

# Seq2Seq Surrogates of Epidemic Models to Facilitate Bayesian Inference

Giovanni Charles<sup>1</sup>, Timothy M. Wolock<sup>1</sup>, Peter Winskill<sup>1</sup>, Azra Ghani<sup>1</sup>, Samir Bhatt<sup>\*2</sup>, Seth Flaxman<sup>3</sup>

<sup>1</sup> Imperial College London

<sup>2</sup> University of Copenhagen

<sup>3</sup> Department of Computer Science, University of Oxford  
s.bhatt@ic.ac.uk

## Abstract

Epidemic models are powerful tools in understanding infectious disease. However, as they increase in size and complexity, they can quickly become computationally intractable. Recent progress in modelling methodology has shown that surrogate models can be used to emulate complex epidemic models with a high-dimensional parameter space. We show that deep sequence-to-sequence (seq2seq) models can serve as accurate surrogates for complex epidemic models with sequence based model parameters, effectively replicating seasonal and long-term transmission dynamics. Once trained, our surrogate can predict scenarios a several thousand times faster than the original model, making them ideal for policy exploration. We demonstrate that replacing a traditional epidemic model with a learned simulator facilitates robust Bayesian inference.

## 1 Introduction

Epidemic models are an essential tool in the study of infectious disease dynamics (Eaton et al. 2019; World Health Organization 2021; Djaafara et al. 2021). By modelling new infections as a function of current infections, they leverage the fundamental dynamics to predict epidemic trajectories accurately with relatively few parameters. Recently, they have risen to unprecedented public prominence through their use during the SARS-CoV-2 (coronavirus) pandemic (Birrell et al. 2021).

In most cases, we do not observe these dynamics or model parameters directly, so we must infer them from what data we do measure (e.g. deaths, positive tests, etc.). In order to produce a set of predictions, we need to infer the unknown parameters in the model, a process that quickly becomes infeasible at the scale needed for methods like Hamiltonian Monte Carlo (HMC) (Neal 2011) for even moderately complex models. Further, because the current state of the model is a function of all past states, complex correlation structures can arise between time-varying parameters, making sampling difficult.

(Griffin et al. 2010) proposed an epidemiologically motivated Individual-Based Model (IBM) of malaria transmission. It simulates infection at an individual-level based on

exposure to three species of infectious mosquito, immunity and a variety of interventions. This model is invaluable for projecting the effects of policy on malaria prevalence, for example, (Winskill et al. 2017), projected the effects of a 44% funding cut from a major global donor. Unfortunately, run times for a single parameter set can take several minutes making it unsuitable for parameter inference where large numbers of repeated function evaluations are required. (Griffin, Ferguson, and Ghani 2014) later showed that a faster, reduced surrogate model can be used to infer these parameters. This surrogate made simplifying assumptions, such as the effects of seasonality and existing interventions being negligible. The observed data had to be carefully curated to minimise those effects. In contrast to a surrogate where a researcher arbitrarily chooses how to reduce model complexity for inference, we will introduce a deep learning solution.

Recent work has demonstrated the utility of learned surrogate models reducing the computational cost of simulation (Cameron et al. 2015; McKinley et al. 2018; Reiker et al. 2021). Data-driven methods could potentially learn more expressive, surrogates which allow researchers to relax their assumptions in parameter inference and use a wider range of data sets. (Cameron et al. 2015) proposed a functional regression method to map the functional space between model parameters and outputs using a library of pre-computed outputs. As the input space becomes large, more sophisticated methods have been employed to explore the sub-spaces most useful for the parameter inference task at hand, as reviewed by (McKinley et al. 2018). (Reiker et al. 2021) showed that a Bayesian optimisation workflow is effective in performing parameter inference on IBM outputs. They proposed training a surrogate and simultaneously performing inference, adaptively sampling parameter sets for training the surrogate which reduce the surrogate’s predictive uncertainty in regions of the input space which are proximal to the best-fit parameters.

However none of the above approaches have developed surrogates designed for sequential inputs. Many epidemiological models use sequences as input, such as transmissibility (Rahimi, Chen, and Gandomi 2021), intervention usage (Griffin, Ferguson, and Ghani 2014) and contact matrices (Watson et al. 2022), and produce sequential outputs such as prevalence trajectories. This format makes them amenable

\*Corresponding Author

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

to a sequence-to-sequence (seq2seq) modelling approach. Advances in seq2seq models have been very promising in machine translation (Vaswani et al. 2017) and time-series forecasting (Wen et al. 2022), but it is yet to be seen if they can be applied to the surrogate modelling of epidemiological models.

We hypothesised that a seq2seq based surrogate architecture could accurately predict IBM outputs for a combination of scalar and sequential parameters. Such a surrogate would make Bayesian inference feasible for long running observational data sets.

### 1.1 Contributions

We created a surrogate for a well known malaria model, allowing researchers to parameterise the surrogate for an arbitrary location and predict the malaria prevalence within milliseconds. Previous approaches have either restricted the input space to scalar parameters, or selected training samples for a specific inference task. As far as we know, no other infectious disease model surrogates are designed to take arbitrary sequence based parameters as input at prediction time.

We propose a formulation for the surrogate in section 3. We outline a surrogate model structure and training method in sections 3.3 and 3.4 where we use synthetic parameter sets designed to explore the parameter space. We evaluated the surrogate on two benchmark parameter sets, in section 4, to show that it generalises to longer parameter sequences and parameter sets generated from real world data.

We then performed parameter inference using a Hamiltonian Monte Carlo algorithm on the learned surrogate. We estimated a parameter for a location in Senegal, from an 18 year sequence of intervention statistics. This is the first approach which would allow for inference using sequential parameters over an arbitrary time horizon and/or location without re-training the surrogate, which can reportedly take over a week per inference task, see (Reiker et al. 2021).

## 2 Model of Malaria Transmission

We first introduce a malaria model to provide context for the formulation of the surrogate model and to illustrate the complexity of an infectious disease model.

We used an implementation of an Individual-Based Model (IBM) of malaria originally published by Griffin et al.. The IBM includes human, mosquito and Insecticide Treated Net (ITN) models. The human model simulates infectious bites, immunity, and the probability of detection by diagnostic tests. This model is used to produce malaria burden statistics and calculate the Force of Infection towards Mosquitoes (FOIM), as explained in section 2.2. The mosquito model simulates the size of the infectious and susceptible mosquito populations given the FOIM and a time-series of rainfall. This model is used to calculate an Entomological Inoculation Rate (EIR), explained in section 2.1, which drives infections in the human model completing the cycle of infection between humans and mosquitoes. The bed net model modifies this cycle by reducing the number of bites in the protected human population and increasing mortality in the mosquito population. These effects are explained in section 2.3.

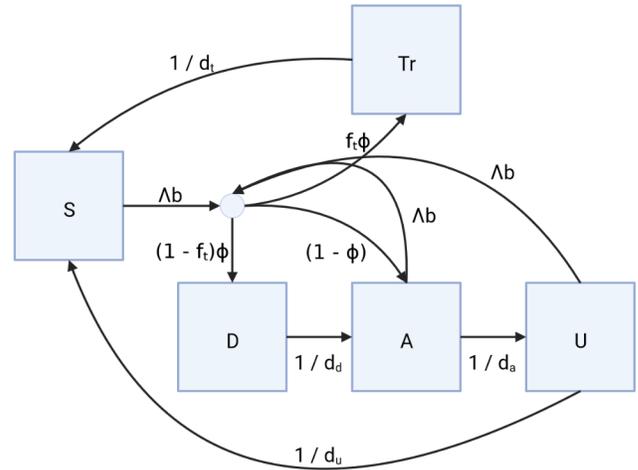


Figure 1: A human state transition diagram. Boxes represent human states and arcs transitions between the states, simulated with Bernoulli trials.  $\Lambda$  represents the EIR,  $f_t$  the probability of treatment,  $b$  the probability of infection and  $\phi$  the probability of clinical disease.

### 2.1 Human Transmission

The human model, shown in figure 1, simulates five infection states: susceptible (S), clinical disease (D), treated (T), asymptomatic infection (A), and subpatent infection (U). Subpatent infections are those not detected using routine malaria testing, instead requiring advanced diagnostic methods. At each time step, Bernoulli trials are simulated to transition individuals between these states at the parameterised rates.

The Entomological Inoculation Rate (denoted  $\Lambda$ ), is calculated as the number of infectious mosquito bites the population experiences over time. However, each individual experiences  $\Lambda$  differently. Older individuals are bitten more often, due to their increased surface area. The age  $a_i$  of each individual  $i$  is sampled from an exponential distribution with mean  $\mu_a$ .

After an individual  $i$  is bitten by an infectious mosquito, we simulate a possible infection.  $b_i(t)$  represents the probability of blood-stage immunity failing, determining whether the individual will develop (at least) an asymptomatic infection.  $\phi_i(t)$  represents likewise probabilities for clinical immunity and infection. For asymptomatic individuals,  $q_i(t, a_i)$  represents the probability of detection by a routine diagnostic test, taking into account the individual's age. Each of these immunity functions are complex estimations based on each individual's previous exposures to bites and infection.

The diagnostic prevalence of malaria in the model  $\text{PfPr}(t, l, u)$  is the proportion of individuals who would test positive in a routine diagnostic test, as defined in equation 1. This statistic is bounded between lower and upper age ranges  $l$  and  $u$ , including all clinically diseased individuals and asymptomatic individuals who would be detected.

$$\text{PfPr}(t, l, u) = \frac{\sum_i \text{detected}(i, t) \times \mathbb{1}_{l < a_i \leq u}}{\sum_i \mathbb{1}_{l < a_i \leq u}} \quad (1)$$

$$\text{detected}(i, t) = \begin{cases} 1 & \text{state}(i, t) = D \\ q_i(t, a) & \text{state}(i, t) = A \\ 0 & \text{otherwise} \end{cases}$$

## 2.2 Mosquito Modelling

Mosquitoes are modelled using a system of delay differential equations with early larval (E), late larval (L), pupal (P), susceptible (Sm), incubating (Em) and infectious (Im) compartments. Early (E) and Late (L) larval compartments represent the number of larvae soon after hatching. The size of these two compartments are dependent on a carrying capacity  $K$  which can vary with rainfall  $R(t)$ , see equation 3. The surviving larvae develop into Pupae (P) and half of the surviving pupae will develop into susceptible female mosquitoes.

$$K^v = K_0^v \frac{R(t)}{\bar{R}} \quad (2)$$

$$R(t) = g_0 + \sum_{i=1}^3 g_i \cos(2\pi t i) + h_i \sin(2\pi t i) \quad (3)$$

Susceptible female mosquitoes (Sm) are infected based on FOIM. FOIM is a driver of infection in the mosquito population, calculated by the number of bites each mosquito would have made on an infectious human. Infected mosquitoes will incubate in the (Em) state for a fixed period before themselves becoming infectious. If they survive, they will transition to the Infected (Im) state where they remain until death. The number of new E state larvae introduced each time step is proportional to the total number of adult female mosquitoes in the simulation.

The mosquito model is replicated for each species of mosquito in the simulation  $v$ . Users can parameterise the proportion of adult female mosquitoes in each species  $\kappa^v$  by setting proportional values of the carrying capacity parameter,  $K_0^v$ . It is often convenient to parameterise  $K_0^v$  using a baseline EIR parameter,  $\Lambda_0$ , or the expected EIR before interventions. (Griffin 2016) explains in detail how to calculate the carrying capacity from baseline EIR.

## 2.3 ITN Modelling

Insecticide treated bednets (ITNs) are an effective malaria intervention which protect individuals while they are in bed. They 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. These responses, for various species and insecticides, have been examined in experimental hut trials, (Sherrard-Smith et al. 2022). On the human side, EIR is reduced, 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 5 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.

## 2.4 Model Fitting Considerations

Typically, malaria researchers would like to infer the unknown  $\Lambda_0$  from observed prevalence data  $\text{PfPr}(t, l, u)$ , given known parameters about the transmission setting, such as  $\mu_a, \kappa^v, \nu(t), g, h, \dots$ . However, simulating each individual, the mosquito population and ITN effects for several years can take several minutes, making parameter inference challenging. We also note that the individuals simulated by this model are synthetic and not measured directly by any real-world data source. Any fitting or inference process for this model will aggregate over individuals to predict population-level averages which can be compared to data.

In the next section we will formulate a model to predict these population-level aggregates directly from the IBM parameters, in order to eliminate a large amount of computation that is, for the purposes of fitting the model to data, unnecessary.

## 3 Surrogate Model

Let  $f : \mathcal{X} \rightarrow \mathcal{Y}$  be a model to be emulated. Assume that  $\mathcal{X} \subset \mathbb{R}^k$  can be bounded to some set of plausible values. Our goal is to train a model  $\tilde{f}_\theta : \mathcal{X} \rightarrow \mathcal{Y}$  such that  $|\tilde{f}_\theta(\mathbf{x}) - f(\mathbf{x})| < \epsilon$  for some suitably small  $\epsilon > 0$ . We refer to  $\tilde{f}_\theta$  as a surrogate model. Let  $\mathcal{L}(\mathbf{x}, \mathbf{y}, \tilde{f}_\theta)$  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  $\hat{\theta}$ :

$$\hat{\theta} = \arg \min_{\theta} \int_{\mathcal{X}} \mathcal{L}(\mathbf{x}, \mathbf{y}, \tilde{f}_\theta) p(\mathbf{x}) d\mathbf{x}. \quad (4)$$

We have included a prior distribution over  $\mathcal{X}$ , but it is not strictly necessary. Here,  $p(\mathbf{x})$  serves as a measure of the relative importance of emulating  $f(\mathbf{x})$  accurately. If  $p$  is uniform, then we are ascribing equal importance to every possible parameter set.

### 3.1 Input Domain

We can translate the malaria IBM framing described in section 2 by first defining  $\mathcal{X}$ .

To parameterise the surrogate for a particular location, we set  $\kappa^v, \mu_a, g_i, h_i$  and  $\nu(t)$  parameters to capture mosquito composition, demography, seasonality and ITN usage respectively. We defined a range for  $\mu_a$  as 14.8 to 55.4 years based on the maximum and minimum average age statistics per country from World Population Review (WorldPopulationReview 2022). Ranges for Fourier coefficients  $g_i, h_i$  are set between -10 and 10, to include the largest values found by fitting to rainfall data in sub-national administrative units in sub-Saharan Africa. Since  $\kappa^v$  are relative proportions of three mosquito species, they are only valid in the 3-simplex in  $\mathbb{R}^3$ .

A realistic range of  $\nu(t)$  is between 0 and 0.8 for each year, as it becomes operationally difficult to distribute to

ITNs to all individuals. We 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. In summary,  $\nu(t)$  is defined as a bounded step function with  $n \geq 0$  annually spaced knots.

One of the main unknowns for a each location is,  $\Lambda_0$ . 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 have ranged between 0.05 and 418 (Griffin, Ferguson, and Ghani 2014), and model outputs suggest that higher values of  $\Lambda_0$  have diminishing marginal effects on PfPr. Therefore we set a range for  $\Lambda_0$  between 0 (exclusive) and 500.

To express  $\mathcal{X}$  as a sequence, we 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.

### 3.2 Output Range

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

$$f(\mathbf{x}) = \left\{ \left\{ \text{PfPr}(d, u, l) \right\}_{d=1}^{365} \right\}_{t=1}^T \quad (5)$$

Note that we have augmented the notation for the IBM run to include parameter sets. With this notation,  $\mathcal{Y}$  is easy to define as the image of  $f$ :

$$\mathcal{Y} = \{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, \quad (6)$$

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

### 3.3 Surrogate Model Structure

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 we chose a recurrent neural network (RNN) for simplicity. The RNN should be able to exploit the relationships between time points to generalise for unseen parameter sets and longer time horizons. We used Mean Squared Error (MSE) as the loss function  $\mathcal{L}$ . We implemented a grid search select hyper-parameters which minimised MSE on the validation set, table 1 lists the considered parameters.

Our 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 3. The resulting inputs were standardised, removing the mean and scaling to unit variance. The final input layer had 383 recurrent cells,

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

Table 1: Hyper-parameters used in the grid search. The selected parameter is emboldened

feeding 365 latent recurrent cells followed by a dense output layer of 365 cells.

### 3.4 Sampling Training Data

Training a surrogate for the malaria IBM requires a large library of simulations covering a wide area of input space. We 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. We 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 3.1. The IBM was run at the sampled points, setting aside a warm-up period before activating the ITN parameters to allow the simulation to reach an equilibrium. We set a large population size for the IBM 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 observational data.

## 4 Surrogate Evaluation

We 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 statistics for a randomly selected country 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  $6.8 \times 10^{-3}$  on un-normalised prevalence values. The prediction error for these benchmarks is shown in figure 2.

We compared execution times for the original IBM, surrogate training, single prediction and batch prediction. The IBM 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  $229 \pm 34$  seconds ( $n = 3$ ). Both single and batch predictions were made on an 8-core Intel CPU. While the IBM required just under 24 hours to run 10,000 samples, the batch surrogate required just over 3 seconds (approximately a 25,000-

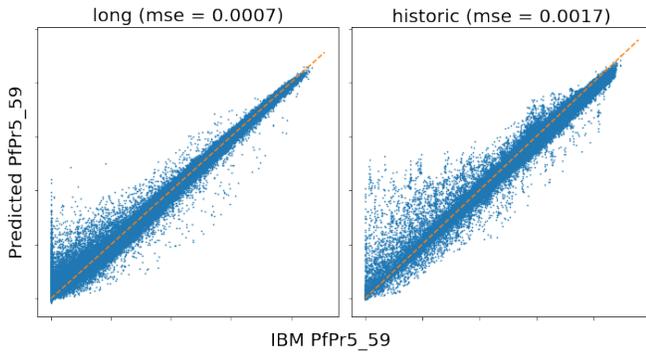


Figure 2: True versus predicted prevalence of malaria in children aged 5-59 months (PfPr5\_59) for the *long* (left) and *historic* (right) datasets where the true estimates come from the IBM model and the predicted estimates come from the surrogate. Each point represents the prevalence for one time 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.

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

Table 2: The timings per simulation for 10,000 simulations on the training and validation parameter sets

fold improvement). The timing statistics for an each run are listed in table 2.

## 5 Parameter Inference

We used Tensorflow Probability (TensorflowProbability 2022) to fit our model to observational data from the Demographic and Health Survey program (DHS 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. We included responses for respondents between 5 and 59 months old and weighted responses by population density data from World Bank data (WorldBank 2022) to get a general picture of parasite prevalence over time in the region,  $D_{\text{PfPr}_{5-59}}$ .

To estimate a value of  $\Lambda_0$  which best explains  $D_{\text{PfPr}_{5-59}}$ , we used a binomial likelihood function, see equation 7. We compared each participant’s infection status  $J^i$ , an indicator variable, to the average predicted prevalence for the month in which they were observed,  $\text{PfPr}(t^i)$ . Our surrogate, parameterised with  $\nu(t), g_i, h_i, \kappa^v$  values for Kolda, and aggregated at month  $t^i$  provides quick estimates for  $\text{PfPr}(t^i)$ . We set a uniform prior for  $\Lambda_0$  between 0 and 500.

$$L(\lambda^* | D_{\text{PfPr}_{5-59}}) = \prod \text{PfPr}(t^i)^{J^i} (1 - \text{PfPr}(t^i))^{1-J^i} \quad (7)$$

We performed Markov Chain Monte Carlo using the HMC algorithm (Neal 2011) with 10 parallel chains, 2,000

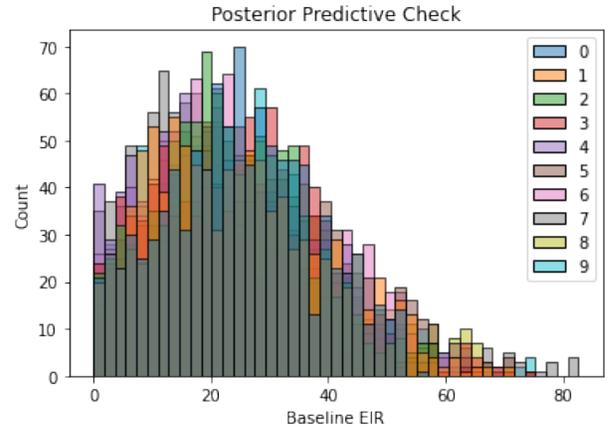


Figure 3: A posterior predictive check plotting  $\Lambda_0$  samples from 10 MCMC chains. A previous rough estimate of baseline EIR in three Senegalese populations was 1, 20 and 200 (infectious bites per person per year) for the cities of Dakar, Ndiop and Dielmo (Trape and Rogier 1996). This estimate of  $23.2 \pm 13.7$  would place Kolda’s baseline transmission alongside Ndiop.

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.

The original IBM was then run with the mean estimate for comparison, see the posterior predictive check in figure 3 and resulting fit in figure 4. 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.

## 6 Discussion

We 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.

### 6.1 A Surrogate for General Analysis

The surrogate could be used as an exploratory tool, allowing researchers to begin modelling malaria transmission for arbitrary locations in seconds. Selective approaches, such as

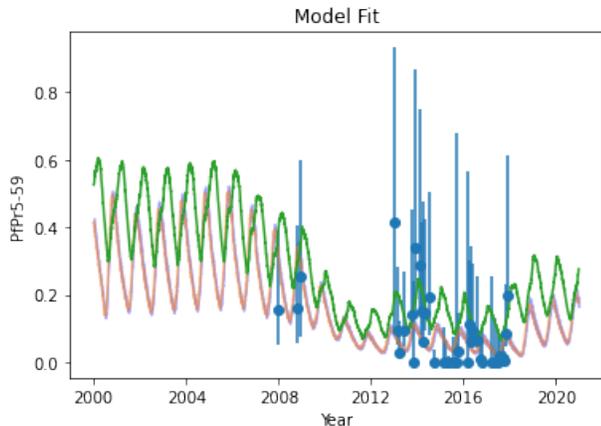


Figure 4: A plot of the final fit including the surrogate trajectory for the mean EIR (in orange). The IBM was then calibrated to the mean EIR using a root finding method and run to simulate the ITN distributions for Kolda, the prevalence for that run is plotted in green. Observations have been plotted 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.

(McKinley et al. 2018; Reiker 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 et al. 2021), training can take between 7 and 12 days.

The benchmarks showed that the surrogate could emulate the IBM for the vast majority of simulation scenarios. 99% and 97% of predictions had an MSE of less than  $1 \times 10^{-2}$  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 5).

Despite the benchmarks showing a small risk in surrogate prediction error, many predictions significantly deviated from the IBM outputs. Manual inspection reveals that the worst predictions for the historic benchmark were in scenarios with low or zero prevalence (see Figure 6). 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  $\Lambda$ .

A more general approach to managing the risk in surrogate prediction error could be uncertainty estimation. Unlike the approach in (Reiker et al. 2021), RNNs provide point estimates of the IBM outputs. Recent advances have shown that RNNs can be adapted to output probability densities, (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).

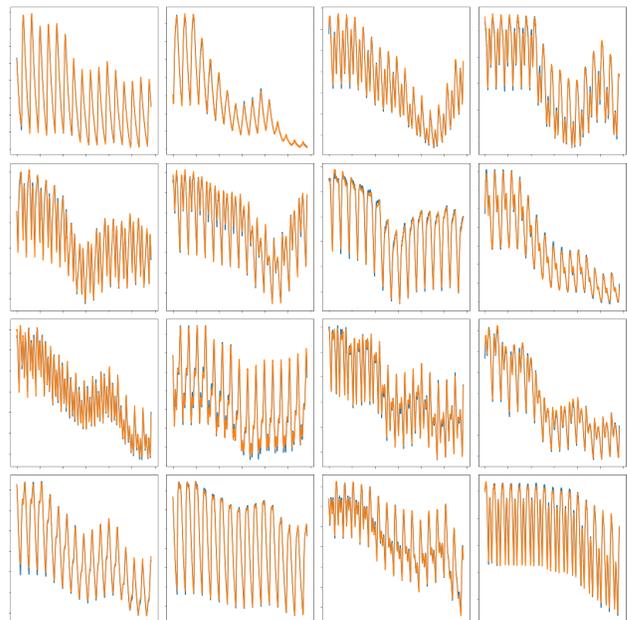


Figure 5: 16 trajectories with the lowest MSE from the historic benchmark. The blue and orange trajectories show daily prevalence in simulation over 16 years for the IBM and surrogate models respectively.

Finally, sharing a surrogate model is more energy efficient and equitable than an IBM. 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 IBM's computational requirements making analyses more feasible in the settings where they are most needed.

## 6.2 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 a 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

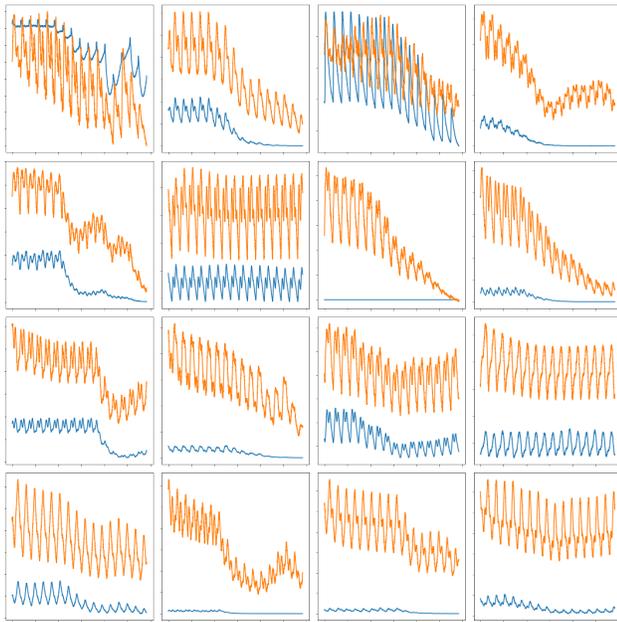


Figure 6: 16 cases with the highest MSE from the historic benchmark. The blue and orange trajectories show daily prevalence in simulation over 16 years for the IBM and surrogate models respectively.

et al. 2017), taking care to avoid the vanishing/exploding gradient problems.

## 7 Conclusion

We believe that the model presented here can be a valuable tool for guiding future malaria modelling, and serves as a proof of concept for the utility of surrogate modelling in infectious disease epidemiology. Given the limitations we have explored, future work will focus on improving the surrogate’s fit and generalising the surrogate structure to adapt it to more varied prediction tasks.

## Acknowledgements

This research was funded in whole, or in part, by the Wellcome Trust [Grant number 220900/Z/20/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

G.C., T.M.W., P.W., A.G. and S.B. acknowledge support from the MRC Centre for Global Infectious Disease Analysis (MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement, and also part of the ED-CTP2 programme supported by the European Union. SB is funded by the National Institute for Health Research (NIHR) Health Protection Research Unit in Modelling and Health Economics, a partnership between the UK Health Security Agency, Imperial College London and LSHTM (grant code NIHR200908). Disclaimer: “The views expressed are those

of the author(s) and not necessarily those of the NIHR, UK Health Security Agency or the Department of Health and Social Care.” S.B. acknowledges support from the Novo Nordisk Foundation via The Novo Nordisk Young Investigator Award (NNF20OC0059309). SB acknowledges support from the Danish National Research Foundation via a chair grant. S.B. acknowledges support from The Eric and Wendy Schmidt Fund For Strategic Innovation via the Schmidt Polymath Award (G-22-63345) which also supports GDC. SF acknowledges the EPSRC (EP/V002910/2). T.M.W. was supported by the Bill and Melinda Gates Foundation (INV-647002606). P.W. is funded by the Bill and Melinda Gates Foundation (INV-043624).

## References

- Bai, T.; Luo, J.; Zhao, J.; Wen, B.; and Wang, Q. 2021. Recent Advances in Adversarial Training for Adversarial Robustness. *ArXiv:2102.01356* [cs].
- Bhatt, S.; Weiss, D.; Cameron, E.; Bisanzio, D.; Mappin, B.; Dalrymple, U.; Battle, K.; Moyes, C.; Henry, A.; Eckhoff, P.; Wenger, E.; Briët, O.; Penny, M.; Smith, T.; Bennett, A.; Yukich, J.; Eisele, T.; Griffin, J.; Fergus, C.; Lynch, M.; Lindgren, F.; Cohen, J.; Murray, C.; Smith, D.; Hay, S.; Cibulskis, R.; and Gething, P. 2015. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*, 526(7572): 207–211.
- Birrell, P.; Blake, J.; van Leeuwen, E.; Gent, N.; and De Angelis, D. 2021. Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1829): 20200279. Publisher: Royal Society.
- Cameron, E.; Battle, K. E.; Bhatt, S.; Weiss, D. J.; Bisanzio, D.; Mappin, B.; Dalrymple, U.; Hay, S. I.; Smith, D. L.; Griffin, J. T.; Wenger, E. A.; Eckhoff, P. A.; Smith, T. A.; Penny, M. A.; and Gething, P. W. 2015. Defining the relationship between infection prevalence and clinical incidence of *Plasmodium falciparum* malaria. *Nature Communications*, 6(1): 8170. Number: 1 Publisher: Nature Publishing Group.
- Cho, K.; van Merriënboer, B.; Bahdanau, D.; and Bengio, Y. 2014. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv:1409.1259* [cs, stat]. *ArXiv*: 1409.1259.
- DHS. 2022. The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs. <https://dhsprogram.com/>. Accessed: 2022-05-12.
- Djaafara, B. A.; Imai, N.; Hamblion, E.; Impouma, B.; Donnelly, C. A.; and Cori, A. 2021. A Quantitative Framework for Defining the End of an Infectious Disease Outbreak: Application to Ebola Virus Disease. *American Journal of Epidemiology*, 190(4): 642–651.
- Eaton, J. W.; Brown, T.; Puckett, R.; Glaubius, R.; Mutai, K.; Bao, L.; Salomon, J. A.; Stover, J.; Mahy, M.; and Hallett, T. B. 2019. The Estimation and Projection Package Age-Sex Model and the r-hybrid model: new tools for estimating HIV incidence trends in sub-Saharan Africa. *AIDS (London, England)*, 33(Suppl 3): S235–S244.

- Griffin, J. T. 2016. Is a reproduction number of one a threshold for *Plasmodium falciparum* malaria elimination? *Malaria Journal*, 15(1): 389.
- Griffin, J. T.; Ferguson, N. M.; and Ghani, A. C. 2014. Estimates of the changing age-burden of *Plasmodium falciparum* malaria disease in sub-Saharan Africa. *Nature Communications*, 5(1): 3136. Number: 1 Publisher: Nature Publishing Group.
- Griffin, J. T.; Hollingsworth, T. D.; Okell, L. C.; Churcher, T. S.; White, M.; Hinsley, W.; Bousema, T.; Drakeley, C. J.; Ferguson, N. M.; Basáñez, M.-G.; and Ghani, A. C. 2010. Reducing *Plasmodium falciparum* Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies. *PLOS Medicine*, 7(8): e1000324. Publisher: Public Library of Science.
- Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. ArXiv:1612.01474 [cs, stat].
- MacKay, D. J. C. 1992. A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4(3): 448–472.
- McKinley, T. J.; Vernon, I.; Andrianakis, I.; McCreesh, N.; Oakley, J. E.; Nsubuga, R. N.; Goldstein, M.; and White, R. G. 2018. Approximate Bayesian Computation and Simulation-Based Inference for Complex Stochastic Epidemic Models. *Statistical Science*, 33(1): 4–18. Publisher: Institute of Mathematical Statistics.
- Neal, R. M. 2011. MCMC using Hamiltonian dynamics. arXiv:1206.1901 [physics, stat]. ArXiv: 1206.1901.
- Rahimi, I.; Chen, F.; and Gandomi, A. H. 2021. A review on COVID-19 forecasting models. *Neural Computing and Applications*.
- Reiker, T.; Golumbeanu, M.; Shattock, A.; Burgert, L.; Smith, T. A.; Filippi, S.; Cameron, E.; and Penny, M. A. 2021. Emulator-based Bayesian optimization for efficient multi-objective calibration of an individual-based model of malaria. *Nature Communications*, 12(1): 7212. Number: 1 Publisher: Nature Publishing Group.
- Sherrard-Smith, E.; Winskill, P.; Hamlet, A.; Ngufor, C.; N’Guessan, R.; Guelbeogo, M. W.; Sanou, A.; Nash, R. K.; Hill, A.; Russell, E. L.; Woodbridge, M.; Tungu, P. K.; Kont, M. D.; Mclean, T.; Fornadel, C.; Richardson, J. H.; Donnelly, M. J.; Staedke, S. G.; Gonahasa, S.; Protopopoff, N.; Rowland, M.; and Churcher, T. S. 2022. Optimising the deployment of vector control tools against malaria: a data-informed modelling study. *The Lancet Planetary Health*, 6(2): e100–e109. Publisher: Elsevier.
- Sinka, M. E.; Bangs, M. J.; Manguin, S.; Rubio-Palis, Y.; Chareonviriyaphap, T.; Coetzee, M.; Mbogo, C. M.; Hemingway, J.; Patil, A. P.; Temperley, W. H.; Gething, P. W.; Kabaria, C. W.; Burkot, T. R.; Harbach, R. E.; and Hay, S. I. 2012. A global map of dominant malaria vectors. *Parasites & Vectors*, 5(1): 69.
- TensorflowProbability. 2022. TensorFlow Probability. <https://www.tensorflow.org/probability/overview>. Accessed: 2022-07-29.
- Trape, J.-F.; and Rogier, C. 1996. Combating malaria morbidity and mortality by reducing transmission. *Parasitology Today*, 12(6): 236–240.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. arXiv:1706.03762 [cs]. ArXiv: 1706.03762.
- Watson, O. J.; Barnsley, G.; Toor, J.; Hogan, A. B.; Winskill, P.; and Ghani, A. C. 2022. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *The Lancet Infectious Diseases*.
- Wen, Q.; Zhou, T.; Zhang, C.; Chen, W.; Ma, Z.; Yan, J.; and Sun, L. 2022. Transformers in Time Series: A Survey. arXiv:2202.07125 [cs, eess, stat]. ArXiv: 2202.07125.
- Winskill, P.; Slater, H. C.; Griffin, J. T.; Ghani, A. C.; and Walker, P. G. T. 2017. The US President’s Malaria Initiative, *Plasmodium falciparum* transmission and mortality: A modelling study. *PLOS Medicine*, 14(11): e1002448. Publisher: Public Library of Science.
- World Health Organization. 2021. *Global Technical Strategy For Malaria - 2016 - 2030*. Geneva: World Health Organization. ISBN 978-92-4-004049-6. Section: liv, 263 p.
- WorldBank. 2022. World Bank Open Data | Data. <https://data.worldbank.org/>. Accessed: 2022-07-29.
- WorldPopulationReview. 2022. 2022 World Population by Country. <https://worldpopulationreview.com/>. Accessed: 2022-07-29.