of information in each cell execution. This diagram was made by <https://tex.stackexchange.com/questions/432312/how-do-i-draw-an-lstm-cell-in-tikz>

4.1.7 Contributions

Development of the “individual” Library for ABMs in R One of my contributions in this chapter has been the development of the “individual” library, a framework for implementing Agent-Based Models (ABMs) in R. This library simplifies the creation of epidemic ABMs by abstracting many common operations and providing optimised C++ implementations for these intensive tasks. It provides the most extensive collection of primitives of its kind in the R ecosystem, including a diverse range of modelling primitives to characterise agents and events. This contribution adds to the R community, providing tools which were previously available only in more complex or less accessible programming environments.

Implementation of Imperial College London’s Malaria Model Another major contribution is the development of an open and extensible implementation of the Imperial College London’s Agent-Based malaria transmission model using the “individual” library. This implementation leverages both R and C++ to offer a flexible system architecture, allowing researchers to explore new malaria interventions and alternative infection models. The design focuses on maintaining simulation times that are competitive with other contemporary malaria models while offering greater flexibility and accessibility for modifications and extensions by researchers familiar with R.

Development of a Reusable Surrogate Model for Site Calibration I also developed a novel surrogate model to produce fast approximations of the ABMs outputs, designed to be reusable across different site calibration problems. This improves on previous surrogate modelling methods which where surrogates are trained for single inference problems, requiring repeated retraining. My proposed surrogate model can adapt to various settings without the need for site-specific retraining, thereby improving the efficiency of site calibration and providing probabilistic estimates where previously researchers may have resorted to optimisation methods.

Evaluation of Surrogate Model Performance I then evaluated the surrogate model’s ability to accurately recreate time-series data that extends beyond the length of those in the training set. This evaluation demonstrates the model’s potential to be trained on shorter sequences, thus saving computational resources and mitigating common issues related to training RNNs on long sequences.

Demonstration of Site Calibration for Kolda, Senegal Finally, I applied the surrogate model to calibrate the ABM to observational data collected in Kolda, Senegal. This application suc-

cessfully provided a plausible posterior estimate for the Baseline Entomological Inoculation Rate (EIR) rapidly and without any site specific training.

4.2 Agent based modelling method

4.2.1 Individual: a framework for ABM development

I designed the “Individual” framework to simplify the development of Agent-Based Models (ABMs) for public health researchers who use R. I identified some useful programming patterns common to epidemic ABM development, particularly when scaling to large populations or modelling complex and competing agent interactions. I then created optimised functions and classes in C++ to provide these patterns to high-level R modellers.

Vectorised computation is encouraged when using the “Individual” framework. Aligning with the Single Instruction, Multiple Data (SIMD) principle, vectorisation allows for the simultaneous processing of multiple data points with a single instruction. By leveraging modern CPU optimisations for SIMD operations, models designed with “Individual” can execute faster than ABMs which perform complex logical operations for each agent in turn. This design philosophy also opens the door to future acceleration with massively parallel computing architectures such as GPUs and TPUs. This is a similar approach to modern machine learning frameworks like JAX, which are specifically built to exploit such hardware capabilities for high-performance scientific computing (see [subsection 2.2.2](#) for an explanation of JAX).

Variables

To steer users towards vectorised computation, “Individual” introduces a set of variable primitives which allow the user to characterise a wide range of agent traits, including *IntegerVariable*, which manages integer-based attributes of agents, suitable for searchable discrete data such as age groups or drug types; *DoubleVariable*, which handles continuous attributes, facilitating calculations involving metrics like immunity level; and *CategoricalVariable*, designed for data that fall into a limited number of categories, such as blood types, genetic traits, or infection states providing an efficient way to subset the population based on characteristics.

These variable types are implemented to represent characteristics across the entire agent population in a vectorised format. Individual also provides common optimised operations such as random sampling, subsetting, and updating of these variables to exploit fast, low-level vectorised operations to improve the performance of large-scale simulations.

Bitsets

One specific optimisation “Individual” provides is efficient subsetting of a population based on specific characteristics, an operation which is required for simulating targeted behaviours such as vaccinating certain age groups or infecting individuals based on their infection status and intervention usage. The “Individual” framework supports efficient population subsetting through bitsets.

In R, typical set operations like `setdiff`, `intersect`, `union`, and `%in%` are implemented using vectors, which do not scale well with set size. The computational complexity of these operations, if naively looping through elements in each vector, would scale polynomially with the population size, i.e. $O(n^2)$ for a population size n . This can become particularly cumbersome and time-consuming with the large population sizes which common in ABM simulations.

Bitsets offer a more efficient alternative. A bitset is a vector of “words”, where each word is a collection of bits, and each bit represents the presence or absence of an agent in a set. This data structure allows for extremely efficient set operations, which can be performed using bitwise logical operations such as AND, OR, and NOT. These operations are supported at the processor level, offering performance proportional to the population size, i.e. $O(n)$. Another advantage of bitsets is the ability to insert and delete elements in constant time. This stands in contrast to operations on vectors R or the binary search tree set provided in C++’s Standard Template Library, where adding or removing elements can involve expensive memory reallocation, which is undesirable at simulation time.

Processes

The “Individual” framework encourages users to implement the dynamics of infectious diseases as “processes”. This is a design choice to help users manage the complexity of interactions and dependencies within ABMs. By defining infection, recovery, vaccination, and other epidemiological events as processes, the framework provides a structured approach to simulate the progression and interaction of disease elements effectively.

In “Individual”, each process is implemented as a function, which can be written in either R or C++. Processes are designed to interact with the state of each agent with minimal side effects, making them testable and maintainable. Each function can access a snapshot of the model variables before any modifications are applied, allowing it to operate on a consistent view of the agent’s state. After evaluating the necessary changes, the process submits them as “transactions” to be applied after all the processes have executed.

This helps users reason through a common challenge in modelling infectious diseases: competing hazards — situations where two or more processes could influence the same characteristics of an agent in different ways. For instance, an agent might simultaneously be at risk of infection and recovery, each altering the agent’s state in divergent directions. By providing a transactional view of the agent’s states, a user is encouraged to consider how competing transactions will resolve.

Transactions submitted by different processes are resolved in a First In, First Out (FIFO) order. This means that transactions submitted later can overwrite the effects of those submitted earlier when simulating the same time step. This transaction management system allows researchers to see how transactions will resolve with a glance at the list of processes and ensure that computations carried out by one process are not adversely affected by the operations of another. However, this system places a responsibility on the users to carefully consider the ordering of processes. The sequence in which processes are executed and transactions are submitted can influence the final state of the simulation.

Events

“Individual” provides event primitives, which allow users to model behaviours which impact future states of agents within the simulation. This functionality is particularly useful for representing infrequent events or events which are targeted toward specific sub-populations, such as booster vaccinations. By using these primitives, researchers can conveniently avoid implementing discrete event schedulers from scratch.

Events in “Individual” are managed through a sparse mapping of time steps to bitsets. Each bitset represents the agents affected by an event at a specific time step. This structure allows the simulation to efficiently schedule events, trigger events, look up sub-populations which are scheduled for any event at any time step, and clear scheduled events for a sub-population.

Simulation loops

“Individual” provides a simulation loop which ensures that all of these components interact together as intended. For each time step, each process is executed in sequence, allowing it to view the current state of the population, submit transactions for updating agent states, and schedule future events; then the event scheduler triggers any scheduled events, calling the appropriate listeners which will submit further transactions for the affected agents; finally, after all events have been handled, all the submitted transactions are processed to update the state of the agents accordingly. The complete flow of the simulation loop, including the interaction between processes, event handling, and transaction management is illustrated in [Figure 4.2](#). By abstracting away the complex and common operations in epidemic ABM simulations, “Individual” allows researchers to devote their efforts to the more critical task of developing and refining the disease dynamics of their models.

4.2.2 Malariasimulation: an Agent-based malaria transmission model

I implemented Imperial College London’s malaria transmission model, which I have called “malar-iasimulation”, using the “Individual” framework. This implementation allowed me to explore the real world applications of the high-dimensional parameter inference methods discussed in this and the next chapter. As an added benefit, transitioning to an R-based codebase offers the research group numerous benefits, including an improved accessibility for extensions and updates to the model, particularly as new malaria interventions emerge.

Individual simulation loop

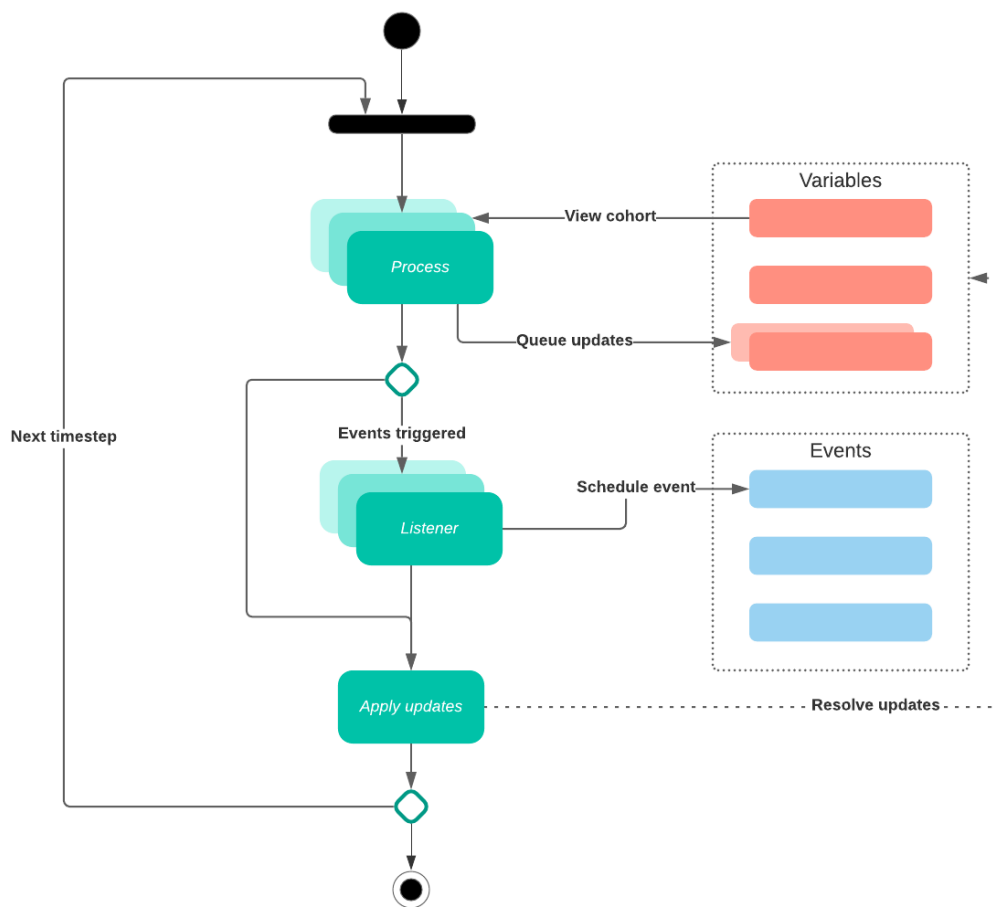


Figure 4.2: Individual Simulation Loop. A flow of operations when simulating an ABM written using Individual. I originally created this diagram for the following publication (G. D. Charles and S. L. Wu, 2021)

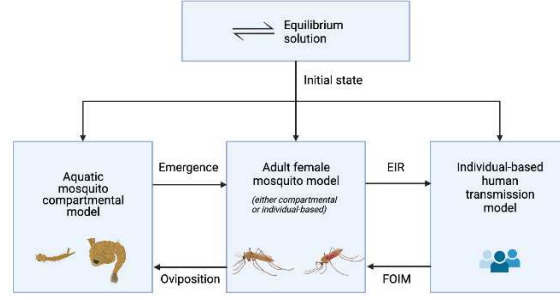


Figure 4.3: A high-level overview of “malariasimulation”. Arrows represent model outputs which are used as inputs for the different components. The equilibrium outputs are used before the simulation begins, however all other outputs are time-varying and are communicated each time step.

The “malariasimulation” model consists of four main components. The human model handles immunity response, disease progression, and intervention dynamics, providing statistics which are important for evaluating malaria burden and calculating the Force Of Infection towards Mosquitoes (FOIM) necessary for the mosquito model. The adult female mosquito model is responsible for tracking the mosquito population density and infectiousness of mosquitoes, which is used to simulate the Entomological Inoculation Rate (EIR). Meanwhile, the aquatic model focuses on the early life-cycle of mosquitoes, particularly from larvae to adult, taking into account the environmental carrying capacity and the simulated reproduction from the adult mosquito population. Lastly, the equilibrium solution establishes initial conditions for all other components based on specified baseline transmission levels, ensuring the simulation starts from a somewhat realistic state. This flow of data is illustrated in Figure 4.3.

The human model

The human population are modelled as agents. There are five human infection states in the model, depicted in Figure 4.4. Individuals default to the Susceptible (S) state. After exposure to an infectious bite, individuals can move to the Clinical Disease (D), Asymptomatic (A) or Treated (T) states. Untreated individuals in the D state will recover to the A state, from the A to the Subpatent (U) state, and from the U to the S state. Infected humans cause infections when exposed to bites from susceptible mosquitoes, which is quantified by the Force Of Infection towards Mosquitoes (FOIM or Λ_M). Individuals in the T state eventually recover to the S state and are protected by a period of prophylaxis to infection. The population in the S, A and U states are susceptible to re-infection on exposure to infectious mosquito bites. The Entomological Inoculation Rate (EIR) governs the number of infectious mosquito bites the population experiences over time.

Individuals experience EIR differently based on their heterogeneity and age variables. Heterogeneity ζ_i is sampled from the log normal distribution, while age a_i is sampled from the exponential distribution. The effective EIR experienced by each individual $\Lambda_i(a, t)$ at time t is modelled relative to the EIR experienced by the complete human population $\Lambda(t)$.

$$\begin{aligned}\Lambda_i(a, t) &= \Lambda(t)\zeta_i(1 - \rho \exp(-\frac{a}{a_0})) \\ \log(\zeta_i) &\sim N(\frac{-\sigma^2}{2}, \sigma^2) \\ a &\sim \exp(\mu_a)\end{aligned}\tag{4.1}$$

There are three functions representing immunity in the model. $b(t)$ and $\phi(t)$, which represent the probability of a blood-stage infection and progression to clinical disease, respectively, influence the infection transition rates, whereas $q_i(t, a)$ represents the probability of detection by a routine diagnostic test. These are calculated from acquired immunity variables I_B, I_{CA}, I_D , which are incremented on exposure to an infectious bite or on progression to clinical disease and later decay exponentially with rates r_b, r_c and r_d . ϕ also takes into account a maternal immunity I_{CM} which is set to a proportion of the individual’s mother’s clinical immunity at the time of birth. Maternal immunity also decays exponentially with rate r_{cm} .

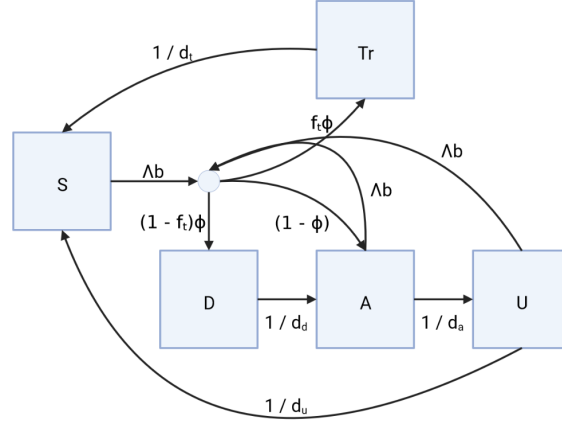


Figure 4.4: Human Infection States. The boxes represent the infection state of a single agent, while arrows represent the transition rates between each state. Here, Λ represents EIR, b pre-erythrocytic immunity, Φ clinical immunity, f_t treatment seeking probability, and d_a, d_u, d_t are static delay parameters which are listed in Table 4.3

$$\begin{aligned}
 b_i(t) &= b_0 \left(b_1 + \frac{1 - b_1}{1 + \frac{I_B}{I_{B0}} k_b} \right) \\
 \phi_i(t) &= \phi_0 \left(\phi_1 + \frac{1 - \phi_1}{1 + \frac{I_{CA}(i,t) + I_{CM}(i,t) k_c}{I_{C0}}} \right) \\
 q_i(t, a) &= d_1 \left(d_1 + \frac{1 - d_{min}}{f_D(i, a) \frac{1 + I_D(i,t) k_D}{I_{D0}}} \right)
 \end{aligned} \tag{4.2}$$

Immunity also plays a part in simulating detection by microscopy. An individual will be detected by microscopy if they are in the D state, or with probability q if they are in the A state. The diagnostic prevalence of malaria in the model $\text{PfPr}(t, l, u)$ is the proportion of individuals between ages l and u who would test positive in a routine diagnostic test at time t , as defined in equation 4.3.

$$\begin{aligned}
 \text{PfPr}(t, l, u) &= \frac{\sum_i \text{detected}(i, t) \times \mathbb{1}_{l < a_i \leq u}}{\sum_i \mathbb{1}_{l < a_i \leq u}} \\
 \text{detected}(i, t) &= \begin{cases} 1 & \text{state}(i, t) = D \\ q_i(t, a) & \text{state}(i, t) = A \\ 0 & \text{otherwise} \end{cases}
 \end{aligned} \tag{4.3}$$

The adult mosquito model

The mosquito population are modelled as a system of delay differential equations. Susceptible female mosquitoes (S_m) are infected at a rate Λ_M , calculated from the human part of the model. Infected mosquitoes will undergo an Extrinsic Incubation Period (EIP or τ) in the E_m state before becoming infectious. If they survive, they will transition to the Infected (I_m) state where they remain until death.

$$\begin{aligned}
 \frac{dS_m^v}{dt} &= -\Lambda_M^v S_m^v + \beta_{em}^v(t) - \mu^v S_m^v \\
 \frac{dE_m^v}{dt} &= \Lambda_M^v S_m^v + \beta_{em}^v(t) - \Lambda_M^v(t - \tau) S_m^v(t - \tau) \exp(-\mu^v \tau) - \mu^v E_m^v \\
 \frac{dI_m^v}{dt} &= \Lambda_M^v(t - \tau) S_m^v(t - \tau) \exp(-\mu^v \tau) - \mu^v I_m^v
 \end{aligned} \tag{4.4}$$

Aquatic mosquito model

The aquatic mosquito model has three states which represent the different stages in larval development. The Early (E) and Late (L) larval states track the number of larvae soon after hatching. These two stages are dependent on the density of larvae and the carrying capacity K which can vary with rainfall $R(t)$. Conversely, the subsequent Pupal (P) stage is influenced by these factors and is the final stage before development into an adult mosquito. These aquatic stages are simulated using an ordinary differential equation model with a system of equations listed in Equation 4.5.

$$\begin{aligned}\frac{dE^v}{dt} &= \beta_{egg}^v(t)M - \frac{E^v}{d_e} - \mu_e E^v \left(1 + \frac{E^v + L^v}{K}\right) \\ \frac{dL^v}{dt} &= \frac{E^v}{d_e} - \frac{L^v}{d_l} - \mu_l L^v \left(1 + \gamma \frac{E^v + L^v}{K}\right) \\ \frac{dP^v}{dt} &= \frac{L^v}{d_l} - \frac{P^v}{d_p} - \mu_p P^v\end{aligned}\tag{4.5}$$

where,

$$K = K_0 \frac{R(t)}{\bar{R}} R(t) = g_0 + \sum_{i=1}^3 g_i \cos(2\pi t i) + h_i \sin(2\pi t i)\tag{4.6}$$

The number of new adult female mosquitoes, $\beta_{em}^v(t)$, which influences growth in the S_m^v compartment is assumed to be $\frac{d_p P}{2}$, since half of the pupae in the simulation will develop into female mosquitoes at rate d_p .

The number of new eggs laid, $\beta_{egg}^v(\mu, f, t)$, which influences the number of new larvae in the aquatic mosquito model, is calculated based on the total number of adult female mosquitoes in the simulation, their mortality rate μ and feeding rate f and egg laying rate β :

$$\beta_{egg}^v(\mu, f, t) = \beta(e^{\frac{\mu}{f}} - 1)e^{-\frac{\mu}{f}}(1 - e^{-\frac{\mu}{f}})^{-1}$$

Insecticide Treated Nets (ITNs)

“malariasimulation” supports several interventions, but in this chapter I will focus on Insecticide Treated Nets (ITNs) for simplicity as they are the most effective intervention (Bhatt et al., 2015). In the subsequent chapter, I will include more interventions including seasonal malaria chemoprevention, front-line treatment, indoor residual spraying to make a more realistic inference.

ITNs are an effective malaria intervention which protect individuals while they are in bed. The nets are treated with an insecticide which repels or kills a mosquito when they attempt to feed on a human while in bed. Mosquitoes which are repelled will take longer to feed, indirectly affecting their mortality and reproduction rates. The effects of ITNs on mortality and reproduction, for various mosquito species and insecticide types, have been examined in experimental hut trials (Sherrard-Smith et al., 2022). For humans, ITNs reduce the EIR an individual experiences, since mosquitoes which have been repelled or killed before feeding no longer contribute towards infectious bites. We assume that bed nets are retained for an average of five years before being thrown away. The model can be parameterised with a sequence, $\nu(t)$, describing the proportion of the population who receive an ITN each year.

Repelling mosquitoes results in a lengthening of the gonotrophic cycle and an increase in the proportion of bites taken on unprotected people and other animals. I express a dynamic mortality rate μ_v as a function of the feeding rate f_R^v and the probabilities of surviving feeding p_1^v and the gonotrophic cycle p_2^v .

$$\mu^v = -f_R^v \log(p_1^v p_2^v)$$

k Without vector control, I model the probability of surviving a feeding attempt as p_{10}^v , a function of the natural time spent in search of a blood meal, δ_{10}^v and the natural mortality rate for a vector, μ_0^v :

$$p_{10}^v = \exp(-\mu_0^v \delta_{10}^v)$$

As vector control interventions are introduced, p_1^v will become less than p_{10}^v . This is because vectors will have a smaller probability of a successful feeding attempt, W^v , and have a larger probability of being repelled Z^v . I express p_1^v as:

$$p_1^v = \frac{p_{10}^v W^v}{1 - Z^v p_{10}^v}$$

p_2^v will remain independent of vector controls as we can express it as a function of μ_0^v and the time for each species' gonotrophic cycle, δ_2^v :

$$p_2^v = \exp(-\mu_0^v \delta_2^v)$$

In the model, vector controls are assumed to reduce the effective biting rate on each human λ_i^v based on their level of protection. I express λ_i^v in terms of the human blood meal rate α^v , the relative proportion of bites taken on each human π_i , the probability of a successful feeding attempt on the i^{th} individual w_i and the probability of feeding *and* surviving on the i^{th} individual y_i .

$$\lambda_i^v = \frac{\alpha^v \pi_i y_i^v}{\sum \pi_i w_i^v}$$

The human blood meal rate is a product of the overall blood meal rate f_R^v and the proportion of blood meals taken on humans Q^v :

$$\alpha^v = f_R^v Q^v$$

As a result of vector control, f_R^v is reduced. This is because mosquitoes are repelled from and require a longer time to successfully find a blood meal, δ_1 . I capture this extension in terms of the natural time to find a blood meal δ_{10} and Z^v .

$$f_R^v = \frac{1}{\delta_1^v + \delta_2^v}$$

Where,

$$\delta_1^v = \frac{\delta_{10}^v}{1 + Z^v}$$

Q^v is also reduced, since fewer bites are successfully made on humans. We can capture that reduction in terms of the natural proportion of bites made on humans Q_0^v :

$$Q^v = 1 - \frac{1 - Q_0^v}{W^v}$$

Probabilities for w_i , y_i and z_i , take into account the level of protection that nets or spraying provided for each individual. When an individual has a net, the probability of a mosquito successfully feeding and being repelled on each feeding attempt of a host i is s_N and r_N respectively. We define these as s_S and r_S for individuals who are protected by indoor residual spraying. The proportion of feeding attempts indoors and on hosts in bed are Φ_I and Φ_B respectively.

$$\begin{aligned} w_i &= 1 - \Phi_I + \Phi_B(1 - r_S)s_N s_S + (\Phi_I - \Phi_B)(1 - r_S)s_S \\ y_i &= 1 - \Phi_I + \Phi_B(1 - r_S)s_N + (\Phi_I - \Phi_B)(1 - r_S) \\ z_i &= \Phi_B(1 - r_S)r_N + \Phi_I r_S \end{aligned} \tag{4.7}$$

For ITNs, r_N and s_N have been estimated from bioassay data (Thomas S Churcher et al., 2016; Nash et al., 2021; Sherrard-Smith et al., 2022). This enables different types of mosquito nets that may incorporate different combinations of insecticide or synergist chemistry and that elicit different protective effects to be simulated. The effectiveness of the net in repelling mosquitoes is modelled as an exponential decay with rate γ_N and scaled between minimum and maximum values r_{NM} and r_{N0} . s_N is modelled using r_N and an intermediate value d_N , which represents the effectiveness of the net in killing mosquitoes. d_N also has an exponential decay with rate γ_N . If an individual does not have a net, I set $\Phi_B = r_N = 0$, $s_N = 1$ to remove the effects of nets from the equations for w_i, y_i, z_i .

$$\begin{aligned}
s_N &= 1 - r_N - d_N \\
r_N &= (r_{N0} - r_{NM}) \exp(-t\gamma_N) + r_{NM} \\
d_N &= d_{N0} \exp(-t\gamma_N)
\end{aligned} \tag{4.8}$$

Susceptible female mosquitoes (Sm) are infected based on FOIM, or Λ_M , expressed in [Equation 4.9](#), which is calculated using the species specific biting rate α^v , the infectiousness of the human population $\sum_i c(i, t)$ and the effective biting rate for each individual π_i .

$$\Lambda_M^v(t) = \alpha^v \sum_i c(i, t) \pi_i \tag{4.9}$$

where,

$$c(i, t) = \begin{cases} 0 & \text{state}(i, t) \in \{S, T\} \\ c_D & \text{state}(i, t) = D \\ c_U & \text{state}(i, t) = U \\ c_A(i, t) & \text{state}(i, t) = A \end{cases}$$

where,

$$c_A(i, t) = c_U + (c_D - c_U)q_i\gamma^1$$

Extended Equilibrium

Setting an equilibrium state for “malariasimulation” is helpful for for starting the simulation from a realistic state. The initial state for the human model is set using the equilibrium model introduced in [chapter 2](#). However, I will now describe how the mosquito models are initialised.

To establish the equilibrium for the adult female mosquito population, I set the derivative of each compartment to zero and factor in the desired baseline EIR, Λ^* , and corresponding FOIM, Λ_M^v , calculated from the human population. This ensures that the mosquito population is balanced with the transmission dynamics in the human model at the start of the simulation. The total number of female adult mosquitoes, M_v , is derived from the following formula:

$$M_v = \frac{\Lambda^*}{\sum_v p_v \alpha_v Q 0_v \frac{\Lambda_M^v \exp(\mu\tau)}{\Lambda_M^v + \mu}}$$

Further, the total numbers of mosquitoes in various aquatic stages— early larval (E), late larval (L), and pupae (P) — are calculated based on the equilibrium number of adult mosquitoes. The relationship is defined by:

$$\begin{aligned}
E &= 2\omega\mu d_L(1 + d_P\mu_P)M \\
L &= 2\mu d_L(1 + d_P\mu_P)M \\
P &= 2d_P\mu M
\end{aligned}$$

Here, ω is a derived factor that accounts for the rate at which larvae develop, taking into consideration the death rates μ_E, μ_L and development durations d_E, d_L for each stage:

$$\begin{aligned}
\omega &= -\frac{1}{2} \left(\gamma \frac{\mu_L}{\mu_E} - \frac{d_E}{d_L} + (\gamma - 1)\mu_L d_E \right) + \\
&\quad \sqrt{\frac{1}{4} \left(\gamma \frac{\mu_L}{\mu_E} - \frac{d_E}{d_L} + (\gamma - 1)\mu_L d_E \right)^2 + \gamma \frac{\beta\mu_L d_E}{2\mu_E\mu d_L(1 + d_P\mu_P)}}
\end{aligned}$$

Implementation within the “Individual” Framework

I encoded the dynamics of malaria transmission, as described in the previous sections, within the “Individual” framework, which allowed for a structured and modular approach to simulation. This implementation involved defining a series of processes, variables, and events which interacted to simulate the interactions of malaria among human and mosquito populations.

Each agent within the simulation is characterised by a set of variables that store the required information about their state and history. These variables include all of the time-varying traits I have mentioned above, from infection status to immune response. The specific variables used in the simulation are detailed in [Table 4.1](#). The model’s dynamics are driven by a series of processes and events which dictate the interactions and transitions within the agent population. These processes are responsible for implementing the core mechanics of malaria transmission, such as infection, recovery, and mosquito biting behaviour. The events handle scheduled actions like the application of ITNs. A detailed listing of these processes and events is provided in [Table 4.2](#).

The model relies on a set of parameters that instantiate the free variables in the above formulae. These parameters have been derived from a combination of empirical data, expert knowledge, and independent parameter inference studies. The complete list of these parameters is listed in [Table 4.3](#).

In the context of site calibration, the “init_EIR” is the free variable which is adjusted to match local prevalence data. Unlike other parameters, which are set based on established research or expert assessment, “init_EIR” will be inferred using Bayesian Inference by calibrating the ABM outputs to an observed time series of prevalence data in a location of interest.

Table 4.3: Default parameter values for “malariasimulation”, set as the mean estimates from Griffin, Ferguson, and Ghani ([2014](#))

Parameter	Description	Value
Fixed state transitions:		
dd	The delay for humans to move from state D to A	5
dt	The delay for humans to move from state T to S	5
da	The delay for humans to move from state A to U	200
du	The delay for humans to move from state U to S	110
del	The delay for mosquitoes to move from state E to L	6.64
dl	The delay for mosquitoes to move from state L to P	3.72
dpl	The delay mosquitoes to move from state P to Sm	0.643
mup	The rate at which pupal mosquitoes die	0.249
mum	The rate at which developed mosquitoes die	0.1253333
Immunity decay rates:		
rm	Decay rate for maternal immunity to clinical disease	67.6952
rvm	Decay rate for maternal immunity to severe disease	76.8365
rb	Decay rate for acquired pre-erythrocytic immunity	3650
rc	Decay rate for acquired immunity to clinical disease	10950
rva	Decay rate for acquired immunity to severe disease	10950
rid	Decay rate for acquired immunity to detectability	3650
Pre-erythrocytic infection:		
b0	Maximum probability due to no immunity	0.59
b1	Maximum reduction due to immunity	0.5
ib0	Scale parameter	43.9
kb	Shape parameter	2.16
Probability of clinical infection:		
phi0	Maximum probability due to no immunity	0.792
phi1	Maximum reduction due to immunity	0.00074
ic0	Scale parameter	18.02366
kc	Shape parameter	2.36949
Probability of detection:		
fd0	Time-scale at which immunity changes with age	0.007055
ad	Scale parameter relating age to immunity	7993.5
gammad	Shape parameter relating age to immunity	4.8183

d1	Minimum probability due to immunity	0.160527
id0	Scale parameter	1.577533
kd	Shape parameter	0.476614
Immunity boost grace periods:		
ub	Period in which blood immunity cannot be boosted	7.2
uc	Period in which clinical immunity cannot be boosted	6.06
uv	Period in which severe immunity cannot be boosted	11.4321
ud	Period in which detection immunity cannot be boosted	9.44512
Infectivity towards mosquitoes:		
cd	Infectivity of clinical disease towards mosquitoes	0.068
γ^1	Parameter for infectivity of asymptomatic humans	1.82425
cu	Infectivity of sub-patent infection	0.0062
ct	Infectivity of treated infection	0.021896
Unique biting rate:		
a0	Age dependent biting parameter	2920
rho	Age dependent biting parameter	0.85
σ^2	Heterogeneity parameter	1.67
Mortality parameters:		
pcm	New-born clinical immunity relative to mother's	0.774368
me	Early stage larval mortality rate	0.0338
ml	Late stage larval mortality rate	0.0348
Carrying capacity parameters:		
γ	effect of density dep. on late relative to early instars	13.25
Incubation periods:		
de	Duration of the human latent period of infection	12
delay_gam	Lag from parasites to infectious gametocytes	12.5
dem	incubation period in mosquito population model	10
Vector biology:		
beta	The average number of eggs laid per mosquito per day	21.2
foraging_time	Time spent taking blood meals	0.69

4.3 Parameter inference method

4.3.1 Surrogate model

Here I extend the surrogate modelling formulation presented in the previous chapter to incorporate an inductive bias geared towards sequential data. Specifically, I introduce an autoregressive function designed to better handle sequential inputs and outputs.

As before, I aim to develop a surrogate model, $\tilde{f}_\phi$, with hyper-parameters ϕ . The goal is to learn the model outputs $\mathbf{y}$ from the model parameters $\mathbf{x} \in \mathcal{X}$ for a space of admissible parameters $\mathcal{X}$ using a loss function $\mathcal{L}$, refer to [subsection 3.2.1](#) for an introduction. The optimisation objective remains:

$$\hat{\phi} = \arg \min_{\phi} \int_{\mathcal{X}} p(\mathbf{x}) \mathcal{L}(\mathbf{x}, \mathbf{y}, \tilde{f}_\phi) d\mathbf{x}$$

However in this context, I assume the model parameters lie within a space of sequences $\mathcal{X} \subset \mathbb{R}^{k \times T}$, where T is the number of time steps, and each element $x_i \in \mathbb{R}^k$ is bounded to plausible inputs. I instead learn an autoregressive function $\tilde{f}_\phi$ which maps a given sequence input and its preceding elements to an output at each time step, i.e. $\tilde{f}_\phi : \mathbf{x}_i \times \mathbf{x}_{<i} \rightarrow \mathbf{y}_i$. Where the surrogate should approximate the true model output at time step i , $f(\mathbf{x})_i$, with a high degree of accuracy. The goal is to ensure $\forall i \quad |\tilde{f}_\phi(\mathbf{x}_i, \mathbf{x}_{<i}) - f(\mathbf{x})_i| < \epsilon$ for some suitably small $\epsilon > 0$.

By adopting an autoregressive approach, the surrogate model should be able to capture the time-varying nature of the outputs. This would come with the benefits of recreating time-varying outputs and predicting outputs beyond the time horizon of the training data, provided the learned relationships hold true over extended periods.

Table 4.1: Model variables for malaria simulation

Variable	Type	Description
state	category	the infection state of a human individual S D A U Tr
birth	integer	an integer representing the time step when this individual was born
last_boosted_*	integer	the last time step at which this individual's immunity was boosted. Used for tracking grace periods in the boost of immunity
ICM	double	Maternal immunity to clinical disease
IVM	double	Maternal immunity to severe disease
IB	double	Pre-erythoctic immunity
ICA	double	Acquired immunity to clinical disease
IVA	double	Acquired immunity to severe disease
ID	double	Acquired immunity to detectability
zeta	double	Heterogeneity of human individuals
zeta_group	category	Discretised heterogeneity of human individuals
rtss_vaccinated	integer	The time step of the last RTSS vaccination (-1 if there have not been any)
rtss_boosted	integer	The time step of the last RTS,S booster (-1 if there have not been any)
rtss_cs	double	peak antibodies
rtss_rho	double	antibody component variable
rtss_ds	double	short-lived antibody delay variable
rtss_dl	double	long-lived antibody delay variable
tbv_vaccinated	integer	The time step of the last TBV vaccination (-1 if there have not been any)
net_time	integer	The time step when a net was last put up (-1 if never)
spray_time	integer	The time step when the house was last sprayed (-1 if never)
infectivity	double	The onward infectiousness to mosquitoes
drug	integer	The last prescribed drug
drug_time	integer	The time step of the last drug
mosquito_state	category	the state of the mosquito, a category Sm Pm Im NonExistent
species	category	the species of mosquito, this is a category gamb fun arab

4.3.2 Sampling strategy

Input domain

I can translate the malaria IBM description in [subsection 4.2.2](#) by first defining $\mathcal{X}$.

To parameterise the surrogate for a particular location, I set κ^v , μ_a , g_i , h_i and $\nu(t)$ parameters to capture mosquito composition, demography, seasonality and ITN usage respectively. I defined a range for μ_a as 14.8 to 55.4 years based on the maximum and minimum average age statistics per country listed in World Population Review (WorldPop, 2022). Ranges for the Fourier coefficients g_i and h_i are limited to $[-10, 10]$. Since κ^v are relative proportions of three mosquito species, they are only valid in the 3-simplexes in $\mathbb{R}^3$.

$\nu(t)$ is defined as a bounded step function with $n \geq 0$ annually spaced knots. A realistic range of $\nu(t)$ is $(0, 0.8]$ for each year, as it becomes operationally difficult to distribute to ITNs to all individuals. I chose not to constrain the relationship between $\nu(t)$ and $\nu(t + \delta)$ as it could limit the effectiveness of the surrogate in exploratory settings. However, parameterisations of $\nu(t)$ based on real world statistics are likely to have a gradual increase from 0 and plateau.

One of the main unknowns for a each location is, Λ_0 . An accurate estimate of this parameter would allow us to simulate a population of humans and mosquitoes which would recreate malaria transmission dynamics before any interventions were implemented. Previous estimates for countries

in sub-Saharan Africa were below 418 (Griffin, Ferguson, and Ghani, 2014), and model outputs suggest that higher values of Λ_0 have diminishing marginal effects on PfPr (Penny et al., 2016). Therefore I set a range for Λ_0 as $[0, 500]$.

To express $\mathcal{X}$ as a sequence, I formulated a sequence of n vectors for each year, t , with time-invariant parameters $(\kappa^v, \mu_a, g_i, h_i, \Lambda_0)$ repeated at each time point, and the $\nu(t)$ selected for the corresponding time point. Each vector belongs to a subset of $\mathbb{R}^{13}$ bounded as described above.

Output range

Let $(1, \dots, T)$ be the sequence of years at which I wish to predict the prevalence of malaria. For any $\mathbf{x} \in \mathcal{X}$, let f represent running the ABM with parameters $\mathbf{x}$ for T years and evaluating PfPr(d, l, u) for each day d in each year t :

$$f(\mathbf{x}) = \left\{ \{\text{PfPr}(d, u, l)\}_{d=1}^{365} \right\}_{t=1}^T \quad (4.10)$$

With this notation, $\mathcal{Y}$ is easy to define as the image of f :

$$\mathcal{Y} = \{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, \quad (4.11)$$

where $\mathcal{X}$ is defined as in the previous section. From inspecting equation 4.3, I get $\mathcal{Y}$ as a space of T -length sequences of vectors $\mathbb{R}^{365}$ bounded between 0 and 1.

4.3.3 Surrogate model architecture

With these definitions of $\mathcal{X}$ and $\mathcal{Y}$, the surrogate modelling task can be reduced to a sequence-to-sequence prediction problem. Sequence-to-sequence prediction is supported by a number of models, however I chose a recurrent neural network (RNN). The RNN should be able to exploit the relationships between time points to generalise for unseen parameter sets and longer time horizons.

I performed a grid search to optimise hyper-parameters for the surrogate, including the loss function, $\mathcal{L}$. Each hyper-parameter set was evaluated using the Mean Squared Error on a validation set, the explored options are listed in Table 4.4. The MSE and log cosh loss functions took the following forms:

$$\mathcal{L}_{\text{MSE}}(\mathbf{x}, \mathbf{y}, \phi) = \left\| \mathbf{y} - \tilde{f}_\phi(\mathbf{x}) \right\|^2 \quad (4.12)$$

$$\mathcal{L}_{\log \cosh}(\mathbf{x}, \mathbf{y}, \phi) = \log \left(\cosh(\mathbf{y} - \tilde{f}_\phi(\mathbf{x})) \right). \quad (4.13)$$

The RNN took each vector from the $\mathcal{X}$ sequence as input to its recurrent cells and output the corresponding vector from $\mathcal{Y}$. The parameters for g_i and h_i were expanded to full rainfall data $R(t)$ using Equation 4.6. The resulting inputs were standardised, removing the mean and scaling to unit variance. The input layer had 383 recurrent cells, feeding 365 latent recurrent cells followed by a dense output layer of 365 cells.

4.3.4 Sampling training data

Training a surrogate for the malaria ABM requires a large library of simulations covering a wide area of input space. I created training and validation data, with an 80/20 split for the surrogate model using a sequence length of $n = 16$. This would allow for the surrogate to learn the long term effects of ITNs on prevalence. I sampled 10,000 values from a 28 dimensional Latin hypercube, 16 for $\nu(t)$ and 12 for the remaining time-invariant parameters, and transformed the margins of each dimension to fit the bounds listed in section 4.3.2. The ABM was executed with the sampled inputs, setting aside a warm-up period before activating the ITN parameters to allow the simulation to reach an equilibrium. I set a large population size for the ABM runs to minimise the stochasticity in PfPr trajectories and treat them as deterministic outputs.

The simulation output was aggregated to calculate the prevalence of malaria among children aged 5 to 59 months for each year after a warm-up period. The age range of 5 to 59 months old was selected to align with statistics commonly found in Demographic and Health Surveys (Demographic and Health Surveys, 2022).

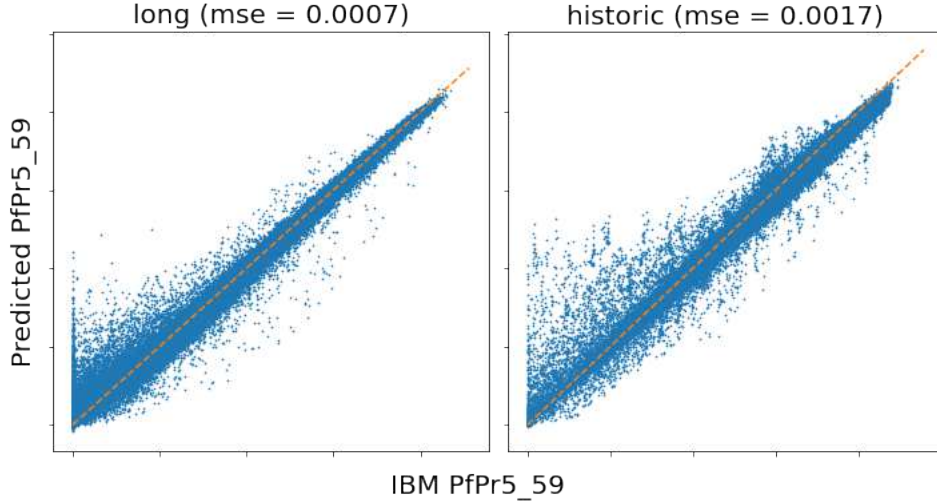


Figure 4.5: True versus predicted prevalence of malaria in children aged 5-59 months (PfPr5_59) for the *long* (left) and *historic* (right) datasets where the true estimates come from the ABM model and the predicted estimates come from the surrogate. Each point represents the prevalence for one time step (a year) of one sample in the corresponding benchmark. Mean Squared Error (mse) was very low, as shown in the titles of the plots. Both axes are range from 0 to 1.

4.3.5 Site calibration formulation

I define the site calibration problem as an inference problem where one attempts to estimate Λ^* given site parameters θ_s .

I used Tensorflow Probability (Abadi et al., 2016) to fit the ABM to observational data from the Demographic and Health Survey (DHS) program (Demographic and Health Surveys, 2022). The DHS responses included geo-location, age, and outcome of Rapid Diagnostic or Microscopy tests for respondents in Kolda, Senegal between 2008 and 2018. I included respondents between 5 and 59 months old and weighted their responses by population density data from World Bank data (World Bank, 2022) to get a representative parasite prevalence over time in the region, $D_{\text{PfPr}_{5-59}}$.

To estimate a value of Λ_0 which best explains $D_{\text{PfPr}_{5-59}}$, I used a binomial likelihood function, see equation 4.14. I compared each participant's infection status J^i , an indicator variable, to the average predicted prevalence for the month in which they were observed, $\text{PfPr}(t^i)$. The surrogate, parameterised with $\nu(t), g_i, h_i, \kappa^v$ values for Kolda, and aggregated at month t^i provided quick estimates for $\text{PfPr}(t^i)$. I set a uniform prior for Λ_0 in $[0, 500)$.

$$L(\Lambda^* | D_{\text{PfPr}_{5-59}}) = \prod \text{PfPr}(t^i)^{J^i} (1 - \text{PfPr}(t^i))^{1-J^i} \quad (4.14)$$

I performed Markov Chain Monte Carlo using the HMC algorithm (Neal, 2011) with 10 parallel chains, 2,000 steps, 1,000 of which were discarded as warm-up, to sample from the resulting posterior. This inference took just over 11 minutes to execute on an 8-core Intel CPU.

4.4 Results

4.4.1 Surrogate evaluation

I evaluated the surrogate on two benchmark parameter sets. *long* contains 5,000 samples generated similarly to the training set except for sequences with $n = 30$, testing the surrogate's ability to generalise to longer time horizons. *historic* contains 5,000 samples where $\nu(t)$ was sampled from historic estimates for randomly selected countries in sub-Saharan Africa, published by (Bhatt et al., 2015), testing the performance on parameters representative of global malaria modelling studies. After 100 epochs of the Adam optimiser, the MSE across the 2,000 out-of-sample simulations was 6.8×10^{-3} on un-standardised prevalence values. The prediction error for these benchmarks is shown in Figure 4.5.

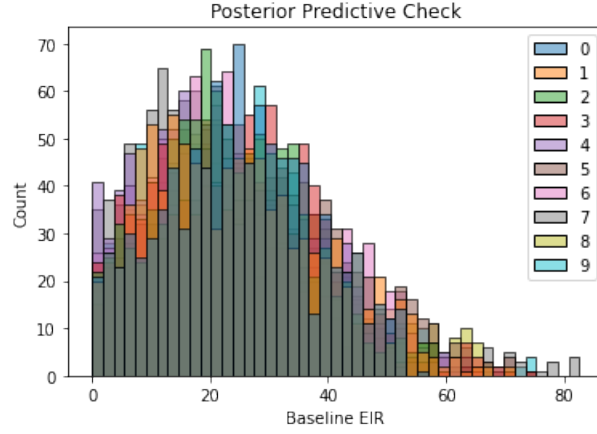


Figure 4.6: A posterior predictive check plotting Λ^* samples from 10 MCMC chains. A previous rough estimate of baseline EIR in three Senegalese populations was 1, 20 and 200 (infectious bites per person per year) for the cities of Dakar, Ndiop and Dielmo (Trape and C. Rogier, 1996). This estimate of 23.2 ± 13.7 would place Kolda's baseline transmission alongside Ndiop.

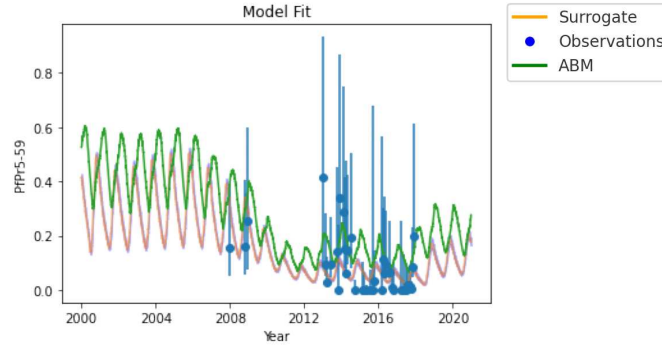


Figure 4.7: A plot of the final fit including the surrogate trajectory for the mean EIR (in orange). The ABM was then calibrated to the mean EIR using a root finding method and run to simulate the ITN distributions for Kolda, the prevalence for that run is plotted in green. Observations have been plotted in blue with their 95% confidence interval, assuming a Binomial distribution. The observational data included samples from low intervention periods (2008) as well as high intervention periods (2013-2017) and was aggregated on a monthly time scale.

I compared execution times for the original ABM, surrogate training, single prediction and batch prediction. The ABM executions were run on a cluster in batches of 10 on 1,000 dedicated cores with a 2GB memory reservation. The surrogate was trained on an Nvidia A1000 GPU, which took 229 ± 34 seconds ($n = 3$). Both single and batch predictions were made on an 8-core Intel CPU. While the ABM required just under 24 hours to run 10,000 samples, the batch surrogate required just over 3 seconds (approximately a 25,000-fold improvement). The timing statistics for an each run are listed in table 4.5.

4.4.2 Parameter inference

The original ABM was then run with the mean estimate for comparison, see the posterior predictive check in Figure 4.6 and resulting fit in Figure 4.7. For this fit, there is a noticeable error between the surrogate and the original model. The seasonal cycle in prevalence is delayed and flattened, while the average prevalence is underestimated. However, the relative effects of ITNs appear similar between the models.

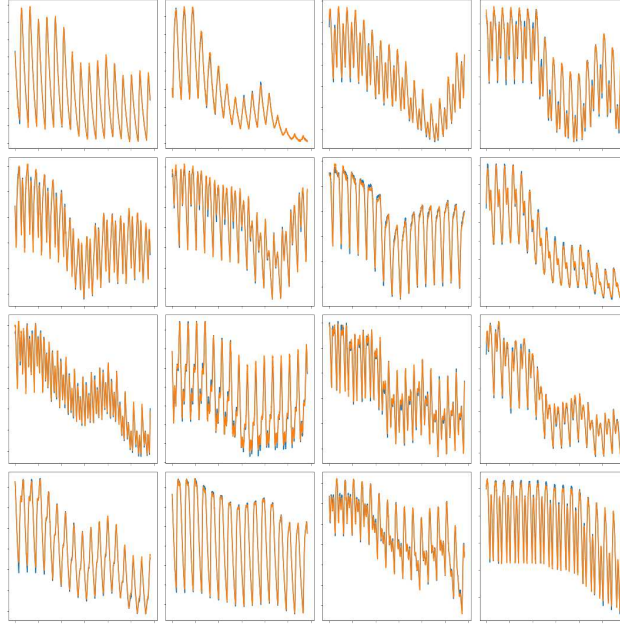


Figure 4.8: 16 cases with the lowest MSE from the historic benchmark

4.5 Discussion

I have demonstrated that it is feasible to perform inference on a highly detailed epidemic model with complex temporal dynamics. Seq2Seq based surrogates can accurately predict trajectories for models with large parameter spaces and generalise transmission dynamics to very long time horizons. This opens up a wider range of data, including longitudinal, from less controlled studies for use in parameter inference in an epidemiological setting.

However, the inference in the previous section should not be used to inform real policy. Simplifying assumptions have been made which ignore the effects of non-ITN interventions and different diagnostic methods. Also, the accuracy of the surrogate varies for different types of model runs. Further work needs to be done to add detail to the surrogate model and ensure a high accuracy for the parameter space which is used for inference.

4.5.1 A surrogate for general analysis

The surrogate presented in this chapter could be used as an exploratory tool, allowing researchers to begin modelling malaria transmission for arbitrary locations in seconds. Selective approaches, such as McKinley et al. (2018) and Reiker, Golumbeanu, et al. (2021), had designed their training set for a particular inference task, thus more training would be required to guarantee their published performance for each use case. In the case of Reiker, Golumbeanu, et al. (2021), training can take between 7 and 12 days. Whereas in this case, inference required 11m31s to execute on a single GPU. This could open up many possibilities for detailed and thorough modelling studies on desktop machines.

In practice, software could be provided for a malaria researcher to minimise disruption to their existing workflow. The software would load the pre-trained weights for the surrogate using their preferred machine learning library, an interface would be provided to validate and vectorise model inputs for use with the surrogate, the surrogate would perform forward inference in milliseconds, and the interface will reformat the outputs into trajectories as expected by the researcher. The researcher could then use the surrogate as a drop in replacement for computationally intensive analyses such as global model runs or site calibration.

The benchmarks showed that the surrogate could emulate the ABM for the vast majority of simulation scenarios. 99% and 97% of predictions had an MSE of less than 1×10^{-2} for the *long* and *historic* benchmarks respectively. On inspection, we could see that short-term seasonal and long-term ITN transmission dynamics could be accurately re-created (see Figure 4.8).

Despite the benchmarks showing a small risk in surrogate prediction error, many predictions

significantly deviated from the ABM outputs. Manual inspection reveals that the worst predictions for the historic benchmark were in scenarios with low or zero prevalence (see [Figure 4.9](#)). The performance of the surrogate would likely be improved by extending the training set with low transmission parameter sets, i.e. samples with smaller values of Λ .

A more general approach to managing the risk in surrogate prediction error could be uncertainty estimation. Unlike the approach in ([ibid.](#)), RNNs provide point estimates of the ABM outputs. However, RNNs can be adapted to produce probabilistic estimates, (MacKay, 1992; Lakshminarayanan, Pritzel, and Blundell, 2017). Such outputs could be used at prediction time to inform the researcher of the trust-worthiness of a prediction. They could also be used at training time to find samples which would reduce surrogate predictive uncertainty, (Bai et al., 2021).

There are several possible methods to incorporate the stochasticity inherent in the ABM outputs to the surrogate. A rudimentary method would be add parametric noise, as is the case in Lamperti, Roventini, and Sani (2018) and in [chapter 5](#). With a parametric noise model, one must take care to ensure that the chosen distribution is expressive enough to capture the underlying model outputs, for example with z-scoring for Gaussian distributions or calibration analyses such as in [Figure 5.5c](#). More recently, researchers have opted for normalising flows, see [subsection 3.1.3](#) - which make weaker assumptions about the underlying distribution.

One could also explore alternative surrogate architectures to improve the accuracy of posterior estimates. As mentioned in [chapter 1](#), there are many alternative surrogates (such as linear Bayes emulators, Gaussian Processes) and neural network based architectures (such as Convolutional Neural Networks, Echo State Networks, Normalising Flows and Transformers), which would introduce different inductive biases into training. Some may produce approximation behaviours which are more acceptable for inference in low burden scenarios. However, the "no free lunch" theorems, see Wolpert and Macready (1997), posit that better approximations in one class of trajectories would be offset by worse approximations in others.

Choosing a surrogate architecture which models autocorrelation would, in principle, improve predictive performance over surrogates which model time points independently. Inductive biases which model dependencies in data with latent structure have shown improved predictive performance in natural language (Vaswani et al., 2017), audio (Oord, Dieleman, et al., 2016) and image data (Oord, Kalchbrenner, et al., 2016). In this case, the Seq2Seq architecture had the added benefit of producing time-varying predictions which adds flexibility to inference.

One could alternatively increase the number of samples in the training set. In the previous chapter, I found that the predictive performance of the surrogate showed diminishing returns after 2,500 samples. It is likely that the optimal number of samples for training the Seq2Seq surrogate is larger given the complexity of the underlying ABM, however an in depth convergence analysis is yet to be done, given the long execution times for the ABM.

Finally, sharing a surrogate model is more energy efficient and equitable than an ABM. Low and middle income countries are disproportionately affected by infectious disease and are less able to finance high performance computing. Surrogate predictions require a small fraction of the ABM's computational requirements making analyses more feasible in the settings where they are most needed.

4.5.2 Generality of the workflow

Despite the overall input domain being large, the vectors for each time point are low-dimensional. Over 50 malaria parameters were fixed to values sources from expert opinion and previous parameter inference studies (Sinka et al., 2012; Sherrard-Smith et al., 2022; Griffin, Ferguson, and Ghani, 2014). If incorporated, it is unlikely that a Latin hypercube sampling strategy would produce the samples sufficient to learn the function f over the larger input space. Developing strategies to more intelligently sample the input space without sacrificing generality poses a challenging research problem.

The sequence structure is also limited in granularity. This surrogate has fixed time-step intervals to a year. However many infectious diseases, including malaria, are seasonal and so there is value in exploring the timing of interventions on a monthly or daily timescale. Reducing the time-step interval would increase the size of training sequences. For example, training this malaria surrogate with the daily time-step would increase the training sequences from $n = 16$ to $n = 5840$. More advanced seq2seq architectures could be considered here, such as those in (Cho et al., 2014; Vaswani et al., 2017), taking care to avoid the vanishing/exploding gradient problems.

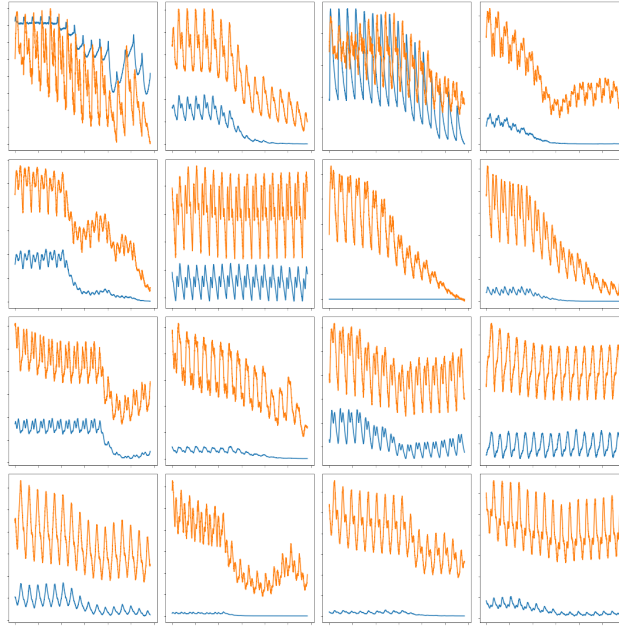


Figure 4.9: 16 cases with the highest MSE from the historic benchmark

4.5.3 Simplified ABM Development with “Individual” and “malariasimulation”

Apart from the Bayesian inference outcomes discussed above, two important contributions of this work has been the development of the “Individual” framework for Agent-Based Model (ABM) development and the new implementation of Imperial College London’s malaria model “malariasimulation”. The framework has helped in the rapid development, testing, and maintenance of a complex malaria model, and enabled researchers to develop timely COVID-19 models, leading to valuable insights on long-term vaccination strategies. The new malaria model provides more functionality than the previous implementation and executes with competitive simulation times.

“malariasimulation” now includes support for agent-based mosquito modelling, which facilitates gene drive modelling; incorporates time-varying carrying capacity parameters, enabling the simulation of different vector control interventions such as Attractive Toxic Sugar Baits (ATSB) and larviciding; and metapopulation modeling, which allows researchers to study the effects of mixing transmission dynamics between spatially adjacent sites.

The integration of surrogate modelling approaches with ABMs, as facilitated by “Individual”, could lead to a future where ABMs are both accessible and applicable to robust Bayesian inference problems. This combination promises to improve the usability of complex models, making them more accessible to a broader range of researchers who may not have expertise in high-performance computing.

One of the primary drawbacks of “Individual” is its execution speed. While most modern malaria models, including OpenMalaria, EMOD, and the original implementation of the Imperial College Model, are written entirely in systems programming languages and optimised for speed, they remain relatively inaccessible to many scientific researchers. In contrast, “malariasimulation” developed using “Individual” offers comparable execution times with the accessibility of R.

Although “malariasimulation” is likely slower than its alternatives in comparable simulations, due to its reliance on interpreted R code, this difference is not significant enough to affect routine modelling tasks. My view is that it is unlikely that any ABM can be optimised to become fast enough for Bayesian inference without the aid of surrogate models. If we can develop robust surrogates, the focus in ABM development can shift from optimisation towards enhancing accessibility and usability.

Table 4.2: Processes for malaria simulation

Name	Type	Description
create_exponential_decay_process	Process	Reduces the value of a variable at an exponential rate. This is applied to all of the immunity variables.
create_mosquito_emergence_process	Process	Move mosquitoes from NonExistent to Sm in line with the number of pupae in the compartmental model
create_biting_process	Process	Samples humans to be bitten by infectious mosquitoes, and leads to calculations for human infection and mosquito infection and mosquito death.
create_mortality_process	Process	This is the process for human mortality, it defines which humans die from natural causes replaces dead individuals with newborns.
create_progression_process	Process	Samples individuals to move towards the later stages of disease and updates their infectivity
create_solver_stepping_process	Process	Steps the differential equation solver for the aquatic and/or adult mosquitoes forward one time step
create_rtss_epi_process	Process	Samples individuals to be vaccinated according to the EPI strategy
categorical_count_renderer_process	Process	Saves a count of the number of individuals with each category for a categorical variable. This is applied to the "state" variable.
create_variable_mean_renderer_process	Process	Saves the mean value of a double variable. This is applied to the immunity variables.
create_prevalance_renderer	Process	Saves the number of individuals who would be detected by microscopy.
create_compartmental_rendering_process	Process	Saves the size of each compartment in the compartmental model.
create_vector_count_renderer_individual	Process	Saves the number of individuals in each of the adult mosquito states (if adult mosquitoes are being simulated individually).
create_total_M_renderer_compartmental	Process	Saves an aggregated total_M figure for every adult mosquito in the compartmental model (if adult mosquitoes are being simulated compartmentally).
distribute_nets	Process	Distributes insecticide treated bed nets according to the model parameters.
indoor_spraying	Process	Activates IRS protection for individuals according to the model parameters.
asymptomatic_progression	Event	Triggered when an individual has been sampled to move from D to A.
subpatent_progression	Event	Triggered when an individual has been sampled to move from A to U.
recovery	Event	Triggered when an individual has been sampled to move from U to S.
clinical_infection	Event	Triggered when an individual has been sampled to move to D.
asymptomatic_infection	Event	Triggered when an individual has been sampled to move to A.
mda_administer	Event	Triggers a Mass Drug Administration.
smc_administer	Event	Triggers a Seasonal Malaria Chemoprevention.
tbv_vaccination	Event	Triggers a distribution of Transmission Blocking Vaccines.
throw_away_net	Event	Triggers an insecticide treated net to be removed for an individual.

Hyper-parameter	Values
Optimiser	Adam , RMSProp
Optimiser Loss	MSE, Log cosh
Batch size	1000, 100 , 50
Dropout	0 , .1
RNN Layer	LSTM , GRU

Table 4.4: Hyper-parameters used in the grid search. The selected parameter is emboldened

Task	Seconds per run	Std
ABM run	1.76×10^3	7.69×10^2
Prediction (single)	1.74×10^{-3}	2.84×10^{-8}
Prediction (batch)	3.74×10^{-4}	6.12×10^{-2}

Table 4.5: The timings per simulation for 10,000 simulations on the training and validation parameter sets

Chapter 5

Bayesian inference of high-dimensional Agent Based Model parameters

In this chapter, I present the latest attempt to infer immunity parameters for Griffin et. al’s malaria transmission Agent Based Model (ABM) since 2014 (Griffin, Ferguson, and Ghani, 2014). This is the first attempt to use a dynamical model for likelihood calculations and account for all WHO suggested interventions since 2000, including insecticide treated nets, indoor residual spraying and anti-malarial treatment. This attempt also used a larger observational data set consisting of prevalence and incidence statistics between 1985 and 2013.

To perform this inference, I combined the insights I gained from the previous chapters. I used the formulation for the Bayesian inference of malaria immunity parameters, tuned priors and diagnostics which I introduced in chapter 2; I trained a surrogate with the annealed Stochastic Variational Inference (SVI) workflow which I introduced in chapter 3; and I used the Seq2Seq architecture for the surrogate, which I introduced in chapter 4, to rapidly simulate ABM outputs while exploiting the temporal nature of transmission dynamics. Further to the previously discussed contributions, I adapted the Seq2Seq surrogate to estimate the density of the ABM outputs to account for the stochasticity in ABM outputs.

5.1 Background

Estimating model parameters for Agent Based Models (ABMs) is a significant challenge. ABMs, due to their complexity and detail, often require substantial computational resources to execute simulations. If they simulate discrete randomness, which is often the case, it is challenging to estimate gradients for their outputs, complicating the use of gradient-based inference methods, this is discussed in detail in Quera-Bofarull et al. (2023). The inherent stochasticity in ABM outputs adds another layer of complexity when deriving a likelihood function, as in most cases a closed form probability density for ABM outputs is not available. Moreover, the intricate dynamics within realistic models can significantly slow down traditional Markov Chain Monte Carlo (MCMC) approaches, which are typically used to sample from the posterior distribution.

ABMs, while advantageous for their detailed dynamics, naturally incorporate processes that are not amenable to auto-differentiation, which is essential for many deep learning and Bayesian inference techniques. For example, modelling infections as Bernoulli trials, as is often the case in epidemic models, does not lend itself to easy gradient estimation. Current research has proposed methods for estimating the gradients for such processes (Chopra et al., 2023; Arya et al., 2023). However these methods come with a trade-off between extra computational overheads and high-variance or biased gradient estimates. These limitations necessitate alternative approaches.

Stochastic ABMs rarely offer closed-form solutions for the likelihood of outputs with respect to model parameters. Researchers have traditionally turned to likelihood-free methods, such as Approximate Bayesian Computation (ABC) (McCreesh et al., 2017), to circumvent these difficulties. More recently, techniques like parametric estimation (Lamperti, Roventini, and Sani, 2018) and neural density estimation (Dyer et al., 2022; Papamakarios, Sterratt, and Murray, 2019) have

been explored to approximate the probability density of model outputs, offering new avenues for inference.

As I found in [chapter 3](#), likelihood functions with high-dimensional input spaces which are approximated by surrogate models can lead to noisy, non-smooth posterior distributions which are not easily sampled with MCMC methods. Even for sophisticated MCMC variants like HMC with the No-U-Turn Sampler (NUTS), the noisy likelihood estimates can hinder the sampler’s ability to efficiently explore the parameter space, leading to prohibitively long convergence times and, in some cases, making inference computationally infeasible. Stochastic Variational Inference (SVI) presents a promising alternative. SVI approximates the posterior distribution through optimisation, typically using Stochastic Gradient Descent (SGD) which can be significantly faster and robust to noisy estimates than MCMC-based methods. However I also found that SVI estimates could be overconfident and susceptible to mode collapse, which has also been reported independently (K. Kim et al., [2023](#); Dhaka and Catalina, [2021](#)).

5.1.1 Inference using stochastic models

The objective of Bayesian methods for parameter inference is to estimate a posterior distribution $p(\theta|\mathbf{y})$ for model parameters θ and observations $\mathbf{y}$ using Bayes’ theorem:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)}{p(\mathbf{y})}p(\theta)$$

Unfortunately, for many useful models, the likelihood, $p(\mathbf{y}|\theta)$, is not tractable. Realistic models often include stochastic processes which have no closed form analytical solution for the probability density of its outputs. This is especially the case with ABMs, where even simple agent behaviours can lead to complex, non-linear, emergent dynamics if agents interact within a shared system.

Parametric density estimation is one method estimating the likelihood. The modeller selects a parametric probability distribution and estimates the hyperparameters of distribution using a pool of model simulations. The success of this method depends on the choice of distribution, as it is important for accurately capturing model dynamics. For example, selecting a diagonal Gaussian distribution is an efficient parametric distribution which is not expressive enough to capture multi-modal, correlated or skewed model outputs, but may be satisfactory under the assumptions of uni-modality, independence and symmetry.

Kernel Density Estimation (KDE) instead estimates density directly from simulations without such restrictive assumptions. KDE works by placing a kernel on each point estimate from a pool of simulations and summing their contributions to form a density estimate for nearby points. A major limitation of this approach is in modelling high dimensional, complex probability spaces. For complex models, the number of samples needed to capture the model dynamics may be computationally infeasible.

More recently Neural Density Estimation (NDE) has been applied to likelihood estimation tasks. These methods involve learning invertible, differentiable mappings from tractable distributions (commonly Gaussian) to the desired, more complex, distribution. Autoregressive flows is one of the most popular techniques within the domain of parameter inference. Importantly, NDE methods do not introduce a restrictive parametric distribution or scale poorly with the sample size as is the case with the previous methods.

5.1.2 Related work

There has been much progress in parameter inference for malaria models. Parameter inference aims to estimate parameters which explain observed malaria transmission patterns and elucidate the uncertain immunity dynamics and vector behaviours. The most recent effort to perform parameter inference on Griffin’s ABM of malaria transmission was published in (Griffin, Ferguson, and Ghani, [2014](#)). The authors employed a Bi-stable equilibrium model to estimate parameters using data from prevalence and incidence studies observed in 15 sites, successfully providing estimates for 21 parameters. However, this approach was limited by its assumption of equilibrium dynamics, constraining the observational data to pre-intervention sites and neglecting variations in transmission due to seasonality.

Soon after, McCarthy et al. ([2015](#)) used incremental mixture importance sampling to perform parameter inference on the EMOD DTK Agent Based Model (ABM) for malaria transmission.

Their methodology achieved good fits for selected data across six sites. Despite this success, importance sampling methods can be of limited effectiveness in complex parameter inference where the posterior space can be concentrated in a small region of the parameter space.

Cameron et al. (2015) explored the relationship between prevalence and incidence by performing parameter inference on three different ABMs. The authors introduced a kernel-weighting method to map model parameters to age-incidence and age-prevalence curves, serving as a surrogate to avoid computationally demanding simulations during the inference process.

More recently, Reiker, Golumbeanu, et al. (2021) undertook a significant effort to estimate 21 parameters for the OpenMalaria model using Bayesian Optimisation. This approach yielded parameters that demonstrated a good fit across a broad spectrum of data, including 61 data points for 11 objectives. However, the execution time, ranging from 7 to 12 days, adds a barrier to experimentation. Additionally, the method produced point estimates for the model parameters, limiting the ability of malaria researchers to incorporate uncertainty into future forecasts.

5.1.3 Contributions

In this chapter, I contribute to methodology for performing parameter inference on complex, stochastic models, specifically within the context of malaria transmission modelling. The key contributions can be summarised as follows:

Introduction of a density estimating surrogate: A novel surrogate modelling approach is presented, which captures the outputs of an Agent Based Model (ABM) as a probability distribution. This method enables sampling using computationally less intensive, low-population simulations. By modelling the ABM outputs in this manner, it becomes possible to approximate the complex dynamics of malaria transmission more reliably than the surrogate introduced in the previous chapter.

Training a high-dimensional density estimating surrogate: This surrogate is capable of accurately capturing model outputs across more than 600 dimensions (prevalence, incidence, population counts and various immunity levels across 100 age groups) for 40-year trajectories at a monthly resolution. The flexibility of this surrogate allows for any model output to be used for inference, diagnostics or experimentation without the need for retraining, facilitating a more versatile and efficient workflow in malaria research.

Bayesian inference for an agent based model of malaria: This is the first probabilistic estimate of parameters for the Griffin model which uses dynamical malaria transmission model outputs to estimate the likelihood. I relaxed the previous assumptions related to equilibrium dynamics and lack of interventions. With this approach, I provide a more comprehensive estimation parameters, offering new insights into malaria immunity dynamics and the effectiveness of potential interventions.

5.2 Methods

5.2.1 An Agent Based Model of malaria transmission

For this chapter, I used a transmission model which simulated the effects of seasonality and interventions on malaria transmission, published in (Griffin, Ferguson, and Ghani, 2014). Humans were modelled individually. Each human had attributes relating to infection state, immunity, age and heterogeneity. Human transmission dynamics were modelled at discrete time steps, usually a day, and updates to each human’s attributes were simulated based on their expected exposure to infected mosquitoes and their time-varying use of interventions (including Insecticide Treated Nets (ITNs), Indoor Residual Spraying (IRS) and Seasonal Malaria Chemoprevention (SMC)). The human, mosquito and ITN dynamics were described in subsection 4.2.2. Below, I will extend the description to include the remaining interventions which I used for inference.

Front-line treatment

Pharmacological treatment, especially Artemisinin-based combination therapies, are the first-line treatment for malaria infection. The challenges for Low and Middle Income Countries (LMICs)

Drug name	eff_d	Λ_d	κ_d	rel_c
Dihydroartemisinin-Piperaquine	0.95	4.4	28.1	0.09434
Artemether-Lumefantrine	0.95	11.3	10.6	0.05094
Sulphadoxine-Pyrimethamine and Amodiaquine	0.9	4.3	38.1	0.32

Table 5.1: Parameters for Anti-malarial drugs. eff_d is the efficacy of each drug; Λ_d and κ_d are the shape and scale parameters describing prophylaxis from infection as a Weibull survival curve; and rel_c is the relative reduction infectiousness towards mosquitoes experienced by each treated individual.

in sub-Saharan Africa are that access to treatment may be limited due to treatment seeking behaviour, distance from a hospital and health system resource constraints (World Health Organization, 2023). Current available treatments can clear clinical infections with high efficacy (Okell et al., 2014), provide short term prophylaxis against future infections and reduce parasitaemia in patients resulting in less infectiousness.

In the model, I simulated first-line treatment for individuals who experience a clinical infection in each time step. I first sampled individuals who would receive treatment using the parameter, f_t , the probability of seeking treatment after infection. The drug which they received was parameterised according to estimates of the drug availability, $\text{cov}_d(t)$ for each drug d . Finally, individuals who receive successful treatment were sampled according to estimates of drug efficacy, eff_d , for the drug they had received. The cohort of successfully treated humans were moved to the Treated (T) state. Prophylaxis to pre-erythrocytic infection for treated individuals was simulated as a Weibull survival curve from treatment time with shape and scale parameters Λ_d and κ_d respectively, i.e $e^{-\frac{t}{\Lambda_d} \kappa_d}$ for a given time since treatment, t . This prophylaxis is assumed to provide extra protection independently of pre-erythrocytic immunity. Treated individuals are also attributed a $\text{rel}_i \in [0, 1]$ value, parameterised by the drug they had received, to reduce their contribution to Λ_M , see Equation 4.9. Estimates of the above parameters for commonly available drugs are listed in Table 5.1.

Indoor Residual Spraying (IRS)

IRS involves spraying the interior walls of a habitat with an insecticide spray. This provides protection by killing or repelling mosquitoes while they are in human homes resting on the walls. As with ITNs, IRS is a long term intervention whose effectiveness is reduced over time. IRS interventions are parameterised in the model by a time series of estimated usage $\text{usage}_{\text{IRS}}(t)$ at time t , representing the proportion of the population who live in homes which are protected.

In the model, the introduction of IRS, similarly to ITNs, has an effect on w_i , y_i and z_i , defined in Equation 4.7. These probabilities decompose feeding attempts into those which are indoors and those which are in bed with probabilities Φ_I and Φ_B respectively. At the time of writing, the most recent model for the probability of successful feeding and repelling in the presence of IRS, s_S and r_S , were been published in (Sherrard-Smith et al., 2022). (ibid.) modelled these probabilities in terms of mosquito behaviours in close proximity to insecticides, i.e. mortality l_s , blood feeding rate k_s and deterrence m_s . These behaviours were in turn modelled using logistic functions with parameters for scale and shape, $l_{S\theta}, l_{S\gamma}, k_{S\theta}, k_{S\gamma}, m_{S\theta}, m_{S\gamma}$, see Equation 5.2. Model parameters were estimated for the latest insecticides, pyrethroid and a piperonyl butoxide (PBO) combination, using observational data from experimental hut trials. I have listed the complete model in Equation 5.1.

$$\begin{aligned}
 s_S &= \frac{k'_S}{k_0} \\
 r_S &= \left(1 - \frac{k'_S}{k_0}\right) \left(\frac{j'_S}{j'_S + l'_S}\right)
 \end{aligned} \tag{5.1}$$

Where,

$$\begin{aligned}
l_s &= \frac{1}{1 + \exp(-(l_{S\theta} + l_{S\gamma}t))} \\
k_s &= \frac{1}{1 + \exp(-(k_{S\theta} + k_{S\gamma}t))} \\
m_s &= \frac{1}{1 + \exp(-(m_{S\theta} + m_{S\gamma}t))} \\
j_s &= 1 - l_s - k_s \\
l'_s &= l_s(1 - m_s) \\
k'_s &= k_s(1 - m_s) \\
j'_s &= j_s(1 - m_s) + m_s
\end{aligned} \tag{5.2}$$

Seasonal Malaria Chemo-prevention

Seasonal Malaria Chemo-prevention (SMC) is a particularly effective intervention in highly seasonal settings. Drugs are distributed to large sub-populations with young demographics typically three months before the projected seasonal peak of transmission. This is in an effort to provide prophylaxis to those most at risk to infection, since immunity to infection is underdeveloped in childhood (World Health Organization, 2023). The WHO recommended drug for SMC use is Sulphadoxine-Pyrimethamine (SP).

I simulated SMC in the model by distributing drugs to a proportion, $\text{dist}_{\text{smc}}(t)$, of the sub-population, irrespective of infection state, between 3 and 59 months old at time t . All selected individuals experience prophylaxis from infection. Any selected individual who had a clinical infection or an asymptomatic infection which would be detectable by microscopy was switched to the Treated (T) state. As with first-line treatment, this led to the individual's contribution to Λ_M being reduced by a factor of rel_{smc} , see Table 5.1. Selected individuals without a detectable infection were set to the Susceptible (S) state.

5.2.2 Site data

I parameterised the model to estimate prevalence and incidence in 31 administrative units in 16 countries in sub-Saharan Africa. The model parameters accounted for the most common interventions over the observational period, i.e. routine treatment, Insecticide Treated Nets (ITNs) and Seasonal Malaria Chemo-prevention (SMC), as well as the demographic and seasonal trends. I used the site and foresite packages developed by Winskill et al, see (Pete Winskill, 2023b; Pete Winskill, 2023a), to appropriately parameterise the ABM.

As observational data I used the database of Malaria prevalence and incidence studies compiled by (Battle, Guerra, et al., 2015). I filtered the incidence data to include active case detection studies, and prevalence data to include cross sectional surveys. I then linked the studies to administrative units as defined by the Malaria Atlas Project border dataset (Pfeffer et al., 2018). I kept studies for administrative units which had both prevalence and incidence data. Each administrative unit had studies which collected data on different age groups and different time periods between 1985 and 2013. The final observational dataset covered 31 sites in sub-Saharan Africa, with 91 prevalence data points and 318 incidence data points. A summary of this reduced dataset is listed in Table 5.2.

Routine treatment for each administrative unit was parameterised by the proportion of people who receive treatment on infection and the proportion of treatments which are Artemisinin Combination Therapies (ACTs). The proportion of people who receive treatment is estimated from the Malaria Atlas Project's "Effective treatment with an Antimalarial drug" raster, (Bhatt et al., 2015). The proportion of ACT treatments is estimated using Demographic and Health Surveys (DHS) using the "Children who took any ACT" indicator. The proportion of ACT before 2006 was assumed to be zero, since this was when the World Health Organisation (WHO) first recommended the intervention.

ITN interventions were parameterised by the number of nets distributed, the efficacy of the type of net used and insecticide resistance in each administrative unit. The number of nets distributed each year was calibrated to match net use as estimated by (Bertozzi-Villa et al., 2021), using the `netz` package, see (Pete Winskill, Schmit, and Gerard-Ursin, 2023). All ITNs were assumed to be treated with a Pyrethroid-based insecticide with efficacy parameters, including the dn_0 , rn_0 ,

iso3c	n_sites	prev_points	inc_points
BEN	1	1	3
BFA	4	10	31
CIV	1	2	12
CMR	2	5	11
GHA	2	3	4
GMB	1	1	3
KEN	3	7	84
MLI	2	12	6
MOZ	1	8	6
MRT	1	1	1
NGA	1	1	1
SEN	3	17	62
SLE	1	1	2
TCD	1	9	9
TZA	6	8	39
UGA	1	3	6

Table 5.2: Representation of countries in sub-Saharan Africa in observational data after pre-processing. `n_sites` refers to the number of administrative units which were included in the respective country. Prevalence points refers to the number of distinct age groups and time periods which were observed in all cross sectional surveys which were conducted in the respective country. Incidence points is analogous for active case detection surveys.

rn_m and γ_n parameters introduced in Equation 4.8, estimated by (Sherrard-Smith et al., 2022). Insecticide resistance was modelled by modifying the efficacy parameters according to estimates published in (Thomas S Churcher et al., 2016).

IRS interventions were parameterised by the coverage in each administrative unit. To estimate coverage, Winskill et al. calibrated estimates of IRS usage from the malaria atlas project (Bhatt et al., 2015) to match the coverages estimated by the World Malaria Report (World Health Organization, 2020). The insecticide efficacy parameters, including the ls , ks and ms introduced in Equation 5.1, were estimated by (Sherrard-Smith et al., 2022) with the insecticide used for spraying assumed to be DDT before 2017 and actellic afterwards.

SMC was parameterised by the coverage in each administrative unit. Initial distributions were identified from Access SMC (SMC Alliance, 2024). The coverage was assumed to linearly increase until 80% of the administrative unit is protected. Distributions were assumed to occur 4 times a year and to target children between 3 months and 5 years old.

The population demographics in the model were parameterised on a country level by annual mortality rates for fixed age bands between 0 and 100 years old, with the “100+” age band modelled with an upper bound of 200. The mortality rates were downloaded from the UN World Population Prospects’ dataset (United Nations, 2022). Since the foresite package only provides data after the year 2000, see (Pete Winskill, 2023a), I assumed the earliest mortality rates for the period between 1985 and 2000. Separate parameter estimates were made for urban and rural regions of each administrative unit. I used the urban parameters where available and the rural parameters only for completely rural administrative units.

The mosquitoes in each administrative unit were parameterised by the relative abundance of each major species. The relative abundance of *An. Arabiensis*, *An. Funestus* and *An. Gambiae* was aggregated from rasters provided by the Malaria Atlas Project derived from models published by (Sinka et al., 2012). A map of these data is shown in Figure 1.2.

The seasonality in each administrative unit is parameterised with the 7 Fourier coefficients, introduced in Equation 4.6, to summarise the average daily rainfall over a year. Firstly, the daily rainfall between 2019 and 2021 for each administrative unit was aggregated from rainfall estimates provided by CHIRPS, see (Funk et al., 2015). The Fourier parameters were fit using L-BFGS-B, see (Zhu et al., 1997) for details, with the objective of minimising the mean squared error between the Fourier projection and the rainfall data.

In a departure from the parameter inference exercise in Griffin, Ferguson, and Ghani (2014), I chose not to add informative priors for Entomological Inoculation Rate (EIR). Griffin et al. found estimates for EIR for each of their observation sites and graded their uncertainty into 5 groups in

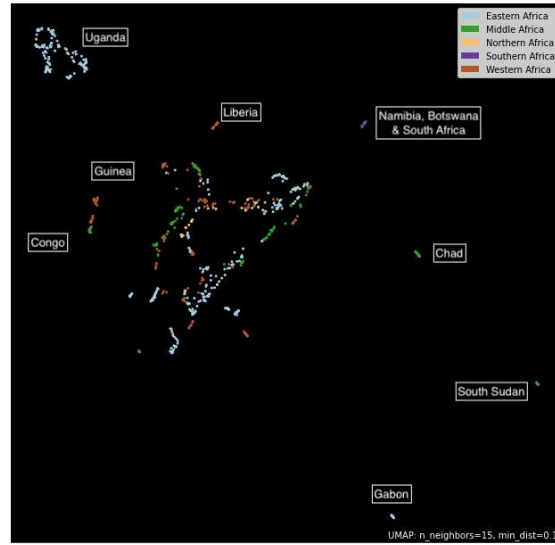


Figure 5.1: Visualisation of Insecticide Treated Net usage in Africa. These are UMAP projections of the time series for Insecticide Treated Net usage in each African administrative unit, coloured by geographic region. This reveals large variability in the usage of bed nets between 2000-2020 even within African regions.

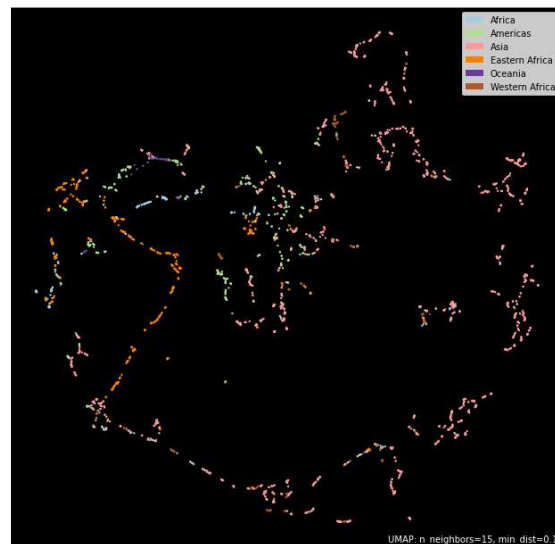


Figure 5.2: Visualisation of global seasonality parameters. A UMAP projection of Fourier coefficients which were fit for global administrative units, coloured by (sub-)continental region. As expected, there is much variation in yearly rainfall patterns, especially in West African administrative units which do not cluster together.

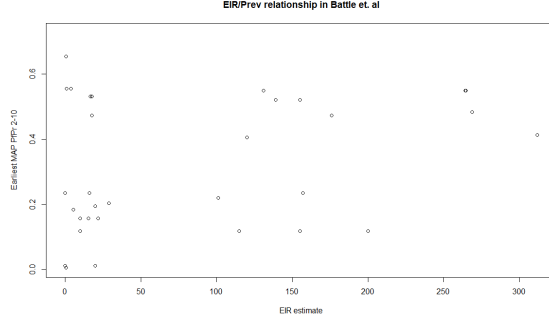


Figure 5.3: Plot of recorded Entomological Inoculation Rate (EIR) against parasite prevalence for 2 to 10 year olds, PfPr₂₋₁₀, which was estimated from Malaria Atlas Project rasters for 2000.

order to set appropriately informative priors. I found the EIR estimates in the (Battle, Guerra, et al., 2015) data set to be too sparse, lacking in temporal meta-data, and unstandardised to produce reliable uncertainty estimates for many of the studies included in the observational data set. The methodology for estimating EIR varied significantly between studies, requiring additional modelling in order to make them comparable. I plotted the relationship between the EIR estimates and estimates of pre-intervention prevalence in the corresponding administrative units to confirm that relationship was not easily discernible, see Figure 5.3.

5.2.3 Surrogate model

Let us model simulations of the Agent Based Model (ABM), parameterised with an arbitrary parameter set $\mathbf{x}$ in some bounded space of parameters $\mathcal{X}$, as discrete stochastic processes $\{Y_{\mathbf{x},t}\}_{t \in \mathbf{T}}$, where $Y_{\mathbf{x},t}$ represents the possible ABM outputs for model parameters $\mathbf{x}$ at time t and are defined on a probability space $(\Omega, \mathcal{F}, P_{\mathbf{x},t})$, where $P_{\mathbf{x},t}$ is a function which assigns a probability to an ABM output at time t given input parameters $\mathbf{x}$, and $\mathbf{T}$ is an index over $\mathbb{Z}$. Thus, the ABM produces trajectories $\mathbf{y} = \{y_t\}_{t \in \mathbf{T}}$ with probability $P_{\mathbf{x}}(\mathbf{y}) = \prod_t P_{\mathbf{x},t}(y_t)$, i.e. the probability of a trajectory is the product of the probabilities of each of its points. My goal is to train a surrogate model $\tilde{f}_{\phi} : \mathcal{X} \rightarrow P$, where P is a probability on $(\Omega, \mathcal{F})$, such that $\forall \mathbf{y} |\tilde{f}_{\phi}(\mathbf{x})(\mathbf{y}) - P_{\mathbf{x}}(\mathbf{y})| < \epsilon$, i.e. the surrogate's output probability distribution is sufficiently close to the ABM output's distribution. Let $\mathcal{L}(\mathbf{x}, \mathbf{y}, \phi)$ be a loss function for the surrogate.

The goal of surrogate modelling is to find a hyper-parameter set $\hat{\phi}$ which replicates the ABM's output distribution for any parameter set in $\mathcal{X}$. That is, we want to find $\hat{\phi}$ such that:

$$\hat{\phi} = \arg \min_{\phi} \int_{\mathcal{X}} p(\mathbf{x}) \mathcal{L}(\mathbf{x}, \mathbf{y}, \tilde{f}_{\phi}) d\mathbf{x}$$

I assumed that the ABM outputs at each time point were Normally distributed with a mean and variance output by the surrogate, $\mu_{\mathbf{x},t}^{\phi}$ and $\sigma_{\mathbf{x},t}^{2\phi}$ respectively, when conditioned on ABM inputs $\mathbf{x}$. To make the outputs valid for inference they were truncated at zero. By introducing in the truncated Gaussian log likelihood, I arrived at the following loss function:

$$\log \mathcal{L}(\mathbf{x}, \mathbf{y}, \phi) = \sum_t \left[\log \left(\Phi \left(\frac{y_t - \mu_{\mathbf{x},t}^{\phi}}{\sigma_{\mathbf{x},t}^{\phi}} \right) - \Phi \left(\frac{-\mu_{\mathbf{x},t}^{\phi}}{\sigma_{\mathbf{x},t}^{\phi}} \right) \right) \right]$$

5.2.4 Surrogate training

I sampled (x, y) pairs from the ABM to generate a training dataset for the surrogate model. The intrinsic parameters were sampled from the prior distribution, as it was empirically the most effective strategy in the previous chapter. Site parameters were generated independently by choosing a site from the observational dataset with replacement. I sampled 90,000 pairs for the initial training step, leaving 10,000 for model validation. The pairs were standardised prior to surrogate training.

I constructed an RNN-based surrogate to learn the conditional probability density of the ABM outputs for each time step. The surrogate had a two layer RNN architecture, with 256 LSTM units per layer, to convert the model parameters into a hidden representation. The RNN outputs

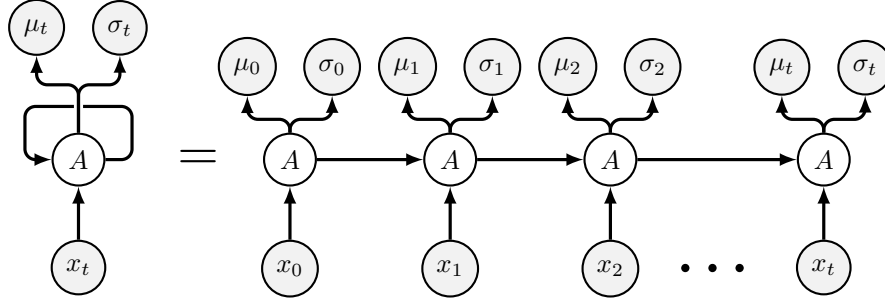


Figure 5.4: A Parametric Density Estimating Recurrent Neural Network (RNN). x_t represents the input for the RNN, which in this case was the time series of input parameters (including interventions, seasonality, demography and immunity parameters for inference). The time-series had a fixed step size of one month. Note that many parameters, including the immunity parameters for inference were kept constant over time. A represents the operation to process inputs for time t based on the input x_t and the hidden state h_{t-1} . In this case I used two layers of 256 units of the Long-Short-Term-Memory (LSTM) cell which is depicted in Figure 4.1. Each hidden state h_t is passed through two densely connected layers to produce estimates for μ_t and $\log(\sigma_t)$.

fed into two independent, densely connected layers to estimate $\mu_{\mathbf{x},t}^\phi$ and $\sigma_{\mathbf{x},t}^\phi$. This architecture is visualised in Figure 5.4. The RNN was optimised by Stochastic Gradient Descent (SGD) using the Adam optimiser. Gradients for the surrogate were estimated on mini-batches of 100 pairs at a time for 100 epochs.

5.2.5 Parameter Inference

The prior distribution from the previous chapter was used with a few modifications for the more complex ABM and realistic observational data. New parameters, c_u , c_d , γ_d were introduced to model the Force Of Infection towards Mosquitoes, Λ_M , which was previously superfluous since no interventions were modelled. Prevalence and Incidence surveys were modelled with BetaBinomial and GammaPoisson, introducing the auxiliary parameters to model overdispersion. Specifically, the BetaBinomial distribution is parameterised by two concentration parameters c_0, c_1 and a count parameter n such that $\text{BetaBinomial}(c_0, c_1, n) \sim \text{Binomial}(p, n)$ where, $p \sim \text{Beta}(c_0, c_1)$. Similarly the GammaPoisson distribution is parameterised by a concentration and rate parameter c_γ and r such that $\text{GammaPoisson}(c_\gamma, r) \sim \text{Poisson}(\lambda)$ where, $\lambda \sim \text{Gamma}(c_\gamma, r)$. Note that the GammaPoisson is equivalent to the Negative Binomial distribution with success and failure rates of $\frac{rc_\gamma}{1+rc_\gamma}$, $\frac{1}{1+rc_\gamma}$ respectively. The updated priors and likelihood function are listed in Table 5.3, and Equation 5.3 respectively.

$$\begin{aligned}
u_{ijt} &\sim \text{BetaBinomial}(\phi_j \text{prev}(i, j, t, \Lambda_j, \theta), \phi_j (1 - \text{prev}(i, j, t, \Lambda_j, \theta)), n_{ijt}) \\
v_{ijt} &\sim \text{GammaPoisson}(q_j^{-1} \text{inc}(i, j, t, \Lambda_j, \theta), q_j^{-1}) \\
\mathcal{L}(\theta) &= P(\theta) \prod_j P(\Lambda_j, \phi_j, q_j) \prod_{ijt} P(u_{ijt} | \phi_j, \theta, \Lambda_j) P(v_{ijt} | q_j, \theta, \Lambda_j)
\end{aligned} \tag{5.3}$$

I used the annealed Stochastic Variational Inference method introduced in section 3.2.2 for sampling from the posterior distribution with some modifications to accommodate processing the larger observational data set and surrogate model outputs on a single Graphics Processing Unit with 16GB of memory. I modified the ELBO loss to use 32 samples from the likelihood per step instead of the previous 128 samples. Also, I used a subsampling scheme where 10 prevalence and 10 incidence points were randomly sampled from the observation per step as opposed to the entire data set. The evidence for the sub-sampled observations were scaled according to their proportion of the overall dataset. As an annealing scheme, I used the previous geometric rate of cooling of $r = 0.45$. However, I performed inference over five rounds instead of ten. After each round, the surrogate was trained with 10,000 further samples from the posterior drawn from the proceeding round.

I have made the code publicly accessible to reproduce this analysis in (G. Charles, 2024c).

Parameter	Prior
initial EIR	Uniform(0, 1000)
Pre-erythrocytic immunity:	
κ_b	LogNormal(0, 0.25)
u_b	Gamma(7, 1)
b_0	Beta(1, 1)
Ik_0	Normal(50, 20) ($25 < IB_0 < 75$)
Clinical immunity:	
κ_c	LogNormal(0, 0.25)
u_c	Gamma(7, 1)
ϕ_0	Beta(5, 1)
ϕ_1	Beta(1, 2)
IC_0	Normal(25, 10) ($5 < IC_0 < 50$)
P_M	Beta(1, 1)
d_m	Normal(50, 20) ($5 < d_m < 100$)
Immunity to detection:	
κ_d	LogNormal(0, 0.25)
u_d	Gamma(7, 1)
d_1	Beta(1, 5)
ID_0	Normal(25, 10) ($5 < ID_0 < 50$)
fd_0	Beta(1, 1)
γ_d	LogNormal(0, 2)
ad_0	Cauchy(70, 1) ($40 < ad_0 < 100$)
r_u	LogNormal(0, 1)
Force Of Infection towards Mosquitoes:	
c_u	Beta(1, 2)
c_d	Beta(1, 5)
γ^1	LogNormal(0, 0.25)
Overdispersion:	
ϕ	Exp(0.05^{-2})
q^{-1}	Exp(0.05^{-2})

Table 5.3: Prior distribution of the Agent Based Model parameters.

5.3 Results

In the following section I present the culmination of my efforts to improve the efficiency and accuracy of parameter estimates for a complex, high-dimensional, stochastic model of malaria. I validated the approximation quality of the surrogate to assess its reliability for inference. I then used the annealed Stochastic Variational Inference (annealed SVI) workflow which I introduced in chapter 3 to reliably infer a posterior distribution over the model parameters. I then validated the inference with diagnostics which I had introduced in chapters 2 and compared the inferred parameters to the previous estimates.

5.3.1 Validation of approximation quality

I firstly assessed the approximation quality of the surrogate's predictions. For this, I used both the Mean Squared Error (MSE) of the mean output and the probability density of the true output to visualise discrepancies in point and distributional estimation respectively. All validation was carried out against out-of-sample model outputs. In my analysis, I stratified the metrics by the type of model output - prevalence, incidence and population count, see Figure 5.5d - and subsequently by both time and age group, see Figure 5.5a and Figure 5.5b. These stratified analyses reflect the surrogate's capacity to capture the temporal evolution and demographic patterns of each aspect of malaria transmission.

I found the Mean Squared Error (MSE) of the mean estimates in the prior distribution to be low relative to the population size, indicating that the surrogate model can effectively estimate the mean values of the parameters. The incidence counts had an MSE within 10^{-4} , the number of

individuals detected by microscopy within 10^{-2} , and population counts in each age group within 3×10^{-2} of the population size.

The heat maps, shown in [Figure 5.5a](#) and [Figure 5.5b](#), revealed that incidence was the most accurately approximated output and all outputs were better approximated in older age groups and during the later months of the simulation. This trend is likely due to the differences in scale, as older populations and later time points tend to have more stable, smaller values. However, the heat maps also showed striations with significant discrepancies along the time axis, suggesting that seasonality in the output trajectories is not consistently captured by the surrogate. These striations were more pronounced in the mean estimate heat maps compared to the likelihood-based ones, indicating that the surrogate often models seasonality as stochastic variations rather than periodic trends. The mean estimate heat map for population counts displayed diagonal patterns, suggesting that the surrogate has effectively learned the temporal relationships between population counts across different demographics as they age in the simulation.

Calibration plots, in [Figure 5.5c](#), supported the reliability of the surrogate model, showing that the probability densities it produced for prevalence and population count outputs were consistent with the empirical probability distributions observed in the validation set. However, for incidence outputs, the surrogate tended to underestimate the density of model trajectories, highlighting a potential area for improvement.

The qualitative comparison of the surrogate trajectories to the actual model trajectories, shown in [Figure 5.6](#), revealed several systematic discrepancies. In a minority of cases, the surrogate struggled to accurately estimate initial prevalence and incidence values, likely due to the initialisation of the hidden state of the Recurrent Neural Network (RNN). Additionally, many low prevalence trajectories were overestimated, similar to findings from chapter 4, though the distinguishing factors for these trajectories remain unclear. Conversely, low incidence trajectories were typically well approximated. The surrogate model had a tendency to simulate either flat or periodic trajectories or trajectories which transition between both. Occasionally, it incorrectly modelled population counts as periodic. Despite these issues, the surrogate model tended to capture the average outputs including the sharp responses of prevalence and incidence to interventions.

5.3.2 Inference of model parameters from synthetic data

To validate the parameter inference method, I performed inference on synthetic data generated using the surrogate. This approach is similar to the evaluation methods used in chapters 2 and 3, where latent *true* parameters were drawn from the prior distribution and observations generated from the posterior predictive distribution. The key difference in this chapter is the use of the surrogate to generate observations for computational efficiency, instead of relying on the underlying model I did previously. A drawback of using the surrogate for generating synthetic data during validation is that it does not account for surrogate approximation error. Additionally, since calculating a reference posterior is intractable for this ABM, it was not possible to assess the distribution of the posterior estimates as comprehensively as in chapter 3. Instead, the posterior estimates were evaluated in relation to their proximity to the true values.

All of the posterior estimates for the immunity parameters had high probability densities close to the true values, see [Figure 5.7](#). However, the posterior estimates for transmission intensity at each site were quite broad, in some cases distributed uniformly from 0 to 800, see [Figure 5.8](#). This suggests that these parameters are not easily identifiable, likely due to insufficient data or small sample sizes at those sites.

The posterior estimates generally fit the observed data well, as evidenced by the Bayesian p-value plot, see [Figure 5.9a](#). The estimates showed higher than expected variance, overestimating the density of rarer observations, but were overall quite balanced, neither over or underestimating observed prevalence or incidence. This balance indicates that the inference model, while not perfect, was reasonably robust in predicting prevalence and incidence in the synthetic scenario.

5.3.3 Inference of model parameters from observational data

I then performed inference on the data from (Battle, Guerra, et al., 2015) to update estimates for the immunity parameters. The prior predictive distribution, which had been tuned based on insights from the previous parameter inference conducted by (Griffin, Ferguson, and Ghani, 2014), did not align well with the observed data, see [Figure 5.9](#), as it tended to overestimate both prevalence and incidence. However, after conducting the inference, the Bayesian p-value plots in

Parameter	Mean	SD	HDI 3%	HDI 97%
κ_{ab}	2.59	0.30	1.98	3.08
u_b	11.88	1.71	8.87	15.03
b_0	0.98	0.01	0.96	1.00
IB_0	67.80	5.30	57.79	74.72
κ_c	1.03	0.17	0.72	1.34
u_c	14.17	0.91	12.54	15.68
IC_0	22.95	6.71	11.97	35.12
ϕ_0	0.86	0.08	0.71	0.97
ϕ_1	0.97	0.02	0.93	0.99
PC_m	0.22	0.14	0.01	0.48
r_m	85.24	10.04	66.81	99.04
κ_d	1.61	0.16	1.31	1.90
u_d	9.19	0.91	7.68	10.72
d_1	0.88	0.04	0.80	0.95
ID_0	27.47	5.14	17.70	36.34
fd_0	0.96	0.03	0.90	1.00
γ_d	1.68	2.75	0.01	6.51
ad	25612.23	278.31	25070.69	26098.87
r_u	175.69	4.28	166.99	183.18
c_d	0.12	0.07	0.01	0.25
c_u	0.04	0.02	0.00	0.08
γ_1	2.35	0.27	1.90	2.88

Table 5.4: Summary of posterior estimates for ABM model parameters. SD refers to the standard deviation and HDI refers to the Highest Density Interval.

Figure 5.9c demonstrated a generally balanced fit to the observational data. This balanced fit, despite some overconfidence, suggests that the model has been adjusted to reflect the observations data more accurately.

The immunity parameter plot, in Figure 5.10, revealed that the new estimates differ significantly from the previous ones. Specifically, the updated estimates indicate a more modest pre-erythrocytic and clinical immunity to infection while a more pronounced increase in immunity to detection with exposure. These differences could be due to changes in the modelling conditions (i.e considering the influences of demographics, seasonality, vector composition and interventions). Alternatively, these differences could be caused by the more extensive observational dataset which reveals a different incidence-prevalence relationship. In order to disentangle these effects, one could perform an ablation study in which they investigate the effects of the alternative data sets and modelling assumptions on the inferred posterior.

Figure 5.11 compares the fits of trajectories sampled from the prior and estimated posterior distributions to observed prevalence and incidence data and demonstrates the significance of this detailed approach to inference. By simulating full ABM outputs, I was able to visualise the effects of impactful interventions that may have occurred at the time of observation. For instance, the significant reduction in background prevalence and incidence in Dodoma, Tanzania, in 2010. As the Bayesian p-value plots previously revealed, many observations were above the prior trajectories. But the full trajectory plots showed that some of these high observations appeared to be outliers, such as the prevalence rate greater than 90% recorded in Morogoro, Tanzania. Such spurious values highlight the necessity for careful data validation and potential outlier mitigation in future modelling efforts. Conversely, other high observations were more accurately fit with the updated, more modest immunity estimates. Some examples include Central Burkina Faso and Brong Ahafo, Ghana. Despite these improvements, many observations still fell outside the 95th percentile of the posterior estimates. This suggests that the posterior estimates could benefit from having a wider variance to better encompass the range of observed data. The narrower posterior distributions may be a common side effect of over confidence in variational models trained using Stochastic Variational Inference (SVI).

To further validate the posterior estimates, I evaluated the squared error between the mean surrogate estimate and ABM outputs for the posterior samples. This check would assess whether the surrogate is a reliable approximation of the underlying ABM model in the parameter space

which is most relevant. Despite annealed SVI reducing the approximation error for posterior samples in chapter 3, I found a larger squared error for the posterior estimates compared to samples from the prior distribution, see the distribution of squared errors in [Figure 5.5d](#). One method to control these errors in future inference would be to investigate appropriate sample sizes, number of rounds and annealing schedules to optimise convergence. These investigations would likely require a lot of computational resources given the demands of executing ABM simulations.

5.3.4 Execution time for inference

The surrogate model significantly reduced the time for forward simulations, with an average of 23m for ABM simulations compared to an average of 2ms for surrogate simulations on a single CPU. This makes it ideal for manual analyses, diagnostics of the parameter space and its impact on prevalence and incidence trajectories, such as those shown in [Figure 5.11](#). These performance statistics are very similar to those I found in chapter 4, see [Table 4.5](#), despite this surrogate estimating densities for a much larger output space, demonstrating the scalability of NN-based surrogates.

The parameter inference took approximately 3 hours and 55 minutes to execute per round. SVI required approximately 3 hours and 48 minutes to train the variational model for 50,000 steps and the remaining time was spent on drawing prior and posterior predictive samples for diagnostics. The SVI training time was much longer than the average of 78 seconds which I recorded in chapter 3. This is likely due to the large number of variables which were introduced for over-dispersion and for 6 times as many sites and the increased amount of data which was used for evaluating the loss function.

The end-to-end inference time for five rounds of annealed SVI was 8 days and 7 hours, with the use of a High-Performance Computing (HPC) cluster and a GPU machine. Most of the inference time was taken simulating $90,000 + 4 \times 10,000$ samples of the ABM with 500 CPUs cores on the HPC. Samples took on average 34 minutes 57 seconds to execute, requiring approximately 6 days. Training the surrogate for 100 epochs required, on average, 2 hours and 40 minutes on a single Nvidia A100 GPU. The remaining time was spent training the variational model and sampling the posterior, which required on average 3 hours 48 minutes, with the same GPU.

5.4 Discussion

In this chapter, I presented a Bayesian inference of parameters for a complex dynamic model of malaria which is a significant advancement over previous attempts. Unlike the last effort which was conducted a decade ago, which assumed malaria transmission was in equilibrium, this study incorporates the temporal dynamics of malaria transmission in the face of now ubiquitous interventions, including ITNs, IRS and SMC. This comprehensive approach facilitates a more realistic representation of the current challenges in malaria control and eradication efforts.

5.4.1 Surrogate approximation quality

The evaluation of the surrogate model has revealed several important insights into its performance and capabilities. Firstly, the surrogate model provided low Mean Squared Error (MSE) of estimates relative to the simulated population size. This suggests that the surrogate model is effective at capturing the underlying dynamics of the Agent Based Model (ABM) and can serve as a reliable tool for producing rapid high-dimensional, probabilistic trajectories for model parameters which are drawn from their prior distributions.

Among the various model outputs, incidence was the best approximated by the surrogate model. The likelihood of ABM incidence values as modelled by the surrogate are especially high in older age groups and in later months of the simulation. The performance in these areas can be attributed to the more straightforward dynamics present in the majority of trajectories for these groups. As individuals age and the simulation progresses, the dynamics often become less complex as malaria becomes endemic or the age groups become very small leading to near zero incidence, making them easier for the surrogate model to capture accurately.

The surrogate had mixed success in learning temporal dynamics. One limitation in this regard was the surrogate model's handling of seasonality in the output trajectories. Instead of accurately capturing periodic dynamics, the model sometimes represented seasonality as variance around the

mean output over time. This limitation could impact the ability to account for seasonal trends in inference. Despite this, the surrogate was able to learn the growth of demographics within the simulated population.

The calibration plots shown in [Figure 5.5c](#) suggest generally well-calibrated probability densities when approximating prevalence and population counts, and miscalibrated approximations for incidence. This discrepancy could be attributed to the surrogate’s (in)ability to recognise temporal dynamics, as previously mentioned, or due to the potentially restrictive assumption of Normally distributed outputs. To investigate this issue, one could test the adherence of ABM outputs to the Gaussian distribution by performing z-tests on simulated trajectories. If the null hypothesis fails, the surrogate could benefit from more flexible density estimation or re-calibration methods such as those presented in Papamakarios (2019) and Kuleshov, Fenner, and Ermon (2018).

5.4.2 Parameter inference

The results from the posteriors estimated using synthetic data show that the surrogate model’s estimates were generally close to the true values for most immunity parameters. Exceptions included the parameters b_0 and r_U , which deviated from the true values. These discrepancies suggest that specific parameters may be more challenging to estimate accurately even after accounting for surrogate approximation error and may require more sophisticated posterior sampling techniques to improve the estimation.

Transmission intensity estimates, represented by the Entomological Inoculation Rate (EIR) parameter, showed a wide variability. This wide range indicates potential identifiability issues, likely due to the lack of sufficient data or small sample sizes. Unlike in (Griffin, Ferguson, and Ghani, 2014), where more informative priors were used for EIR, I found that data on EIR were unreliable and inconsistent in the data published in (Battle, Guerra, et al., 2015). The variability in measurement methods and the lack of standardisation in recording the number of infectious bites led to my decision to model EIR with an uninformative prior.

For the (ibid.) data, the initial prior predictive distributions, informed by previous estimates, tended to overestimate prevalence and incidence. The posterior estimates for immunity parameters showed significant differences from previous estimates. The new estimates suggested a more modest effect of pre-erythrocytic and clinical immunity, aligning better with the empirical data and the current scientific consensus.

The Bayesian p-value plots indicated that the new estimates fit the observed data well, provided that the surrogate model maintained a reasonable approximation quality. The surrogate posterior trajectories captured various malaria burden trends over time, including stark declines likely resulting from intervention measures. These results suggest that the model can effectively reflect the impact of interventions for parameter estimates.

I identified some observations, such as the prevalence of $>90\%$ in Morogoro, Tanzania, as outliers. Finding outliers in large databases of highly variable prevalence and incidence data remains a challenge and it is important to develop robust outlier detection and handling mechanisms. Nevertheless, overall high observations were better fit with the new immunity estimates, demonstrating the model’s improved performance with revised parameter estimates.

However, many observations fell outside the 95th percentile of posterior estimates, suggesting that the posterior estimates could benefit from a wider variance to capture the full range of real-world variability. Future work should focus on refining the model to increase its robustness and accuracy, particularly in capturing the more extreme observations.

5.4.3 Computational performance

The surrogate model’s fast execution time was important in making parameter inference feasible but also allowing me to perform diagnostics and explore the parameter space in a timely manner. This rapid simulation allowed for extensive analyses that would be impractical with less flexible surrogates, facilitating a deeper understanding of the model dynamics and the influence of prevalence and incidence trajectories on parameter estimates.

The overall parameter inference process required under 9 days, with single rounds of inference requiring less than 4 hours. Such time frames are acceptable given that these parameters need not be estimated frequently. Once published, these parameter estimates can be reused for site calibration and planning intervention scenarios.

One of the major bottlenecks to inference is the sampling of the ABM, which requires substantial HPC resources to simulate enough samples within a reasonable time frame. To address this bottleneck, it would be worthwhile investigating methods for more efficient sampling of the ABM, particularly focusing on areas of the parameter space which would be most informative for posterior sampling.

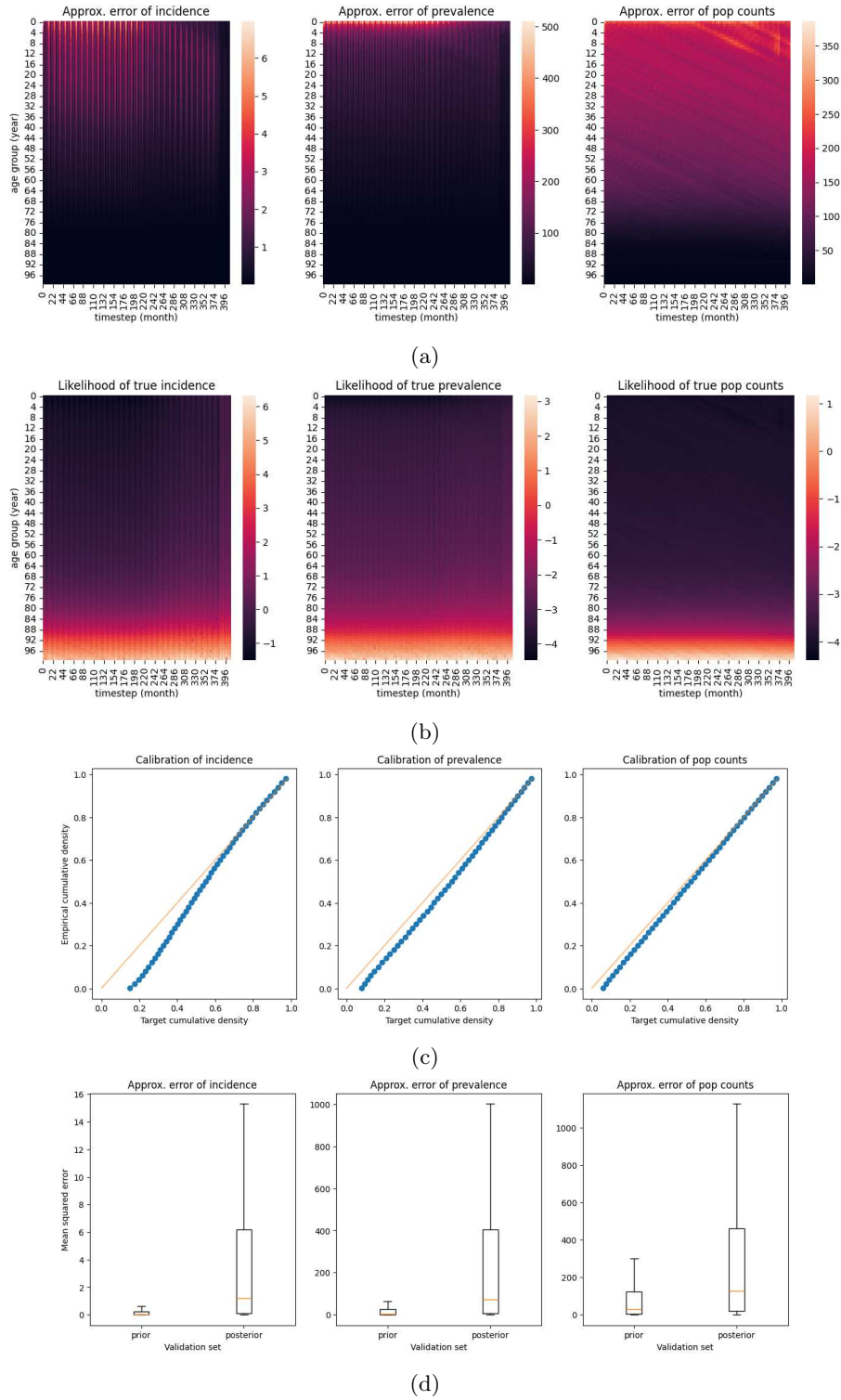


Figure 5.5: Approximation quality of density estimating ABM surrogate. Incidence outputs represent new clinical cases per month. Prevalence outputs represents the number of agents who would be detected by a microscopy test. "Pop count" outputs represent the number of agents in the corresponding age group. All outputs were collected for 1,000 out-of-sample model parameterisations with a simulation population size of 10,000. The samples for Figure 5.5a, Figure 5.5b and Figure 5.5c were drawn from the prior distribution. Figure 5.5a shows the average Mean Squared Error (MSE) of each point in the mean surrogate trajectories. Figure 5.5b shows the average probability density of each point in the out-of-sample trajectories. Figure 5.5c shows the calibration error of the empirical cumulative density of surrogate outputs. Figure 5.5d shows the validation MSE of the prior distribution compared to the estimated posterior.

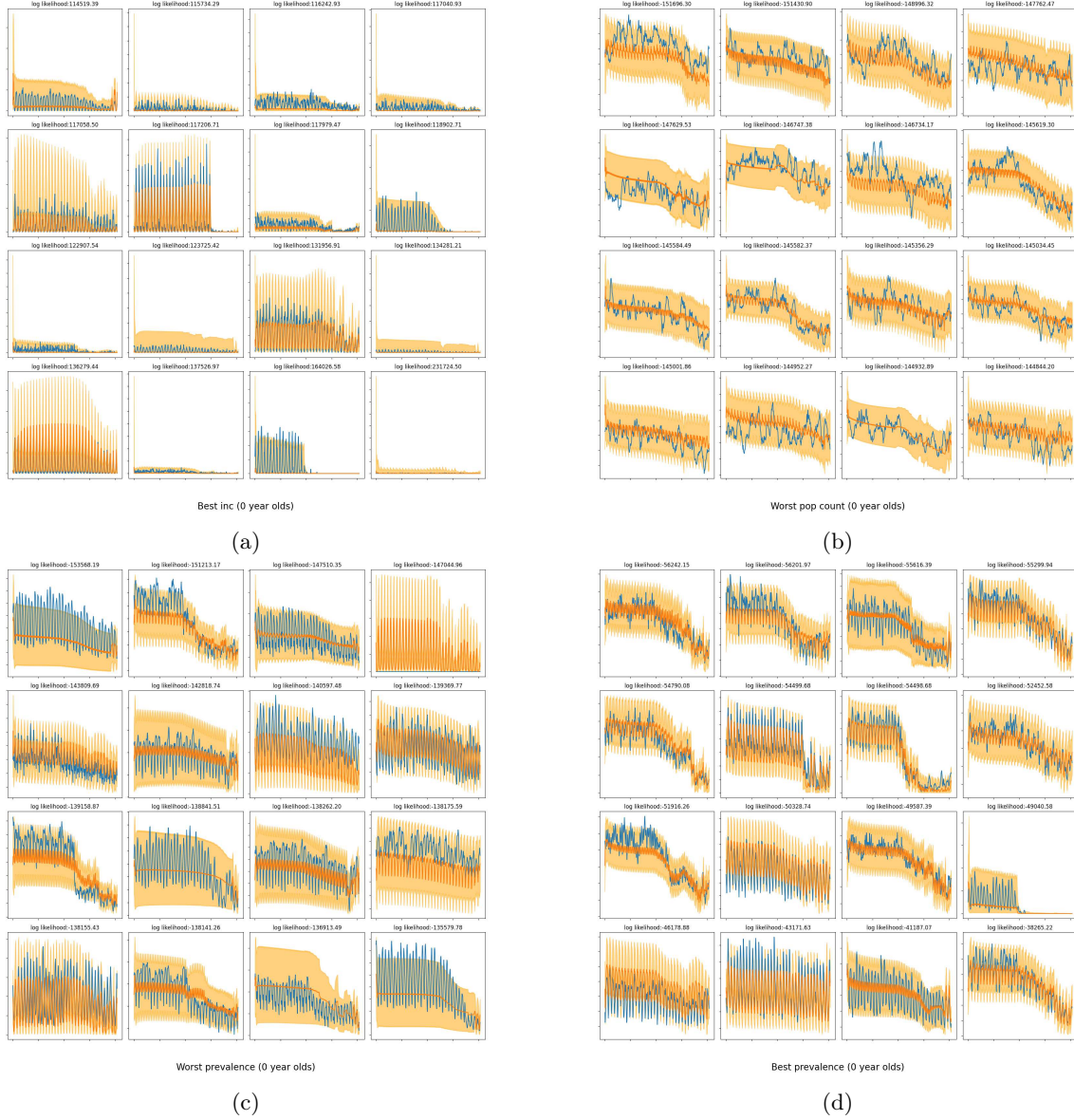


Figure 5.6: Qualitative analysis of density estimating surrogate trajectories. Each panel shows a comparison of model trajectories in blue with the surrogate estimate represented by the mean with the solid orange line and the 95% confidence interval shaded in orange. Each trajectory was taken from a validation set constructed of 1000 out-of-sample draws from the ABM model parameters' prior distribution and corresponding output trajectories. The "Best" and "Worst" trajectories were ranked by MSE or likelihood, as indicated by the metrics above each plot.

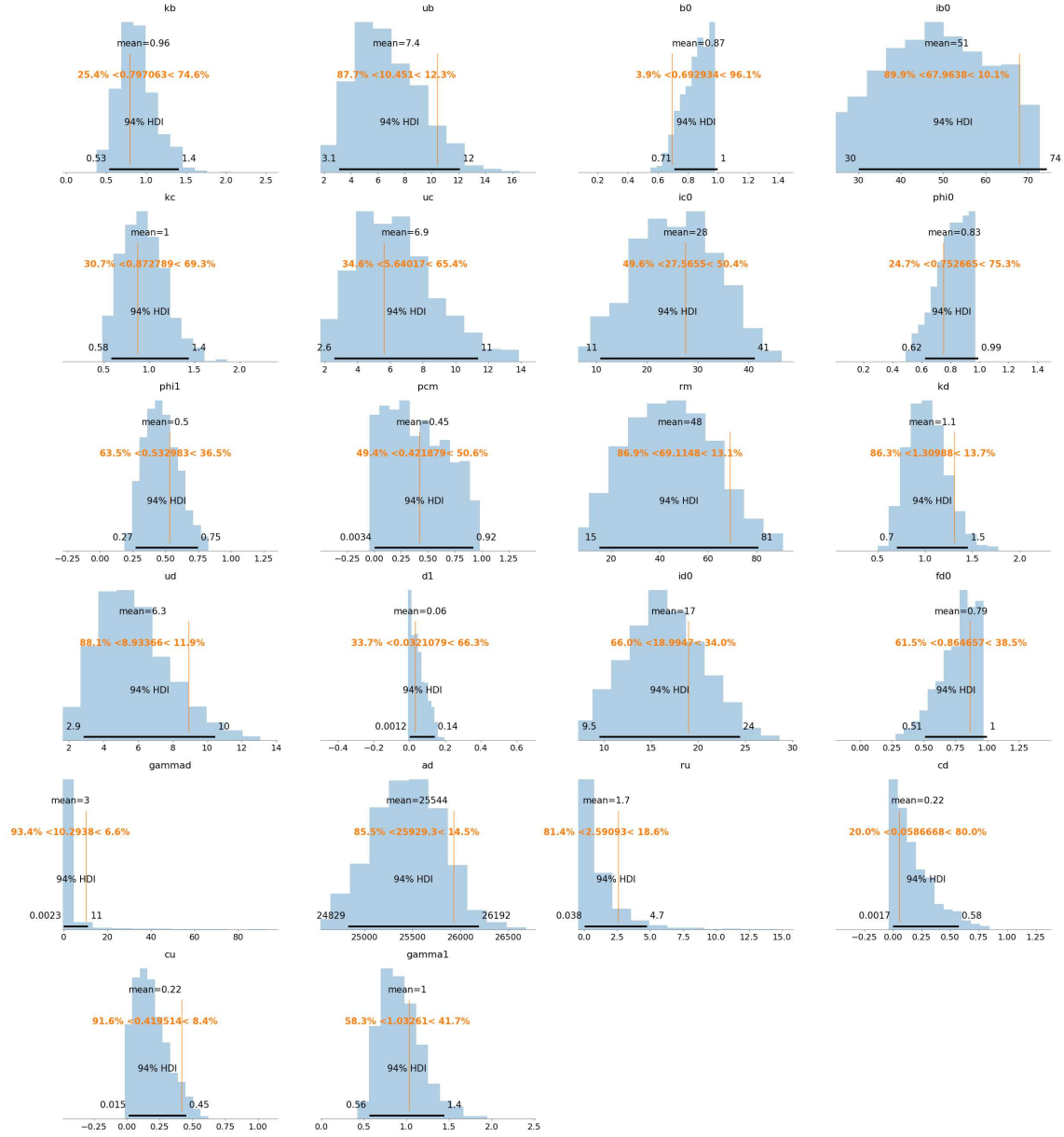


Figure 5.7: Posterior estimates for immunity parameters based on synthetic observations. These are histograms of the marginal density of each immunity parameter from 1000 posterior draws. The true parameter is marked with the orange vertical line if it is within the scale of the plot.

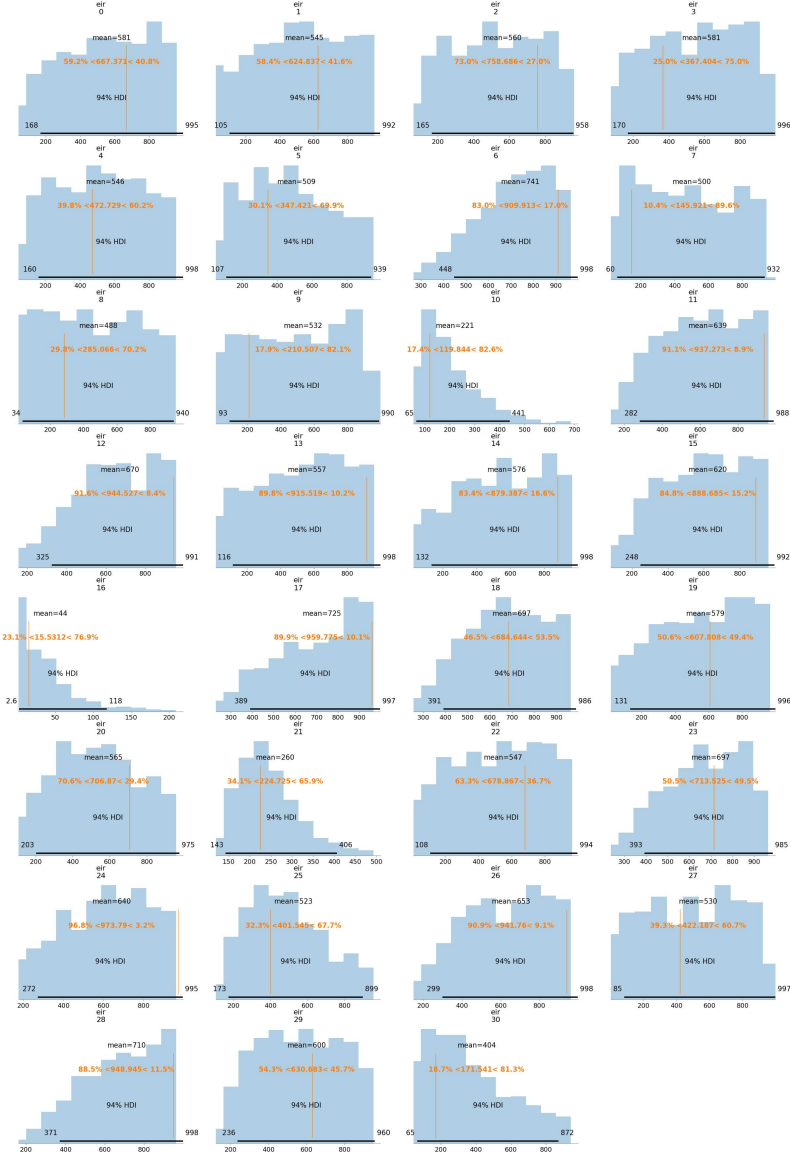
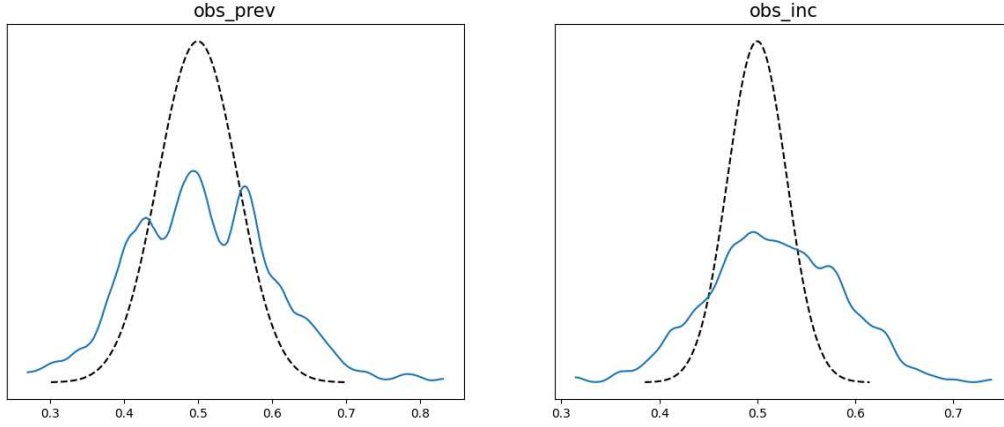
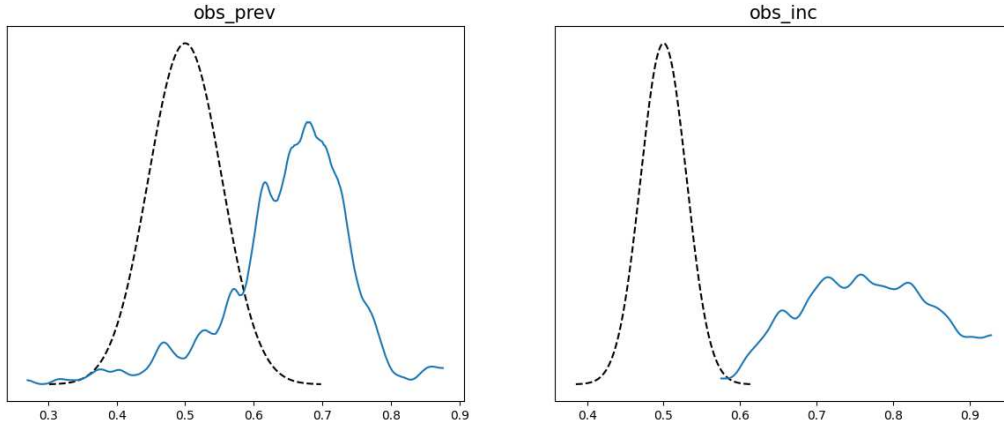


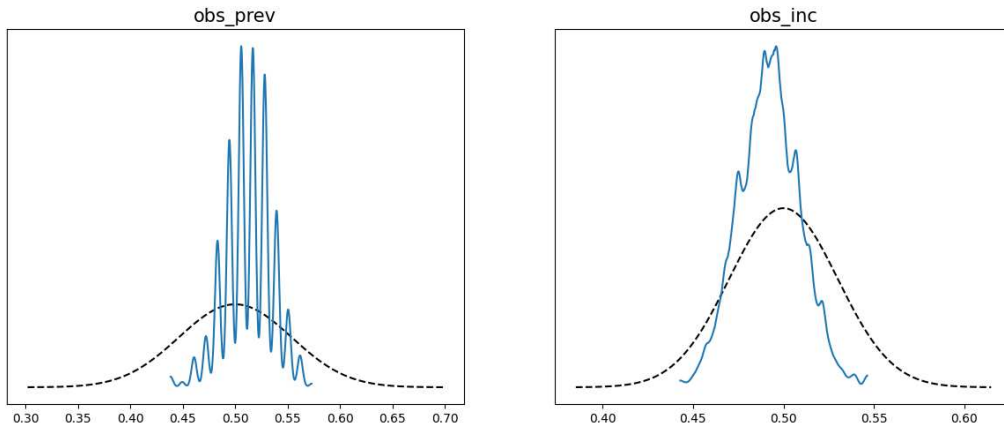
Figure 5.8: Posterior estimates for site transmission parameters based on synthetic observations. The "eir" parameter for each site represents the Entomological Inoculation Rate (EIR) which is assumed when initialising the ABM to an equilibrium state as is described in [section 4.2.2](#). Since this does not account for seasonality or interventions, it does not relate to the EIR which is realised during a simulation but is instead a proxy for baseline malaria transmission.



(a) Bayesian p-value plot of estimated posterior on synthetic data



(b) Bayesian p-value plot of prior distribution on data from (Battle, Guerra, et al., 2015)



(c) Bayesian p-value plot of posterior estimate on data from (Battle, Guerra, et al., 2015)

Figure 5.9: Bayesian p-value plots for surrogate based inference of ABM parameters. The blue line plots the empirical probability that the observations are less than the predictive distribution given the corresponding parameter estimates. The ideal curve is plotted in black.

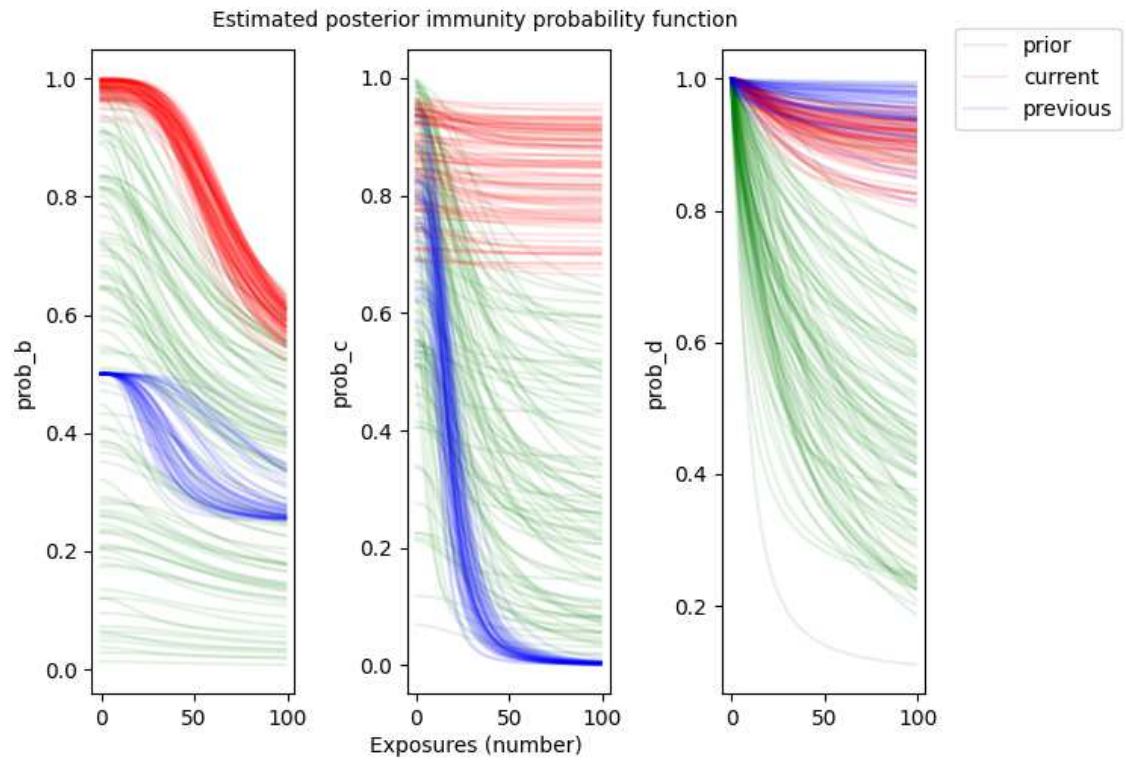


Figure 5.10: The immunity functions for prior and posterior estimates of ABM parameters. prob_b, prob_c and prob_d refer to the probability of pre-erythrocytic infection, clinical infection and detection by microscopy respectively. The green lines represent response curves drawn from the prior distribution, blue from posterior estimates published in (Griffin, Ferguson, and Ghani, 2014), and red from my estimated posterior.

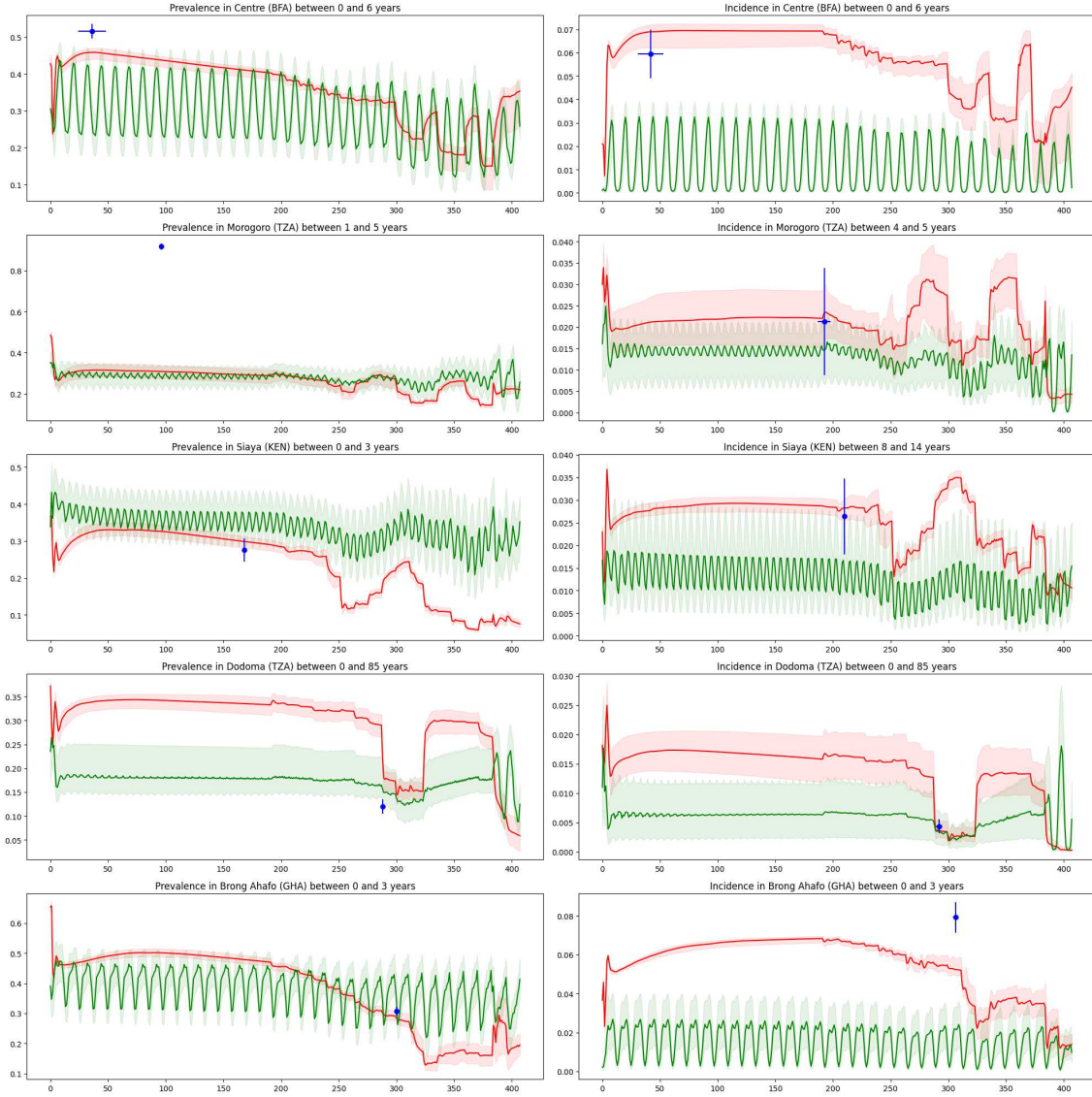


Figure 5.11: Prevalence (left) and incidence (right) trajectories for selected sites sampled from the prior (in green) and estimated posterior (in red). The solid lines represent the median value and shaded regions represent the range between the 5th% and 95th% quantiles. The blue points represent prevalence and incidence estimates from the related observation. The vertical error bars for each observation represents the 95% confidence interval, approximated using a Bernoulli distribution for prevalence and Poisson distribution for incidence. The horizontal error bars represent the duration of the study.

Chapter 6

Discussion

Parameter inference for complex models, especially agent-based models, poses significant challenges. These models often require long simulation times and may not be auto-differentiable, limiting researchers from some gradient-based inference methods. Increasingly, researchers would like to perform inference on models with high-dimensional parameter spaces and intricate dynamics which can confound many inference methods. Models with stochastic outputs with intractable densities are all too common and make inference even more challenging, as researchers have to consider how to approximate their likelihood function.

To rise to these challenges, it is important to develop methods that improve the speed, robustness, and reliability of inference. Speed makes the inference of detailed, long-running models tractable, allowing researchers to obtain results in a reasonable time frame. Robustness mitigates poor approximations which lead to estimates which either fail to fit existing data adequately or lack predictive performance. Methods should produce probabilistic parameter estimates, capturing the uncertainty in the observations and the model parameters, and include diagnostics to explain complex estimates and ensure reliability.

In this thesis, I have explored various methods to improve parameter inference for a complex malaria model. These methods include derivative-based Markov Chain Monte Carlo (MCMC), vanilla, autoregressive, and density-estimating neural network (NN) based surrogates, as well as normalising flow-based variational models for Stochastic Variational Inference (SVI) applied to real-world inference problems. Each method has been evaluated to understand its implications in the context of a complex malaria dynamical transmission model, providing insights into their effectiveness and limitations in the field of infectious disease modelling.

To address the shortcomings I found, I investigated various mitigation strategies. These included annealed Stochastic Variational Inference (SVI) to improve convergence, subsampling techniques to scale SVI for larger datasets, and computational methods to balance the trade-offs between the accessibility and efficiency of forward simulations, gradient estimation, and posterior estimation. I also have developed diagnostics to enhance the reliability and interpretation of the results.

6.1 Summary of contributions and findings

In [chapter 2](#), I investigated the recurring problem of estimating model parameters for modern malaria transmission models from prevalence and inference data (Griffin, Hollingsworth, et al., 2010; Griffin, Ferguson, and Ghani, 2014; Cameron et al., 2015; P. A. Eckhoff, 2011; Reiker, Golumbeanu, et al., 2021). These estimates shape our understanding of the relationship between malaria prevalence and incidence and the dynamics of immunity.

Through a parameter tuning exercise, I demonstrated the impact it has on capturing up-to-date knowledge and its effects on posterior estimates. In large parameter spaces, careful tuning can reveal issues in parameter estimates which standard diagnostics might miss. I developed diagnostics which help interpret the parameter space for prior tuning and understand the implications of posterior estimates. These diagnostics also can also diagnose potential issues during inference. The diagnostics included prevalence and immunity response curves, gradient-based sensitivity and correlation analyses. Though implemented on a synthetic inference problem, these diagnostics were integral for validating inference when a source of truth is not available or in surrogate based

inference scenarios where it is less feasible to diagnose inference issues with numerous forward simulations, as is the case in chapters 3 and 5.

JAX proved to be an effective framework for implementing an auto-differentiable model and optimising forward simulation, inference, and sensitivity analysis. JAX’s design for general numerical programming simplified the implementation of the complex and recursive bi-stable equilibrium solution for malaria transmission. The primitives for vectorisation, looping, and control over auto-differentiation methods, particularly forward-mode, helped me use computational resources efficiently. Furthermore, I found the No-U-Turn Sampler (NUTS) to be an effective posterior sampler, showing strong convergence after only 100 warmup samples. This is the first instance of NUTS being applied to a complex malaria model, setting a precedent for future research in this area.

In [chapter 3](#), I explored various surrogate-based inference methods for the same parameter inference problem. These approaches aim to allow researchers to use a NUTS sampler by training a Neural Network (NN)-based surrogate instead of requiring an auto-differentiable implementation for gradient estimation, which is less feasible for large and complex ABMs. However, I also considered Stochastic Variational Inference (SVI) methods, instead of NUTS, to sample the posterior from a likelihood derived from surrogate outputs.

Among the surrogate-based methods, I found that Annealed SVI was the most effective strategy. Annealed SVI produced more reliable approximations after each round of inference and yielded a more accurate fit and better predictive performance after the final round. The training process for the SVI methods in general converged rapidly, within approximately 20 minutes and 50,000 samples of the surrogate. My diagnostics consistently indicated a strong agreement between the posterior estimates and the latent true parameters.

Conversely, SVI without annealing may be unreliable in large parameter spaces due to approximation errors in the surrogate model combined with an overconfident variational model, which can lead to mode collapse. This makes sequential sampling ineffective as the posterior estimates in each round fail to cover the true posterior, causing erratic convergence behaviour. NUTS also showed limitations when sampling from NN-based surrogates which emulate high-dimensional and complex model dynamics. The surrogate produced non-smooth and inconsistently approximated outputs, and therefore NUTS was less robust than SVI to noisy posterior estimates, often getting stuck in local optima. Extending the warmup period did not ameliorate this issue, and although increasing the sample size somewhat improved NUTS convergence, it remained far less sample-efficient than the other methods considered.

An informative prior proposal distribution greatly influenced inference quality compared to commonly used Latin Hypercube Sampling (LHS) proposals which often appear in the literature. Prior proposals led to greater agreement with the reference posterior, lower losses with (Annealed) SVI methods, and better approximations on validation samples drawn from the prior and estimated posterior. However, this performance is predicated on priors being consistent with the observed data, as was the case in the synthetic scenario I considered. If not, the surrogate would be trained using a proposal distribution that is not informative for explaining the observed data, which can be caught in prior predictive checks. In such cases, training using an LHS proposal can be beneficial for tuning priors.

The formulation of the surrogate can also enhance training efficiency, flexibility, and diagnostics. By decomposing the likelihood function’s evidence into independent sites, the surrogate can be trained to recreate single simulations, reducing the parameter space it needs to emulate. Neural network surrogates were effective in producing high-dimensional outputs - when trained on the prior proposal - including contextual outputs which were not necessarily relevant to a specific inference problem, thereby aiding diagnostics. This formulation allows the same surrogate to be reused when new observational data becomes available or applied to different inference problems, reducing the computational demands of surrogate training. High-dimensional outputs are less feasible with surrogates previously applied to malaria model inference, such as Gaussian Processes (GPs) and kernel methods.

In [chapter 4](#), I applied Neural Network (NN)-based surrogates to a simpler inference problem, the site calibration problem, and successfully sampled the posterior using the No-U-Turn Sampler (NUTS), contrary to the findings in the previous chapter. Site calibration involves inferring far fewer parameters, making the surrogate approximation more reliable. This is the first time NUTS has been applied to an agent-based model (ABM) of malaria transmission.

Consequently, researchers can align observed malaria prevalence estimates with an ABM of

malaria transmission within two minutes, produce a probabilistic estimate of their model parameters and rely on good MCMC convergence. This could dramatically change global malaria modelling research, as the calibration of multiple sites could be inferred on a desktop machine in minutes instead of hours on a high-performance computing (HPC) cluster. Additionally, the probabilistic estimates can be used to communicate the uncertainty in the ABM model outputs, which is preferable to the point estimates typically inferred in large-scale ABM studies due to computational constraints.

The NN surrogate used in this study had an autoregressive formulation, which allowed it to simulate time-varying outputs. This makes it more flexible than other surrogates which are typically designed for a single inference problem, as it could potentially be used for any site calibration task. This approach could offer further computational savings as the surrogate can be trained on shorter trajectories to still learn the general temporal dynamics of the ABM. This is supported by my benchmarking results, which indicated that the surrogate could extrapolate to estimate trajectories twice as long as those in the training set.

For the underlying ABM, I developed the first modern malaria model with an extensible, high-level R interface, "malariasimulation". Malariasimulation supports similar features to other malaria models regarding interventions and site parameterisation. Despite a significant portion of the execution taking place within the R interpreter, it maintains a competitive execution time—30 minutes per simulation in Chapter 4 and 23 minutes per simulation in Chapter 5. This efficiency is largely due to the ABM development package I created, called "individual," which optimises common ABM operations in C++ while providing an R interface for rapid development and experimentation.

The "individual" package simplifies the development of time-critical ABMs, making the process easier and more accessible to a broader range of researchers. Its potential for encouraging rapid ABM development is made clear by its application to modelling the implications of vaccine distributions for COVID-19 (Hogan et al., 2023). The combination of an accessible, powerful and efficient computational tools paves the way for more effective and timely global health research, potentially leading to better-informed intervention strategies and improved health outcomes.

In chapter 5, I developed the most sophisticated surrogate presented in this thesis. This surrogate recreated time-varying trajectories similar to those in Chapter 4 but included a much broader range of outputs. These included prevalence and incidence for 100 age groups, as well as univariate trajectories for immunity, Entomological Inoculation Rate (EIR), and population counts. Moreover, unlike the point estimates provided in Chapter 4, the outputs in this chapter were probability distributions designed to account for the stochastic nature of ABM outputs.

The additional outputs made the surrogate more flexible, allowing for inference on observations of prevalence and incidence across any age range (as long as the boundaries are integers and less than 100). This flexibility also enables in-depth diagnostics and parameter exploration by allowing one to rapidly simulate rich multivariate trajectories, which can be invaluable for prior tuning.

I used an SVI workflow to estimate parameters for the ABM from real world observations. This approach relaxed several assumptions from the most recent parameter inference on this ABM by (Griffin, Ferguson, and Ghani, 2014). Griffin et al. assumed that all sites were in bi-stable equilibrium state of transmission and that no interventions were taking place, apart from the treatment of clinical malaria. These assumptions limited the available data to pre-intervention scenarios, which are older, geographically limited and not reflective of the current state of malaria transmission. In contrast, I used a more comprehensive dataset which included recent prevalence and incidence estimates and accounted for various interventions. By incorporating modern data and intervention effects, the surrogate provides a more realistic and applicable model for current malaria scenarios.

The flexibility and scalability of neural networks (NNs) are particularly advantageous for inference on modern transmission models, where the complexity and volume of our models and data continue to grow. By using NNs, we can perform more robust and comprehensive analyses and diagnostics, without suffering from the computational bottlenecks associated with traditionally used surrogates including kernel methods and Gaussian Processes.

The resulting estimates from the NN surrogate updates our understanding of the transmission dynamics in the agent-based model (ABM) with respect to real-world data. These estimates fit well with the updated observational data, suggesting it could improve the accuracy and relevance of ABM predictions in current malaria transmission scenarios.

The estimates suggest a more modest increase in acquired immunity to pre-erythrocytic and

clinical infection with exposure and a steeper increase in immunity to detection, which aligns with recent scientific findings on the development of immunity in malaria. This refined understanding has implications for our understanding of the relationship between prevalence and incidence of malaria.

The workflow developed in Chapter 5 demonstrates significant progress in inference capabilities. Its ability to generate detailed probabilistic outputs for a wide range of parameters and scenarios represents a substantial improvement over previous methods. This progress could not only improve the model’s predictive power but also provide tools for researchers to conduct more thorough and informed analyses of malaria dynamics and interventions. Beyond malaria, this approach could be applied to other complex infectious disease models for timely and robust parameter inference.

6.2 Limitations and future directions

In [chapter 2](#), I found certain limitations to the feasibility of Markov Chain Monte Carlo (MCMC) for inference. Drawing 100 samples from the posterior took an hour, despite the bi-stable equilibrium solution itself requiring only milliseconds for an execution. The No-U-Turn Sampler (NUTS) required several executions and derivatives of the model to construct its binary tree of samples for each step. For this specific inference problem, the complex posterior likely required a large binary tree before a U-turn was detected, leading to the long execution time.

This does not bode well for agent-based models (ABMs), which typically require hours for forward simulations and are not usually auto-differentiable. Nevertheless, extending auto-differentiation frameworks to support ABMs, as is currently being researched ([Chopra et al., 2023](#); [Quera-Bofarull et al., 2023](#); [Krieken, Tomczak, and Teije, 2021](#); [Arya et al., 2023](#)), remains a worthwhile endeavour. Despite the long execution times, the exact inference presented in this chapter proved to be the most robust method in this thesis.

My inference diagnostics, particularly sensitivity analysis, revealed that many model parameters were not sensitive to incidence or prevalence. This suggests that a reformulation of the model, either by reducing the number of parameters or improving the identifiability of these parameters, could simplify the model or improve the quality of inference.

Other popular MCMC kernels in this research space, such as Random Walk Metropolis-Hastings (RWMH), Parallel Tempering (PT), and Metropolis-Adjusted Langevin Algorithm (MALA), were not considered in this study as they were not natively supported by NumPyro, the framework I used for probabilistic inference. Although some kernels, like RWMH and PT, had limited support through TensorFlow Probability, I encountered numerous technical issues when attempting to use them. Given that the inference was successful and NUTS is regarded as a robust kernel, I did not pursue a comparative analysis of different kernels. However, for future research, one might consider PT for handling multi-modal posteriors.

In [chapter 3](#), despite some surrogate-based inference methods successfully recovering the true parameters and demonstrating a degree of predictive capability, all approximate posterior exhibited noticeable discrepancies when compared to the reference posterior. Even Annealed SVI showed large differences in the highly sensitive clinical immunity parameters. The approximated posterior predictive samples were at most half as likely than those drawn from the model using the reference posterior. Additionally, none of the methods showed clear improvement in predictive performance with sequential sampling, even after five rounds.

I suspect that the primary cause of these issues is the quality of the surrogate’s approximation in the estimated posterior distribution. To address this, it would be worthwhile investigating SVI workflows without the use of a surrogate. This would involve training the variational model directly on samples from the underlying model, as the derivatives required for NUTS are not necessary for training a variational model. However, this approach may require a greater number of samples from the underlying model, potentially increasing computational demands. Additionally, exploring surrogate methods that account for surrogate approximation error or uncertainty, similar to approaches used in Bayesian Optimisation ([Gramacy, 2020](#)), could lead to better convergence behaviour for complex inference problems like the one studied here. Research into uncertainty quantification for neural networks is progressing ([Lakshminarayanan, Pritzel, and Blundell, 2017](#); [Kristiadi, Hein, and Hennig, 2021](#); [Gal and Ghahramani, 2016](#)), but it has not yet been applied to surrogate based inference. Integrating these advancements could improve the robustness and accuracy of these methods.

The annealing schedule used for the Annealed SVI method was determined through trial and error. Manually devising these schedules is notoriously challenging (Huang et al., [n.d.](#); Mandt et al., [2016](#)), as the hyper-parameter space could consider all possible temperatures per parameter per training step and optimal schedules may also be influenced by the learning rate schedule for Stochastic Gradient Descent - which in my experiments was kept constant. Automated methods for determining annealing schedules exist, though they have not yet been applied in surrogate-based inference methods. Implementing these automated methods could potentially lead to improvement through sequential sampling.

Despite the challenges presented by noisy surrogates, it is still possible for MCMC methods to work in high-dimensional NN-surrogate based inference workflows. Overcoming these issues may involve experimenting with methods for learning smooth functions with neural networks (Anil, Lucas, and Grosse, [2019](#)), Lipschitz networks, or applying kernels which are robust to non-smooth functions (Chaaari et al., [2015](#)). However, Lipschitz networks are known to be more challenging to train than traditional neural networks since they constrain the gradients of NN operations either architecturally or with optimisation constraints. These extensions could lead to viable alternatives to SVI.

The workflows suggested in Chapter 3 are analogous to Sequential Neural Likelihood Estimation (SNLE) and Sequential Neural Variational Inference (SNVI) workflows, which have recently been published for likelihood-free inference. It is plausible that these methods would encounter similar convergence issues when applied to more complex models beyond the relatively simple benchmarks, such as dual moon, M/G/1, Lotka-Volterra, which they are typically tested with. To confirm these potential issues and advance research, benchmarks need to be developed that specifically target more complex models to guide the development of more robust black-box inference methods.

Sequential Neural Posterior Estimation (SNPE), for instance, is less flexible and offers different trade-offs to likelihood-estimating alternatives. SNPE offers no control over the formulation of the likelihood, which in the case of epidemiological models is attractive, researchers often have an observational model derived from the study design which they may want to introduce into the likelihood formulation, even if they do not have a formulation for the probability density of model outputs. For example, in chapter 5, it is intuitive to assume that my prevalence and incidence observations are distributed according to over-dispersed Binomial and Poisson distributions, even though it is intractable to calculate the probability of my ABM's prevalence and incidence outputs given a specific parameterisation. SNPE is also reported to struggle with biases which are introduced by the proposal distribution. In each round of posterior sampling, researchers must take care to adjust the loss function during the training of the variational model to account for the fact that the proposal distribution will stray further and further from the prior (Papamakarios, Sterratt, and Murray, [2019](#)).

I also did not consider many other simulation-based inference methods in this chapter. I did not compare my methods to more commonly used Gaussian Process (GP) based surrogates due to their reported scalability issues (Reiker, Golumbeanu, et al., [2021](#)), particularly when dealing with high-dimensional parameter spaces. Also, I did not compare sequential sampling to history matching (McKinley et al., [2018](#)). History matching may be more robust to surrogate approximation errors as it uses an independent "implausibility" metric to assess the fit of the model to data. However, it remains unexplored how well history matching will scale to high-dimensional problems, which is an consideration models used in malaria transmission studies.

There are countless ways to formulate a surrogate model, and exploring these alternatives could lead to improvements over those which I considered in this chapter. For example, one might devise alternative representations of model outputs or apply transfer learning or contrastive techniques to address the issues associated with learning different types of outputs (in this case prevalence, incidence, population proportions, and various types of immunity values). Surrogates could benefit from learning these outputs jointly, however using a homogeneous loss function for these heterogeneously distributed outputs with varying impacts on the final inference could have contributed to the poor approximation quality which was exhibited.

Before using surrogates in [chapter 4](#) for site calibration more broadly, it is advisable to extensively test the quality of inference across multiple sites. I selected the Kolda, Senegal site as a case study due to large number of prevalence surveys available in the Demographic Health Surveys (DHS) data set. However, alternative datasets, such as randomised controlled trials (RCTs) or other prevalence data repositories, may contain different types of time series and sample sizes. Additionally, various sites will require different parameterisations, as visualised in [subsection 5.2.2](#).

My validation tests which consider the approximation quality suggest that the surrogate should perform well in most scenarios, with the exception of low transmission sites where it tends to overestimate prevalence. Interestingly, this overestimation did not affect the inference for Kolda, which has relatively low transmission. Distinguishing which low transmission scenarios are well approximated remains challenging due to the inherent opaqueness of neural networks. To address this, one could investigate methods to visualise the loss landscape (H. Li et al., 2018), apply regularisation techniques to improve generalisation, or analyse the effect of increasing the number of samples. Investigating the approximation quality with surrogates of ABMs is computationally intensive, which is why I carried out the convergence analyses on the bi-stable equilibrium model in chapter 3.

Another drawback of the proposed method is the assumption that all of the model parameters which are not specific to the particular site are assumed to be the mean estimates. As discussed in chapters 2, 3, and 5, considerable effort is devoted to generating probabilistic estimates for these parameters. Currently, there is limited research on combining or modularising estimates from different inference problems using MCMC. Exploring alternative methods, such as Sequential Monte Carlo (SMC) or combining variational models for SVI, could be viable alternatives. These approaches could help quantify the predictive uncertainty of all model parameters jointly after site calibration.

The framework I created for agent-based model (ABM) development is still in its early stages and may not provide all its benefits across all epidemiological models. For example, one challenge is that simulations can experience slowdowns when variables are copied from the C++ runtime into the R interpreter for manipulation. This occurs when a user needs to implement an operation in R which does not yet have an optimised C++ implementation in the 'individual' package. For example, I have not provided optimised operations for network-based modelling, such as calculating contacts within a sub-population.

There are also many opportunities for improving the usability of 'individual'. Currently, it lacks high-level operations analogous to those provided by 'dplyr' for dataframes, which make programming ABMs more accessible to beginners. For instance, creating a seamless interface between 'individual's bitsets and native R vectors could improve the user experience. Some users struggle with non-native data structures and the reference semantics I introduced for performance reasons. A more opaque interface could automatically translate between bitsets and vectors as needed, accommodating users who prefer not to concern themselves with performance implications.

Additionally, the framework could benefit from primitives for resolving competing hazards, which would reduce boilerplate code and discourage monolithic processes. Processes in 'individual' are designed to run independently, making it challenging to modularise dynamics which influence the same variables, which are called competing hazards. An example of a competing hazard is the conflict between re-infection and recovery processes, since both may manipulate the health state of an agent in the same time step. Designing primitives which indicate competing hazards could allow 'individual' to manage conflicting updates to the same variable, allowing users to define more fine-grained and maintainable processes.

Malariasimulation has since evolved into a community project, with ongoing work to add functionality for various interventions and model extensions. These include spatial meta-population models and P. Vivax modelling. With such significant and impactful changes, careful consideration and design are required to maintain performance and encourage further contributions in future versions of the model.

In chapter 5, while the surrogate approximation showed significant advancements, there remains room for improvement to make parameter estimates and diagnostics more robust. One important area is in training the surrogate to distinguish between seasonal variation and the stochasticity in the ABM's outputs. Addressing this limitation would impact parameter estimates as the model would accurately reflect seasonal patterns in malaria transmission in the observational period for the data used in inference.

The parameter estimates could also benefit from a more robust sequential inference approach. As observed in Chapter 3, sequential inference can fail for complex inference problems due to a poorly approximated posterior. There are several potential avenues for resolving this issue. Investigating appropriate annealing schedules could improve approximation between rounds lead better convergence towards the true posterior. Alternatively, incorporating uncertainty into surrogate predictions using Bayesian Neural Networks (BNNs) (Lakshminarayanan, Pritzel, and Blundell, 2017; Gal and Ghahramani, 2016; Kristiadi, Hein, and Hennig, 2021) could account for approxima-

tion error more explicitly. Another promising avenue is to explore methods for targeting sampling towards informative areas of the parameter space which are robust to surrogate approximation errors. Techniques such as History Matching (McKinley et al., 2018) have not yet been tested on high-dimensional inference problems.

Many parameters for the ABM are still assumed as point estimates, which could limit the degrees of freedom for parameter inference, which is especially important since the posterior predictive often failed to cover observations. Specifically, all ABM site parameters were treated as point estimates, which does not account for the variability and uncertainty inherent in real-world conditions. For instance, the site parameterisation for rainfall patterns which drive the ABM’s intra-year variation in mosquito density is generally far from certain.

To address this limitation, new data and modelling approaches should be investigated to better capture and incorporate this uncertainty. By moving beyond point estimates and incorporating probabilistic estimates for site parameters, one could potentially improve the model’s predictions and allow for more nuanced and informed decision-making in malaria control efforts.

Additionally, only the mean surrogate outputs are used for the likelihood calculation in the current framework, despite evidence that the variation in outputs capture seasonality. This decision was made for computational reasons, as modelling probability distributions over entire trajectories would significantly increase the computational burden, especially for the variational model which was used - a Block Neural Autoregressive Flow. It is therefore worth investigating workflows which could incorporate the large number of probabilistic outputs from the surrogate, perhaps either through parallelism or a more efficient formulation.

Though I have made reasonable efforts to pre-process observational data, more rigorous outlier detection, management, and investigation would be beneficial. While filtering the large dataset by provided fields helped make it suitable for inference, my diagnostics revealed the presence of outliers. These outliers can affect the model’s performance and the accuracy of parameter estimates, highlighting the need for more advanced data cleaning techniques.

The large size of the dataset makes it challenging to manually review all observational studies. Automated outlier detection methods could be implemented to systematically identify and manage outliers. Or, as is not uncommon in systematic literature reviews, a large group of researchers could re-curate the dataset to make sure each data point is appropriate for this particular inference problem. Investigating the causes of these outliers could provide valuable insights into data collection processes and potential biases and understanding why certain data points deviate from expected patterns can help refine data collection methodologies and improve future datasets.

The workflows presented in this thesis are potentially applicable to computationally expensive models of infectious disease or other natural phenomena. In this thesis, I made a few modelling assumptions when developing the surrogates which could restrict their use cases. I assume a regular discretisation of the model inputs and outputs for training; deterministic outputs or a parametric distribution for model outputs; and an autoregressive relationship in model inputs and outputs. These assumptions make them suitable for agent based or differential equation based models, such as disease transmission, phylo-dynamic, predator-prey, and increasingly used economic agent based models as long as the chosen discretisation facilitates an accurate likelihood estimation. Also, for many tractable or computationally inexpensive models, such as the SIR or Lotka-Volterra models, the trade off between inference speed and approximation error due to the surrogate may no longer be favourable for researchers. Concurrent research on scale free surrogates which operate on infinite dimensional spaces may be more appropriate for differential equation based models where one would like to exploit simulations at different scales or require highly irregular model parameters, see Kovachki et al. (2022) as an example.

6.3 Conclusions

In this thesis I introduced scalable and flexible NN surrogates, some of which were very suitable for inference and surpassed the limitations identified for previously used surrogates, particularly in handling large datasets and high-dimensional parameter spaces. The NN surrogates were used with NUTS or in SVI to facilitate efficient site calibration, significantly reducing the time and computational resources required, or large scale probabilistic parameter estimation, to provide more realistic parameter estimates. I developed detailed diagnostics to improve model interpretation and parameter tuning. Furthermore, robust sequential inference techniques were suggested, to address issues where inference leads to mode collapse or gets trapped in local optima. I then updated

parameter estimates for an agent based model of malaria transmission incorporating modern data and accounting for interventions.

I have shown that NN-based surrogates for parameter inference on complex epidemiological models is viable, paving the way for future research to build on these methods to further improve the transparency, flexibility and robustness of our models and uncover new insights into disease dynamics. I would encourage the application of NN-surrogates and surrogate based inference for complex models in other scientific domains, such as economic, climate and biological and physics modelling, as their unique dynamics would offer new challenges and could drive the next wave of methodological innovation.

Bibliography

- Abadi, Martín et al. (Nov. 2, 2016). “TensorFlow: a system for large-scale machine learning”. In: *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation*. OSDI’16. USA: USENIX Association, pp. 265–283. ISBN: 978-1-931971-33-1. (Visited on 08/30/2024).
- Adegbite, Gbenga et al. (Dec. 1, 2023). “Mathematical modeling of malaria transmission dynamics in humans with mobility and control states”. In: *Infectious Disease Modelling* 8.4, pp. 1015–1031. ISSN: 2468-0427. DOI: [10.1016/j.idm.2023.08.005](https://doi.org/10.1016/j.idm.2023.08.005). URL: <https://www.sciencedirect.com/science/article/pii/S2468042723000817> (visited on 09/02/2024).
- Adegboye, Oyelola et al. (Aug. 2021). “Bayesian spatial modelling of Ebola outbreaks in Democratic Republic of Congo through the INLA-SPDE approach”. In: *Zoonoses and Public Health* 68.5, pp. 443–451. ISSN: 1863-2378. DOI: [10.1111/zph.12828](https://doi.org/10.1111/zph.12828).
- Andelfinger, Philipp (Mar. 23, 2021). *Differentiable Agent-Based Simulation for Gradient-Guided Simulation-Based Optimization*. arXiv: [2103.12476\[cs,eess\]](https://arxiv.org/abs/2103.12476). URL: <http://arxiv.org/abs/2103.12476> (visited on 05/15/2024).
- Andrianakis, Ioannis et al. (Jan. 8, 2015). “Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda”. In: *PLoS Computational Biology* 11.1, e1003968. ISSN: 1553-734X. DOI: [10.1371/journal.pcbi.1003968](https://doi.org/10.1371/journal.pcbi.1003968). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4288726/> (visited on 05/19/2023).
- Anil, Cem, James Lucas, and Roger Grosse (June 11, 2019). *Sorting out Lipschitz function approximation*. arXiv: [1811.05381\[cs,stat\]](https://arxiv.org/abs/1811.05381). URL: <http://arxiv.org/abs/1811.05381> (visited on 01/23/2023).
- Arras, Philipp et al. (Feb. 2022). “Variable structures in M87* from space, time and frequency resolved interferometry”. In: *Nature Astronomy* 6.2. Publisher: Nature Publishing Group, pp. 259–269. ISSN: 2397-3366. DOI: [10.1038/s41550-021-01548-0](https://doi.org/10.1038/s41550-021-01548-0). URL: <https://www.nature.com/articles/s41550-021-01548-0> (visited on 03/21/2024).
- Arya, Gaurav et al. (Jan. 9, 2023). *Automatic Differentiation of Programs with Discrete Randomness*. DOI: [10.48550/arXiv.2210.08572](https://doi.org/10.48550/arXiv.2210.08572). arXiv: [2210.08572\[cs,math\]](https://arxiv.org/abs/2210.08572). URL: <http://arxiv.org/abs/2210.08572> (visited on 05/25/2023).
- Babuschkin, Igor et al. (2020). *The DeepMind JAX Ecosystem*. URL: <http://github.com/deepmind>.
- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (May 19, 2016). *Neural Machine Translation by Jointly Learning to Align and Translate*. DOI: [10.48550/arXiv.1409.0473](https://doi.org/10.48550/arXiv.1409.0473). arXiv: [1409.0473](https://arxiv.org/abs/1409.0473). URL: <http://arxiv.org/abs/1409.0473> (visited on 11/25/2024).
- Bai, Tao et al. (Apr. 20, 2021). *Recent Advances in Adversarial Training for Adversarial Robustness*. arXiv: [2102.01356\[cs\]](https://arxiv.org/abs/2102.01356). URL: <http://arxiv.org/abs/2102.01356> (visited on 08/10/2022).
- Barua, Priyanka et al. (Jan. 18, 2019). “The impact of early life exposure to Plasmodium falciparum on the development of naturally acquired immunity to malaria in young Malawian children”. In: *Malaria Journal* 18.1, p. 11. ISSN: 1475-2875. DOI: [10.1186/s12936-019-2647-8](https://doi.org/10.1186/s12936-019-2647-8). URL: <https://doi.org/10.1186/s12936-019-2647-8> (visited on 08/20/2024).
- Battle, Katherine E., Ewan Cameron, et al. (May 7, 2015). “Defining the relationship between Plasmodium vivax parasite rate and clinical disease”. In: *Malaria Journal* 14.1, p. 191. ISSN: 1475-2875. DOI: [10.1186/s12936-015-0706-3](https://doi.org/10.1186/s12936-015-0706-3). URL: <https://doi.org/10.1186/s12936-015-0706-3> (visited on 09/01/2023).
- Battle, Katherine E., Carlos A. Guerra, et al. (Aug. 18, 2015). “Global database of matched Plasmodium falciparum and P. vivax incidence and prevalence records from 1985–2013”. In: *Scientific Data* 2.1. Number: 1 Publisher: Nature Publishing Group, p. 150012. ISSN: 2052-4463. DOI:

- 10.1038/sdata.2015.12. URL: <https://www.nature.com/articles/sdata201512> (visited on 07/21/2023).
- Bertozzi-Villa, Amelia et al. (June 11, 2021). “Maps and metrics of insecticide-treated net access, use, and nets-per-capita in Africa from 2000-2020”. In: *Nature Communications* 12.1. Publisher: Nature Publishing Group, p. 3589. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23707-7. URL: <https://www.nature.com/articles/s41467-021-23707-7> (visited on 06/19/2024).
- Bhatt, S. et al. (Oct. 8, 2015). “The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015”. In: *Nature* 526.7572, pp. 207–211. ISSN: 0028-0836. DOI: 10.1038/nature15535. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4820050/> (visited on 05/12/2022).
- Bingham, Eli et al. (Oct. 18, 2018). *Pyro: Deep Universal Probabilistic Programming*. DOI: 10.48550/arXiv.1810.09538. arXiv: 1810.09538. URL: <http://arxiv.org/abs/1810.09538> (visited on 11/27/2024).
- Bower, Richard G., Michael Goldstein, and Ian Vernon (Dec. 2010). “Galaxy formation: a Bayesian uncertainty analysis”. In: *Bayesian Analysis* 5.4. Publisher: International Society for Bayesian Analysis, pp. 619–669. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/10-BA524. URL: <https://projecteuclid.org/journals/bayesian-analysis/volume-5/issue-4/Galaxy-formation-a-Bayesian-uncertainty-analysis/10.1214/10-BA524.full> (visited on 11/20/2024).
- Bradbury, James et al. (2018). *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. URL: <http://github.com/google/jax>.
- Cameron, Ewan et al. (Sept. 8, 2015). “Defining the relationship between infection prevalence and clinical incidence of *Plasmodium falciparum* malaria”. In: *Nature Communications* 6.1. Number: 1 Publisher: Nature Publishing Group, p. 8170. ISSN: 2041-1723. DOI: 10.1038/ncomms9170. URL: <https://www.nature.com/articles/ncomms9170> (visited on 06/10/2022).
- Cappello, Lorenzo, Jaehee Kim, and Julia A. Palacios (Mar. 20, 2023). “adaPop: Bayesian inference of dependent population dynamics in coalescent models”. In: *PLOS Computational Biology* 19.3. Publisher: Public Library of Science, e1010897. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010897. URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010897> (visited on 03/21/2024).
- Chaari, Lotfi et al. (Jan. 30, 2015). *A Hamiltonian Monte Carlo Method for Non-Smooth Energy Sampling*. DOI: 10.48550/arXiv.1401.3988. arXiv: 1401.3988[stat]. URL: <http://arxiv.org/abs/1401.3988> (visited on 06/19/2024).
- Charles, Giovanni (Dec. 3, 2024a). *Code submission for "Seq2Seq Surrogates of Epidemic Models to Facilitate Bayesian Inference"*. DOI: 10.5281/zenodo.14267273. URL: <https://zenodo.org/records/14267273> (visited on 12/03/2024).
- (Dec. 3, 2024b). *dmeq*. Version 0.0.2. DOI: 10.5281/zenodo.14267131. URL: <https://zenodo.org/records/14267131> (visited on 12/03/2024).
- (Dec. 5, 2024c). *Fastms (Bayesian inference of high-dimensional Agent Based Model parameters)*. DOI: 10.5281/zenodo.14283422. URL: <https://zenodo.org/records/14283422> (visited on 12/05/2024).
- (Dec. 3, 2024d). “Notebook for "An auto-differentiable malaria model for efficient parameter inference"”. In: Publisher: Zenodo. URL: <https://zenodo.org/records/14267351> (visited on 12/03/2024).
- (Dec. 3, 2024e). “Pipeline and notebooks for "Fast black-box Bayesian inference using neural-network based surrogates"”. In: Publisher: Zenodo. URL: <https://zenodo.org/records/14267453> (visited on 12/03/2024).
- Charles, Giovanni, Peter Winskill, et al. (2022). *malariasimulation: An individual based model for malaria*.
- Charles, Giovanni, Timothy M. Wolock, et al. (June 26, 2023). “Seq2Seq Surrogates of Epidemic Models to Facilitate Bayesian Inference”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 37.12. Number: 12, pp. 14170–14177. ISSN: 2374-3468. DOI: 10.1609/aaai.v37i12.26658. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/26658> (visited on 10/02/2023).
- Charles, Giovanni D. and Sean L. Wu (Oct. 7, 2021). “individual: An R package for individual-based epidemiological models”. In: *Journal of Open Source Software* 6.66, p. 3539. ISSN: 2475-9066. DOI: 10.21105/joss.03539. URL: <https://joss.theoj.org/papers/10.21105/joss.03539> (visited on 05/20/2024).

- Charniga, Kelly et al. (Nov. 30, 2022). “Estimating Zika virus attack rates and risk of Zika virus-associated neurological complications in Colombian capital cities with a Bayesian model”. In: *Royal Society Open Science* 9.11. Publisher: Royal Society, p. 220491. DOI: [10.1098/rsos.220491](https://doi.org/10.1098/rsos.220491). URL: <https://royalsocietypublishing.org/doi/full/10.1098/rsos.220491> (visited on 03/21/2024).
- Chen, Ingrid et al. (Apr. 21, 2018). “The Lancet Commission on malaria eradication”. In: *The Lancet* 391.10130. Publisher: Elsevier, pp. 1556–1558. ISSN: 0140-6736, 1474-547X. DOI: [10.1016/S0140-6736\(18\)30911-5](https://doi.org/10.1016/S0140-6736(18)30911-5). URL: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(18\)30911-5/abstract](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)30911-5/abstract) (visited on 06/14/2021).
- Cho, Kyunghyun et al. (Oct. 7, 2014). “On the Properties of Neural Machine Translation: Encoder-Decoder Approaches”. In: *arXiv:1409.1259 [cs, stat]*. arXiv: [1409.1259](https://arxiv.org/abs/1409.1259). URL: <http://arxiv.org/abs/1409.1259> (visited on 05/12/2022).
- Chopra, Ayush et al. (May 30, 2023). “Differentiable Agent-based Epidemiology”. In: *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. AAMAS ’23. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 1848–1857. ISBN: 978-1-4503-9432-1. (Visited on 10/02/2023).
- Christiansen-Jucht, Céline D. et al. (Sept. 17, 2015). “Larval and adult environmental temperatures influence the adult reproductive traits of *Anopheles gambiae* s.s.” In: *Parasites & Vectors* 8.1, p. 456. ISSN: 1756-3305. DOI: [10.1186/s13071-015-1053-5](https://doi.org/10.1186/s13071-015-1053-5). URL: <https://doi.org/10.1186/s13071-015-1053-5> (visited on 08/29/2024).
- Churcher, Thomas S et al. (Aug. 22, 2016). “The impact of pyrethroid resistance on the efficacy and effectiveness of bednets for malaria control in Africa”. In: *eLife* 5. Ed. by Simon I Hay. Publisher: eLife Sciences Publications, Ltd, e16090. ISSN: 2050-084X. DOI: [10.7554/eLife.16090](https://doi.org/10.7554/eLife.16090). URL: <https://doi.org/10.7554/eLife.16090> (visited on 05/12/2022).
- Collins, O.C. and K.J. Duffy (Nov. 5, 2022). “A mathematical model for the dynamics and control of malaria in Nigeria”. In: *Infectious Disease Modelling* 7.4, pp. 728–741. ISSN: 2468-2152. DOI: [10.1016/j.idm.2022.10.005](https://doi.org/10.1016/j.idm.2022.10.005). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9661649/> (visited on 09/02/2024).
- Collins, W. E. and G. M. Jeffery (July 1999). “A retrospective examination of the patterns of recrudescence in patients infected with *Plasmodium falciparum*”. In: *The American Journal of Tropical Medicine and Hygiene* 61.1, pp. 44–48. ISSN: 0002-9637. DOI: [10.4269/tropmed.1999.61-044](https://doi.org/10.4269/tropmed.1999.61-044).
- Collins, William E. and Geoffrey M. Jeffery (May 2002). “A retrospective examination of sporozoite-induced and trophozoite-induced infections with *Plasmodium ovale*: development of parasitologic and clinical immunity during primary infection”. In: *The American Journal of Tropical Medicine and Hygiene* 66.5, pp. 492–502. ISSN: 0002-9637. DOI: [10.4269/ajtmh.2002.66.492](https://doi.org/10.4269/ajtmh.2002.66.492).
- Collins, William E., Geoffrey M. Jeffery, and Jacquelin M. Roberts (June 2004). “A retrospective examination of the effect of fever and microgametocyte count on mosquito infection on humans infected with *Plasmodium vivax*”. In: *The American Journal of Tropical Medicine and Hygiene* 70.6, pp. 638–641. ISSN: 0002-9637.
- Cowman, Alan F. et al. (Oct. 20, 2016). “Malaria: Biology and Disease”. In: *Cell* 167.3. Publisher: Elsevier, pp. 610–624. ISSN: 0092-8674, 1097-4172. DOI: [10.1016/j.cell.2016.07.055](https://doi.org/10.1016/j.cell.2016.07.055). URL: [https://www.cell.com/cell/abstract/S0092-8674\(16\)31008-X](https://www.cell.com/cell/abstract/S0092-8674(16)31008-X) (visited on 03/21/2024).
- Craig, Peter S. et al. (1997). “Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments”. In: *Case Studies in Bayesian Statistics*. Ed. by Constantine Gatsonis et al. New York, NY: Springer, pp. 37–93. ISBN: 978-1-4612-2290-3. DOI: [10.1007/978-1-4612-2290-3_2](https://doi.org/10.1007/978-1-4612-2290-3_2).
- Cranmer, Kyle, Johann Brehmer, and Gilles Louppe (Dec. 2020). “The frontier of simulation-based inference”. In: *Proceedings of the National Academy of Sciences* 117.48. Publisher: Proceedings of the National Academy of Sciences, pp. 30055–30062. DOI: [10.1073/pnas.1912789117](https://doi.org/10.1073/pnas.1912789117). URL: <https://www.pnas.org/doi/10.1073/pnas.1912789117> (visited on 03/21/2024).
- Datseris, George, Ali R. Vahdati, and Timothy C. DuBois (Jan. 2022). “Agents.jl: a performant and feature-full agent-based modeling software of minimal code complexity”. In: *SIMULATION* 0.0. Publisher: SAGE Publications, p. 003754972110688. DOI: [10.1177/00375497211068820](https://doi.org/10.1177/00375497211068820). URL: <https://doi.org/10.1177/00375497211068820>.
- De Cao, Nicola, Ivan Titov, and Wilker Aziz (Apr. 9, 2019). *Block Neural Autoregressive Flow*. DOI: [10.48550/arXiv.1904.04676](https://doi.org/10.48550/arXiv.1904.04676). arXiv: [1904.04676\[cs, stat\]](https://arxiv.org/abs/1904.04676). URL: <http://arxiv.org/abs/1904.04676> (visited on 05/15/2024).

- Deloron, P. and C. Chougnet (Nov. 1992). “Is immunity to malaria really short-lived?” In: *Parasitology Today (Personal Ed.)* 8.11, pp. 375–378. ISSN: 0169-4758. DOI: [10.1016/0169-4758\(92\)90174-z](https://doi.org/10.1016/0169-4758(92)90174-z).
- Demographic and Health Surveys (May 12, 2022). *The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs*. URL: <https://dhsprogram.com/> (visited on 05/12/2022).
- Dhaka, Akash Kumar and Alejandro Catalina (Feb. 12, 2021). “Challenges and Opportunities in High-dimensional Variational Inference”. In.
- Doan, N. A. K., W. Polifke, and L. Magri (Nov. 1, 2020). “Physics-informed echo state networks”. In: *Journal of Computational Science* 47, p. 101237. ISSN: 1877-7503. DOI: [10.1016/j.jocs.2020.101237](https://doi.org/10.1016/j.jocs.2020.101237). URL: <https://www.sciencedirect.com/science/article/pii/S1877750320305408> (visited on 11/20/2024).
- Duffy, Patrick E. (Feb. 2021). “Transmission-Blocking Vaccines: Harnessing Herd Immunity for Malaria Elimination”. In: *Expert review of vaccines* 20.2. Publisher: NIH Public Access, p. 185. ISSN: 1476-0584. DOI: [10.1080/14760584.2021.1878028](https://doi.org/10.1080/14760584.2021.1878028). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11127254/> (visited on 08/29/2024).
- Dyer, Joel et al. (Feb. 1, 2022). *Black-box Bayesian inference for economic agent-based models*. arXiv: [2202.00625\[cs,econ,stat\]](https://arxiv.org/abs/2202.00625). URL: <http://arxiv.org/abs/2202.00625> (visited on 11/30/2023).
- Eckhoff, Philip (Sept. 19, 2012). “P. falciparum Infection Durations and Infectiousness Are Shaped by Antigenic Variation and Innate and Adaptive Host Immunity in a Mathematical Model”. In: *PLoS ONE* 7.9, e44950. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0044950](https://doi.org/10.1371/journal.pone.0044950). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3446976/> (visited on 08/27/2024).
- Eckhoff, Philip A (Oct. 17, 2011). “A malaria transmission-directed model of mosquito life cycle and ecology”. In: *Malaria Journal* 10, p. 303. ISSN: 1475-2875. DOI: [10.1186/1475-2875-10-303](https://doi.org/10.1186/1475-2875-10-303). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3224385/> (visited on 05/15/2024).
- (Dec. 15, 2012). “Malaria parasite diversity and transmission intensity affect development of parasitological immunity in a mathematical model”. In: *Malaria Journal* 11, p. 419. ISSN: 1475-2875. DOI: [10.1186/1475-2875-11-419](https://doi.org/10.1186/1475-2875-11-419). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3557182/> (visited on 08/27/2024).
- Flaxman, Seth et al. (Aug. 2020). “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe”. In: *Nature* 584.7820. Number: 7820 Publisher: Nature Publishing Group, pp. 257–261. ISSN: 1476-4687. DOI: [10.1038/s41586-020-2405-7](https://doi.org/10.1038/s41586-020-2405-7). URL: <https://www.nature.com/articles/s41586-020-2405-7> (visited on 11/15/2023).
- Funk, Chris et al. (Dec. 8, 2015). “The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes”. In: *Scientific Data* 2.1. Publisher: Nature Publishing Group, p. 150066. ISSN: 2052-4463. DOI: [10.1038/sdata.2015.66](https://doi.org/10.1038/sdata.2015.66). URL: <https://www.nature.com/articles/sdata201566> (visited on 06/19/2024).
- Gal, Yarin and Zoubin Ghahramani (June 11, 2016). “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning”. In: *Proceedings of The 33rd International Conference on Machine Learning*. International Conference on Machine Learning. ISSN: 1938-7228. PMLR, pp. 1050–1059. URL: <https://proceedings.mlr.press/v48/gal16.html> (visited on 10/02/2023).
- Ge, Hong, Kai Xu, and Zoubin Ghahramani (2018). “Turing: a language for flexible probabilistic inference”. In.
- Gelman, Andrew, John B Carlin, et al. (2021). “Bayesian Data Analysis Third edition (with errors fixed as of 15 February 2021)”. In.
- Gelman, Andrew, Aki Vehtari, et al. (Nov. 3, 2020). *Bayesian Workflow*. DOI: [10.48550/arXiv.2011.01808](https://doi.org/10.48550/arXiv.2011.01808). arXiv: [2011.01808\[stat\]](https://arxiv.org/abs/2011.01808). URL: <http://arxiv.org/abs/2011.01808> (visited on 05/31/2024).
- Glockler, Manuel, Michael Deistler, and Jakob H Macke (2022). “VARIATIONAL METHODS FOR SIMULATION-BASED INFERENCE”. In.
- Goldstein, Michael and Nathan Huntley (2017). “Bayes Linear Emulation, History Matching, and Forecasting for Complex Computer Simulators”. In: *Handbook of Uncertainty Quantification*. Ed. by Roger Ghanem, David Higdon, and Houman Owhadi. Cham: Springer International Publishing, pp. 9–32. ISBN: 978-3-319-12385-1. DOI: [10.1007/978-3-319-12385-1_14](https://doi.org/10.1007/978-3-319-12385-1_14). URL: https://doi.org/10.1007/978-3-319-12385-1_14 (visited on 11/20/2024).

- Gorinova, Maria I., Dave Moore, and Matthew D. Hoffman (June 7, 2019). *Automatic Reparameterisation of Probabilistic Programs*. DOI: [10.48550/arXiv.1906.03028](https://doi.org/10.48550/arXiv.1906.03028). arXiv: [1906.03028\[cs, stat\]](https://arxiv.org/abs/1906.03028). URL: <http://arxiv.org/abs/1906.03028> (visited on 05/30/2024).
- Gramacy, Robert B. (Jan. 8, 2020). *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. 1st edition. Boca Raton: Chapman and Hall/CRC. 560 pp. ISBN: 978-0-367-41542-6.
- Greenberg, David S., Marcel Nonnenmacher, and Jakob H. Macke (May 17, 2019). *Automatic Posterior Transformation for Likelihood-Free Inference*. DOI: [10.48550/arXiv.1905.07488](https://doi.org/10.48550/arXiv.1905.07488). arXiv: [1905.07488\[cs, stat\]](https://arxiv.org/abs/1905.07488). URL: <http://arxiv.org/abs/1905.07488> (visited on 05/15/2024).
- Griffin, Jamie T. (July 26, 2016). “Is a reproduction number of one a threshold for Plasmodium falciparum malaria elimination?” In: *Malaria Journal* 15.1, p. 389. ISSN: 1475-2875. DOI: [10.1186/s12936-016-1437-9](https://doi.org/10.1186/s12936-016-1437-9). URL: <https://doi.org/10.1186/s12936-016-1437-9> (visited on 06/21/2021).
- Griffin, Jamie T., Neil M. Ferguson, and Azra C. Ghani (Feb. 11, 2014). “Estimates of the changing age-burden of Plasmodium falciparum malaria disease in sub-Saharan Africa”. In: *Nature Communications* 5.1. Number: 1 Publisher: Nature Publishing Group, p. 3136. ISSN: 2041-1723. DOI: [10.1038/ncomms4136](https://doi.org/10.1038/ncomms4136). URL: <https://www.nature.com/articles/ncomms4136> (visited on 06/13/2022).
- Griffin, Jamie T., T. Deirdre Hollingsworth, et al. (Aug. 10, 2010). “Reducing Plasmodium falciparum Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies”. In: *PLOS Medicine* 7.8. Publisher: Public Library of Science, e1000324. ISSN: 1549-1676. DOI: [10.1371/journal.pmed.1000324](https://doi.org/10.1371/journal.pmed.1000324). URL: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000324> (visited on 02/12/2021).
- Heek, Jonathan et al. (2023). *Flax: A neural network library and ecosystem for JAX*. Version 0.6.6. URL: <http://github.com/google/flax>.
- Hellewell, Joel et al. (2023). *ICDMM: Deterministic malaria odin model*.
- Hochreiter, Sepp and Jürgen Schmidhuber (Nov. 15, 1997). “Long Short-Term Memory”. In: *Neural Computation* 9.8, pp. 1735–1780. ISSN: 0899-7667. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735). URL: <https://doi.org/10.1162/neco.1997.9.8.1735> (visited on 05/12/2022).
- Hoffman, Matthew D (Jan. 5, 2013). “Stochastic Variational Inference”. In: — (n.d.). “Improving Explorability in Variational Inference with Annealed Variational Objectives”. In: *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*. arXiv: [1111.4246\[cs, stat\]](https://arxiv.org/abs/1111.4246). URL: <http://arxiv.org/abs/1111.4246> (visited on 03/05/2024).
- Hogan, Alexandra B. et al. (Nov. 28, 2023). “Long-term vaccination strategies to mitigate the impact of SARS-CoV-2 transmission: A modelling study”. In: *PLOS Medicine* 20.11. Publisher: Public Library of Science, e1004195. ISSN: 1549-1676. DOI: [10.1371/journal.pmed.1004195](https://doi.org/10.1371/journal.pmed.1004195). URL: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1004195> (visited on 08/28/2024).
- Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (Jan. 1, 1989). “Multilayer feedforward networks are universal approximators”. In: *Neural Networks* 2.5, pp. 359–366. ISSN: 0893-6080. DOI: [10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8). URL: <https://www.sciencedirect.com/science/article/pii/0893608089900208> (visited on 03/21/2024).
- Huang, Chin-Wei et al. (Feb. 12, 2018). “Improving Explorability in Variational Inference with Annealed Variational Objectives”. In: — (n.d.). “Improving Explorability in Variational Inference with Annealed Variational Objectives”. In: ().
- Jabot, Franck et al. (Dec. 16, 2014). *A comparison of emulation methods for Approximate Bayesian Computation*. DOI: [10.48550/arXiv.1412.7560](https://doi.org/10.48550/arXiv.1412.7560). arXiv: [1412.7560\[q-bio, stat\]](https://arxiv.org/abs/1412.7560). URL: <http://arxiv.org/abs/1412.7560> (visited on 03/06/2023).
- Jang, Eric, Shixiang Gu, and Ben Poole (Aug. 5, 2017). *Categorical Reparameterization with Gumbel-Softmax*. DOI: [10.48550/arXiv.1611.01144](https://doi.org/10.48550/arXiv.1611.01144). arXiv: [1611.01144\[cs, stat\]](https://arxiv.org/abs/1611.01144). URL: <http://arxiv.org/abs/1611.01144> (visited on 05/15/2024).
- Jirak, Doreen et al. (Sept. 1, 2023). “Echo State Networks and Long Short-Term Memory for Continuous Gesture Recognition: a Comparative Study”. In: *Cognitive Computation* 15.5, pp. 1427–1439. ISSN: 1866-9964. DOI: [10.1007/s12559-020-09754-0](https://doi.org/10.1007/s12559-020-09754-0). URL: <https://doi.org/10.1007/s12559-020-09754-0> (visited on 11/20/2024).
- Jong, Sanne E. de et al. (May 2021). “Systems analysis and controlled malaria infection in Europeans and Africans elucidate naturally acquired immunity”. In: *Nature Immunology* 22.5.

- Publisher: Nature Publishing Group, pp. 654–665. ISSN: 1529-2916. DOI: [10.1038/s41590-021-00911-7](https://doi.org/10.1038/s41590-021-00911-7). URL: <https://www.nature.com/articles/s41590-021-00911-7> (visited on 08/20/2024).
- Kazil, Jackie, David Masad, and Andrew Crooks (2020). “Utilizing Python for Agent-Based Modeling: The Mesa Framework”. In: *Social, Cultural, and Behavioral Modeling*. Ed. by Robert Thomson et al. Cham: Springer International Publishing, pp. 308–317. ISBN: 978-3-030-61255-9.
- Kennedy, Marc C. and Anthony O’Hagan (Sept. 1, 2001). “Bayesian Calibration of Computer Models”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 63.3, pp. 425–464. ISSN: 1369-7412, 1467-9868. DOI: [10.1111/1467-9868.00294](https://doi.org/10.1111/1467-9868.00294). URL: <https://academic.oup.com/jrssb/article/63/3/425/7083367> (visited on 11/20/2024).
- Kim, Kyurae et al. (Feb. 12, 2023). “On the Convergence of Black-Box Variational Inference”. In: Kim, Yongku, L. Mark Berliner, and Dal Ho Kim (June 1, 2015). “Approximated sensitivity analysis in posterior predictive distribution”. In: *Journal of the Korean Statistical Society* 44.2, pp. 261–270. ISSN: 1226-3192. DOI: [10.1016/j.jkss.2014.09.002](https://doi.org/10.1016/j.jkss.2014.09.002). URL: <https://www.sciencedirect.com/science/article/pii/S122631921400074X> (visited on 11/27/2024).
- Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi (May 13, 1983). “Optimization by Simulated Annealing”. In: *Science* 220.4598. Publisher: American Association for the Advancement of Science, pp. 671–680. DOI: [10.1126/science.220.4598.671](https://doi.org/10.1126/science.220.4598.671). URL: <https://www.science.org/doi/10.1126/science.220.4598.671> (visited on 08/01/2024).
- Knight, Gwenan M. et al. (Dec. 30, 2015). “The Impact and Cost-Effectiveness of a Four-Month Regimen for First-Line Treatment of Active Tuberculosis in South Africa”. In: *PLoS ONE* 10.12, e0145796. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0145796](https://doi.org/10.1371/journal.pone.0145796). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4696677/> (visited on 05/20/2024).
- Kovachki, Nikola et al. (Oct. 6, 2022). *Neural Operator: Learning Maps Between Function Spaces*. DOI: [10.48550/arXiv.2108.08481](https://doi.org/10.48550/arXiv.2108.08481). arXiv: [2108.08481](https://arxiv.org/abs/2108.08481)[cs,math]. URL: <http://arxiv.org/abs/2108.08481> (visited on 02/20/2023).
- Krieken, Emile van, Jakub M. Tomczak, and Annette ten Teije (Oct. 26, 2021). *Stochastic: A Framework for General Stochastic Automatic Differentiation*. DOI: [10.48550/arXiv.2104.00428](https://doi.org/10.48550/arXiv.2104.00428). arXiv: [2104.00428](https://arxiv.org/abs/2104.00428)[cs,stat]. URL: <http://arxiv.org/abs/2104.00428> (visited on 05/15/2024).
- Kristiadi, Agustinus, Matthias Hein, and Philipp Hennig (June 7, 2021). *Learnable Uncertainty under Laplace Approximations*. DOI: [10.48550/arXiv.2010.02720](https://doi.org/10.48550/arXiv.2010.02720). arXiv: [2010.02720](https://arxiv.org/abs/2010.02720)[cs]. URL: <http://arxiv.org/abs/2010.02720> (visited on 06/01/2023).
- Kuleshov, Volodymyr, Nathan Fenner, and Stefano Ermon (June 30, 2018). *Accurate Uncertainties for Deep Learning Using Calibrated Regression*. DOI: [10.48550/arXiv.1807.00263](https://doi.org/10.48550/arXiv.1807.00263). arXiv: [1807.00263](https://arxiv.org/abs/1807.00263)[cs,stat]. URL: <http://arxiv.org/abs/1807.00263> (visited on 08/19/2022).
- Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell (Nov. 3, 2017). *Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles*. arXiv: [1612.01474](https://arxiv.org/abs/1612.01474)[cs,stat]. URL: <http://arxiv.org/abs/1612.01474> (visited on 08/10/2022).
- Lam, Remi et al. (Dec. 22, 2023). “Learning skillful medium-range global weather forecasting”. In: *Science*. Publisher: American Association for the Advancement of Science. DOI: [10.1126/science.adi2336](https://doi.org/10.1126/science.adi2336). URL: <https://www.science.org/doi/10.1126/science.adi2336> (visited on 07/31/2024).
- Lamperti, Francesco, Andrea Roventini, and Amir Sani (May 1, 2018). “Agent-based model calibration using machine learning surrogates”. In: *Journal of Economic Dynamics and Control* 90, pp. 366–389. ISSN: 0165-1889. DOI: [10.1016/j.jedc.2018.03.011](https://doi.org/10.1016/j.jedc.2018.03.011). URL: <https://www.sciencedirect.com/science/article/pii/S0165188918301088> (visited on 05/25/2023).
- Langhorne, Jean et al. (July 2008). “Immunity to malaria: more questions than answers”. In: *Nature Immunology* 9.7. Publisher: Nature Publishing Group, pp. 725–732. ISSN: 1529-2916. DOI: [10.1038/ni.f.205](https://doi.org/10.1038/ni.f.205). URL: <https://www.nature.com/articles/ni.f.205> (visited on 03/21/2024).
- Li, Hao et al. (Nov. 7, 2018). *Visualizing the Loss Landscape of Neural Nets*. DOI: [10.48550/arXiv.1712.09913](https://doi.org/10.48550/arXiv.1712.09913). arXiv: [1712.09913](https://arxiv.org/abs/1712.09913)[cs,stat]. URL: <http://arxiv.org/abs/1712.09913> (visited on 06/19/2024).
- Li, Zhaohui and Matthias Hwai Yong Tan (July 3, 2022). “A Gaussian Process Emulator Based Approach for Bayesian Calibration of a Functional Input”. In: *Technometrics* 64.3. Publisher: ASA Website _eprint: <https://doi.org/10.1080/00401706.2021.1971567>, pp. 299–311. ISSN: 0040-1706.

- DOI: [10.1080/00401706.2021.1971567](https://doi.org/10.1080/00401706.2021.1971567). URL: <https://doi.org/10.1080/00401706.2021.1971567> (visited on 11/20/2024).
- Liepe, Juliane et al. (Feb. 2014). “A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation”. In: *Nature Protocols* 9.2. Publisher: Nature Publishing Group, pp. 439–456. ISSN: 1750-2799. DOI: [10.1038/nprot.2014.025](https://doi.org/10.1038/nprot.2014.025). URL: <https://www.nature.com/articles/nprot.2014.025> (visited on 03/21/2024).
- Lopez-Paz, David and Maxime Oquab (Mar. 13, 2018). *Revisiting Classifier Two-Sample Tests*. DOI: [10.48550/arXiv.1610.06545](https://doi.org/10.48550/arXiv.1610.06545). arXiv: [1610.06545\[stat\]](https://arxiv.org/abs/1610.06545). URL: <http://arxiv.org/abs/1610.06545> (visited on 08/08/2024).
- Lueckmann, Jan-Matthis et al. (Sept. 4, 2021). “Benchmarking Simulation-Based Inference”. In: Lukoševičius, Mantas and Herbert Jaeger (Aug. 1, 2009). “Reservoir computing approaches to recurrent neural network training”. In: *Computer Science Review* 3.3, pp. 127–149. ISSN: 1574-0137. DOI: [10.1016/j.cosrev.2009.03.005](https://doi.org/10.1016/j.cosrev.2009.03.005). URL: <https://www.sciencedirect.com/science/article/pii/S1574013709000173> (visited on 11/21/2024).
- Luo, Wei et al. (June 29, 2018). “Development of an Agent-Based Model to Investigate the Impact of HIV Self-Testing Programs on Men Who Have Sex With Men in Atlanta and Seattle”. In: *JMIR public health and surveillance* 4.2, e58. ISSN: 2369-2960. DOI: [10.2196/publichealth.9357](https://doi.org/10.2196/publichealth.9357).
- MacKay, David J. C. (May 1, 1992). “A practical Bayesian framework for backpropagation networks”. In: *Neural Computation* 4.3, pp. 448–472. ISSN: 0899-7667. DOI: [10.1162/neco.1992.4.3.448](https://doi.org/10.1162/neco.1992.4.3.448). URL: <https://doi.org/10.1162/neco.1992.4.3.448> (visited on 08/15/2022).
- Madkhali, Aymen M., Ahmad Hassn Ghzwani, and Hesham M. Al-Mekhlafi (June 17, 2022). “Comparison of Rapid Diagnostic Test, Microscopy, and Polymerase Chain Reaction for the Detection of Plasmodium falciparum Malaria in a Low-Transmission Area, Jazan Region, Southwestern Saudi Arabia”. In: *Diagnostics* 12.6, p. 1485. ISSN: 2075-4418. DOI: [10.3390/diagnostics12061485](https://doi.org/10.3390/diagnostics12061485). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9222139/> (visited on 08/20/2024).
- Makenga, Geoffrey et al. (Nov. 1, 2020). “Prevalence of malaria parasitaemia in school-aged children and pregnant women in endemic settings of sub-Saharan Africa: A systematic review and meta-analysis”. In: *Parasite Epidemiology and Control* 11, e00188. ISSN: 2405-6731. DOI: [10.1016/j.parepi.2020.e00188](https://doi.org/10.1016/j.parepi.2020.e00188). URL: <https://www.sciencedirect.com/science/article/pii/S240567312030057X> (visited on 08/27/2024).
- Mandt, Stephan et al. (May 28, 2016). *Variational Tempering*. DOI: [10.48550/arXiv.1411.1810](https://doi.org/10.48550/arXiv.1411.1810). arXiv: [1411.1810\[cs, stat\]](https://arxiv.org/abs/1411.1810). URL: <http://arxiv.org/abs/1411.1810> (visited on 05/15/2024).
- Marmin, Sébastien and Maurizio Filippone (Dec. 2022). “Deep Gaussian Processes for Calibration of Computer Models (with Discussion)”. In: *Bayesian Analysis* 17.4. Publisher: International Society for Bayesian Analysis, pp. 1301–1350. ISSN: 1936-0975, 1931-6690. DOI: [10.1214/21-BA1293](https://doi.org/10.1214/21-BA1293). URL: <https://projecteuclid.org/journals/bayesian-analysis/volume-17/issue-4/Deep-Gaussian-Processes-for-Calibration-of-Computer-Models-with-Discussion/10.1214/21-BA1293.full> (visited on 11/20/2024).
- McCarthy, Kevin A et al. (Jan. 7, 2015). “Calibration of an intrahost malaria model and parameter ensemble evaluation of a pre-erythrocytic vaccine”. In: *Malaria Journal* 14, p. 6. ISSN: 1475-2875. DOI: [10.1186/1475-2875-14-6](https://doi.org/10.1186/1475-2875-14-6). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4326442/> (visited on 08/31/2023).
- McCreesh, Nicky et al. (May 3, 2017). “Universal test, treat, and keep: improving ART retention is key in cost-effective HIV control in Uganda”. In: *BMC Infectious Diseases* 17.1, p. 322. ISSN: 1471-2334. DOI: [10.1186/s12879-017-2420-y](https://doi.org/10.1186/s12879-017-2420-y). URL: <https://doi.org/10.1186/s12879-017-2420-y> (visited on 09/18/2023).
- McDermott, Patrick L. and Christopher K. Wikle (2019). “Deep echo state networks with uncertainty quantification for spatio-temporal forecasting”. In: *Environmetrics* 30.3. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.2553>, e2553. ISSN: 1099-095X. DOI: [10.1002/env.2553](https://doi.org/10.1002/env.2553). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2553> (visited on 11/20/2024).
- McKinley, Trevelyan J. et al. (2018). “Approximate Bayesian Computation and Simulation-Based Inference for Complex Stochastic Epidemic Models”. In: *Statistical Science* 33.1. Publisher: Institute of Mathematical Statistics, pp. 4–18. ISSN: 0883-4237. URL: <https://www.jstor.org/stable/26770975> (visited on 07/20/2022).

- Mohamed, Shakir et al. (Sept. 29, 2020). *Monte Carlo Gradient Estimation in Machine Learning*. DOI: [10.48550/arXiv.1906.10652](https://doi.org/10.48550/arXiv.1906.10652). arXiv: [1906.10652\[cs,math,stat\]](https://arxiv.org/abs/1906.10652). URL: <http://arxiv.org/abs/1906.10652> (visited on 05/25/2023).
- Müller, Ivo et al. (July 14, 2009). “Three different Plasmodium species show similar patterns of clinical tolerance of malaria infection”. In: *Malaria Journal* 8, p. 158. ISSN: 1475-2875. DOI: [10.1186/1475-2875-8-158](https://doi.org/10.1186/1475-2875-8-158).
- Müller, Nicola F. and Remco R. Bouckaert (Apr. 9, 2019). *Adaptive parallel tempering for BEAST 2*. DOI: [10.1101/603514](https://doi.org/10.1101/603514). URL: <http://biorxiv.org/lookup/doi/10.1101/603514> (visited on 08/29/2024).
- Nash, Rebecca K. et al. (Jan. 1, 2021). “Systematic review of the entomological impact of insecticide-treated nets evaluated using experimental hut trials in Africa”. In: *Current Research in Parasitology & Vector-Borne Diseases* 1, p. 100047. ISSN: 2667-114X. DOI: [10.1016/j.crpvbd.2021.100047](https://doi.org/10.1016/j.crpvbd.2021.100047). URL: <https://www.sciencedirect.com/science/article/pii/S2667114X21000418> (visited on 07/05/2022).
- Neal, Radford M. (May 10, 2011). “MCMC using Hamiltonian dynamics”. In: *arXiv:1206.1901 [physics, stat]*. DOI: [10.1201/b10905](https://doi.org/10.1201/b10905). arXiv: [1206.1901](https://arxiv.org/abs/1206.1901). URL: <http://arxiv.org/abs/1206.1901> (visited on 05/12/2022).
- NetLogo Home Page (2024). URL: <https://ccl.northwestern.edu/netlogo/> (visited on 05/20/2024).
- Nguyen, Tran Dang et al. (July 29, 2023). “Preventing antimalarial drug resistance with triple artemisinin-based combination therapies”. In: *Nature Communications* 14.1. Publisher: Nature Publishing Group, p. 4568. ISSN: 2041-1723. DOI: [10.1038/s41467-023-39914-3](https://doi.org/10.1038/s41467-023-39914-3). URL: <https://www.nature.com/articles/s41467-023-39914-3> (visited on 08/28/2024).
- Ohm, Johanna R. et al. (Mar. 12, 2018). “Rethinking the extrinsic incubation period of malaria parasites”. In: *Parasites & Vectors* 11.1, p. 178. ISSN: 1756-3305. DOI: [10.1186/s13071-018-2761-4](https://doi.org/10.1186/s13071-018-2761-4). URL: <https://doi.org/10.1186/s13071-018-2761-4> (visited on 08/28/2024).
- Okell, Lucy C. et al. (Nov. 26, 2014). “Contrasting benefits of different artemisinin combination therapies as first-line malaria treatments using model-based cost-effectiveness analysis”. In: *Nature Communications* 5.1. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Combination drug therapy;Health care economics;Malaria;Policy and public health in microbiology Subject_term_id: combination-drug-therapy;health-care-economics;malaria;policy-and-public-health-in-microbiology, p. 5606. ISSN: 2041-1723. DOI: [10.1038/ncomms6606](https://doi.org/10.1038/ncomms6606). URL: <https://www.nature.com/articles/ncomms6606> (visited on 08/09/2021).
- Oord, Aaron van den, Sander Dieleman, et al. (Sept. 19, 2016). *WaveNet: A Generative Model for Raw Audio*. DOI: [10.48550/arXiv.1609.03499](https://doi.org/10.48550/arXiv.1609.03499). arXiv: [1609.03499\[cs\]](https://arxiv.org/abs/1609.03499). URL: <http://arxiv.org/abs/1609.03499> (visited on 10/02/2023).
- Oord, Aaron van den, Nal Kalchbrenner, et al. (June 18, 2016). *Conditional Image Generation with PixelCNN Decoders*. DOI: [10.48550/arXiv.1606.05328](https://doi.org/10.48550/arXiv.1606.05328). arXiv: [1606.05328\[cs\]](https://arxiv.org/abs/1606.05328). URL: <http://arxiv.org/abs/1606.05328> (visited on 10/02/2023).
- Oshagbemi, Olorunfemi A. et al. (Dec. 5, 2023). “Estimated distribution of malaria cases among children in sub-Saharan Africa by specified age categories using data from the Global Burden of Diseases 2019”. In: *Malaria Journal* 22.1, p. 371. ISSN: 1475-2875. DOI: [10.1186/s12936-023-04811-z](https://doi.org/10.1186/s12936-023-04811-z). URL: <https://doi.org/10.1186/s12936-023-04811-z> (visited on 08/27/2024).
- Paananen, Topi et al. (Apr. 11, 2019). “Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution”. In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. The 22nd International Conference on Artificial Intelligence and Statistics. ISSN: 2640-3498. PMLR, pp. 1743–1752. URL: <https://proceedings.mlr.press/v89/paananen19a.html> (visited on 11/27/2024).
- Papamakarios, George (Oct. 29, 2019). *Neural Density Estimation and Likelihood-free Inference*. DOI: [10.48550/arXiv.1910.13233](https://doi.org/10.48550/arXiv.1910.13233). arXiv: [1910.13233](https://arxiv.org/abs/1910.13233). URL: <http://arxiv.org/abs/1910.13233> (visited on 10/19/2024).
- Papamakarios, George, Theo Pavlakou, and Iain Murray (June 14, 2018). *Masked Autoregressive Flow for Density Estimation*. DOI: [10.48550/arXiv.1705.07057](https://doi.org/10.48550/arXiv.1705.07057). arXiv: [1705.07057\[cs,stat\]](https://arxiv.org/abs/1705.07057). URL: <http://arxiv.org/abs/1705.07057> (visited on 05/15/2024).
- Papamakarios, George, David C. Sterratt, and Iain Murray (Jan. 21, 2019). *Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows*. DOI: [10.48550/arXiv](https://doi.org/10.48550/arXiv).

- 1805.07226. arXiv: 1805.07226[cs, stat]. URL: <http://arxiv.org/abs/1805.07226> (visited on 05/30/2024).
- Paszke, Adam et al. (Dec. 3, 2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. DOI: [10.48550/arXiv.1912.01703](https://doi.org/10.48550/arXiv.1912.01703). arXiv: 1912.01703[cs, stat]. URL: <http://arxiv.org/abs/1912.01703> (visited on 09/13/2023).
- Penny, Melissa A. et al. (Jan. 23, 2016). “Public health impact and cost-effectiveness of the RTS,S/AS01 malaria vaccine: a systematic comparison of predictions from four mathematical models”. In: *The Lancet* 387.10016. Publisher: Elsevier, pp. 367–375. ISSN: 0140-6736, 1474-547X. DOI: [10.1016/S0140-6736\(15\)00725-4](https://doi.org/10.1016/S0140-6736(15)00725-4). URL: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)00725-4/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)00725-4/fulltext) (visited on 08/28/2024).
- Pfeffer, Daniel A. et al. (Oct. 5, 2018). “malariaAtlas: an R interface to global malariometric data hosted by the Malaria Atlas Project”. In: *Malaria Journal* 17.1, p. 352. ISSN: 1475-2875. DOI: [10.1186/s12936-018-2500-5](https://doi.org/10.1186/s12936-018-2500-5). URL: <https://doi.org/10.1186/s12936-018-2500-5> (visited on 06/18/2024).
- Phan, Du, Neeraj Pradhan, and Martin Jankowiak (Dec. 24, 2019). *Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro*. DOI: [10.48550/arXiv.1912.11554](https://doi.org/10.48550/arXiv.1912.11554). arXiv: 1912.11554[cs, stat]. URL: <http://arxiv.org/abs/1912.11554> (visited on 03/06/2023).
- Pohl, Kai and Ian A. Cockburn (Aug. 19, 2022). “Innate immunity to malaria: The good, the bad and the unknown”. In: *Frontiers in Immunology* 13. Publisher: Frontiers. ISSN: 1664-3224. DOI: [10.3389/fimmu.2022.914598](https://doi.org/10.3389/fimmu.2022.914598). URL: <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2022.914598/full> (visited on 08/20/2024).
- Quera-Bofarull, Arnau et al. (July 3, 2023). *Some challenges of calibrating differentiable agent-based models*. arXiv: 2307.01085[cs, q-fin, stat]. URL: <http://arxiv.org/abs/2307.01085> (visited on 10/02/2023).
- Raftery, Adrian E. and Le Bao (Dec. 2010). “Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling”. In: *Biometrics* 66.4, pp. 1162–1173. ISSN: 0006-341X. DOI: [10.1111/j.1541-0420.2010.01399.x](https://doi.org/10.1111/j.1541-0420.2010.01399.x). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2925068/> (visited on 09/05/2023).
- Reiker, Theresa, Nakul Chitnis, and Thomas Smith (July 30, 2019). “Modelling reactive case detection strategies for interrupting transmission of Plasmodium falciparum malaria”. In: *Malaria Journal* 18.1, p. 259. ISSN: 1475-2875. DOI: [10.1186/s12936-019-2893-9](https://doi.org/10.1186/s12936-019-2893-9). URL: <https://doi.org/10.1186/s12936-019-2893-9> (visited on 08/28/2024).
- Reiker, Theresa, Monica Golumbeanu, et al. (Dec. 10, 2021). “Emulator-based Bayesian optimization for efficient multi-objective calibration of an individual-based model of malaria”. In: *Nature Communications* 12.1. Number: 1 Publisher: Nature Publishing Group, p. 7212. ISSN: 2041-1723. DOI: [10.1038/s41467-021-27486-z](https://doi.org/10.1038/s41467-021-27486-z). URL: <https://www.nature.com/articles/s41467-021-27486-z> (visited on 05/12/2022).
- Renner, J. et al. (Nov. 1980). “Anti-gamete monoclonal antibodies synergistically block transmission of malaria by preventing fertilization in the mosquito.” In: *Proceedings of the National Academy of Sciences of the United States of America* 77.11. Publisher: National Academy of Sciences, p. 6797. DOI: [10.1073/pnas.77.11.6797](https://doi.org/10.1073/pnas.77.11.6797). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC350376/> (visited on 08/29/2024).
- Riebesell, Janosh and Stefan Bringuier (Aug. 9, 2020). *Collection of standalone TikZ images*. Version 0.1.0. DOI: [10.5281/zenodo.7486911](https://doi.org/10.5281/zenodo.7486911). URL: <https://github.com/janosh/tikz> (visited on 01/01/2023).
- Rios-Zertuche, Diego et al. (Apr. 30, 2021). “Performance of passive case detection for malaria surveillance: results from nine countries in Mesoamerica and the Dominican Republic”. In: *Malaria Journal* 20.1, p. 208. ISSN: 1475-2875. DOI: [10.1186/s12936-021-03645-x](https://doi.org/10.1186/s12936-021-03645-x). URL: <https://doi.org/10.1186/s12936-021-03645-x> (visited on 03/21/2024).
- Roberts, Leslie and Martin Enserink (Dec. 7, 2007). “Did They Really Say ... Eradication?” In: *Science* 318.5856. Publisher: American Association for the Advancement of Science, pp. 1544–1545. DOI: [10.1126/science.318.5856.1544](https://doi.org/10.1126/science.318.5856.1544). URL: <https://www.science.org/doi/10.1126/science.318.5856.1544> (visited on 06/09/2022).
- Sambridge, Malcolm (Jan. 1, 2014). “A Parallel Tempering algorithm for probabilistic sampling and multimodal optimization”. In: *Geophysical Journal International* 196.1, pp. 357–374. ISSN: 0956-540X. DOI: [10.1093/gji/ggt342](https://doi.org/10.1093/gji/ggt342). URL: <https://doi.org/10.1093/gji/ggt342> (visited on 03/06/2023).

- Sanchez-Gonzalez, Alvaro et al. (Sept. 14, 2020). “Learning to Simulate Complex Physics with Graph Networks”. In: *arXiv:2002.09405 [physics, stat]*. arXiv: [2002.09405](https://arxiv.org/abs/2002.09405). URL: <http://arxiv.org/abs/2002.09405> (visited on 04/22/2022).
- Sauboin, Christophe J. et al. (Dec. 23, 2015). “Potential public health impact of RTS,S malaria candidate vaccine in sub-Saharan Africa: a modelling study”. In: *Malaria Journal* 14, p. 524. ISSN: 1475-2875. DOI: [10.1186/s12936-015-1046-z](https://doi.org/10.1186/s12936-015-1046-z).
- Sawe, Sharon Jepkorir et al. (Jan. 27, 2024). “Gaussian process emulation to improve efficiency of computationally intensive multidisease models: a practical tutorial with adaptable R code”. In: *BMC Medical Research Methodology* 24.1, p. 26. ISSN: 1471-2288. DOI: [10.1186/s12874-024-02149-x](https://doi.org/10.1186/s12874-024-02149-x). URL: <https://doi.org/10.1186/s12874-024-02149-x> (visited on 11/20/2024).
- Schmit, Nora, Hillary M. Topazian, H. Magloire Natama, et al. (Feb. 8, 2024). “The public health impact and cost-effectiveness of the R21/Matrix-M malaria vaccine: a mathematical modelling study”. In: *The Lancet Infectious Diseases* 0.0. Publisher: Elsevier. ISSN: 1473-3099, 1474-4457. DOI: [10.1016/S1473-3099\(23\)00816-2](https://doi.org/10.1016/S1473-3099(23)00816-2). URL: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(23\)00816-2/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(23)00816-2/fulltext) (visited on 03/20/2024).
- Schmit, Nora, Hillary M. Topazian, Matteo Pianella, et al. (Nov. 20, 2023). “Resource allocation strategies for insecticide-treated bed nets to achieve malaria eradication”. In: *eLife* 12. Publisher: eLife Sciences Publications Limited. DOI: [10.7554/eLife.88283.2](https://doi.org/10.7554/eLife.88283.2). URL: <https://elifesciences.org/reviewed-preprints/88283> (visited on 08/27/2024).
- Sherrard-Smith, Ellie et al. (Feb. 2022). “Optimising the deployment of vector control tools against malaria: a data-informed modelling study”. In: *The Lancet. Planetary Health* 6.2, e100–e109. ISSN: 2542-5196. DOI: [10.1016/S2542-5196\(21\)00296-5](https://doi.org/10.1016/S2542-5196(21)00296-5).
- Sinka, Marianne E. et al. (Apr. 4, 2012). “A global map of dominant malaria vectors”. In: *Parasites & Vectors* 5.1, p. 69. ISSN: 1756-3305. DOI: [10.1186/1756-3305-5-69](https://doi.org/10.1186/1756-3305-5-69). URL: <https://doi.org/10.1186/1756-3305-5-69> (visited on 09/27/2021).
- SMC Alliance (2024). *SMC / Optimising Seasonal Malaria Chemoprevention in West and Central Africa*. SMC. URL: <https://www.smc-alliance.org/> (visited on 09/03/2024).
- Smith, David L., Katherine E. Battle, et al. (Apr. 5, 2012). “Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens”. In: *PLOS Pathogens* 8.4. Publisher: Public Library of Science, e1002588. ISSN: 1553-7374. DOI: [10.1371/journal.ppat.1002588](https://doi.org/10.1371/journal.ppat.1002588). URL: <https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002588> (visited on 08/13/2024).
- Smith, David L., Justin M. Cohen, et al. (Aug. 5, 2013). “A sticky situation: the unexpected stability of malaria elimination”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1623. Publisher: Royal Society, p. 20120145. DOI: [10.1098/rstb.2012.0145](https://doi.org/10.1098/rstb.2012.0145). URL: <https://royalsocietypublishing.org/doi/10.1098/rstb.2012.0145> (visited on 09/01/2023).
- Smith, Thomas, Amanda Ross, Nicolas Maire, Nakul Chitnis, et al. (Jan. 17, 2012). “Ensemble Modeling of the Likely Public Health Impact of a Pre-Erythrocytic Malaria Vaccine”. In: *PLoS Medicine* 9.1, e1001157. ISSN: 1549-1277. DOI: [10.1371/journal.pmed.1001157](https://doi.org/10.1371/journal.pmed.1001157). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3260300/> (visited on 09/06/2023).
- Smith, Thomas, Amanda Ross, Nicolas Maire, Christophe Rogier, et al. (Aug. 1, 2006). “AN EPIDEMIOLOGIC MODEL OF THE INCIDENCE OF ACUTE ILLNESS IN PLASMODIUM FALCIPARUM MALARIA”. In: *The American Journal of Tropical Medicine and Hygiene* 75.2. Publisher: American Society of Tropical Medicine and Hygiene Section: The American Journal of Tropical Medicine and Hygiene, pp. 56–62. DOI: [10.4269/ajtmh.2006.75.56](https://doi.org/10.4269/ajtmh.2006.75.56). URL: https://www.ajtmh.org/view/journals/tpmd/75/2_suppl/article-p56.xml (visited on 03/21/2024).
- Smith, Thomas A. et al. (Dec. 5, 2019). “Resurgence of malaria infection after mass treatment: a simulation study”. In: *Malaria Journal* 18.1, p. 409. ISSN: 1475-2875. DOI: [10.1186/s12936-019-3019-0](https://doi.org/10.1186/s12936-019-3019-0). URL: <https://doi.org/10.1186/s12936-019-3019-0> (visited on 08/28/2024).
- Snow, Robert W. et al. (Oct. 2017). “The prevalence of Plasmodium falciparum in sub-Saharan Africa since 1900”. In: *Nature* 550.7677. Publisher: Nature Publishing Group, pp. 515–518. ISSN: 1476-4687. DOI: [10.1038/nature24059](https://doi.org/10.1038/nature24059). URL: <https://www.nature.com/articles/nature24059> (visited on 04/03/2024).
- Stan Development Team (2018). *Stan Modeling Language Users Guide and Reference Manual, Version 2.18.0*. URL: <http://mc-stan.org/>.

- Stopard, Isaac J., Thomas S. Churcher, and Ben Lambert (Feb. 2021). “Estimating the extrinsic incubation period of malaria using a mechanistic model of sporogony”. In: *PLoS computational biology* 17.2, e1008658. ISSN: 1553-7358. DOI: [10.1371/journal.pcbi.1008658](https://doi.org/10.1371/journal.pcbi.1008658).
- Stuckey, Erin M. et al. (2014). “Modeling the cost effectiveness of malaria control interventions in the highlands of western Kenya”. In: *PloS One* 9.10, e107700. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0107700](https://doi.org/10.1371/journal.pone.0107700).
- Syed, Saifuddin et al. (Apr. 1, 2022). “Non-Reversible Parallel Tempering: A Scalable Highly Parallel MCMC Scheme”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.2, pp. 321–350. ISSN: 1369-7412. DOI: [10.1111/rssb.12464](https://doi.org/10.1111/rssb.12464). URL: <https://doi.org/10.1111/rssb.12464> (visited on 04/03/2024).
- Talts, Sean et al. (Oct. 21, 2020). *Validating Bayesian Inference Algorithms with Simulation-Based Calibration*. arXiv: [1804.06788\[stat\]](https://arxiv.org/abs/1804.06788). URL: <http://arxiv.org/abs/1804.06788> (visited on 05/31/2024).
- Thompson, Hayley A. et al. (Dec. 1, 2022). “Seasonal use case for the RTS,S/AS01 malaria vaccine: a mathematical modelling study”. In: *The Lancet Global Health* 10.12. Publisher: Elsevier, e1782–e1792. ISSN: 2214-109X. DOI: [10.1016/S2214-109X\(22\)00416-8](https://doi.org/10.1016/S2214-109X(22)00416-8). URL: [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(22\)00416-8/fulltext](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(22)00416-8/fulltext) (visited on 08/27/2024).
- Thüring, Tobias et al. (Apr. 27, 2016). *openmalaria: Schema 36*. DOI: [10.5281/zenodo.50570](https://doi.org/10.5281/zenodo.50570). URL: <https://zenodo.org/record/50570> (visited on 03/27/2023).
- Topazian, Hillary M. et al. (May 11, 2023). “Modelling the relative cost-effectiveness of the RTS,S/AS01 malaria vaccine compared to investment in vector control or chemoprophylaxis”. In: *Vaccine* 41.20, pp. 3215–3223. ISSN: 0264-410X. DOI: [10.1016/j.vaccine.2023.04.011](https://doi.org/10.1016/j.vaccine.2023.04.011). URL: <https://www.sciencedirect.com/science/article/pii/S0264410X23003948> (visited on 08/27/2024).
- Trape, J-F. and C. Rogier (June 1, 1996). “Combating malaria morbidity and mortality by reducing transmission”. In: *Parasitology Today* 12.6, pp. 236–240. ISSN: 0169-4758. DOI: [10.1016/0169-4758\(96\)10015-6](https://doi.org/10.1016/0169-4758(96)10015-6). URL: <https://www.sciencedirect.com/science/article/pii/S0169475896100156> (visited on 08/04/2022).
- Tsallis, Constantino and Daniel A. Stariolo (Nov. 15, 1996). “Generalized simulated annealing”. In: *Physica A: Statistical Mechanics and its Applications* 233.1, pp. 395–406. ISSN: 0378-4371. DOI: [10.1016/S0378-4371\(96\)00271-3](https://doi.org/10.1016/S0378-4371(96)00271-3). URL: <https://www.sciencedirect.com/science/article/pii/S0378437196002713> (visited on 05/15/2024).
- United Nations (Aug. 23, 2022). *World Population Prospects 2022: Summary of Results*. Statistical Papers - United Nations (Ser. A), Population and Vital Statistics Report. United Nations. ISBN: 978-92-1-001438-0. DOI: [10.18356/9789210014380](https://doi.org/10.18356/9789210014380). URL: <https://www.un-ilibrary.org/content/books/9789210014380> (visited on 06/18/2024).
- University of Oxford News (July 15, 2024). *Côte d’Ivoire makes history as first nation to deploy R21/Matrix-M™ Malaria Vaccine | University of Oxford*. URL: <https://www.ox.ac.uk/news/2024-07-15-c-te-d-ivoire-makes-history-first-nation-deploy-r21matrix-m-malaria-vaccine> (visited on 08/27/2024).
- Vapnik, V. (1991). “Principles of Risk Minimization for Learning Theory”. In: *Advances in Neural Information Processing Systems*. Vol. 4. Morgan-Kaufmann. URL: <https://proceedings.neurips.cc/paper/1991/hash/ff4d5fbbafdf976cfdc032e3bde78de5-Abstract.html> (visited on 08/08/2024).
- Vaswani, Ashish et al. (Dec. 5, 2017). “Attention Is All You Need”. In: *arXiv:1706.03762 [cs]*. arXiv: [1706.03762](https://arxiv.org/abs/1706.03762). URL: <http://arxiv.org/abs/1706.03762> (visited on 05/12/2022).
- Vernon, I. et al. (Aug. 15, 2022). “Bayesian emulation and history matching of JUNE”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380.2233. Publisher: Royal Society, p. 20220039. DOI: [10.1098/rsta.2022.0039](https://doi.org/10.1098/rsta.2022.0039). URL: <https://royalsocietypublishing.org/doi/10.1098/rsta.2022.0039> (visited on 11/20/2024).
- Warrell, D. A. (David Alan) and Herbert M. (Herbert Michael) Gilles (2002). *Essential malariology*. 4th ed. London: Arnold. ISBN: 978-0-340-81622-6.
- White, Michael T. et al. (Aug. 17, 2018). “Mathematical modelling of the impact of expanding levels of malaria control interventions on *Plasmodium vivax*”. In: *Nature Communications* 9.1. Number: 1 Publisher: Nature Publishing Group, p. 3300. ISSN: 2041-1723. DOI: [10.1038/s41467-018-05860-8](https://doi.org/10.1038/s41467-018-05860-8). URL: <https://www.nature.com/articles/s41467-018-05860-8> (visited on 03/06/2023).

- Winskill, Pete (2023a). *foresite: Access to Malaria Site Files*. URL: <https://mrc-ide.github.io/foresite/>.
- (2023b). *site: Site specific malariasimulation modelling*. URL: <https://mrc-ide.github.io/site/>.
- Winskill, Pete, Nora Schmit, and Ines Gerard-Ursin (2023). *netz: All Things Bed Nets*. URL: <https://mrc-ide.github.io/netz/>.
- Winskill, Peter et al. (Nov. 21, 2017). “The US President’s Malaria Initiative, Plasmodium falciparum transmission and mortality: A modelling study”. In: *PLOS Medicine* 14.11. Publisher: Public Library of Science, e1002448. ISSN: 1549-1676. DOI: [10.1371/journal.pmed.1002448](https://doi.org/10.1371/journal.pmed.1002448). URL: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002448> (visited on 03/15/2021).
- Wolpert, D.H. and W.G. Macready (Apr. 1997). “No free lunch theorems for optimization”. In: *IEEE Transactions on Evolutionary Computation* 1.1. Conference Name: IEEE Transactions on Evolutionary Computation, pp. 67–82. ISSN: 1941-0026. DOI: [10.1109/4235.585893](https://doi.org/10.1109/4235.585893). URL: <https://ieeexplore.ieee.org/document/585893/?arnumber=585893> (visited on 12/03/2024).
- World Bank (July 29, 2022). *World Bank Open Data / Data*. URL: <https://data.worldbank.org/> (visited on 07/29/2022).
- World Health Organisation (2022). “WHO Guidelines for malaria - 3 June 2022”. In: p. 396.
- World Health Organization (2020). *Malaria eradication: benefits, future scenarios and feasibility: a report of the Strategic Advisory Group on Malaria Eradication*. Section: xxv, 116 p. Geneva: World Health Organization. ISBN: 978-92-4-000367-5. URL: <https://apps.who.int/iris/handle/10665/331795> (visited on 06/09/2022).
- (2021). *Global Technical Strategy For Malaria - 2016 - 2030*. Section: liv, 263 p. Geneva: World Health Organization. ISBN: 978-92-4-004049-6. URL: <https://apps.who.int/iris/handle/10665/350147> (visited on 06/13/2022).
- (2023). *World malaria report 2023*. Type: Publications. World Health Organization. lxvii, 283 p.
- WorldPop (July 29, 2022). *2022 World Population by Country*. URL: <https://worldpopulationreview.com/> (visited on 07/29/2022).
- Wu, Lindsey et al. (Dec. 2015). “Comparison of diagnostics for the detection of asymptomatic Plasmodium falciparum infections to inform control and elimination strategies”. In: *Nature* 528.7580. Publisher: Nature Publishing Group, S86–S93. ISSN: 1476-4687. DOI: [10.1038/nature16039](https://doi.org/10.1038/nature16039). URL: <https://www.nature.com/articles/nature16039> (visited on 08/20/2024).
- Yao, Yuling, Aki Vehtari, and Andrew Gelman (Nov. 18, 2021). *Stacking for Non-mixing Bayesian Computations: The Curse and Blessing of Multimodal Posteriors*. DOI: [10.48550/arXiv.2006.12335](https://doi.org/10.48550/arXiv.2006.12335). arXiv: [2006.12335\[stat\]](https://arxiv.org/abs/2006.12335). URL: <http://arxiv.org/abs/2006.12335> (visited on 03/21/2024).
- Zhu, Ciyu et al. (Dec. 1, 1997). “Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization”. In: *ACM Transactions on Mathematical Software* 23.4, pp. 550–560. ISSN: 0098-3500. DOI: [10.1145/279232.279236](https://doi.org/10.1145/279232.279236). URL: <https://dl.acm.org/doi/10.1145/279232.279236> (visited on 06/19/2024).