

TOPICAL REVIEW

Bayesian Calibration to Address the Challenge of Antimicrobial Resistance: A Review

CONOR ROSATO¹, PETER L. GREEN², JOHN HARRIS³, SIMON MASKELL⁴,
WILLIAM HOPE¹, ALESSANDRO GERADA¹, AND ALEX HOWARD¹

¹Department of Pharmacology and Therapeutics, University of Liverpool, L69 7BE Liverpool, U.K.

²Department of Mechanical Engineering, University of Liverpool, L69 7BE Liverpool, U.K.

³United Kingdom Health Security Agency (UKHSA), SW1P 3JR London, U.K.

⁴Department of Electrical Engineering and Electronics, University of Liverpool, L69 7BE Liverpool, U.K.

Corresponding author: Conor Rosato (cmrosa@liverpool.ac.uk)

This work was supported in part by the Wellcome Trust under Grant 226691/Z/22/Z. The work of Conor Rosato, William Hope, Alessandro Gerada, and Alex Howard was supported by the Wellcome CAMO-Net U.K. under Grant 226691/Z/22/Z. The work of Simon Maskell was supported by EPSRC through the Big Hypotheses under Grant EP/R018537/1.

ABSTRACT Antimicrobial resistance (AMR) emerges when disease-causing microorganisms develop the ability to withstand the effects of antimicrobial therapy. This phenomenon is often fueled by the human-to-human transmission of pathogens and the overuse of antibiotics. Over the past 50 years, increased computational power has facilitated the application of Bayesian inference algorithms. In this comprehensive review, the basic theory of Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) methods are explained. These inference algorithms are instrumental in calibrating complex statistical models to the vast amounts of AMR-related data. Popular statistical models include hierarchical and mixture models as well as discrete and stochastic epidemiological compartmental and agent based models. Studies encompassed multi-drug resistance, economic implications of vaccines, and modeling AMR in vitro as well as within specific populations. We describe how combining these topics in a coherent framework can result in an effective antimicrobial stewardship. We also outline recent advancements in the methodology of Bayesian inference algorithms and provide insights into their prospective applicability for modeling AMR in the future.

INDEX TERMS Antimicrobial resistance, antimicrobial stewardship, approximate Bayesian computation, Bayesian inference, epidemiology, Markov chain Monte Carlo, sequential Monte Carlo.

I. INTRODUCTION

The routine application of Bayesian algorithms, including Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) methods, in the field of health sciences has been instrumental in combining disparate data sources to make disease-related inferences when closed-form solutions are not known. An advantageous feature of using Bayesian methods is the incorporation of prior knowledge through probability distributions associated with each data source, distinguishing Bayesian methods from other prediction-based methods. Calibrating models using Bayesian methods allows the propagation of uncertainty through the model, providing probability distributions over predictions. Despite the utility

of Bayesian methods, a potential obstacle to widespread usage is the required computational demand, which can be addressed by deploying Bayesian algorithms across multiple computers in a distributed framework. As healthcare-related data continues to grow in both magnitude and complexity, the computational resources required for gaining insights into challenges like AMR also increase. A number of review articles outlining the theory, principles, challenges and working examples of performing Bayesian analysis in healthcare-related disciplines can be found in [1], [2], and [3].

The World Health Organization (WHO) has recently classified AMR as a significant global threat to both health and society at large [4]. While the discovery of new antimicrobial agents remains crucial in combating AMR, the effectiveness of this strategy is limited without a comprehensive understanding and optimization of both existing agents and

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko¹.

future agents yet to be licensed. Certain challenges particular to AMR are especially suited to modelling using Bayesian techniques. For example, imperfect diagnostic criteria and laboratory tests can lead to uncertainty in the underlying patient disease and colonisation state. Furthermore, the impact of interventions extends beyond the duration of most clinical studies, since the generation of AMR in a population can take years to manifest.

Antimicrobial Stewardship (AMS) constitutes a comprehensive initiative aimed at fostering the judicious utilization of antimicrobials based on data obtained from the delivery of routine clinical care [5]. The primary goals of AMS include reducing AMR and curbing the dissemination of infections stemming from multidrug-resistant organisms, all while taking into account the individual's infection risk [6]. Effective AMS relies on a combination of measures, including infection control, precise detection of AMR in diagnostic laboratory settings and population-level surveillance. Quantifying these measures in isolation is challenging for several reasons, including insufficient quantities of data, unmonitored resistance rates, the complexity of required models, and the dynamics of co-infection with sensitive and resistant strains [7]. The Bayesian analyses conducted in the reviewed articles have the potential to bridge the gap between Artificial Intelligence (AI) and AMR, as outlined in [8].

The contribution of this paper is to provide a comprehensive and up-to-date overview of the current state of knowledge in the combined topics of AMR and Bayesian modelling. Some of the key points are summarised below:

- **Integration of Bayesian modelling in AMR research:** This review aims to bridge the gap between AMR and Bayesian modeling by exploring and reviewing studies and applications where Bayesian methods have been employed to analyse and model AMR data. Several review articles have detailed the applications of Machine Learning (ML) and AI in tackling antimicrobial resistance (AMR) [9], [10], [11], [12]. However, this review article is the first to exclusively focus on Bayesian methods.
- **Uncertainty Quantification:** This paper highlights the importance of Bayesian modelling in quantifying uncertainty. Bayesian methods allow for the incorporation of prior knowledge, which is especially crucial when dealing with incomplete or heterogeneous data sources.
- **Prediction and Forecasting:** This paper outlines how Bayesian modelling is an effective method for predicting and forecasting trends in AMR. Selected studies have shown how the evolution of resistance patterns within certain populations are affected by interventions over time.
- **Methodological Overview:** A detailed methodological overview of the development and progression of Bayesian inference algorithms is provided, with a specific emphasis on efficiently deploying such models on distributed computing platforms. A description of commonly used models, such as hierarchical models,

TABLE 1. A list of pathogens and corresponding studies that were included in this review. The process of selecting studies is outlined in Section I-A.

Pathogen	Studies
<i>Mycobacterium tuberculosis</i>	[13]–[22]
Methicillin-resistant <i>Staphylococcus aureus</i> and methicillin-sensitive <i>Staphylococcus aureus</i>	[23]–[33]
<i>Escherichia coli</i>	[34]–[38]
<i>Neisseria gonorrhoea</i>	[39]–[44]
<i>Klebsiella pneumoniae</i>	[36], [37], [45]
<i>Acinetobacter baumannii</i>	[46], [47]
<i>Streptococcus pneumoniae</i>	[48], [49]
<i>Pseudomonas aeruginosa</i>	[50]
<i>Salmonella</i> spp	[51], [52]
<i>Enterococcus</i> spp	[53], [54]

and their application in understanding the dynamics of AMR is provided.

- **Identifying Gaps and Challenges:** Finally, this paper identifies gaps in the current literature and the challenges associated with the application of Bayesian modelling in AMR research.

The structure of this review is organised as follows: Section IV defines Bayes' theorem, while Sections IV-A and IV-E outline MCMC and SMC methods, respectively. Section V covers statistical methods for evaluating Bayesian algorithms. In Sections VI, a comprehensive list of techniques to model AMR is provided. Finally, Section VIII discusses potential directions for future work, including recent advancements in both modeling and sampling algorithms.

A. SEARCH CRITERIA

This review encompasses studies examining a diverse array of causative pathogens, as identified in internationally-recognised lists which highlight their critical importance in AMR. These lists include the WHO Priority Pathogens List and the ESKAPE pathogens [55], [56]. The search criteria for this review did not identify Bayesian modelling studies of AMR in *Helicobacter pylori*, *Campylobacter* spp and *Enterobacter* spp, which are pathogens identified in the prior lists. As outlined in Table 1, the most commonly studied pathogens are *Mycobacterium tuberculosis* and methicillin-resistant *Staphylococcus aureus* (MRSA), highlighting their significant impact on health outcomes. For instance, *M. tuberculosis* can develop resistance to treatment and, due to its mode of transmission, can lead to community outbreaks of multidrug-resistant *M. tuberculosis*. Similarly, MRSA can have severe consequences within healthcare settings. Since both of these pathogens are communicable, Bayesian calibration is a widely used approach for epidemiological transmission models.

The search engines considered for the advanced search of this review included Web of Science, PubMed-MEDLINE and Scopus. The search was not restricted by date and after publications with no full text available, review articles with no novelty and duplicates were eliminated, a total of 567 articles were considered. To be considered, an article had

to include Bayesian modelling of bacterial AMR in humans (veterinary and environmental studies were excluded). Viral and vector-borne diseases such as HIV and malaria were not considered. Population pharmacokinetic (popPK) models and phylogenetic tree analyses were considered out of scope and excluded. However, a brief introduction to popPK and phylogenetic trees are provided in Section VIII-D. In terms of Bayesian modelling, studies had to use Approximate Bayesian Computation (ABC), MCMC or SMC methods. Therefore studies that employed, for example naïve Bayes ML classifiers and Bayesian information criterion, were not included. This resulted in 43 studies. Additionally, references of selected papers and Google Scholar were manually searched and this identified 22 research articles that were not captured by the advanced search queries.

II. CONTEXTUAL EXAMPLE

Let's consider the scenario where we aim to estimate the prevalence of AMR, characterized by parameters of the model that has D dimensions, $\theta = \{\theta_1, \theta_2, \dots, \theta_D\}$, within a population using observed laboratory test outcomes denoted by $\mathbf{y} = \{y_1, y_2, \dots, y_N\}$. Each element in \mathbf{y} signifies whether the bacterium in the sample exhibits resistance ($y = 1$) or susceptibility ($y = 0$) to the antibiotic being studied.

III. BAYES' THEOREM

Bayes' theorem, outlined in (1), is a mathematical formula used to determine the probability of θ conditional on \mathbf{y} :

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})}, \quad (1)$$

where $p(\mathbf{y}|\theta)$ is the probability of \mathbf{y} given θ and $p(\theta)$ and $p(\mathbf{y})$ are the independent probabilities of θ and \mathbf{y} , respectively. Equation (1) provides a method for revising predictions after considering newly obtained evidence.

IV. BAYESIAN INFERENCE

Bayesian inference uses Bayes' theorem to update ones personal belief after observing the data. The parameters of a statistical model are often unknown and difficult to measure directly and need to be inferred from the data. Prior information can be attributed to θ in the form of a probability distribution, $p(\theta)$. This prior distribution may encapsulate our initial beliefs about the prevalence of AMR within the population, representing uncertainty regarding θ before encountering any data. The likelihood is given by the conditional probability of \mathbf{y} given the prevalence of resistance, $p(\mathbf{y}|\theta)$. The likelihood describes the probability that the data was produced from the statistical model whose parameters, θ , are fixed to a particular value. If the prior and likelihood can be defined, the parameter posterior distribution, $p(\theta|\mathbf{y})$, can be calculated using Bayes' theorem:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\theta)p(\theta)}{\int_{\theta} p(\mathbf{y}|\theta)p(\theta)d\theta} \propto p(\mathbf{y}|\theta)p(\theta). \quad (2)$$

Bayes' theorem expresses a probability distribution over θ which is conditional on \mathbf{y} and allows us to update our belief

about the prevalence of resistance based on the observed test results.

If the likelihood and prior can be easily defined, the calculations are often algebraic or can be performed using statistical software. This explicitness allows for easy computation of the posterior distribution. When a closed-form solution for the integral in equation (2) cannot be found, approximating it can pose challenges, particularly in high dimensions where it may become exceedingly difficult or even impossible. In more complex scenarios, where analytical solutions are not feasible, numerical methods like MCMC may be employed to draw samples from the posterior. The posterior is typically estimated up to a normalisation constant, given by the integral which marginalises out θ .

Table 2 outlines a number of Bayesian inference algorithms as well as corresponding research studies in which they have been applied when modelling AMR. The algorithms include MCMC, ABC and SMC based algorithms which are described in Sections IV-A, IV-C and IV-E, respectively.

A. MARKOV CHAIN MONTE CARLO

MCMC methods provide a mechanism for sampling from an arbitrary probability distribution $\pi(\theta)$ that, in the context of Bayesian inference, is set proportional to the posterior:

$$\pi(\theta) \propto p(\mathbf{y}|\theta)p(\theta). \quad (3)$$

MCMC methods generate a Markov chain that forms a sequence of correlated samples from $\pi(\theta)$. A Markov chain is Markovian in that the current sample only depends on the previous. The transition operator $T(\theta, \theta')$ describes the probability of going from the current sample θ to the proposed sample θ' . For a Markov chain to generate samples from $\pi(\theta)$, it must meet specific criteria. These criteria encompass the chain being ergodic and possessing a stationary distribution. The stationary distribution is the distribution that the Markov chain converges to after running for a sufficiently long time. It represents the desired distribution from which we want to sample.

Detailed balance is a condition that the transition probabilities of the Markov chain must satisfy to ensure that the stationary distribution is indeed the desired target distribution. Mathematically, detailed balance is expressed as:

$$\pi(\theta)T(\theta, \theta') = \pi(\theta')T(\theta', \theta). \quad (4)$$

The time until the Markov chain has converged and reached its stationary distribution is dependent on its initial starting point in the parameter space and choice of transition operator $T(\theta, \theta')$. Popular MCMC methods are described in the subsequent sections.

1) METROPOLIS-HASTINGS RANDOM WALK

As seen in Table 2, the Metropolis-Hastings Random Walk (MHRW) is the most commonly used MCMC algorithm when modelling AMR. It splits $T(\theta, \theta')$ into two distinct steps: the proposal step and accept/reject step.

TABLE 2. A list of Bayesian algorithms and corresponding studies that were selected in the review article. The process of selecting studies is outlined in Section I-A.

Sampling Algorithm	Proposal	Studies
ABC	-	[13], [20], [21], [30], [48]
MCMC	MH	[14]–[18], [27], [34], [35], [40], [46], [47], [54], [57], [58]
Adaptive MCMC	MH	[18], [19], [57]
MCMC	Gibbs	[26], [28], [38], [50], [52]–[54]
MCMC	Langevin	[38]
MCMC	NUTS	[31], [37], [41], [44], [45], [51]
SMC Sampler	ABC proposal	[43]
Particle-MCMC	MH	[39], [59]
SMC ²	MH	[36]

The proposal step draws a vector of states θ' from a proposal distribution $q(\theta'|\theta)$. A simple example of a MHRW proposal is the Normal distribution:

$$q(\theta'|\theta) = \mathcal{N}(\theta'; \theta, \Sigma), \quad (5)$$

where $\Sigma \in \mathbb{R}^{D \times D}$ is the covariance, which is selected by the user.

The accept/reject step ensures that the Markov chain explores the state space efficiently by probabilistically accepting proposed moves, while detailed balance ensures that the stationary distribution of the Markov chain is the desired target distribution. Together, these components ensure the correctness and convergence of the Metropolis-Hastings algorithm. The acceptance probability, $\alpha(\theta, \theta')$, is given by

$$\alpha(\theta, \theta') = \min \left\{ 1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right\}. \quad (6)$$

If the proposal is symmetric $q(\theta|\theta') = q(\theta'|\theta)$, as is with the MHRW, the acceptance probability simplifies to

$$\alpha(\theta, \theta') = \min \left\{ 1, \frac{\pi(\theta')}{\pi(\theta)} \right\}. \quad (7)$$

The acceptance rate is a measure of the proportion of proposed moves that are accepted during the sampling process. Proposing steps that are too small can result in slow exploration which can lead to θ' being proposed too close to θ , causing high autocorrelation within the Markov chain. Taking steps that are too large can result in low acceptance rates and inefficient sampling as θ' may fall outside high probability regions of $\pi(\theta)$.

Figure 1 illustrates the importance of selecting an appropriate σ . It is evident that using a smaller value of σ in Figure 1(a) takes longer to reach the probability mass when compared to using a larger value in Figure 1(c). The acceptance rate for Figures 1(a), (b) and (c) are 0.906, 0.602 and 0.084, respectively. To implement the acceptance step in (6) algorithmically, a random variable u is drawn from a Uniform distribution on $[0, 1]$. If $u < \alpha(\theta, \theta')$ then θ' is accepted as the next state of the Markov chain. If $u > \alpha(\theta, \theta')$, θ' is rejected and the Markov chain's state remains equal to θ . This process is repeated for M MCMC iterations.

The steps performed by the MHRW algorithm are outlined below:

- 1) Assign and draw initial values for θ from respective prior distributions.
- 2) Following (3), calculate the parameter posterior distribution by multiplying the prior and the likelihood.
- 3) Propose a new vector of θ' via the proposal distribution in (5).
- 4) Accept θ' using the acceptance step in (6).
- 5) Repeat steps 2-4 for M iterations to produce a Markov chain consisting of M samples.
- 6) Perform diagnostic checks (see Section V) on the resulting Markov chain to determine if the sampling process is producing samples that are representative of $\pi(\theta)$.

The MHRW sampler may encounter the ‘‘curse of dimensionality’’; as the dimensionality of θ increases, efficiently sampling from $\pi(\theta)$ becomes increasingly challenging.

2) GIBBS

Similarly to the MHRW sampler, Gibbs sampling randomly walks through the parameter space. Constructing a Markov chain using Gibbs sampling [60] can occur iteratively on each dimension of θ , (or a partition of dimensions (see Section IV-A3)), conditional on the most recent values of the other dimensions. Denoted by the superscript, at the m th iteration of the MCMC simulation, random samples are proposed from all conditional distributions:

$$p(\theta_1^{(m)} | \theta_2^{(m-1)}, \dots, \theta_D^{(m-1)}, \mathbf{y}), \quad (8a)$$

$$p(\theta_2^{(m)} | \theta_1^{(m)}, \dots, \theta_D^{(m-1)}, \mathbf{y}), \quad (8b)$$

$$\dots, \quad (8c)$$

$$p(\theta_D^{(m)} | \theta_1^{(m)}, \theta_2^{(m)}, \dots, \mathbf{y}). \quad (8d)$$

If the conditional distributions are known, a high-dimensional problem can be broken down into a sequence of smaller, low-dimensional conditional simulations. If, however, the conditional distributions are not known, other methods such as the MHRW will need to be used. The Gibbs sampler is a special case of MHRW where θ' is always accepted due to the acceptance probability ratio in (6) being one.

Two advantages of Gibbs sampling over MHRW are not having to define a proposal distribution and proposals

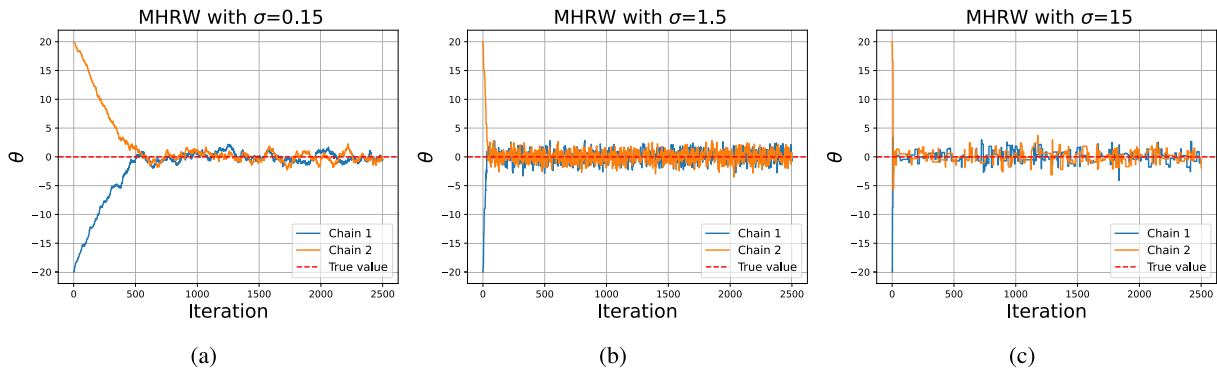


FIGURE 1. Two Markov chains with different initial starting points when sampling from a $\mathcal{N}(0, 1)$ when using (a) $\sigma = 0.15$, (b) $\sigma = 1.5$ and $\sigma = 15$ in the proposal distribution. The horizontal dashed red line is the true value. The average acceptance rate for both chains in subplots (a), (b) and (c) are 0.906, 0.602 and 0.084, respectively.

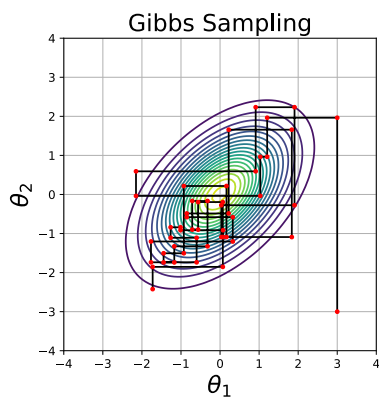


FIGURE 2. Sampling a 2-dimensional Gaussian distribution with Gibbs with 0 means for θ_1 and θ_2 for 50 MCMC iterations. The solid black lines show the trajectory between two successive accepted samples represented by the red dots.

always being accepted. However, to employ Gibbs sampling, the conditional distributions need to be derived which can be problematic. The efficiency of Gibbs sampling can also be a concern if the dimensions of θ are highly correlated. Figure 2 shows an example where the Gibbs sampler struggles to sample efficiently from a 2-dimensional Gaussian distribution.

Gibbs sampling is particularly suitable for calibrating complex hierarchical Bayesian models, as discussed in Section VI-B. These models can be broken down into smaller sub-models, and Gibbs sampling facilitates the iterative sampling of these sub-models. If conjugacy exists for a subset of parameters within the model, Gibbs sampling can be employed for those conjugate parameters. An example illustrating this can be found in [24], where the transmission parameter for *Staphylococcus aureus* has a Gamma prior, which is conjugate to the Binomial likelihood.

3) METROPOLIS WITHIN GIBBS

As explained in Section IV-A2, the Gibbs sampler is suitable when the conditional distributions of the posterior are

known. However, when dealing with challenging conditional distributions, a practical approach is to employ a MH step to sample a subset of parameters for which the conditionals are not known, resulting in block updates. This method is known as the Metropolis Within Gibbs sampler [61] and proves to be an attractive option when modeling AMR [26], [28], [33], [38], [50], [53].

4) DELAYED REJECTION ADAPTIVE METROPOLIS

The need to fine-tune the proposal e.g. Σ in Equation (5) can be time-consuming, given its problem-specific nature. Adaptive MCMC [62] and the Delayed Rejection Adaptive Metropolis (DRAM) algorithm [63], [64] address this challenge by enabling automatic adjustments of exploration strategies. This adaptability, grounded in the characteristics of the data and model complexity, significantly improves the efficiency and effectiveness of parameter estimation. This is particularly valuable in scenarios like AMR, where disease dynamics are influenced by various factors and exhibit nonlinear behavior [18], [19], [57].

5) HAMILTONIAN MONTE CARLO

As discussed in previous sections, MHRW, Gibbs and DRAM samplers explore $\pi(\theta)$ by randomly proposing a new state θ' from the current state θ using some proposal distribution. In high-dimensional problems, this random exploration can be inefficient. Hamiltonian Monte Carlo (HMC) addresses this issue by generating new proposal samples based on gradient information about $\pi(\theta)$, encouraging greater exploration of the parameter space. HMC was first developed in the late 1980s [65] and has gained popularity in the last decade as an effective approach for implementing MCMC [66], [67].

HMC simulates a trajectory from θ to θ' by employing Hamilton's equations. Conceptually, Hamilton's equations represent a frictionless puck moving on a surface, such that the potential energy is defined as the negative logarithm of $\pi(\theta)$, $U(\theta) = -\log(\pi(\theta))$. Figure 3 outlines this phenomenon. The surface is analogous to $\pi(\theta)$ which can have hills and valleys. The height of the surface corresponds

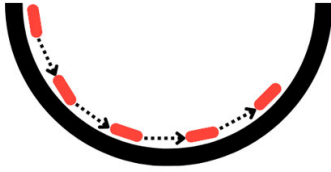


FIGURE 3. HMC analogy: The starting position of the left most puck (red) in a frictionless bowl (black solid line) will have potential energy but no kinetic energy. As the puck follows the trajectory of the arrows, potential energy is changed to kinetic energy.

to the potential energy function of the system. The higher the point on the surface, the higher the potential energy. The puck's kinetic energy, $K(\mathbf{r})$, corresponds to the energy associated with the momentum of the "puck" which is denoted \mathbf{r} . The faster the puck is moving, the higher its kinetic energy. The total energy of the system is the sum of its potential and kinetic energies $H(\boldsymbol{\theta}, \mathbf{r}) = K(\mathbf{r}) + U(\boldsymbol{\theta})$. In Hamiltonian dynamics, the total energy remains constant over time, as long as there are no external forces acting on the system (i.e., in our frictionless scenario). Hamiltonian dynamics describe how the puck's position and momentum change over time. These changes are governed by Hamilton's equations, which depend on the gradient of the potential energy (surface) $\nabla U(\boldsymbol{\theta})$ and the momentum of the system. Hamilton's equations are defined as:

$$\frac{d\boldsymbol{\theta}}{dt} = \frac{\partial H(\boldsymbol{\theta}, \mathbf{r})}{\partial \mathbf{r}}, \quad \frac{d\mathbf{r}}{dt} = -\frac{\partial H(\boldsymbol{\theta}, \mathbf{r})}{\partial \boldsymbol{\theta}}. \quad (9)$$

The joint density over $\boldsymbol{\theta}$ and \mathbf{r} is:

$$p(\boldsymbol{\theta}, \mathbf{r}) \propto \exp(-H(\boldsymbol{\theta}, \mathbf{r})) \quad (10)$$

$$= \exp(-K(\mathbf{r})) \cdot \exp(-U(\boldsymbol{\theta})) \quad (11)$$

$$= p(\boldsymbol{\theta})p(\mathbf{r}). \quad (12)$$

As $\boldsymbol{\theta}$ and \mathbf{r} are independent in (12), \mathbf{r} can be sampled from any distribution. For simplicity, this is often chosen to be Gaussian with zero mean.

A numerical integrator is necessary to discretize the differential equations in (9). The leapfrog numerical integrator is widely used as it is reversible and therefore maintains detailed balance. The leapfrog step alternates between updating the position of the puck based on its momentum and updating its momentum based on the gradient of the potential energy. Simulating a leapfrog trajectory with some momentum \mathbf{r} causes $\boldsymbol{\theta}$ to transition to $\boldsymbol{\theta}'$. When starting at $\boldsymbol{\theta}'$, if the same momentum is negated to $-\mathbf{r}$, $\boldsymbol{\theta}'$ follows the same trajectory back to $\boldsymbol{\theta}$. A single iteration of HMC consists of L leapfrog steps:

$$\mathbf{r}' = \mathbf{r} + 0.5 \cdot \epsilon \cdot \nabla U(\boldsymbol{\theta}), \quad (13)$$

$$\boldsymbol{\theta}' = \boldsymbol{\theta} + \epsilon \cdot \mathbf{r}', \quad (14)$$

$$\mathbf{r}' = \mathbf{r}' + 0.5 \cdot \epsilon \cdot \nabla U(\boldsymbol{\theta}'), \quad (15)$$

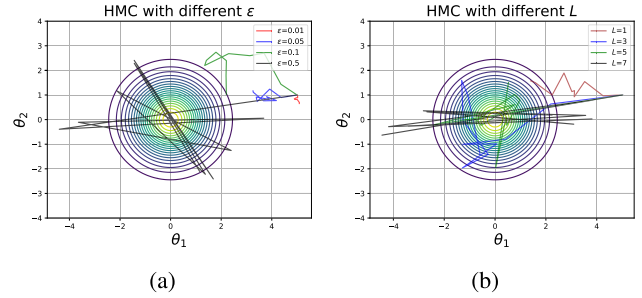


FIGURE 4. Sampling a 2-dimensional Gaussian distribution with HMC when (a) changing ϵ and $L = 5$ and (b) changing parameter L with $\epsilon = 0.1$ for 10 MCMC iterations.

where ϵ is the step size. The acceptance criterion for HMC is defined as the following special case of (6):

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min \left\{ 1, \frac{\exp \left\{ U(\boldsymbol{\theta}') - \frac{1}{2} \mathbf{r}' \cdot \mathbf{r}' \right\}}{\exp \left\{ U(\boldsymbol{\theta}) - \frac{1}{2} \mathbf{r} \cdot \mathbf{r} \right\}} \right\}. \quad (16)$$

Simulating the dynamics of the system using Hamiltonian dynamics and leapfrog integration and employing the acceptance step in (16), we can sample from $\pi(\boldsymbol{\theta})$. The sampled positions of the puck correspond to samples from $\pi(\boldsymbol{\theta})$. The samples generated are governed by a predefined number of steps L of size ϵ , determined by the user. The choice of L and ϵ is problem specific as HMC is sensitive to the choice of these parameters. If L is too large, computation time can be wasted, as the trajectory might end up close to where it started. On the other hand, if L is too small, the proposal may exhibit random-walk behavior. Figure 4 illustrates an example when sampling from a 2-dimensional Gaussian distribution. Figure 4(a) and (b) depict how the choice of step-size and parameter L , respectively, can impact the efficiency of the sampler.

6) METROPOLIS HASTINGS LANGEVIN ALGORITHM

The Metropolis Adjusted Langevin Algorithm (MALA) [68] can be regarded as a specific instance of HMC when $L = 1$. In MALA, the gradient of the posterior $\nabla \pi(\boldsymbol{\theta})$ is integrated into the Metropolis proposal, serving as a drift term that proposes $\boldsymbol{\theta}'$ in regions of higher probability within $\pi(\boldsymbol{\theta})$. An illustration of this approach can be found in [38], where parameters are estimated using both a Metropolis Within Gibbs step and a standalone Metropolis step through the modified Langevin-Hastings algorithm when modeling MIC distributions of *E. coli* tested against ampicillin. The proposal is defined to be

$$\boldsymbol{\theta}' = \mathcal{N} \left(\boldsymbol{\theta}'; \boldsymbol{\theta} + \frac{1}{2} \Gamma \nabla \pi(\boldsymbol{\theta}), \Gamma \right), \quad (17)$$

where $\Gamma = \gamma^2 \mathbf{I}_d$, for step size γ . Selecting an appropriate step size for MALA involves a trade-off between exploration and exploitation, and it often requires experimentation and fine-tuning based on the characteristics of $\pi(\boldsymbol{\theta})$ and the problem at hand.

7) NO-U-TURN SAMPLER

The No-U-Turn Sampler (NUTS) [69] is an extension of HMC which adaptively estimates L . NUTS underpins the probabilistic programming languages (ppls) Stan [70], PyMC3 [71] and Pyro [72] and has been used in a variety of AMR studies [37], [41], [44], [45], [51].

8) REVERSIBLE JUMP - MARKOV CHAIN MONTE CARLO

Sampling with MHRW, Gibbs, HMC and NUTS assumes that the probability distribution $\pi(\theta)$ has a fixed number of dimensions. In some scenarios, there might be multiple competing models that could be applicable within the defined framework. These models could differ in aspects such as dimensions or interpretations of the likelihood function. For instance, consider count data like the number of vancomycin-resistant *enterococci* in a hospital ward [73]; such data could be modeled using a Negative Binomial or Poisson likelihood, depending on the over-dispersion of the count data. Employing MHRW would require conducting independent MCMC runs for each candidate model. Reversible Jump MCMC (RJ-MCMC) [74] is a MHRW-based algorithm designed to sample from $\pi(D, \theta_D | y)$ by jumping between competing models as part of the sampling process. The jump is crafted to maintain detailed balance, ensuring the algorithm's validity. An application of RJ-MCMC can be found in the analysis of nosocomial infection data [53]. In this model, the unobservable times at which patients acquire a drug-resistant organism and the total number of acquisitions are both unknown. Since the dimensionality of the model can change, RJ-MCMC proves to be a suitable method for sampling from the joint distribution of all the model unknowns.

B. REJECTION AND IMPORTANCE SAMPLING

The MCMC methods discussed in Section IV-A are Monte Carlo techniques that build a Markov chain by drawing samples from $\pi(\theta)$. Characteristics of $\pi(\theta)$, such as its mean and variance, can be approximated using the Markov chain. However, when sampling directly from $\pi(\theta)$ is challenging, other Monte Carlo methods like importance sampling and rejection sampling can be employed. Both importance and rejection sampling involve transforming samples drawn from a proposal distribution $q(\theta)$ into samples from $\pi(\theta)$. The choice of $q(\theta)$ can be a practical challenge, especially in high-dimensional spaces, but it can be any distribution that is evaluable at different points.

1) REJECTION SAMPLING

In rejection sampling, a constant c and proposal distribution $q(\theta)$ needs to be chosen such that:

$$cq(\theta) > \pi(\theta), \forall \theta. \quad (18)$$

The idea is that a sample from $q(\theta)$ can be used when accepting a sample from $\pi(\theta)$ with probability $\frac{\pi(\theta)}{cq(\theta)}$. Samples are generated from $q(\theta)$ until θ is accepted. This should

lead to more samples around θ being accepted when $\pi(\theta)$ is relatively large and $q(\theta)$ is relatively small. In contrast, a relatively small $\pi(\theta)$ and large $q(\theta)$ will result in more samples of θ being rejected. The accept/reject step is similar to that as described in the MHRW algorithm (see Section IV-A1). A random variable u is drawn from a Uniform distribution on $[0, 1]$ such that if u is smaller than the acceptance probability, $u < \frac{\pi(\theta)}{cq(\theta)}$, θ will be rejected.

It can, however, be problematic finding a suitable c and $q(\theta)$ such that a high percentage of samples are accepted and $cq(\theta) > \pi(\theta), \forall \theta$.

2) IMPORTANCE SAMPLING

Importance sampling overcomes this by weighting samples and providing an estimator for the expectation of $\pi(\theta)$ by

$$\mathbb{E}[\pi(\theta)] \approx \frac{1}{N} \sum_{i=1}^N \frac{\pi(\theta_i)}{q(\theta_i)}. \quad (19)$$

The weights in IS represent a scaled version of the rejection probabilities and serve to quantify the likelihood that sampled values from $q(\theta)$ originate from $\pi(\theta)$.

Although these two methods are not used to calibrate models to data in the studies identified in this review, they form the basis of the calibration methods described in the subsequent sections.

C. APPROXIMATE BAYESIAN CALIBRATION

The definition of the likelihood in Bayes' theorem (2) is crucial as it encapsulates the probability of the data originating from the model. In certain scenarios, defining the likelihood may be challenging or computationally intensive, especially when calibrating stochastic compartmental models (CMs) or ABMs [13], [20], [21], [30], [48]. Approximate Bayesian Computation (ABC) involves simulating data from the model with a set of parameters θ . A summary statistic, such as the root mean square error (RMSE) or Kullback-Leibler (KL) divergence, quantifies the closeness between the simulated and observed data. Numerous sets of parameters θ are simulated from the prior distribution and assessed against the true observations. The summary statistic acts as a pseudo-likelihood, helping distinguish between different parameter sets. ABC-rejection sampling [13], [20], [30], [48] follows a similar procedure as rejection sampling in Section IV-B. A user-defined tolerance or threshold determines whether a particular θ is accepted or rejected.

An example of ABC applied to a simple AMR related problem can be seen in Figure 5. The example considers estimating the relationship between antibiotic concentration and bacterial growth inhibition, which is a linear relationship. Firstly, data is generated from the simple linear model with slope and intercept parameters fixed at 0.5 and 1.5, with Uniform $[0, 1]$ and $[0, 2]$ priors, respectively. The ABC algorithm samples from the parameter space by simulating data using sampled parameters from the prior and calculates the summary statistic, which is the Mean Absolute

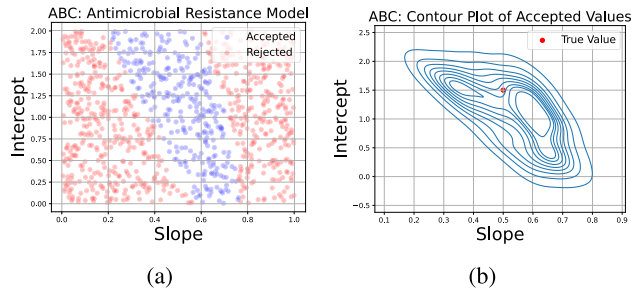


FIGURE 5. Toy example of using approximate Bayesian computation to estimate the parameters governing the linear relationship between antibiotic concentration and bacterial growth inhibition. (a) The blue and red points are the accepted and rejected values, respectively. (b) Is the contour plot of the accepted values with the true set of values indicated by the red dot.

Error (MAE), between the distance between simulated and observed data. Parameters are then accepted if the distance is below a certain tolerance level. Figure 5 (a) depicts the parameter sets that are within the tolerance threshold (blue) and the rejected parameter sets (red). A contour plot of the accepted values can then be seen in Figure 5 (b).

Examples of threshold determination in the context of AMR include simulated *M. tuberculosis* incidence falling within the 95% confidence interval of the true observations [21] or a value resulting in the closest 2.5 or 5% of simulations to the targets not being rejected [48]. Parameters leading to pseudo-likelihoods smaller than a given tolerance constitute the posterior distribution, while those exceeding it are rejected. ABC-rejection sampling may be inefficient if the prior does not closely resemble the posterior (e.g., due to a very uninformative prior) or if the model is high-dimensional, requiring numerous simulations for a high acceptance rate. Intelligent methods for selecting θ , such as Latin Hypercube Sampling [21], exist to address this issue.

D. STATE-SPACE MODELS

State-space models (SSMs) or Hidden Markov Models (HMMs) (often used interchangeably) are commonly employed in time-series problems to represent the probabilistic dependence at time-step t between the hidden (or unobservable) states of a model, denoted as \mathbf{X}_t , and observed variables, denoted as \mathbf{Y}_t . SSMs model how the state \mathbf{X}_{t-1} is transitioned to \mathbf{X}_t through the linear or nonlinear function $f_t(\cdot)$. An observation function, $g_t(\cdot)$, illustrates the relationship between the dynamical model and the observations. A stochastic formulation is presented as follows:

$$\mathbf{X}_t \sim f_t(\mathbf{X}_{t-1}, V_t), \quad (20)$$

$$\mathbf{Y}_t \sim g_t(\mathbf{X}_t, W_t), \quad (21)$$

where the initial state at time 0 is defined to be x_0 and is drawn from the distribution $\mu(\cdot)$, and V_t and W_t are independent and identically distributed process noise. A deterministic representation of an SSM is equivalent to (20) and (21) without the process noise.

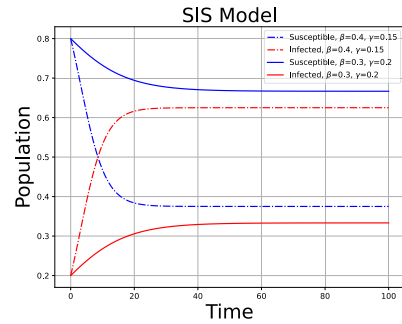


FIGURE 6. Two simulations of the discrete time approximation of the ODEs in (22) and (23) when $\beta = 0.4$, $\gamma = 0.15$ (dashed/dot lines) and $\beta = 0.3$, $\gamma = 0.2$ (solid lines).

1) EXAMPLE OF A DETERMINISTIC SSM

An example of an SSM, which is related to AMR epidemiology, is the Susceptible, Infected, Susceptible (SIS) disease model, a variant of the more widely known Susceptible, Infected, Recovered (SIR) model [75]. The key distinction lies in the assumption that individuals do not gain immunity after recovering from infection in the SIS model. This makes the SIS model particularly relevant when modeling scenarios related to AMR, where individuals can become repeatedly infected with the same pathogen or variant. Some examples include: *P. aeruginosa* [50]; beta-lactamase producing *E. coli* and *K. pneumoniae* [36]; *A. baumannii* [46] and *N. gonorrhoea* [41]. A common approach to representing SSMs are Ordinary differential equation (ODE) [16], [18], [35]. A discrete time approximation of ODEs for the SIS model are defined as

$$S_{t+1} = S_t - \frac{\beta I_t S_t}{P_{op}} - \gamma I_t \Delta t, \quad (22)$$

$$I_{t+1} = I_t + \frac{\beta I_t S_t}{P_{op}} + \gamma I_t \Delta t, \quad (23)$$

where β and γ are the effective transmission rate of the disease and the mean recovery rate, respectively, Δt is the interval between timesteps and P_{op} is the total population. In the context of AMR, a population could signify individuals in a homogeneous population that originate in a susceptible compartment and become colonised then infected by a drug resistant pathogen [16], or genetic plasmids encoding enzymes that confer resistance to antibiotics [35]. The parameters β and γ govern how much of the population get infected at each increment of time. This is summarised by the reproductive number, R , which is calculated by $R = \beta/\gamma$. Figure 6 exemplifies how the R number can effect the total number of individuals in the susceptible and infected compartments when $R = 2.66$ (dashed/dot lines) and $R = 1.5$ (solid lines). Equations (22) and (23) are analogous to (20) such that $\mathbf{X}_t = \{S_t, I_t\}$. In the context of AMR modelling, examples of the observation equation in (21), often referred to as the likelihood, are the Binomial distribution [46], Bernoulli distribution [36] or beta-binomial and dirichlet-multinomial distributions [41].

E. SEQUENTIAL MONTE CARLO METHODS

1) PARTICLE FILTER

To formulate the problem, consider an SSM that is simulated for t timesteps, where data is acquired at each increment of time. For brevity, the sequence of states and observations up to time t can be defined as $\mathbf{x}_{0:t} = \{x_0, x_1, \dots, x_t\}$ and $\mathbf{y}_{0:t} = \{y_0, y_1, \dots, y_t\}$, respectively. In filtering problems, the posterior distribution $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}, \boldsymbol{\theta})$ can be estimated recursively by prediction-update steps. Assuming Markovianity, the posterior can be defined as $p(\mathbf{x}_t|\mathbf{y}_{1:t}, \boldsymbol{\theta})$. The posterior from the previous timestep, $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})$, and the state equation in (20) can be used to predict \mathbf{x}_t . This is done by evaluating $p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})$ by using the Chapman-Kolmogorov equation such that

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \boldsymbol{\theta})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})d\mathbf{x}_{t-1} \quad (24)$$

$$= \int p(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})d\mathbf{x}_{t-1}, \quad (25)$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta})$ is equal to $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \boldsymbol{\theta})$ because (20) is a Markov process of order one. As data at timestep t becomes available, the prediction step can be updated using Bayes' theorem in (1) such that the posterior is defined to be

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}, \boldsymbol{\theta}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{\theta})p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})} \quad (26)$$

$$= \frac{p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{\theta})p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})}{\int p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{\theta})p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta})d\mathbf{x}_t}. \quad (27)$$

If the integrals in (25) and (26) are high dimensional, they can be hard to solve analytically. Section IV-A outlines how MCMC can sample from complex distributions that are static. However, as the posterior distribution is dynamic and needs to be estimated recursively, MCMC methods are unsuitable. One method to overcome this issue is to approximate (25) and (26) recursively using a particle filter.

Particle filters are an SMC method which uses the IS principles outlined in Section IV-B in a recursive process to infer the time-dependent hidden states in SSMs that are nonlinear and non-Gaussian [76]. At every timestep t , the particle filter draws N samples (particles) from a proposal distribution, $q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$, which is parameterised by the sequence of states and observations. The samples are statistically independent and each represents a different hypothesis of the sequence of states of the system. The i th sample has an associated weight, \mathbf{w}_t^i , which indicates the relative importance of each of the corresponding samples. A set of N particles can then be represented as $\{\mathbf{x}_{1:t}^i, \mathbf{w}_t^i\}_{i=1}^N$.

By following (19), the weights of the particles can be calculated by

$$\mathbf{w}_{1:t}^i = \frac{p(\mathbf{x}_{0:t}^i, \mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}^i|\mathbf{y}_{1:t})}. \quad (28)$$

As is with the case for (26), the state at timestep t is required (not the full state sequence), such that (28) can be rewritten

as

$$\mathbf{w}_{1:t}^i = \mathbf{w}_{1:t-1}^i \frac{p(\mathbf{y}_t|\mathbf{x}_t^i, \boldsymbol{\theta})p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \boldsymbol{\theta})}{q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \mathbf{y}_t)}, \quad (29)$$

where $\mathbf{w}_{1:t-1}^i$ is the weight from the previous timestep. For $t = 1$,

$$\mathbf{w}_1^i = \frac{p(\mathbf{y}_1|\mathbf{x}_1^i, \boldsymbol{\theta})p(\mathbf{x}_1^i|\boldsymbol{\theta})}{q(\mathbf{x}_1^i|\mathbf{y}_1)}. \quad (30)$$

The normalised weights can be computed by

$$\tilde{\mathbf{w}}_{1:t}^i = \frac{\mathbf{w}_{1:t}^i}{\sum_{i=1}^N \mathbf{w}_{1:t}^i}, \quad (31)$$

which can be used to calculate an estimate of the true state \mathbf{x}_t via a weighted sum:

$$\mathbb{E}(\mathbf{x}_t) = \sum_{i=1}^{N_x} \mathbf{x}_t^i \tilde{\mathbf{w}}_{1:t}^i \quad (32)$$

The algorithm described up to now is termed Sequential Importance Sampling (SIS). As time evolves, the normalised weights in (31) can suffer from the phenomenon particle degeneracy. This occurs when one weight becomes close to one while the remainder tend to 0. To overcome this, a process reminiscent to the survival of the fittest can be employed. Calculating the number of effective samples:

$$N_{eff} = \frac{1}{\sum_{i=1}^N (\tilde{\mathbf{w}}_{1:t}^i)^2}, \quad (33)$$

at each iteration can help diagnose when particle degeneracy is occurring. Once N_{eff} drops below a certain threshold, resampling can be used to replicate particles with higher weights and kill off particles with lower weights. Multinomial resampling is commonly used which draws a new set of N particles based on the current set and proportionally to their normalised weights

$$\tilde{\mathbf{w}}_{1:t}^i \text{ for } i = 1, \dots, N. \quad (34)$$

After resampling, the normalised weights and unnormalised weights are set to $\frac{1}{N}$ and $\frac{1}{N} \sum_{i=1}^N \mathbf{w}_{1:t}^i$, respectively.

Stochastic elements within a model can result in particles that have slightly different representations of the model. This makes the PF a suitable candidate when modelling stochastic SSMs. Table 2 outlines the limited number of published studies that use a PF to model stochastic SSMs relating to AMR [36], [39], [59]. A potential reason for this low number is that the PF can be computationally expensive to run if there are lots of particles. An estimate of the log-likelihood from the particle filter can be used within the inference algorithms p-MCMC [39], [59] and SMC² [36]. The log-likelihood calculation is outlined in Section IV-E2.

2) PARTICLE - MARKOV-CHAIN MONTE CARLO

P-MCMC obtains numerical estimates related to $\pi(\theta)$ by combining two Monte Carlo methods. In the original contribution of [77], an outer MCMC layer is used to estimate the parameters of an SSM while an inner PF layer estimates the states of the dynamical system. The outer MCMC layer runs in a similar fashion to the MHRW algorithm in Section IV-A1, but the log-likelihood is estimated by the PF in Section IV-E1. For $t = 1, \dots, T$ an unbiased estimate of the log-likelihood can be estimated recursively by summing the log-weights:

$$p(\mathbf{y}_{1:T}|\theta) \approx \frac{1}{N} \sum_{i=1}^N \mathbf{w}_{1:t}^i. \quad (35)$$

This is a byproduct of running the PF and so no additional calculations are required. The resulting log-likelihood estimate can be used within the acceptance probability in (6) such that

$$\alpha(\theta', \theta) = \min \left\{ 1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right\}, \quad (36)$$

where $\pi(\theta)$ is the product of the prior density $p(\theta)$ and the approximate likelihood $p(\mathbf{y}_{1:T}|\theta)$. As with the MHRW algorithm, when a symmetric proposal for θ is used, (36) cancels to

$$\alpha(\theta', \theta) = \min \left\{ 1, \frac{\pi(\theta')}{\pi(\theta)} \right\}. \quad (37)$$

Reference [77] proves that the Markov chain converges to $\pi(\theta)$ when using a fixed number of particles within the PF.

Two examples of utilizing P-MCMC to infer the parameters of models related to antibiotic resistance in *N. gonorrhoea* can be found in [39] and [59]. A disadvantage of P-MCMC is that it can be impractical to run if one iteration of the Particle Filter (PF) is computationally expensive. Due to the sequential nature of P-MCMC and MCMC algorithms, parallelization over the M MCMC iterations is not readily available. SMC samplers and SMC², described in Sections IV-E3 and IV-E4, respectively, are two alternatives to MCMC and P-MCMC, which can easily exploit high-performance computing architectures. Although IS sampling in the PF can be readily parallelised, we note that the computational bottleneck occurs when estimating the log-likelihood from a PF. Therefore, parallelising the instances of the PFs is more desirable.

3) SEQUENTIAL MONTE CARLO SAMPLERS

Sequential Monte Carlo (SMC) samplers [78], [79] leverage the principles of both MCMC and IS to sequentially sample from $\pi(\theta)$. At each iteration k , SMC samplers provide approximations for intermediate target distributions of interest, $\pi_1(\theta_1), \dots, \pi_K(\theta_K)$, ultimately converging to the terminal distribution $\pi_K(\theta_K)$, which accurately represents $\pi(\theta)$. The joint distribution encompassing all intermediate

target distributions is formulated as follows:

$$\pi_K(\theta_{1:K}) = \pi_K(\theta_K) \prod_{k=2}^K L(\theta_{k-1}|\theta_k), \quad (38)$$

where $L(\theta_{k-1}|\theta_k)$ denotes the L-kernel. The L-kernel is a probability distribution defined by the user and plays a crucial role in determining the efficiency of the sampler [80].

The initialization step of the Sequential Monte Carlo (SMC) sampler involves using importance sampling. At $k = 1, N$ statistically independent samples are drawn from a prior distribution $q_1(\cdot)$ according to:

$$\theta_1^i \sim q_1(\cdot), \quad \forall i. \quad (39)$$

Following (19), each sample is assigned an importance weight using the following formula:

$$\mathbf{w}_1^i = \frac{\pi(\theta_1^i)}{q_1(\theta_1^i)}, \quad \forall i. \quad (40)$$

where $\pi_1(\theta_1^i)$ is the prior distribution, and $q_1(\theta_1^i)$ is the proposal distribution at time step $k = 1$.

As time evolves, samples are proposed based on the previous iteration. Equation (5) represents a commonly chosen proposal distribution for SMC samplers:

$$\theta_k^i \sim q(\theta_k^i|\theta_{k-1}^i). \quad (41)$$

After new samples are proposed, they are assigned weights according to:

$$\mathbf{w}_k^i = \mathbf{w}_{k-1}^i \frac{\pi(\theta_k^i) L(\theta_{k-1}^i|\theta_k^i)}{\pi(\theta_{k-1}^i) q(\theta_k^i|\theta_{k-1}^i)}, \quad \forall i. \quad (42)$$

Similar to the PF, following the importance sampling step, the weights are normalized according to (31), and resampling is performed if the number of effective samples N_{eff} in (33) falls below a certain threshold. This process is repeated for K iterations, and in contrast to PFs, it is assumed with SMC samplers that all data is known before each run. In a similar fashion to (32), estimates of the expectations of functions are obtained by

$$\mathbb{E}(f(\theta_k^i)) = \sum_{i=1}^N \tilde{\mathbf{w}}_k^i \theta_{1:k}^i. \quad (43)$$

In contrast to using samples from the previous iteration only, as is with (43), a process termed recycling [80], [81] can be employed which provides estimates using all samples from all iterations. Let

$$\mathbf{c}_k = \frac{\mathbf{l}_k}{\sum_{k=1}^K \mathbf{l}_k}, \quad \forall k, \quad (44)$$

where

$$\mathbf{l}_k = \frac{\left(\sum_{i=1}^N \mathbf{w}_k^i \right)^2}{\sum_{i=1}^N (\mathbf{w}_k^i)^2}. \quad \forall k. \quad (45)$$

The expectation of functions can then be realised by

$$\mathbb{E}(f(\theta_k^i)) = \sum_{k=1}^K \mathbf{c}_k \mathbb{E}(f(\theta_k^i)), \quad (46)$$

where $\sum_{k=1}^K \mathbf{c}_k = 1$.

In situations where evaluating the likelihood in a SMC sampler is challenging or computationally expensive, it may be appropriate to utilize methods outlined in Section IV-C for simulating a pseudo-likelihood. This pseudo-likelihood can be employed in place of the actual likelihood. An illustration of this approach can be found in [43] when simulating the emergence of drug-resistant *N. gonorrhoea*.

4) SEQUENTIAL MONTE CARLO²

Sequential Monte Carlo² (SMC²) integrates the Particle Filter (PF) and Sequential Monte Carlo (SMC) samplers, as outlined in Sections IV-E1 and IV-E3, respectively, to estimate $\pi(\theta)$. Similar to p-MCMC, the PF yields an unbiased estimate of the likelihood (35), which can be incorporated into the target calculation of the SMC sampler. In (40), the target $\pi(\theta_k^i)$ is the product of the prior $p(\theta_k^i)$ and the likelihood $p(\mathbf{y}_{1:T}|\theta_k^i)$ (provided by the PF).

An example of SMC² is presented in [36]. In this example, the likelihood and states of the model, which outline individuals colonized by drug-resistant bacteria, are estimated using the PF, while the SMC sampler estimates the model parameters.

V. EVALUATING BAYESIAN ALGORITHMS

When performing Bayesian analysis, it is essential to understand how well the proposed sampler is performing in terms of correctness, accuracy and efficiency. Critically evaluating outputs from different sampling techniques can help determine which sampler to use. This section outlines some of the methods provided in [82] on how to do this.

A. RECOVERING THE CORRECT DISTRIBUTIONS FROM THE MODEL

Simulating data from a model with predefined parameters and subsequently testing the ability of the sampler to recover the correct parameters (posterior distribution) used in the data-generating process is a validated method to assess the accuracy of the sampling process [26], [28], [36], [50], [53].

Simulating data from the model and assessing their consistency with true observations through posterior predictive checks is a valuable method to determine if the model accurately represents reality [16], [19], [25], [27], [28], [30], [39], [46], [47], [49], [53], [59]. With MCMC methods, the posterior predictive distributions (PPD) manifest as a set of samples. To validate the credibility of the PPD, the true observations are expected to fall within the 95% confidence intervals of the samples, and the mean sample should follow the trend of the true observations. A goodness-of-fit test, such as the chi-squared test, can be employed for a visual inspection of the accuracy between the mean sample and true observations. An illustrative example is provided in [53], where the chi-squared test is used to compare the mean predicted numbers of positive swabs of vancomycin-resistant *enterococci* with the true observed swabs.

The practice of utilizing data to estimate the model and then comparing the results with the same data, often

referred to as *using the data twice*, should be approached with caution to avoid potential biases. To address this concern, generating forecasts of true observations into the future, using data the model has not encountered, can be employed. An example of this approach is illustrated in Figure 2 of [39], where the model is fitted to annual *N. gonorrhoea* cases from 2008 to 2017, and forecasts are made for the period between 2018 and 2030. Simple scoring rules, such as those proposed by [83], can effectively summarize the statistical consistency between the PPD and the true observations. Notably, refinement and improvement of predictive performance for MCMC algorithms in forecasting COVID-19-related deaths are explored in [84] and [85].

Scoring rules are applicable when evaluating observable variables, such as positive swabs or colonized patients. However, for assessing latent quantities like the growth rate and reproductive number R_t , simulation-based calibration (SBC) provides a valuable alternative [86]. In SBC, data is simulated from a model using parameters drawn from the prior distribution, and the posterior calibration over independent simulated datasets is then tested against the inference algorithm. An example of this approach is demonstrated in [59], where the total number of *N. gonorrhoea* cases and the number of cefixime-resistant infections are compared with simulated datasets generated using parameters sampled from their posterior distributions. This enables the assessment of the model's goodness of fit to the data, with one simulation performed for each of the one thousand sampled parameter sets.

B. EVALUATING MARKOV CHAINS

The MCMC sampling methods outlined in Section IV-A generate a Markov chain. It is essential to assess how effectively and efficiently a sampler is performing to ensure that the samples are correctly drawn from the posterior distribution. Evaluating Markov chains facilitates comparisons between different samplers and aids in deciding when to terminate the sampling process.

Using PPLs like Stan offers the advantage of automatic provision of summary statistics once the sampling process concludes. However, there are also standalone software packages that can be employed post-sampling. Two examples in the R programming language are CODA [87] and Bayesian Output Analysis (BOA) [88]. CODA has been utilized to evaluate chains sampled using MCMC for multidrug-resistant *A. baumannii* [47] and *N. gonorrhoea* [40], as well as p-MCMC for *N. gonorrhoea* [59]. Similarly, BOA has been applied in the analysis of *enterococci* [53]. These tools provide summary statistics such as effective sample size (ESS), autocorrelations, Gelman-Rubin statistics, among others. A number of commonly used statistics are described in the subsequent sections.

1) EFFECTIVE SAMPLE SIZE

The effective sample size (ESS) serves as an explicit approximation of the number of independent samples necessary

for the Markov chain to possess the same estimation power as the set of auto-correlated samples. Higher ESS values are desirable, indicating a more efficient sampler. Comparing ESS values per second of computation time can provide insights into the efficiency of different samplers. For example, MHRW runs may take less computation time than HMC and NUTS because MH does not involve gradient evaluations. The efficiency of a sampler can be assessed by calculating the ESS per second based on the time taken for the sampler to complete its run.

2) INTEGRATED AUTO-CORRELATED TIME

The Integrated Auto-Correlated Time (IACT) provides an estimate of the number of samples, on average, required to draw an independent sample from a continuous Markov chain. This measure is associated with the concept of “mixing” and lower IACT values are more desirable as they indicate faster mixing and better exploration of the parameter space.

3) GELMAN-RUBIN DIAGNOSTIC

The Gelman-Rubin diagnostic [89] is a numerical method that determines if multiple chains have converged by comparing the variances between chains. This diagnostic is commonly used in ppls (where it is referred to as \hat{R}) to ascertain if the sampler has correctly sampled from the posterior. Stan’s documentation states that an \hat{R} value below 1.05 passes their internal diagnostic check.

C. SOFTWARE PACKAGES

Table 3 outlines software packages that implement Bayesian inference algorithms that have been used when modelling AMR. Software packages for performing Bayesian analysis using MCMC methods with random walk proposals include WinBUGS [90], BayesianTools [91], BEAST [92] and DRAM [63]. As explained in Section IV-A1, random walk proposals can suffer from the problem of *curse of dimensionality*.

HMC and NUTS, outlined in Sections IV-A5 and IV-A7, respectively, overcome this problem by using gradient information to generate efficient proposals. Stan [70] is a ppl that performs Bayesian inference on user defined statistical models with NUTS. Stan has been extensively developed since its creation and has variants written in the programming languages C++, Python and R. Undertaking calibration using Stan requires statistical models to be deterministic [37], [41], [45], [51].

The particle filter outlined in Section IV-E1 can be described in the software package BayesianTools and the R package: pomp [93] which has the functionality of performing Bayesian inference with p-MCMC in Section IV-E2.

When the likelihood evaluation is intractable the ABC methods outlined in Section IV-C can be used. The EasyABC software package, created in R, implements an SMC sampler that uses ABC to estimate the likelihood.

VI. POPULAR METHODS FOR MODELING AMR

The purpose of this section is to present well-known approaches for incorporating Bayesian principles in modeling AMR. These approaches encompass mixture models, hierarchical Bayesian models, as well as deterministic and stochastic epidemiological models. Examples of AMR applications are specified, where available. In Section VIII, we explore potential advancements and developments for these models in the future.

A. MIXTURE MODELS

It is often the case that complex data has visible underlying sub-populations or clusters which cannot be adequately described by a one-dimensional Gaussian distribution. A Gaussian Mixture Model (GMM) is a probabilistic model that assigns and defines the probability of each datum coming from a specific cluster. A GMM can be defined as:

$$p(\theta) = \sum_{i=1}^K w_i \cdot \mathcal{N}(\theta; \mu_i, \Sigma_i), \quad (47)$$

where $p(\theta)$ is the probability density function of the GMM at θ , K is the number of Gaussian components and w_i are the mixture weights, which are probabilities that sum to 1. GMMs in a Bayesian framework require prior distributions to be placed over θ . The mixture weights and parameters will therefore be random variables and in this context, (47) can be defined as

$$p(\theta|\mathbf{y}) = \sum_{i=1}^K \hat{w}_i \cdot \mathcal{N}(\hat{\mu}_i, \hat{\Sigma}_i), \quad (48)$$

where \mathbf{y} is the data and \hat{w} , $\hat{\mu}$ and $\hat{\Sigma}$ are estimated using expectation-maximization (EM). Mixture models have been successfully applied to in vitro studies [38], [51], [52] (see Section VII-C for more information) and phylogenetic trees [14], [15] (see Section VII-D for more information) when modelling AMR.

B. HIERARCHICAL BAYESIAN MODEL

Hierarchical or multi-level Bayesian modeling is a powerful approach that involves breaking down complex statistical models into a series of interconnected sub-models arranged in a hierarchical structure. This framework allows for the combination of information from various sub-models, introducing additional uncertainty that can be beneficial in the inference process. The interdependence of priors among parameters contributes to this added uncertainty. Multiple examples of hierarchical models and AMR can be found in [50], [53], and [57].

In (2), the likelihood function, $p(\mathbf{y}|\theta)$, is a function of θ . However, in hierarchical Bayesian models the likelihood is a function of θ and hyper-parameters, ϕ , such that it is defined to be

$$p(\mathbf{y}|\theta, \phi), \quad (49)$$

TABLE 3. Software packages that have been used to perform Bayesian inference on models relating to AMR.

Software Package	Ref.	Description	Studies
Stan	[70]	MCMC with NUTS	[31], [37], [41], [44], [45], [51]
WinBUGS	[90]	MCMC with Gibbs	[14], [15], [58]
R package: pomp	[93]	particle filters, p-MCMC, ABC	[39], [59]
BayesianTools	[91]	MCMC with RW, particle filters	[16], [57]
BEAST	[92]	MCMC with RW	[22], [32], [42], [94]
DRAM	[63]	Adaptive MCMC	[18]
EasyABC	[95]	SMC sampler with ABC	[43]

with

$$p(\boldsymbol{\theta}, \phi), \quad (50)$$

as its prior. Using the definition of conditional probability, (50) can be defined to be

$$p(\boldsymbol{\theta}, \phi) = p(\boldsymbol{\theta}|\phi)p(\phi). \quad (51)$$

Using Bayes theorem, the posterior distribution can then be defined to be

$$p(\boldsymbol{\theta}, \phi|y) \propto p(y|\boldsymbol{\theta}, \phi)p(\boldsymbol{\theta}, \phi), \quad (52)$$

$$\propto p(y|\boldsymbol{\theta}, \phi)p(\boldsymbol{\theta}|\phi)p(\phi). \quad (53)$$

As the dependence of parameters are modelled conditionally, the Gibbs sampler, outlined in section IV-A2, is a suitable MCMC algorithm to perform inference.

Hierarchical models may account for variability at different levels, such as variations between different patient populations, geographic regions, or strains of microorganisms. Data from hospital wards or separate ICUs within hospitals are often highly interdependent. Therefore, complex mathematical transmission models can be split into sub-models. For example, the colonisation status of infected patients in one ward can affect patients in another [50], [53], [57]. Each hospital ward or ICU has different parameters which are intrinsically linked.

C. EPIDEMIOLOGICAL ANTIMICROBIAL RESISTANCE

Two common approaches for representing and simulating AMR are Compartmental Models (CM) and population-based Agent-Based Models (ABMs).

CMs divide the population into distinct compartments, each representing a different health state. The transitions between these compartments are governed by a set of differential equations (see Section IV-D1).

Incorporating stochasticity within the SIS model (see Section IV-D1) can better reflect real life scenarios [21], [23], [28], [30], [36], [39], [45], [47], [48], [59]. Stochastic fluctuations to the disease dynamics can be added to the discrete ODEs in (22) by adding a noise term, ϵ_x , for each time-varying parameter. The noise mimics random interactions between individuals within the population. The

stochastic SIS model is then defined as

$$S_{t+1} = S_t - \frac{\beta I_t S_t}{P_{op}} - \gamma I_t - \epsilon_\beta + \epsilon_\gamma \Delta t, \quad (54)$$

$$I_{t+1} = I_t + \frac{\beta I_t S_t}{P_{op}} + \gamma I_t + \epsilon_\beta - \epsilon_\gamma \Delta t, \quad (55)$$

where $\epsilon_\beta \sim \mathcal{N}(0, \sqrt{\beta}/P_{op})$ and $\epsilon_\gamma \sim \mathcal{N}(0, \sqrt{\gamma}/P_{op})$.

References [21], [24], and [47] stress the importance of stochastic models when working with data from small populations as stochastic effects can become highly important. Reference [47] collects samples of multidrug-resistant *A. baumannii* from 15 patients in a hospital ward and report a Stochastic Differential Equation (SDE) is necessary, over an ODE, to obtain reliable estimates. They do this by adding a continuous-time Markov process to the ODE which results in an SDE. In a study of cefixime resistance in *N. gonorrhoea* [59], it was shown that another advantage of using a stochastic model over the deterministic equivalent is that the small number of cases in the early and late stages of the outbreak would not be captured by their deterministic model.

Individual ABMs are another method for simulating interactions with individuals within a population based on a set of rules [30], [36], [45]. ABMs are useful for modeling individual-level behaviors, such as antibiotic usage and transmission dynamics. They can capture heterogeneity in the population and simulate the emergence and spread of AMR at the individual level. ABMs relax the homogeneity assumption, which underpins traditional CMs. Although CMs do simulate interactions with individuals in a population, they don't represent them explicitly like ABMs. This has been argued as a major advantage when making inferences about nosocomial transmission of MRSA [30]. ABMs can easily express the heterogeneity of patient and health care worker contacts within different wards of a hospital. The flexibility of individual ABMs is also an advantage when compared to ODE or SDE based models. For example, it is shown in [36] that the colonisation rate of extended spectrum beta-lactamase producing *E. coli* and *K. pneumoniae* is affected by factors including time-of year, coinhabitant information and individual related information such as gender, income and age, which can easily be modelled using ABMs.

1) METHODS TO PREVENT AND CONTROL AMR

In epidemiological studies, identifying and implementing effective measures to prevent and control AMR within specific populations is a critical area of focus. Healthcare

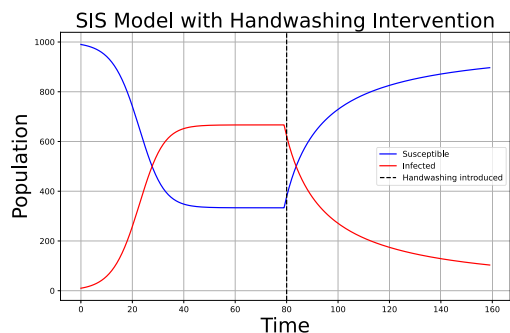


FIGURE 7. Simulation of the discrete time approximation of the ODEs in (22) and (23). Hand hygiene by health-care workers is introduced at $t = 80$ (black dashed vertical line) with $\beta = 0.3$ and $\beta = 0.1$ pre and post hand hygiene intervention, respectively.

settings, in particular, have garnered significant attention due to the heightened risk of patients acquiring infections caused by drug-resistant bacteria [23], [24], [25], [29], [30], [54], [57]. Assessing whether a patient is colonised with resistant bacteria upon arrival at a healthcare facility is crucial to prevent the potential spread to other inpatients. Methicillin resistant *S. aureus*, in particular, is a focus of concern given the severity of infections it can cause [23], [24], [25], [29], [30]. The use of antimicrobials to manage MRSA rates may lead to unintended collateral AMR, with manifestations occurring over months or years, extending beyond the typical follow-up period of clinical studies. Simulation studies, particularly those employing Bayesian calibration, can effectively model long-term AMR outcomes. For instance, a simulation study revealed that the long-term consequence of a “universal” mupirocin usage policy was the development of mupirocin resistance [23]. Interventions, such as incorporating improved hand hygiene by health-care workers [24], [29], [57], decolonisation measures e.g., enhanced cleaning of medical equipment and the environment [29], [54], [57] and limiting patient exposures and contact precautions via isolation rooms for colonised patients [25], [54], [57] could be beneficial. Figure 7 shows the impact of implementing improved hand-hygiene measures when simulating the discrete time approximation of the ODEs in (22) and (23). The change of the transmissibility parameter β from 0.3 to 0.1, when the measures are introduced at $t = 80$, has an effect on the number of individuals in the infected compartment. However, employing blanket package interventions of isolation, testing and decolonisation treatments to all patients can consume valuable resources [25]. Therefore, an approach targeted to high risk patients could be more desirable [30]. Employing individual, patient-level compartmental transmission models that depict Intensive Care Units (ICUs) or hospital wards allows for the monitoring of each patient’s status while intervention and control strategies are enacted [23], [24], [25], [30]. The implementation of control strategies in real-world scenarios during data collection [23], [25], [29], [54], [57] or through sensitivity analysis [24], [30], such

as adjusting parameters to explore various scenarios (e.g., modifying a hand hygiene parameter to assess the impact of healthcare workers’ adherence to handwashing), enables the identification of changes in bacteria transmission dynamics [24].

Enhancing adherence to hand hygiene is identified as the most effective measure in preventing the transmission of MRSA, as evidenced by Bayesian estimates of the transmission count per uncolonized patient, referred to as the attack rate in [24]. An MCMC algorithm found that including isolation and decolonisation measures reduced the transmission dynamics of MRSA by 64% in [25].

VII. ANTIMICROBIAL RESISTANCE

This section describes how studies have examined AMR. As explained in Section I, an effective AMS is multifaceted and aims to integrate various initiatives. Among these initiatives are programs addressing multi-drug resistance, economic implications, and modeling AMR in vitro as well as within specific populations. The subsequent paragraphs provide an overview of these programs.

A. DRUG/MULTI-DRUG RESISTANCE

The global trend of multi-drug resistance is a cause for concern. In 2019, an estimated 1.27 million deaths worldwide were attributed to drug and multi-drug resistance, with AMR-resistant infections contributing to 4.95 million deaths [96]. A significant contributor to these fatalities, numbering approximately 50,000 to 100,000, is multi-drug-resistant *M. tuberculosis*, a form of *M. tuberculosis* resistant to isoniazid and rifampin, the two most potent anti *M. tuberculosis* medications. Bayesian modeling and the estimation of anticipated cases, incidence, and mortality related to multi-drug-resistant *M. tuberculosis* constitute an active and ongoing area of research [13], [14], [16], [17], [20]. Bayesian analysis of other pathogens resistant to multiple drugs include: *Salmonella* spp [51], [52]; *E. coli* [34], [35]; *N. gonorrhoea* [41] and *A. baumannii* [47].

B. ECONOMIC COST

As stated in [97], the rise of AMR is associated with an annual increase of nine billion euros and 20 billion dollars in healthcare costs in Europe and the United States, respectively. With diminishing effectiveness of antimicrobials against diseases, these costs are expected to increase. Two noteworthy studies, focusing on the epidemiological and public health impact, as well as the cost-effectiveness of implementing vaccine programs for *N. gonorrhoea* [40] and *M. tuberculosis* [20], exist. Both studies employ integrated transmission-dynamic compartmental health-economic models to project cost savings based on vaccine rollout. According to [20], *M. tuberculosis* vaccination is poised to significantly reduce future case burdens, while [40] advocates for a targeted vaccination approach for better economic outcomes. Utilizing Bayesian methods for calibration is well-suited for this

analysis, given the inherent uncertainty in model parameters and cost inputs, as highlighted in [20].

C. MODELLING OF IN VITRO PHENOMENA

Bayesian calibration of in vitro studies has been an active area of research [18], [31], [35], [38], [41], [51], [52], [58], [98]. The Minimal Inhibitory Concentration (MIC) establishes the susceptibility or resistance of antibiotics when tested against antimicrobials in vitro. It signifies the minimum antibiotic concentration at which detectable bacterial growth is observed. A lower MIC value suggests that a smaller drug concentration is needed to inhibit bacterial growth, making it more desirable. MIC serves as a benchmark for comparing antimicrobial agents across various studies. For example, *N. gonorrhoea* was tested against ciprofloxacin, cefixime and azithromycin in [41] and ciprofloxacin, penicillin, and tetracycline in [58]. The same classes of antibiotics can also be applied to different bacteria for example, tetracycline tested against *E. coli* [35] and *N. gonorrhoea* [58] and amoxicillin tested against *E. coli* [35] and *Salmonella* spp [51].

MIC experiments can be resource-intensive and expensive due to the necessary equipment. The potential for human error in recording MIC readings is also a problem. Recent work, named AIgarMIC [99], has attempted use ML to automate the readings of MIC experiments. The observation of bacterial growth frequently leads to censored data, where observations fall within fixed time intervals between dilution experiments. This limitation arises because the exact time of growth occurs within the interval from the last dilution inhibiting growth to the first dilution where no growth is visible. Approaches to address censored data within a Bayesian framework are discussed in [31], [38], [51], [52], and [58].

Among the in vitro studies discussed previously, two prevalent methods for modeling MIC are noteworthy. Examples include ODE based models, such as those describing how single or multiple drug resistance mutations influence MIC distribution [41]. These models also explore noninherited resistance levels in bacteria when exposed to antibiotics [35], as well as the emergence of antibiotic-resistant subpopulations in *M. tuberculosis* cells treated with isoniazid [18]. The second method involves hierarchical mixture models, exemplified by their ability to represent subpopulations of antibiotic resistance within an overall population of isolates [38], [51], [52].

D. SPECIALISED AREAS OF AMR

This review revealed two specialized areas of research related to AMR, briefly discussed here. Firstly, several studies incorporated the analysis of molecular sequences associated with drug resistance using phylogenetic trees [14], [15], [22], [32], [33], [42], [44], [94]. As seen in Table 3, the software Bayesian Evolutionary Analysis Sampling Trees (BEAST) [92] is widely used for conducting Bayesian analysis on the evolution of drug-resistant genomes. Additionally, the

software package Tracer [100] is commonly employed to visualize the MCMC traceplots generated by the BEAST software. The second specialized area of research pertains to popPK models. These models aim to depict relationships between patient characteristics and drug exposure, with a focus on inferring the susceptibility of the drug to the microorganism and the exposure to the drug [101], [102].

VIII. RECENT ADVANCEMENTS AND SCOPE FOR FUTURE WORK

This section outlines potential avenues for future work with regards the applicability of the latest advancements and applications in Bayesian modeling of AMR. As is with healthcare in general, AMR is a field that is dynamic and continuously evolving.

A. A COMBINATION APPROACH TO AMR

This review has outlined studies that have attempted to address the measures outlined in Section I which constitute an effective AMS. According to [8], the challenges arise when combining these measures in a coherent framework which is interpretable by clinicians and healthcare workers and adheres to the legal requirements needed to utilise in the healthcare system. Examples of how future research could integrate multiple areas discussed in this review include:

- Extend the use of hierarchical Bayesian models to integrate data from multiple scales e.g. combining MIC laboratory testing of antimicrobials with broader surveillance of AMR in target populations to better understand the transmission dynamics of AMR.
- Integrating epidemiological and genomic data to identify genetic markers associated with resistance and track the evolution of resistance genes could follow a similar approach to that described by [103] for HIV. In their study, they calibrate time series prevalence data and dated phylogeny reconstructed from sequences to infer the reproduction number using p-MCMC. Given the potential computational expense of this approach, especially with long time-series data, employing SMC² could be advantageous.
- Economic and policy models could be combined using Bayesian methods to quantify the economic impact of AMR and evaluate the cost-effectiveness of different intervention strategies.

B. SAMPLING ALGORITHMS

The progression from the MHRW proposal in MCMC to gradient-based methods is clearly outlined in this review. It is evident from Table 2 that the majority of studies identified in the search criteria are random walk based algorithms. In the following paragraphs, we intend to describe recent advancements in p-MCMC, SMC², and SMC samplers.

As detailed in Section IV-E2, the standard proposal utilized in p-MCMC and SMC samplers is MHRW, which can face challenges associated with the *curse of dimensionality*.

Recent advancements have explored the use of MALA, HMC, and NUTS as proposals within p-MCMC [104], [105]. Similarly, recent advancements with gradient-based proposals within SMC samplers can be found in [106] and [107]. One notable challenge in p-MCMC and MCMC methods lies in their inherent sequential nature, limiting efficient parallelization. Multichain MCMC is one method that can efficiently explore $\pi(\theta)$ with [108] presenting a framework for running on GPUs. To overcome this constraint, a potential solution involves replacing MCMC with an SMC sampler and p-MCMC with SMC². Recent studies have shown how to implement these algorithms on high-end computing facilities [36], [109]. A specific example can be found in [36] where they implement SMC² on GPU nodes when calibrating transmission models pertinent to the spread of *E. coli*.

C. DIFFERENTIABLE AGENT BASED MODELS

ABMs and equation-based CMs represent two frameworks employed for modeling the spread of communicable diseases within populations. ABMs diverge from the homogeneity assumption inherent in traditional CMs by simulating heterogeneous agents with predefined behaviors. The challenges of scalability when simulating millions of individuals and the non-differentiable nature of ABMs have been perceived as hindrances to widespread use. Recent efforts, such as those highlighted in [110], have focused on developing differentiable ABMs. A notable example is GradABM [111], which introduces a scalable, differentiable ABM capable of simulating populations in the millions within a few seconds and accommodating heterogeneous data sources related to COVID-19. A promising avenue for future research would involve calibration of differentiable ABMs to AMR data using gradient-based MCMC and SMC samplers.

D. SPECIALISED AREAS OF AMR

In Section VII-D, two specialized areas of AMR are discussed: phylogenetic trees and popPK models. The software package BEAST is a prominent choice for Bayesian analysis of molecular sequences using MCMC. A recent review has outlined practical guidelines for Bayesian phylogenetic inference using MCMC [112]. As the size and complexity of data and models increase, so does the runtime required to obtain meaningful results. Recent advancements, as discussed in [113], introduce an SMC sampler specifically designed for decision trees, and this implementation is compatible with shared memory architectures. The study demonstrates that SMC samplers are as accurate as MCMC but faster, offering potential benefits in this context. Concerning popPK models, the use of gradient based MCMC algorithms, such as HMC, could be a beneficial alternative to likelihood statistical methods, as outlined in the review article [114]. A software package named Torsten [115] utilizes functions derived from the ppl Stan to tackle popPK related problems. While both areas necessitate further exploration, they represent intriguing directions for future research.

IX. CONCLUSION

This review article explores the application of Bayesian methods in understanding AMR. It examines various studies that employ Bayesian statistical techniques to model and predict the dynamics of AMR. The paper outlines the evolution of sampling algorithms, detailing the shift from traditional random walk methods to more advanced gradient-based approaches. These advancements have improved the accuracy and efficiency of AMR modeling, offering deeper insights into the spread and control of resistance.

It is evident from this review that a combination approach to antimicrobial AMR stewardship should be employed to optimize the use of antimicrobial agents, reduce the spread of resistance, and improve patient outcomes. As computational resources increase, the ability to analyse AMR related data in real time will become hugely important. One such example is phylogenetic analysis and the ability to detect mutations that are resistant to antibiotics.

As AMR continues to pose a global health threat, the application of Bayesian techniques will be crucial in developing effective strategies for monitoring and mitigating resistance. Future research should focus on further refining these methods and exploring their integration with other computational tools to enhance our understanding and management of AMR.

ACKNOWLEDGMENT

Alessandro Gerada and Alex Howard are joint senior authors.

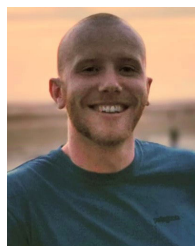
REFERENCES

- [1] D. J. Spiegelhalter, J. P. Myles, D. R. Jones, and K. R. Abrams, "An introduction to Bayesian methods in health technology assessment," *Bmj*, vol. 319, no. 7208, pp. 508–512, 2008.
- [2] D. Ferreira, M. Barthoulot, J. Pottecher, K. D. Torp, P. Diemunsch, and N. Meyer, "Theory and practical use of Bayesian methods in interpreting clinical trial data: A narrative review," *Brit. J. Anaesthesia*, vol. 125, no. 2, pp. 201–207, Aug. 2020.
- [3] A. Gelman and D. B. Rubin, "Markov chain Monte Carlo methods in biostatistics," *Stat. Methods Med. Res.*, vol. 5, no. 4, pp. 339–355, Dec. 1996.
- [4] (2023). *Antimicrobial Resistance*. Accessed: Jul. 3, 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
- [5] K. Abel, P. Dark, T. Felton, A. Ustianowski, T. Van Staa, P. Wilson, and P. Bower, "System-wide approaches to antimicrobial therapy and antimicrobial resistance in the U.K.: The AMR-X framework," *Lancet Microbe*, vol. 5, no. 5, pp. e500–e507, 2024.
- [6] (2023). *Antimicrobial Stewardship*. Accessed: Nov. 16, 2023. [Online]. Available: <https://apic.org/professional-practice/practice-resources/antimicrobial-stewardship/>
- [7] A. M. Niewiadomska, B. Jayabalasingham, J. C. Seidman, L. Willem, B. Grenfell, D. Spiro, and C. Viboud, "Population-level mathematical modeling of antimicrobial resistance: A systematic review," *BMC Med.*, vol. 17, no. 1, pp. 1–20, Dec. 2019.
- [8] A. Howard, S. Aston, A. Gerada, N. Reza, J. Bincalar, H. Mwandumba, T. Butterworth, W. Hope, and I. Buchan, "Antimicrobial learning systems: An implementation blueprint for artificial intelligence to tackle antimicrobial resistance," *Lancet Digit. Health*, vol. 6, no. 1, pp. e79–e86, Jan. 2024.
- [9] A. Sakagianni, C. Koufopoulou, G. Feretzakis, D. Kalles, V. S. Vrykiotis, P. Myrianthefs, and G. Fildisis, "Using machine learning to predict antimicrobial resistance—A literature review," *Antibiotics*, vol. 12, no. 3, p. 452, Feb. 2023.

- [10] J. I. Kim, F. Maguire, K. K. Tsang, T. Gouliouris, S. J. Peacock, T. A. McAllister, A. G. McArthur, and R. G. Beiko, "Machine learning for antimicrobial resistance prediction: Current practice, limitations, and clinical perspective," *Clin. Microbiology Rev.*, vol. 35, no. 3, Sep. 2022, Art. no. e0017921.
- [11] M. N. Anahitar, J. H. Yang, and S. Kanjilal, "Applications of machine learning to the problem of antimicrobial resistance: An emerging model for translational research," *J. Clin. Microbiology*, vol. 59, no. 7, pp. 10–1128, Jun. 2021.
- [12] J. Lv, S. Deng, and L. Zhang, "A review of artificial intelligence applications for antimicrobial resistance," *Biosafety Health*, vol. 3, no. 1, pp. 22–31, Feb. 2021.
- [13] M. O. Fofana, S. Shrestha, G. M. Knight, T. Cohen, R. G. White, F. Cobelens, and D. W. Dowdy, "A multistrain mathematical model to investigate the role of pyrazinamide in the emergence of extensively drug-resistant tuberculosis," *Antimicrobial Agents Chemotherapy*, vol. 61, no. 3, pp. 10–1128, Mar. 2017.
- [14] A. Izu, T. Cohen, and V. DeGruttola, "Bayesian estimation of mixture models with prespecified elements to compare drug resistance in treatment-Naïve and experienced tuberculosis cases," *PLoS Comput. Biol.*, vol. 9, no. 3, Mar. 2013, Art. no. e1002973.
- [15] A. Izu, T. Cohen, C. Mitnick, M. Murray, and V. De Gruttola, "Bayesian methods for fitting mixture models that characterize branching tree processes: An application to development of resistant TB strains," *Statist. Med.*, vol. 30, no. 22, pp. 2708–2720, Sep. 2011.
- [16] W. M. Han, W. Mahikul, T. Pouplin, S. Lawpoolsri, L. J. White, and W. Pan-Ngum, "Assessing the impacts of short-course multidrug-resistant tuberculosis treatment in the Southeast Asia region using a mathematical modeling approach," *PLoS One*, vol. 16, no. 3, Mar. 2021, Art. no. e0248846.
- [17] J. Kang, N. Zhang, and R. Shi, "A Bayesian nonparametric model for spatially distributed multivariate binary data with application to a multidrug-resistant tuberculosis (MDR-TB) study," *Biometrics*, vol. 70, no. 4, pp. 981–992, Dec. 2014.
- [18] D. Ayabina, C. Hendon-Dunn, J. Bacon, and C. Colijn, "Diverse drug-resistant subpopulations of mycobacterium tuberculosis are sustained in continuous culture," *J. Roy. Soc. Interface*, vol. 13, no. 124, Nov. 2016, Art. no. 20160745.
- [19] H. Fu, J. A. Lewnard, I. Frost, R. Laxminarayan, and N. Arinaminpathy, "Modelling the global burden of drug-resistant tuberculosis avertable by a post-exposure vaccine," *Nature Commun.*, vol. 12, no. 1, p. 424, Jan. 2021.
- [20] C. K. Weerasuriya, R. C. Harris, C. F. McQuaid, F. Bozzani, Y. Ruan, R. Li, T. Li, K. Rade, R. Rao, A. M. Ginsberg, G. B. Gomez, and R. G. White, "The epidemiologic impact and cost-effectiveness of new tuberculosis vaccines on multidrug-resistant tuberculosis in India and China," *BMC Med.*, vol. 19, no. 1, pp. 1–13, Dec. 2021.
- [21] G. M. Knight, M. Zimic, S. Funk, R. H. Gilman, J. S. Friedland, and L. Grandjean, "The relative fitness of drug-resistant mycobacterium tuberculosis : A modelling study of household transmission in Peru," *J. Roy. Soc. Interface*, vol. 15, no. 143, Jun. 2018, Art. no. 20180025.
- [22] Y. Xu, J. E. Stockdale, V. Naidu, H. Hatherell, J. Stimson, H. R. Stagg, I. Abubakar, and C. Colijn, "Transmission analysis of a large tuberculosis outbreak in London: A mathematical modelling study using genomic data," *Microbial Genomics*, vol. 6, no. 11, pp. 1–23, Nov. 2020.
- [23] S. R. Deeny, C. J. Worby, O. Tosas Auguet, B. S. Cooper, J. Edgeworth, B. Cookson, and J. V. Robotham, "Impact of mupirocin resistance on the transmission and control of healthcare-associated MRSA," *J. Antimicrobial Chemotherapy*, vol. 70, no. 12, pp. 3366–3378, Sep. 2015, Art. no. dkv249.
- [24] E. S. McBryde, A. N. Pettitt, and D. L. S. McElwain, "A stochastic mathematical model of methicillin resistant *Staphylococcus aureus* transmission in an intensive care unit: Predicting the impact of interventions," *J. Theor. Biol.*, vol. 245, no. 3, pp. 470–481, Apr. 2007.
- [25] C. J. Worby, D. Jeyaratnam, J. V. Robotham, T. Kypraios, P. D. O'Neill, D. De Angelis, G. French, and B. S. Cooper, "Estimating the effectiveness of isolation and decolonization measures in reducing transmission of methicillin-resistant *Staphylococcus aureus* in hospital general wards," *Amer. J. Epidemiology*, vol. 177, no. 11, pp. 1306–1313, Jun. 2013.
- [26] E. Verykoui, T. Kypraios, and P. D. O'Neill, "Modelling the effect of antimicrobial treatment on carriage of hospital pathogens with application to MRSA," *Biostatistics*, vol. 17, no. 1, pp. 65–78, Jan. 2016.
- [27] R. Cassidy, T. Kypraios, and P. D. O'Neill, "Modelling, Bayesian inference, and model assessment for nosocomial pathogens using whole-genome-sequence data," *Statist. Med.*, vol. 39, no. 12, pp. 1746–1765, May 2020.
- [28] M. Laager, B. S. Cooper, and D. W. Eyre, "Probabilistic modelling of effects of antibiotics and calendar time on transmission of healthcare-associated infection," *Sci. Rep.*, vol. 11, no. 1, p. 21417, Nov. 2021.
- [29] X. J. Lee, A. N. Pettitt, and S. J. Dancer, "Quantifying the relative effect of environmental contamination on surgical ward MRSA incidence: An exploratory analysis," *Infection, Disease Health*, vol. 23, no. 3, pp. 127–136, Sep. 2018.
- [30] S. Pei, F. Morone, F. Liljeros, H. Makse, and J. L. Shaman, "Inference and control of the nosocomial transmission of methicillin-resistant *Staphylococcus aureus*," *ELife*, vol. 7, Dec. 2018, Art. no. e40977.
- [31] F. Ojala, M. R. A. Sater, L. G. Miller, J. A. McKinnell, M. K. Hayden, S. S. Huang, Y. H. Grad, and P. Marttinen, "Bayesian modeling of the impact of antibiotic resistance on the efficiency of MRSA decolonization," *PLoS Comput. Biol.*, vol. 19, no. 10, Oct. 2023, Art. no. e1010898.
- [32] S. Manara, E. Pasolli, D. Dolce, N. Ravenni, S. Campana, F. Armanini, F. Asnicar, A. Mengoni, L. Galli, C. Montagnani, E. Venturini, O. Rota-Stabelli, G. Grandi, G. Taccetti, and N. Segata, "Whole-genome epidemiology, characterisation, and phylogenetic reconstruction of *Staphylococcus aureus* strains in a paediatric hospital," *Genome Med.*, vol. 10, no. 1, pp. 1–19, Dec. 2018.
- [33] E. M. Volz and X. Didelot, "Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance," *Systematic Biol.*, vol. 67, no. 4, pp. 719–728, Jul. 2018.
- [34] A. Vellinga, S. Tansey, B. Hanahoe, K. Bennett, A. W. Murphy, and M. Cormican, "Trimethoprim and ciprofloxacin resistance and prescribing in urinary tract infection associated with *Escherichia Coli*: A multilevel model," *J. Antimicrobial Chemotherapy*, vol. 67, no. 10, pp. 2523–2530, Oct. 2012.
- [35] M. C. J. Bootsma, M. A. van der Horst, T. Guryeva, B. H. T. Kuile, and O. Diekmann, "Modeling non-inherited antibiotic resistance," *Bull. Math. Biol.*, vol. 74, no. 8, pp. 1691–1705, Aug. 2012.
- [36] L. Rimella, S. Alderton, M. Sammarro, B. Rowlingson, D. Cocker, N. Feasey, P. Fearnhead, and C. Jewell, "Inference on extended-spectrum beta-lactamase *Escherichia Coli* and *Klebsiella pneumoniae* data through SMC²," *J. Roy. Stat. Soc. Ser. C, Appl. Statist.*, vol. 72, no. 5, pp. 1435–1451, Dec. 2023.
- [37] R. León-Sampedro, J. DelaFuente, C. Díaz-Agero, T. Crellen, P. Musicha, J. Rodríguez-Beltrán, C. de la Vega, M. Hernández-García, N. López-Fresneña, P. Ruiz-Garrajosa, R. Cantón, B. S. Cooper, and Á. S. Millán, "Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients," *Nature Microbiology*, vol. 6, no. 5, pp. 606–616, Mar. 2021.
- [38] S. Jaspers, P. Lambert, and M. Aerts, "A Bayesian approach to the semiparametric estimation of a minimum inhibitory concentration distribution," *Ann. Appl. Statist.*, vol. 10, no. 2, pp. 906–924, Jun. 2016.
- [39] L. K. Whittles, P. J. White, and X. Didelot, "Assessment of the potential of vaccination to combat antibiotic resistance in gonorrhoea: A modeling analysis to determine preferred product characteristics," *Clin. Infectious Diseases*, vol. 71, no. 8, pp. 1912–1919, Nov. 2020.
- [40] L. K. Whittles, X. Didelot, and P. J. White, "Public health impact and cost-effectiveness of gonorrhoea vaccination: An integrated transmission-dynamic health-economic modelling analysis," *Lancet Infectious Diseases*, vol. 22, no. 7, pp. 1030–1041, Jul. 2022.
- [41] J. Riou, C. L. Althaus, H. Allen, M. J. Cole, Y. H. Grad, J. C. M. Heijne, M. Unemo, and N. Low, "Projecting the development of antimicrobial resistance in neisseria gonorrhoeae from antimicrobial surveillance data: A mathematical modelling study," *BMC Infectious Diseases*, vol. 23, no. 1, p. 252, Apr. 2023.
- [42] L. Tazi, M. Pérez-Losada, W. Gu, Y. Yang, L. Xue, K. A. Crandall, and R. P. Viscidi, "Population dynamics of neisseria gonorrhoeae in Shanghai, China: A comparative study," *BMC Infectious Diseases*, vol. 10, no. 1, pp. 1–12, Dec. 2010.
- [43] A. Tsoumanis, C. Van Dijk, N. Hens, and C. Kenyon, "Rethinking screening intensity in terms of reducing prevalence or increasing selection pressure for the emergence of resistant gonorrhoea: A modeling study of men who have sex with men in Belgium," in *Open Forum Infectious Diseases*, vol. 10. London, U.K.: Oxford Univ. Press, 2023.

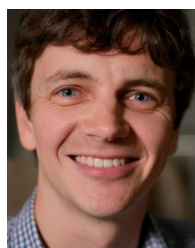
- [44] D. Helekal, M. Keeling, Y. H. Grad, and X. Didelot, "Estimating the fitness cost and benefit of antimicrobial resistance from pathogen genomic data," *J. Roy. Soc. Interface*, vol. 20, no. 203, Jun. 2023, Art. no. 20230074.
- [45] T. Crellen, P. Turner, S. Pol, S. Baker, T. N. T. Nguyen, N. Stoesser, N. P. Day, C. Turner, and B. S. Cooper, "Transmission dynamics and control of multidrug-resistant klebsiella pneumoniae in neonates in a developing country," *eLife*, vol. 8, Dec. 2019, Art. no. e50468.
- [46] T. N. Doan, D. C. M. Kong, C. Marshall, C. M. J. Kirkpatrick, and E. S. McBryde, "Characterising the transmission dynamics of *Acinetobacter Baumannii* in intensive care units using hidden Markov models," *PLoS One*, vol. 10, no. 7, Jul. 2015, Art. no. e0132037.
- [47] M. Domenech de Cellès, J. Salomon, A. Marinier, C. Lawrence, J.-L. Gaillard, J.-L. Herrmann, and D. Guillemot, "Identifying more epidemic clones during a hospital outbreak of multidrug-resistant *Acinetobacter Baumannii*," *PLoS One*, vol. 7, no. 9, Sep. 2012, Art. no. e45758.
- [48] J. Lourenço, Y. Daon, A. Gori, and U. Obolski, "Pneumococcal competition modulates antibiotic resistance in the pre-vaccination era: A modelling study," *Vaccines*, vol. 9, no. 3, p. 265, Mar. 2021.
- [49] F. Blanquart, S. Lehtinen, and C. Fraser, "An evolutionary model to predict the frequency of antibiotic resistance under seasonal antibiotic use, and an application to streptococcus pneumoniae," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 284, no. 1855, May 2017, Art. no. 20170679.
- [50] T. M. Pham, M. Kretzschmar, X. Bertrand, and M. Bootsma, "Tracking *Pseudomonas Aeruginosa* transmissions due to environmental contamination after discharge in ICUs using mathematical models," *PLOS Comput. Biol.*, vol. 15, no. 8, Aug. 2019, Art. no. e1006697.
- [51] M. Zhang, C. Wang, and A. O'Connor, "A Bayesian approach to modeling antimicrobial multidrug resistance," *PLoS One*, vol. 16, no. 12, Dec. 2021, Art. no. e0261528.
- [52] M. Zhang, C. Wang, and A. O'Connor, "A hierarchical Bayesian latent class mixture model with censorship for detection of linear temporal changes in antibiotic resistance," *PLoS ONE*, vol. 15, no. 1, Jan. 2020, Art. no. e0220427.
- [53] B. S. Cooper, G. F. Medley, S. J. Bradley, and G. M. Scott, "An augmented data method for the analysis of nosocomial infection data," *Amer. J. Epidemiology*, vol. 168, no. 5, pp. 548–557, Sep. 2008.
- [54] Y. Wei, T. Kyraios, P. D. O'Neill, S. S. Huang, S. L. Rifas-Shiman, and B. S. Cooper, "Evaluating hospital infection control measures for antimicrobial-resistant pathogens using stochastic transmission models: Application to vancomycin-resistant enterococci in intensive care units," *Stat. Methods Med. Res.*, vol. 27, no. 1, pp. 269–285, Jan. 2018.
- [55] S. Shrivastava, P. Shrivastava, and J. Ramasamy, "World health organization releases global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics," *J. Med. Soc.*, vol. 32, no. 1, p. 76, 2018.
- [56] L. B. Rice, "Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE," *J. Infectious Diseases*, vol. 197, no. 8, pp. 1079–1081, Apr. 2008.
- [57] D. A. Martínez, J. Cai, G. Lin, K. E. Goodman, R. Paul, J. Lessler, S. R. Levin, M. Toerper, P. J. Simmer, A. M. Milstone, and E. Y. Klein, "Modelling interventions and contact networks to reduce the spread of carbapenem-resistant organisms between individuals in the ICU," *J. Hospital Infection*, vol. 136, pp. 1–7, Jun. 2023.
- [58] J. van de Kasstele, M. G. van Santen-Verheuvél, F. D. H. Koedijk, A. P. van Dam, M. A. B. van der Sande, and A. J. de Neeling, "New statistical technique for analyzing MIC-based susceptibility data," *Antimicrobial Agents Chemotherapy*, vol. 56, no. 3, pp. 1557–1563, Mar. 2012.
- [59] L. K. Whittles, P. J. White, and X. Didelot, "Estimating the fitness cost and benefit of cefixime resistance in neisseria gonorrhoeae to inform prescription policy: A modelling study," *PLOS Med.*, vol. 14, no. 10, Oct. 2017, Art. no. e1002416.
- [60] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
- [61] P. Müller, *A Generic Approach to Posterior Integration and Gibbs Sampling*. West Lafayette, IN, USA: Purdue Univ., Department of Statistics, 1991. [Online]. Available: https://books.google.co.uk/books?id=u8D_HAAACA AJ
- [62] H. Haario, E. Saksman, and J. Tamminen, "An adaptive metropolis algorithm," *Bernoulli*, vol. 7, no. 2, p. 223, Apr. 2001.
- [63] H. Haario, M. Laine, A. Mira, and E. Saksman, "DRAM: Efficient adaptive MCMC," *Statist. Comput.*, vol. 16, no. 4, pp. 339–354, Dec. 2006.
- [64] L. Tierney, "Markov chains for exploring posterior distributions," *Ann. Statist.*, vol. 22, no. 4, pp. 1701–1728, Dec. 1994.
- [65] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Phys. Lett. B*, vol. 195, pp. 216–222, Sep. 1987.
- [66] R. M. Neal, "MCMC using Hamiltonian dynamics," in *Handbook Markov Chain Monte Carlo*, vol. 2, no. 11, Jun. 2011, p. 2.
- [67] M. Betancourt, "A conceptual introduction to Hamiltonian Monte Carlo," 2017, *arXiv:1701.02434*.
- [68] G. O. Roberts and R. L. Tweedie, "Exponential convergence of Langevin distributions and their discrete approximations," *Bernoulli*, vol. 2, no. 4, p. 341, Dec. 1996.
- [69] M. D. Hoffman and A. Gelman, "The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1593–1623, 2014.
- [70] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A probabilistic programming language," *J. Stat. Softw.*, vol. 76, no. 1, pp. 1–20, 2017.
- [71] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC3," *PeerJ Comput. Sci.*, vol. 2, p. e55, Apr. 2016.
- [72] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman, "Pyro: Deep universal probabilistic programming," *J. Mach. Learn. Res.*, vol. 20, no. 1, pp. 973–978, 2019.
- [73] E. S. McBryde, A. N. Pettitt, B. S. Cooper, and D. L. S. McElwain, "Characterizing an outbreak of vancomycin-resistant enterococci using hidden Markov models," *J. Roy. Soc. Interface*, vol. 4, no. 15, pp. 745–754, Aug. 2007.
- [74] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, p. 711, Dec. 1995.
- [75] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. Roy. Soc. London. Ser. A*, vol. 115, pp. 700–721, Aug. 1927, doi: [10.1098/rspa.1927.0118](https://doi.org/10.1098/rspa.1927.0118).
- [76] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Aug. 2002.
- [77] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 72, no. 3, pp. 269–342, 2010.
- [78] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 68, no. 3, pp. 411–436, Jun. 2006.
- [79] C. Dai, J. Heng, P. E. Jacob, and N. Whiteley, "An invitation to sequential Monte Carlo samplers," *J. Amer. Stat. Assoc.*, vol. 117, no. 539, pp. 1587–1600, Jul. 2022.
- [80] P. L. Green, L. J. Devlin, R. E. Moore, R. J. Jackson, J. Li, and S. Maskell, "Increasing the efficiency of sequential Monte Carlo samplers through the use of approximately optimal L-kernels," *Mech. Syst. Signal Process.*, vol. 162, Jan. 2022, Art. no. 108028.
- [81] T. L. T. Nguyen, F. Septier, G. W. Peters, and Y. Delignon, "Efficient sequential Monte-Carlo samplers for Bayesian inference," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1305–1319, Mar. 2016.
- [82] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Boca Raton, FL, USA: CRC Press, 1995.
- [83] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *J. Amer. Stat. Assoc.*, vol. 102, no. 477, pp. 359–378, Mar. 2007.
- [84] R. E. Moore, C. Rosato, and S. Maskell, "Refining epidemiological forecasts with simple scoring rules," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 380, no. 2233, Oct. 2022, Art. no. 20210305.
- [85] C. Rosato, R. E. Moore, M. Carter, J. Heap, J. Harris, J. Storopoli, and S. Maskell, "Extracting self-reported COVID-19 symptom tweets and Twitter movement mobility origin/destination matrices to inform disease models," *Information*, vol. 14, no. 3, p. 170, Mar. 2023.
- [86] S. Talts, M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman, "Validating Bayesian inference algorithms with simulation-based calibration," 2018, *arXiv:1804.06788*.

- [87] M. Plummer, N. Best, K. Cowles, and K. Vines, "CODA: Convergence diagnosis and output analysis for MCMC," *R News*, vol. 6, no. 1, pp. 7–11, 2006.
- [88] B. J. Smith, "Boa: AnRPackage for MCMC output convergence assessment and posterior inference," *J. Stat. Softw.*, vol. 21, no. 11, pp. 1–37, 2007.
- [89] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Stat. Sci.*, vol. 7, no. 4, pp. 457–472, Nov. 1992.
- [90] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, "Winbugs-a Bayesian modelling framework: Concepts, structure, and extensibility," *Statist. Comput.*, vol. 10, pp. 325–337, Jul. 2000.
- [91] (2023). *Bayesiantools: General-Purpose Mcmc and Smc Samplers and Tools for Bayesian Statistics*. Accessed: Jul. 4, 2023. [Online]. Available: <https://cran.r-project.org/web/packages/BayesianTools/index.html>
- [92] A. J. Drummond and A. Rambaut, "BEAST: Bayesian evolutionary analysis by sampling trees," *BMC Evol. Biol.*, vol. 7, no. 1, p. 214, 2007.
- [93] A. A. King, D. Nguyen, and E. L. Ionides, "Statistical inference for partially observed Markov processes via theRPackagepomp," *J. Stat. Softw.*, vol. 69, no. 12, p. 1, 2016.
- [94] R. A. Gladstone et al., "Emergence and dissemination of antimicrobial resistance in Escherichia Coli causing bloodstream infections in Norway in 2002–17: A nationwide, longitudinal, microbial population genomic study," *Lancet Microbe*, vol. 2, no. 7, pp. e331–e341, Jul. 2021.
- [95] F. Jabot, T. Faure, and N. Dumoulin, "EasyABC: Performing efficient approximate Bayesian computation sampling schemes using R," *Methods Ecology Evol.*, vol. 4, no. 7, pp. 684–687, Jul. 2013.
- [96] C. J. L. Murray et al., "Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis," *Lancet*, vol. 399, no. 10325, pp. 629–655, 2022.
- [97] P. Dadgostar, "Antimicrobial resistance: Implications and costs," *Infection drug resistance*, vol. 1, pp. 3903–3910, Jun. 2019.
- [98] S. Jaspers, A. Komárek, and M. Aerts, "Bayesian estimation of multivariate normal mixtures with covariate-dependent mixing weights, with an application in antimicrobial resistance monitoring," *Biometrical J.*, vol. 60, no. 1, pp. 7–19, Jan. 2018.
- [99] A. Gerada, N. Harper, A. Howard, N. Reza, and W. Hope, "Determination of minimum inhibitory concentrations using machine-learning-assisted agar dilution," *Microbiology Spectr.*, vol. 12, no. 5, May 2024, Art. no. e0420923.
- [100] A. Rambaut, A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard, "Posterior summarization in Bayesian phylogenetics using tracer 1.7," *Systematic Biol.*, vol. 67, no. 5, pp. 901–904, Sep. 2018.
- [101] M. A. Lyons, "Modeling and simulation of pretomanid pharmacodynamics in pulmonary tuberculosis patients," *Antimicrobial Agents Chemotherapy*, vol. 63, no. 12, pp. 10–1128, Dec. 2019.
- [102] K. Iwata, N. Morishita, M. Nishiwaki, and C. Miyakoshi, "Use of rifampin compared with isoniazid for the treatment of latent tuberculosis infection in Japan: A Bayesian inference with Markov chain Monte Carlo method," *Internal Med.*, vol. 59, no. 21, pp. 2687–2691, 2020.
- [103] A. Gill, J. Koskela, X. Didelot, and R. G. Everitt, "Bayesian inference of reproduction number from epidemiological and genetic data using particle MCMC," 2023, *arXiv:2311.09838*.
- [104] C. Rosato, L. Devlin, V. Beraud, P. Horridge, T. B. Schön, and S. Maskell, "Efficient learning of the parameters of non-linear models using differentiable resampling in particle filters," *IEEE Trans. Signal Process.*, vol. 70, pp. 3676–3692, 2022.
- [105] J. Dahlin, F. Lindsten, and T. B. Schön, "Particle Metropolis–Hastings using gradient and Hessian information," *Statist. Comput.*, vol. 25, no. 1, pp. 81–92, Jan. 2015.
- [106] A. Buchholz, N. Chopin, and P. E. Jacob, "Adaptive tuning of Hamiltonian Monte Carlo within sequential Monte Carlo," *Bayesian Anal.*, vol. 16, no. 3, pp. 745–771, Sep. 2021.
- [107] L. Devlin, M. Carter, P. Horridge, P. L. Green, and S. Maskell, "The No-U-Turn sampler as a proposal distribution in a sequential Monte Carlo sampler without accept/reject," *IEEE Signal Process. Lett.*, vol. 31, pp. 1089–1093, 2024.
- [108] M. Hoffman, A. Radul, and P. Sountsov, "An adaptive-MCMC scheme for setting trajectory lengths in Hamiltonian Monte Carlo," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2021, pp. 3907–3915.
- [109] C. Rosato, A. Varsi, J. Murphy, and S. Maskell, "An $O(\log_2 N)$ SMC₂ algorithm on distributed memory with an approx. Optimal L-Kernel," in *Proc. IEEE Symp. Sensor Data Fusion Int. Conf. Multisensor Fusion Integr. (SDF-MFI)*, Nov. 2023, pp. 1–8.
- [110] P. Andelfinger, "Towards differentiable agent-based simulation," *ACM Trans. Model. Comput. Simul.*, vol. 32, no. 4, pp. 1–26, Oct. 2022.
- [111] A. Chopra, A. Rodríguez, J. Subramanian, A. Quera-Bofarull, B. Krishnamurthy, B. A. Prakash, and R. Raskar, "Differentiable agent-based epidemiology," in *Proc. Int. Conf. Auto. Agents Multiagent Syst.*, 2023, pp. 18–48.
- [112] J. Barido-Sottani, O. Schwery, R. C. M. Warnock, C. Zhang, and A. M. Wright, "Practical guidelines for Bayesian phylogenetic inference using Markov chain Monte Carlo (MCMC)," *Open Res. Eur.*, vol. 3, p. 204, Jun. 2024.
- [113] E. Drouiotis, A. Varsi, P. G. Spirakis, and S. Maskell, "A shared memory SMC sampler for decision trees," in *Proc. IEEE 35th Int. Symp. Comput. Archit. High Perform. Comput. (SBAC-PAD)*, Oct. 2023, pp. 209–218.
- [114] K. Choi, "A review of the Bayesian approach with the MCMC and the HMC as a competitor of classical likelihood statistics for pharmacometricians," *Transl. Clin. Pharmacol.*, vol. 31, no. 2, p. 69, 2023.
- [115] C. C. Margossian, Y. Zhang, and W. R. Gillespie, "Flexible and efficient Bayesian pharmacometrics modeling using stan and torsten, Part I," *CPT, Pharmacometrics Syst. Pharmacol.*, vol. 11, no. 9, pp. 1151–1169, Sep. 2022.



CONOR ROSATO received the B.Sc. degree in computer technology from Liverpool John Moores University, Liverpool, U.K., in 2015, and the M.Sc. degree in big data and high performance computing and the M.Res. degree in decision making under risk and uncertainty from the University of Liverpool, Liverpool, in 2016 and 2018, respectively. He is currently a Postdoctoral Researcher with the Department of Pharmacology and Therapeutics, University of

Liverpool, focusing on Bayesian inference and data fusion using Markov chain Monte Carlo and sequential Monte Carlo methods and machine learning, with specific application to healthcare and epidemiological related problems.



PETER L. GREEN received the M.Eng. degree in mechanical engineering and the Ph.D. degree in nonlinear structural dynamics from the University of Sheffield, in 2009 and 2012, respectively. Before moving to the University of Liverpool, in 2015, he received the Ph.D. degree. He is currently a Senior Lecturer with the Department of Mechanical Engineering, University of Liverpool, and the Director of Engineering Data Analytics Ltd. His research interests include the application

of machine learning solutions in risk-averse sections, including engineering and healthcare.



JOHN HARRIS received the B.Sc. degree in science, technology and society from Middlesex University, in 1996, the M.Sc. degree in epidemiology from London School of Hygiene and Tropical Medicine, University of London, in 2000, and the Ph.D. degree in epidemiology from the University of Liverpool, in 2014. He has been working in both non-infectious and infectious disease epidemiology, since 1997. His research interests include using statistical methods applied

to important health questions and improving public health.



SIMON MASKELL received the M.A., M.Eng., and Ph.D. degrees in engineering from the University of Cambridge, Cambridge, U.K., in 1998, 1999, and 2003, respectively. In 2013, he was the Technical Manager of Command, Control and Information Systems with QinetiQ, U.K. Since 2013, he has been a Professor in autonomous systems with the University of Liverpool, Liverpool, U.K. His research interests include Bayesian inference applied to signal processing, multitarget

tracking, data fusion, and decision support, with a particular emphasis on the application of sequential Monte Carlo methods in challenging data science contexts.



WILLIAM HOPE received the Ph.D. degree in antimicrobial pharmacology, in 2006. While undertaking fellowships with The University of Manchester, U.K., and the National Institutes of Health, Bethesda, USA. He is currently the Dame Sally Davies Chair of AMR Research with the University of Liverpool, U.K. He qualified in medicine, in 1991, before undertaking specialist training in infectious diseases and clinical microbiology. He was a NIHR Clinician Scientist and this

award focused on individualized antimicrobial therapy. He leads the Centre of Excellence in Infectious Diseases Research (CEIDR) which focuses on infection therapeutics. His research interests include antimicrobial pharmacokinetics and pharmacodynamics, antimicrobial drug development, and individualization of antimicrobial therapy. He is a fellow of the Royal Australasian College of Physicians (RACP), the Royal College of Pathologists of Australasia (RCPA), American Academy of Microbiology, European Society of Clinical Microbiology and Infectious Diseases, and NIHR National Specialty Co-Lead for Infectious Diseases. He is also an OBE (BMBS).



ALESSANDRO GERADA graduated from the University of Malta Medical School, and underwent microbiology specialty training in Liverpool, U.K. He is currently the Ph.D. degree with the Department of Pharmacology and Therapeutics, University of Liverpool, focusing in artificial intelligence and digital health. He is a consultant microbiologist in Liverpool Clinical Laboratories, where he also leads the national Brucella Reference Unit. His research interests include machine

learning and computer simulation.



ALEX HOWARD is a Consultant in medical microbiology with Liverpool University Hospitals NHS Foundation Trust and the Clatterbridge Cancer Centre, previously working as the Antimicrobial Stewardship Lead and the Infection Control Lead for these respective sites, until 2023, when he began a Ph.D. at the Department of Pharmacology and Therapeutics, University of Liverpool, investigating data science approaches to antimicrobial resistance. His clinical interests

are in immunocompromise and medical mycology.

...