bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

An Efficient Moments-Based Inference Method for Within-Host Bacterial Infection Dynamics

David J. Price[1] Alexandre Breuzé[1,2] Richard Dybowski[1]

Piero Mastroeni[1] Olivier Restif[1*]

April 13, 2017

[1] Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK

[2] ENSTA-ParisTech, 828 Boulevard des Maréchaux, 91120 Palaiseau, France

* Author for correspondence: or226@cam.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Over the last ten years, isogenic tagging (IT) has revolutionised the study of bacterial infection dynamics in laboratory animal models. However, quantitative analysis of IT data has been hindered by the piecemeal development of relevant statistical models. The most promising approach relies on stochastic Markovian models of bacterial population dynamics within and among organs. Here we present an efficient numerical method to fit such stochastic dynamic models to *in vivo* experimental IT data. A common approach to statistical inference with stochastic dynamic models relies on producing large numbers of simulations, but this remains a slow and inefficient method for all but simple problems, especially when tracking bacteria in multiple locations simultaneously. Instead, we derive and solve the systems of ordinary differential equations for the two lower-order moments of the stochastic variables (mean, variance and covariance). For any given model structure, and assuming linear dynamic rates, we demonstrate how the model parameters can be efficiently and accurately estimated by divergence minimisation. We then apply our method to an experimental dataset and compare the estimates and goodness-of-fit to those obtained by maximum likelihood estimation. This flexible framework can easily be applied to a range of experimental systems. Its computational efficiency paves the way for model comparison and optimal experimental design.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Author Summary:

Recent advancements in technology have meant that microbiologists are producing vast amounts of experimental data. However, statistical methods by which we can analyse that data, draw informative inference, and test relevant hypotheses, are much needed. Here, we present a new, efficient inference tool for estimating parameters of stochastic models, with a particular focus on models of within-host bacterial dynamics. The method relies on matching the two lower-order moments of the experimental data (i.e., mean, variance and covariance), to the moments from the mathematical model. The method is verified, and particular choices justified, through a number of simulation studies. We then use this method to estimate models that have been previously estimated using a "gold-standard" maximum likelihood procedure.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

List of symbols:

- A: number of animals
- T: number of tagged strains
- n: number of organs
- N_i: number of bacteria in organ i
- m_{ij}: migration rate from organ i to organ j
- k_i: killing rate in organ i
- r_i: replication rate in organ i
- τ_i: observation time i
- $\mathbf{A}, \mathbf{B}, \mathbf{C}$: matrices
- $\boldsymbol{\lambda}$: vector of transition rates
- B: Number of bootstrap samples
- $\boldsymbol{\theta}^*$: MDE parameter estimate

Abbreviations:

- ABC: approximate Bayesian computation
- IT: isogenic tagging
- LV: live vaccine
- MARE: mean absolute relative error
- MDE: minimum divergence estimate
- MLE: maximum likelihood estimate
- qPCR: quantitative polymerase chain reaction
- WITS: wildtype isogenic tagged strain

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 Introduction

The elucidation of basic kinetic rates governing bacterial growth during infection (such as division and death rates) has been recognised as an important challenge for over 60 years [1]. Thanks to recent technological developments in microbiology, and pushed by growing concern over antimicrobial resistance and the need for new vaccines, the last decade has witnessed rapid progress in the quantification of *in vivo* dynamics of bacterial infection in animal models. Two experimental approaches in particular have shown great promise across multiple pathogen species: isogenic tagging, the focus of this report, and fluorescence dilution, a term encompassing several techniques from which bacterial repication can be inferred [2]. Isogenic tagging (IT) consists in generating an arbitrary number of sub-clones of a given bacterial strain, each defined by a unique genetic tag (a predetermined nucleotide sequence) inserted in a non-coding region of the chromosome. When grown together *in vitro* or *in vivo*, every tagged strain behaves identically to the original strain. Their relative frequencies within a bacterial culture can be measured by quantitative qPCR or sequencing of the tagged region. Taken together with the absolute number of viable bacteria (e.g., by plating colonies), changes in the frequencies of the tags within the bacterial population can reveal underlying variations in the rates at which bacteria divide, die and disperse. For example, a constant number of viable bacteria accompanied by a loss of some of the tags in a closed population would indicate that a certain proportion of bacteria have died and been replaced by replication. Likewise, when monitoring tag frequencies in two or more anatomical compartments within animals, a gradual homogenisation among organs can reveal the transfer of bacteria. While some studies have stopped at qualitative interpretations of such empirical patterns [3, 4, 5, 6], it is possible to quantify underlying processes with the help of mathematical models. Two different types have been used: stochastic population dynamic models to estimate bacterial division, death and migration rates [7, 8, 9, 10, 11, 12], and population genetic models to estimate bottleneck sizes [13, 14]. Our aim is to develop efficient inference methods to deal with the former type of models.

Stochastic birth–death–migration models (a canonical class of Markovian processes [15]) are a common choice to analyse IT experiments, and naturally lead to likelihood–based

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

inference, using either maximum likelihood [9, 11] or Bayesian estimation [12]. In a given experiment, assuming that all of the A animals sampled at a given time and in given conditions are identical, and that all of the T tagged strains infecting each animal act independently of each other and are governed by identical rates, we can treat the $A \times T$ observed strain abundances as independent realisations of a stochastic birth–death–migration process, and calculate the likelihood of any model of interest accordingly. In most published IT studies, bacteria can grow at different rates in different locations within an animal, and migrate from one location to another, generating a network of subpopulations (or metapopulation). This increases the dimensions of both the state variable space (as the model must keep track of the multivariate distribution of bacterial abundance) and the parameter space. Calculating the likelihood of such a model given an experimental dataset requires solving a complex stochastic model, which will rarely be possible analytically. Even a linear birth–death process with a non–Poisson immigration process (representing the transfer of a finite inoculum dose) is sufficient to prevent a fully analytical treatment, and results in a computationally intensive estimation process [12]. Alternatively, approximate–likelihood (e.g., iterative filtering [16]) or likelihood–free (e.g., approximate Bayesian computation or ABC [17]) methods involve the generation of a large number of stochastic simulations of the model of interest, which can be equally time–consuming, even when taking advantage of parallel computation. Although this may not be a problem when fitting a single model to a single dataset, it limits our ability to compare multiple models across complex datasets (typically involving multiple experimental treatments) and, beyond that, use these inference tools for the purpose of optimising experimental design [18]. Hence, there is a need for alternative inference methods using suitable approximations to achieve greater gains in computational efficiency.

The dynamics of multivariate Markovian processes can be approximated using moment–closure methods [19]. Mathematically, a system of differential equations for the moments of the state variables can be derived analytically from the governing equation of any stochastic model [20]. By effectively ignoring the higher moments, a closed, small-dimension system can be derived, allowing fast numerical solution of the lower moments at any time point. Parameter estimation can then be achieved by fitting the first and second

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

order moments of the model to the mean, variance and covariance of the corresponding variables in the data. Apart from a few proof–of–principle studies using simulated chemical reaction data [21, 22, 23] that show great promise, application to statistical inference from biological data remain scarce. As a rare example, Buchholz et al. [24] implemented a moment–based method to solve a multiple T–cell differentiation pathway problem, fitting the moments of a large number of alternative stochastic models to experimental data using a χ^2 statistic. This suggests that efficient moment–based inference methods should be made more readily available to unleash the full potential of stochastic models in experimental biology.

Our objective is to provide a functional and flexible computational framework to estimate the parameters of stochastic metapopulation models for the within–host dynamics of infection, and demonstrate its application and value to analyse IT studies. The model tracks the probability distribution of the number of copies of a tagged strain of bacteria across a network of anatomical compartments within an animal. The goal is to estimate the bacterial division and death rates within each organ, and migration rates between each pair of compartments. First, we present an algorithm that evaluates the first- and second-order moments of the state variables for arbitrary network structures, and assess its accuracy and speed against a gold standard for stochastic models: the exact Gillespie algorithm. We then compare the accuracy of several inference options against simulated data, and finally apply the most promising method to a recent dataset on the dynamics of *Salmonella enterica* serovar Typhimurium in the blood, liver and spleen of vaccinated mice [11]. The massive gain in speed compared to likelihood-based inference allows us to use parametric bootstrap to quantify parameter uncertainty and goodness-of-fit. We also demonstrate how empirically derived noise terms (e.g., caused by imprecise data collection) can be taken into account.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2 Methodology

2.1 Biological context

We consider the general case of a bacterial pathogen inoculated into an animal host where it can potentially reach n anatomical compartments—which can be distinct organs, tissues, lumens, or predefined sections thereof. All our examples are motivated by IT experiments in which a set of identical animals receive the same initial inoculum dose in one compartment (e.g., mouth, nose, blood, peritoneum, etc) at time $t = 0$. The inoculum is composed of an even mix of T tagged strains. At given times τ_1, τ_2 etc, a subset of A_1, A_2, etc animals are chosen at random and euthanised. The abundance of each tagged strain in each of the n anatomical compartments of interest is measured. Thus, at a given time τ_i, the data consist of a matrix $\mathbf{D}_i$ with n rows by $A_i T$ columns, filled with observed bacterial numbers. From this matrix, we can calculate the observed moments, namely the mean and variance of strain abundance within each compartment, and the covariance between each pair of compartments.

Depending on the experimental procedures, these observations are usually subject to some degree of uncertainty, due to observational error. In general we assume that this error is random with a mean of zero, so that there is no systematic bias; this should be assessed by the researchers who conducted the experiments. As a result, we assume that the observed means are unbiased, but the observed (co)variances may be incorrect. In Section 2.7, we describe how known sources of error can be accounted for as part of the data processing procedure. In addition, there usually is some uncertainty about the actual inoculum dose received by each animal. Variations in the abundance of each strain should be assessed experimentally by testing several inoculum doses: this provides estimates for the initial mean and variance of the number of bacteria present in the target compartment at $t = 0$.

We emphasise a few key assumptions and caveats of the present study, which we review in further detail in the Discussion. First, the variable of interest from a modelling perspective is the abundance of a single tagged strain, rather than the total bacterial load per animal (as the latter can be obtained by adding up individual strains). Indeed,

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

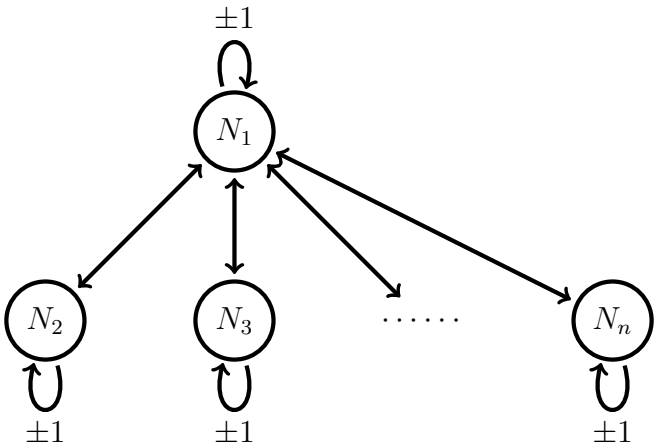
our model framework assumes that, over the time period considered and for a given set of initial conditions, the rates of bacterial division, death and migration *per capita* are independent of the total bacterial load. Second, we assume that all the bacterial cells are governed by identical probabilities of division, death and migration. While this excludes the case of so–called persister cells (i.e., a subset of bacteria with a much lower division rate than the rest) or similar discrete partition, it is worth noting that the stochastic model we describe below does generate continuous variations in the times of events, consistent with empirical distributions of bacterial replication in vivo [25].

2.2 Stochastic model framework

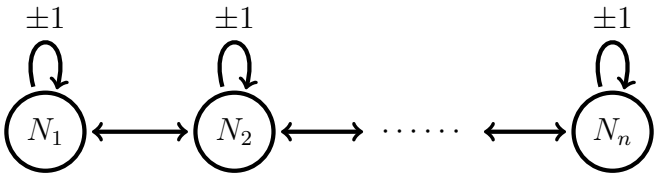
As a function of time t since inoculation, the vector of positive integer state variables $\boldsymbol{N}(t) = \{N_1(t), ..., N_n(t)\}$ represents the simultaneous abundance of bacteria in compartments 1 to n. In the context of IT studies, this represents a single tagged strain. Three types of stochastic events drive the bacterial dynamics: division (which adds one bacterium to a given compartment), death (which removes one bacterium from a given compartment) and migration (which moves one bacterium from one compartment to another). Assuming linear transition rates, we have a total of n division rates $r_i N_i$ and death rates $k_i N_i$ within organ i, and $n(n-1)$ migration rates $m_{i,j} N_i$ from compartment i to j. Note that specific models may assume that some of the parameters are equal to zero, for example if there is no physical connection between given pairs of compartments.

In particular, we consider two geometries, illustrated in Figure 1, corresponding to two typical anatomical topologies of relevance to bacterial infection: a radial network with a central compartment (e.g., bloodstream supplying every organ), and a linear network (e.g., digestive track).

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



(a) Radial network.



(b) Linear network.

Figure 1: Diagram illustrating the two types of network structure we consider.

2.3 Computation of the first- and second-order moments

The method we propose for parameter inference relies on the first two moments of the stochastic system. That is, we use only the expected number of bacteria within each compartment, the variance of the number of bacteria within each compartment, and the pair-wise covariances. A simple approach to generating these moments for a particular stochastic system in terms of the model parameters, is given by [26]. Letting $\boldsymbol{\lambda} = \{r_1 N_1, ..., r_n N_n, k_1 N_1, ..., k_n N_n, m_{1,2} N_1, ...\}$ be the vector of transition rates, we can write the hth non-central moment of the state of the ith compartment as:

$$\frac{\partial}{\partial t}\mathbb{E}\left[N_i^h(t)\right] = \mathbb{E}\left[\sum_j \lambda_j(t)\left[\phi_j(N_i(t))^h - N_i(t)^h\right]\right], \tag{1}$$

where ϕ_j is a function describing the change of the state for the jth transition. In Supplementary Materials (S1.7), we show that this leads to a closed, linear system of differential equations for the first moments. Letting $\boldsymbol{M}_1(t) = \{\mathbb{E}\left[N_i(t)\right], 1 \leq i \leq n\}$ be the vector of first moments as a function of time, we can express these differential equations in matrix

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

form as:

$$\frac{\partial}{\partial t}\boldsymbol{M}_1(t) = \mathbf{A} \times \boldsymbol{M}_1(t), \tag{2}$$

which leads to the solution,

$$\boldsymbol{M}_1(t) = \exp(\mathbf{A}t) \times \boldsymbol{M}_1(0), \tag{3}$$

where $\boldsymbol{M}_1(0)$ are the initial conditions of the system, $\mathbf{A}$ is a time-independent matrix containing the model parameters, and *exp* is the matrix exponential function.

Next, let $\boldsymbol{M}_2(t) = \{\mathbb{E}\left[N_i(t)N_j(t)\right], 1 \leq i \leq n, 1 \leq j \leq n\}$ be the vector of second moments. By applying Duhamel's formula to the differential equations obtained from equation (1), we obtain the following expression for the second-order moments:

$$\boldsymbol{M}_2(t) = \exp(\mathbf{C}t) \times \boldsymbol{M}_2(0) + \exp(\mathbf{C}t) \times \left[\int_0^t \exp(-\mathbf{C}s) \times \mathbf{B} \times \exp(\mathbf{A}s)ds\right] \boldsymbol{M}_1(0), \tag{4}$$

where $\mathbf{B}$ and $\mathbf{C}$ are time-independent matrices containing model parameters, and $\boldsymbol{M}_1(0)$ and $\boldsymbol{M}_2(0)$ are vectors containing the initial moments. Using the numerical method for matrix exponential in [27], we can evaluate the first- and second-order moments at any time point. Remarkably, no moment-closure approach is required as the expressions for the second-order moments are independent of higher-order moments. See Supplementary Information for a full derivation.

2.4 Parameter Inference by Divergence Minimisation

Given a dataset consisting of one or more matrices $\mathbf{D}$ of bacterial counts (as per section 2.1), and a stochastic model (as per section 2.2), we now describe methods to estimate the parameter values of the model that minimise the divergence between the predicted and observed distributions of bacterial abundance, using only the lower moments of those distributions. Specifically, we evaluate the means, variances and covariances of the N_i variables at a given time t. From the corresponding matrix $\mathbf{D}$, we calculate the vector of observed means $\boldsymbol{\mu}^{(\mathbf{D})}$ and the matrix of observed variance-covariance $\mathbf{V}^{(\mathbf{D})}$; and from the model's solution given a set of parameters $\boldsymbol{\theta}$, we compute the vector of predicted means $\boldsymbol{\mu}^{(\theta)}$ and the matrix of predicted variances-covariances $\mathbf{V}^{(\theta)}$ which can be derived from $\boldsymbol{M}_2(t)$. We compared four common divergence measures: a Chi-Squared metric, the Mahalanobis distance, the Hellinger distance, and the Kullback-Leibler divergence.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Note that none of the measures below is designed to deal with the particular situation when any of the organs is reported void of bacteria (i.e., $N_i = 0$) in all replicates at a given time point, i.e. if all the observed moments related to that organ are equal to zero. In some experimental systems, this may occur as an artefact of the observation method, e.g., when only a small sample is measured: in this case, it is possible to "correct" the data for sampling biases (see Section 2.7). Otherwise, a simple solution would be to remove the moments relative to that organ from the inference procedure. In some cases, it may make sense to completely remove the empty organ from the model if no meaningful inference can be expected from its inclusion, as illustrated in Section 3.3.

The Chi-Squared metric adds up the squared pairwise-differences between each predicted moment and its corresponding observed moment, each term being scaled by the magnitude of the observed moment. As a result, all moments are effectively treated equally. The expression for this divergence is:

$$\Delta_\chi = \sum_i \frac{(\boldsymbol{\mu}_i^{(\theta)} - \boldsymbol{\mu}_i^{(D)})^2}{|\boldsymbol{\mu}_i^{(D)}|} + \sum_{i,j} \frac{(\mathbf{V}_{i,j}^{(\theta)} - \mathbf{V}_{i,j}^{(D)})^2}{\mathbf{V}_{i,j}^{(D)}} \tag{5}$$

The other three divergence expressions we tested make use of the fact that our chosen statistics $\boldsymbol{\mu}$ and $\mathbf{V}$ are moments of distributions. Since we do not establish complete characterisations of the distributions, we decided to borrow the expressions of well-known divergences measures for multivariate normal distributions, as these only require the knowledge of their first and second-order moments. In other words, we compute the divergence between two multivariate normal distributions with respective moments $(\boldsymbol{\mu}^{(\mathbf{D})}, \mathbf{V}^{(\mathbf{D})})$ and $(\boldsymbol{\mu}^{(\boldsymbol{\theta})}, \mathbf{V}^{(\boldsymbol{\theta})})$.

The Mahalanobis distance is measured between each point in the observed data, and the distribution described by the predicted moments. That is, by estimating the parameters we are trying to find the distribution that these data are most likely to have been sampled from. Specifically, each observation $\boldsymbol{X_j}$ is given by column j of the data matrix $\mathbf{D}$. The divergence is obtained by summing the distances to the predicted distribution from every observation:

$$\Delta_M = \sum_j \sqrt{(\boldsymbol{X_j} - \boldsymbol{\mu}^{(\theta)})^T \ \mathbf{V}^{(\theta)^{-1}} (\boldsymbol{X_j} - \boldsymbol{\mu}^{(\theta)})} \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The squared Hellinger distance is given by:

$$\Delta_H^2 = 1 - \exp\left(-\frac{1}{8}\,\boldsymbol{\mu}^T\,\mathbf{V}^{-1}\,\boldsymbol{\mu} - \frac{1}{2}\log\frac{|\mathbf{V}|}{\sqrt{|\mathbf{V}^{(\theta)}||\mathbf{V}^{(\mathbf{D})}|}}\right), \tag{7}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}^{(\theta)} - \boldsymbol{\mu}^{(\mathbf{D})}$ and $\mathbf{V} = \frac{\mathbf{V}^{(\theta)}+\mathbf{V}^{(\mathbf{D})}}{2}$.

The Kullback-Leibler divergence from the predicted distribution to the observed one is given by:

$$\Delta_K = \frac{1}{2}\left[\mathrm{tr}(\mathbf{V}^{(\mathbf{D})^{-1}}\,\mathbf{V}^{(\theta)}) - n + \boldsymbol{\mu}^T\,\mathbf{V}^{(\mathbf{D})^{-1}}\,\boldsymbol{\mu} + \log\frac{|\mathbf{V}^{(\mathbf{D})}|}{|\mathbf{V}^{(\theta)}|}\right], \tag{8}$$

where $\mathrm{tr}(\cdot)$ is the trace operator. The derivations of the Hellinger and Kullback-Leibler divergences are given in Supplementary Materials S1.2.

Even though these expressions may not provide correct estimates of the actual "Mahalanobis distance", "Hellinger distance" and "Kullback-Leibler divergence" between the data and the predicted distributions (as these are not normally distributed), they still provide adequate divergence measures: they all return positive values which are only equal to zero (hence are minimised) when $\boldsymbol{\mu}^{(\theta)} = \boldsymbol{\mu}^{(\mathbf{D})}$ and $\mathbf{V}^{(\theta)} = \mathbf{V}^{(\mathbf{D})}$, which occurs when the data are drawn from the predicted distribution. We use that property for the purpose of parameter inference, given a dataset and a model. The set of parameter values $\boldsymbol{\theta}$ that minimises the divergence measure is termed the "minimum divergence estimate" (MDE).

All code was written in R [28]. We used the `UObyQA` optimisation routine in the `powell` package [29].

2.5 Bootstrap Variance Estimate

In order to quantify the uncertainty in our parameter estimates conditional on the dynamic model considered, we utilise the parametric bootstrap method (e.g., [30]), which can be exploited simply here due to the computational efficiency of the inference approach. Having obtained an MDE, $\boldsymbol{\theta}^*$ for a given data matrix $\mathbf{D}$ and a given model, we simulate B data sets from the model at these parameters (i.e., $\boldsymbol{x}^b \sim f(x \mid \boldsymbol{\theta}^*)$, $b = 1, \ldots, B$). For each simulation, we estimate the corresponding parameters using the MDE technique ($\boldsymbol{\theta}^b$, $b = 1, \ldots, B$). These B estimated parameters are used to estimate the variance-covariance matrix of the parameter estimates. Subsequently, one can use this matrix to estimate

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

confidence intervals in a number of ways [31]. For simplicity, the analysis in Section 3.3 uses the `ellipse` package in R [32] to obtain approximate 95% confidence intervals of the model parameters (that is, assuming the bootstrapped parameter estimates follow a multivariate normal distribution).

2.6 Model Goodness-of-fit

In order to assess the model goodness-of-fit, we once again utilise the parametric bootstrap. Concurrent to calculating the uncertainty estimates using the MDE method, the divergences corresponding to each simulated data set at their estimated parameter values are recorded. These bootstrapped divergences can thus be used to represent the null distribution of divergences for the model at the estimated parameter values – giving a representation of the divergences we should expect from the model at these values. The divergence estimated for the observed data is then compared to the null distribution to obtain a p-value for the hypothesis that the data could have been generated by the model.

2.7 Data analysis with observational noise

A common source of error when fitting a model to experimental data is the observation process. In the type of microbiology experiments considered here, where the data represent bacterial loads in infected animals, there are at least two steps that affect the accuracy of the measurements: sampling (when only part of the bacterial loads are recovered) and and quantitation (the process by which the number of bacterial cells in the samples is measured). For example, in a recent IT experiment [11], sampling error was modelled as a binomial process (as known fractions of each homogenised organ were plated) and quantitation error was modelled using a log-normal distribution which was estimated empirically using an independent control experiment (in which known numbers of bacterial colonies were processed by qPCR in the same way as bacterial samples extracted from animals in the main experiment). Although both error distributions were centred on the true bacterial numbers (i.e. the mean numbers of bacteria were not biased), the variance and covariance in the reported data would not have been accurate estimates of the variance and covariance within the animals: hence our MDE could be biased if we did not account

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

for observational errors.

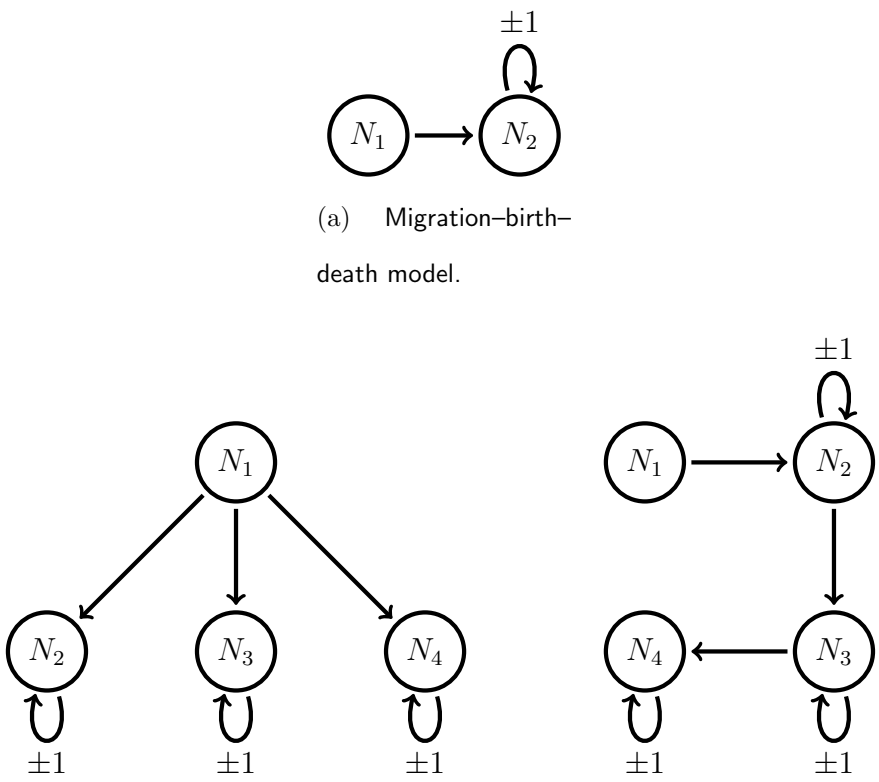
In our present reanalysis of those data, we integrate both error terms before performing parameter inference, using the following procedure. Our goal is to propose a simple heuristic which could be applied to any experimental dataset with known (or assumed) observational error distributions. First, we generate a large number of stochastic simulations of the model under a biologically reasonable range of parameter values (i.e. using uniform prior distributions across sensible ranges) to generate "perfect" observations. We then calculate the corresponding "perfect moments" from every simulated dataset. Next, we apply the observation noise to the simulated data, and calculate the corresponding "observed moments". We then use linear regression models to establish a relationship between each of the perfect and observed moments (with transformations where appropriate), with weights given by the simulations proximity to the actual observed moments from the experiment. This calibrated regression model is then applied to the moments of the experimental data of interest, in order to estimate the moments of the true, unobserved bacterial loads in the animals. We eventually compute MDE using these corrected moments.

We note that for some models of the observation process, it may be possible to establish analytic relationships between the "perfect" and "observed" moments using conditional expectation and variance theory (e.g., [33]). The choice of suitable correction methods for a given system will depend on both the model complexity and the level of empirical quantification of observation noise.

3 Results

We begin by demonstrating the speed and accuracy of the computation of the first- and second-order moments for arbitrary network structures, compared to Gillespie simulations. We then assess the four previously mentioned divergence measures to validate our minimum divergence estimation procedure with each model, across a range of parameter values. Finally, we apply our inference method to reanalyse a published experimental dataset, and compare the results with the previous maximum likelihood estimates.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



(a) Migration–birth–death model.

(b) Radial network.

(c) Linear network.

Figure 2: Illustrations of three representative model structures.

3.1 Moment computation

As a proof-of-concept, we compare our proposed computation method of the means, variances and covariances of bacterial loads, with the values derived from large numbers of Gillespie simulations, across a range of model structures and parameter values. Our results illustrate the computational effort required to obtain the same level of accuracy with each method. For each model and each parameter set, we simulated 100 experiments, each consisting of 10, 50, 100, 250, 500, or 1000 observations (representing the product of the number of animals A by the number of tagged strains T as per section 2.1) at a given time, without observational error. The initial bacterial loads at $t = 0$ in each replicate experiment were drawn at random from a Poisson distribution, to mimic typical variability in inoculum doses in experiments [11]. We considered three model structures, illustrated in Fig. 2: (a) basic migration-birth-death model, (b) four-compartment radial network, and (c) four-compartment linear network. For each virtual experiment, we calculated the

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

first- and second-order moments of the bacterial loads in two ways at a given time point ($t = 6$ for the basic migration-birth-death model, and $t = 4$ for the four-compartment radial and linear network).

We randomly drew the stated number of initial conditions (ranging from 10 to 1000), and evaluated the moments at the future time point in the following two ways. First, we used Gillespie simulations to progress each initial condition forward to the stated observation time, and then calculated the moments from the collection of simulations. Second, we took the stated number of initial conditions to estimate the moments at time zero, and used these to evaluate the moments at the observation time using equations (3) and (4); we refer to this as the direct moment-calculation method.

Figure 3 shows that there is still a greater amount of variation in the moments from 1000 Gillespie simulations, compared to the direct approach. The direct method produces less variable moment estimates as experiment size increases, as a result of more accurate estimates of the initial distribution of bacteria. Furthermore, the computation time of the Gillespie approach increases steadily with the number of simulations, while the direct method is consistently more efficient, independent of the size of the experiment (Figure 4). Similar patterns are shown in the Supplementary Materials (S1.3) for the four-compartment linear and and radial network models.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

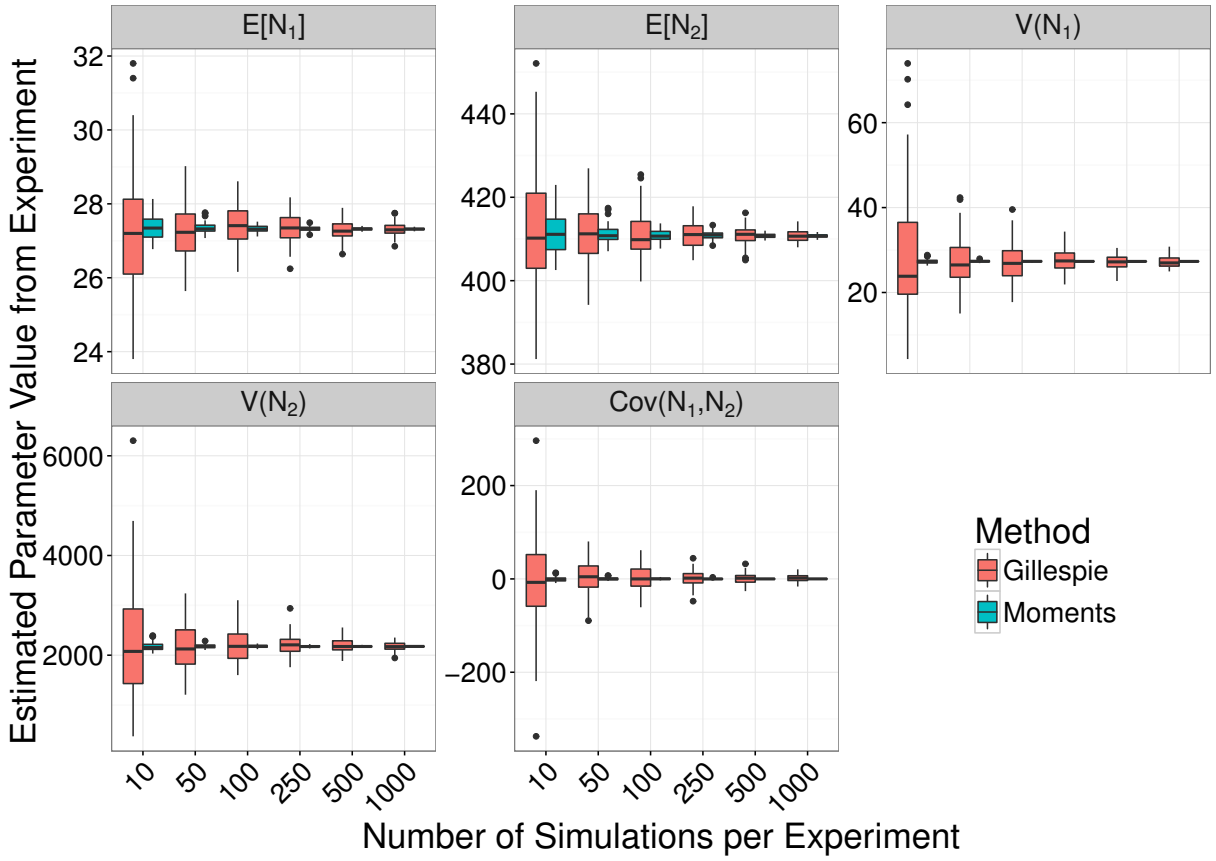


Figure 3: Box plots of estimated moments from 100 simulated experiments using the two-compartment birth-death-migration model (Figure 2a). For each experiment (of size 10, 50, 100, 250, 500 or 1000), the lower moments of variables N_1 and N_2 at time $t = 6$ were calculated from Gillespie simulations (red boxes) or from the moments equations ("direct method", blue boxes), starting from Poisson-distributed initial conditions. The stated number of simulations per experiments is equivalent to the total number of replicate observation ($A_i T$ at time point τ_i as per section 2.1).

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

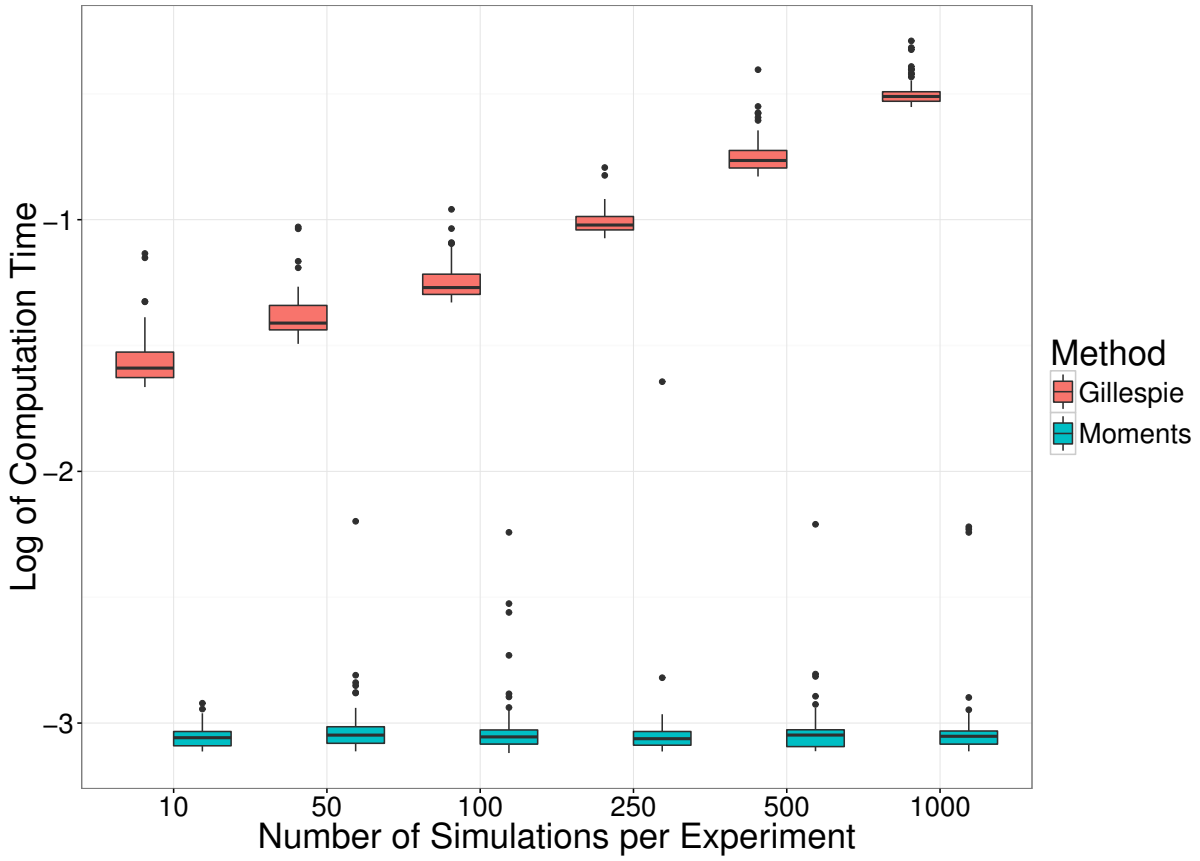


Figure 4: Box plots of computation time to evaluate the moments from 100 simulated experiments of the two-compartment birth-death-migration model, as per Figure 3. The vertical axis shows the logarithm in base 10 of running times in seconds.

3.2 Divergence Measures: Simulation Study

Next, we compare the accuracy of MDE among the four candidate divergence measures, using data simulated from the same three model structures as in the previous section and across a range of parameter values. For each model structure and each parameter combination, we ran 100 series of 100 Gillespie simulations from $t = 0$ to each of 8 observation times, representing 800 experiments (each of size 100) with "perfect" observations of the system (i.e., without experimental noise). This process was repeated for a set of different parameter values for each model to represent different scenarios – three representative values for each parameter in the basic model ($3^3 = 27$ scenarios), and five randomly generated sets of parameters for the four compartment linear and radial networks with parameters. The initial conditions for each simulation, in every scenario, were randomly

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

generated from a Poisson distribution with mean parameter 200. The optimisation routine was initiated at randomly generated conditions each time, with each parameter values drawn independently from a Uniform(0,1) distribution.

From the results of each simulated experiment, we computed the MDE using each of the Chi-Squared, Mahalanobis, Hellinger, and Kullback-Leibler divergences. The performance of each divergence was measured by the mean absolute relative error (MARE). That is, under scenario s, with p target parameters $\boldsymbol{\theta}_s = (\theta_{1s}, \theta_{2s}, \ldots, \theta_{ps})$, and estimated parameters $\hat{\boldsymbol{\theta}}_{sj} = (\hat{\theta}_{1sj}, \hat{\theta}_{2sj}, \ldots, \hat{\theta}_{psj})$ for the jth simulation, the MARE is given by:

$$\text{MARE}_{s,j} = \frac{1}{p} \sum_{i=1}^{p} \frac{|\theta_{is} - \hat{\theta}_{isj}|}{\theta_{is}}$$

Figure 5 displays the average error across each scenario for the four divergence measures, at a range of observation times, for each of the three models under consideration. We can see that both the Hellinger and Kullback-Leibler divergence measures perform considerably better than the Chi-Squared and Mahalanobis divergences for the Basic, Linear and Radial networks. The similar performance of the Hellinger and Kullback-Leibler divergences is not unexpected, due to their relationship. We observe a marginal advantage in favour of the Kullback-Leibler divergence for the basic and radial network, and thus we use the Kullback-Leibler divergence in analysing experimental data in Section 3.3. The results aggregated by scenario for each model are shown in Supplementary Materials S1.4.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

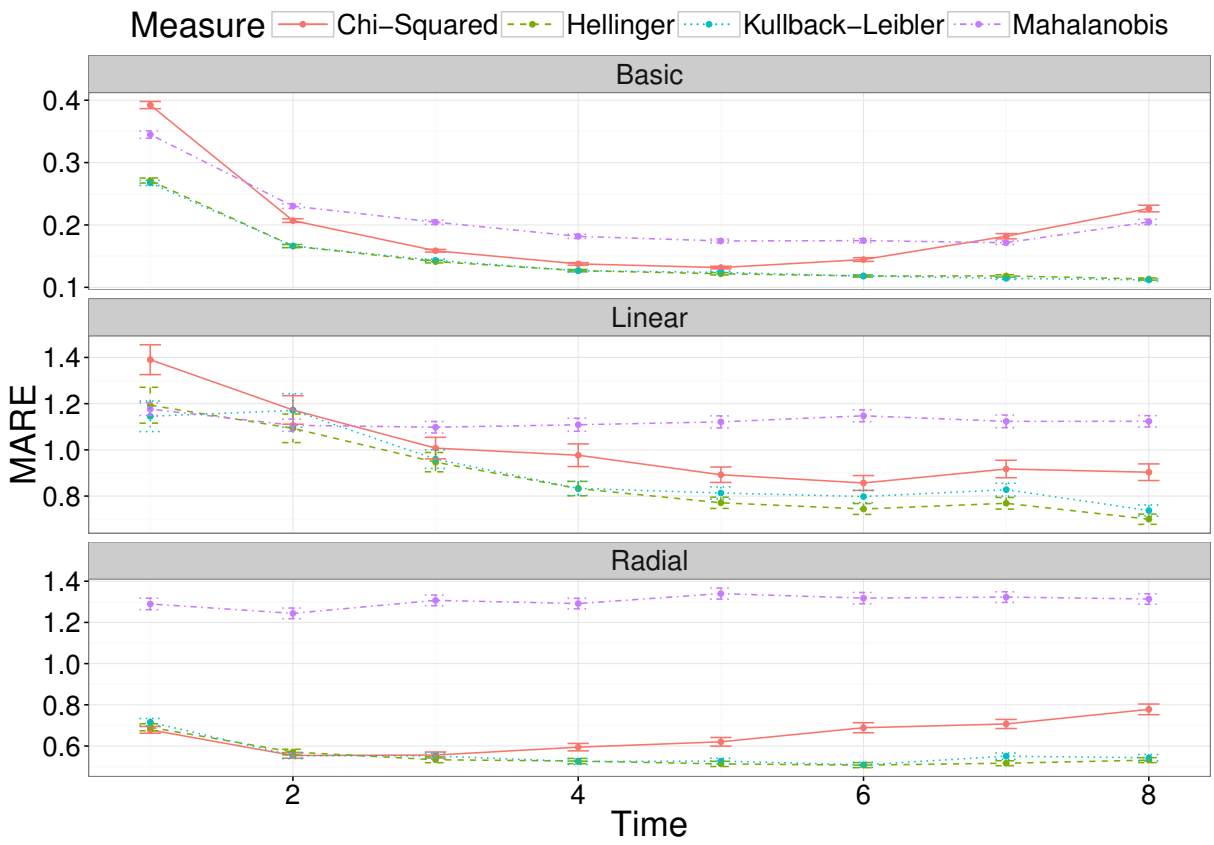


Figure 5: The average mean absolute relative error (MARE) across a number of observation times, for the two-compartment network (top panel), the linear four-compartment network (middle panel) and the radial four-compartment network (bottom panel). Each point is the average for that measure, at that observation time, with error bars representing the standard error.

Figure 9 further illustrates the distribution of parameter estimates for a few selected scenarios and time points. For each time point, the variation was generated by the stochastic birth-death-migration process, demonstrating the level of uncertainty in parameter estimates due to the biological process itself. In this case, each simulated experiment was made up of 100 simulations; fewer replicates would increase the parameter uncertainty. At each time point, the estimates of the migration rate m_{12} are much less variable than hose of the killing rate k_2 and replication rate r_2: this is actually caused by a very strong positive correlation between the latter two parameters (Figure 7).

This can be understood intuitively, as the net growth rate $r_2 - k_2$ is the main determinant of the mean bacterial load in compartment 2. We also remark that the earlier

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

observation time contains more information about the clearance rate, whereas the later observation time better captures the replication and killing rates. Results of all scenarios are shown in Supplementary Materials S1.6.

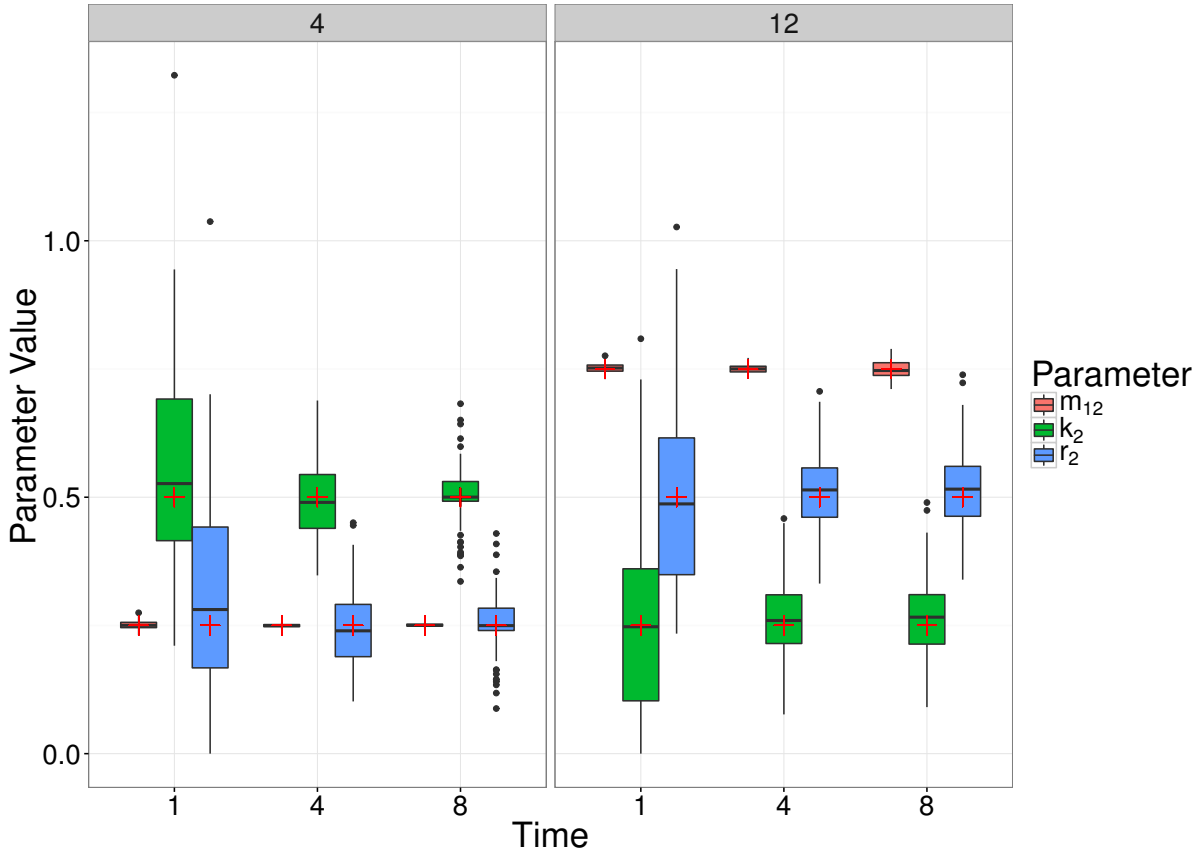


Figure 6: Box plots of MDE's corresponding to simulated data for two different choices of parameter values (Scenario 4 and 12), at a number of different observation times (1, 4 and 8) of the simple model. The box plots represent 100 parameter estimates corresponding to the 100 simulations, and the red crosses denote the true parameter value.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

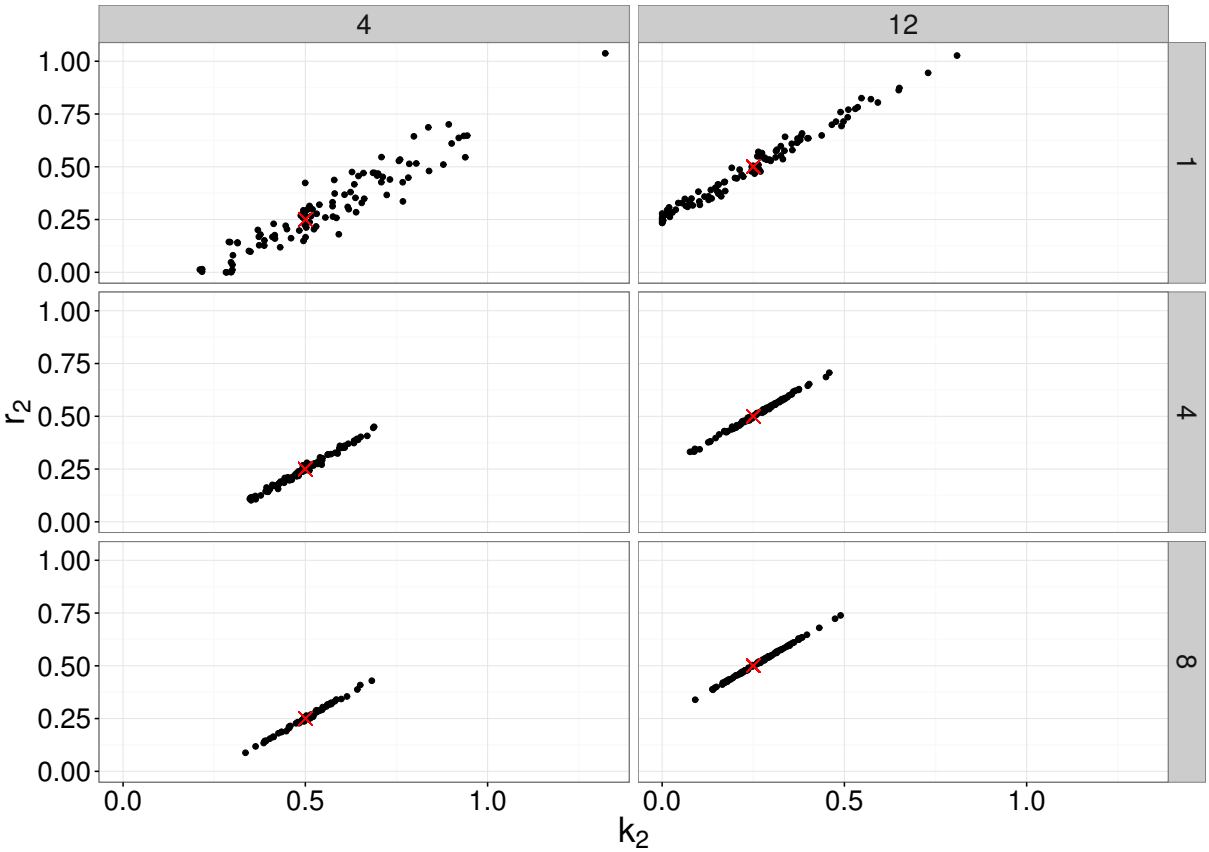


Figure 7: Bivariate plot of MDE's corresponding to simulated data for two different choices of parameter values (Scenario 4 and 12), at a number of different observation times (1, 4 and 8) of the simple model. The points represent 100 parameter estimates of r_2 and k_2 corresponding to the 100 simulations, and the red crosses denote the true parameter value.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

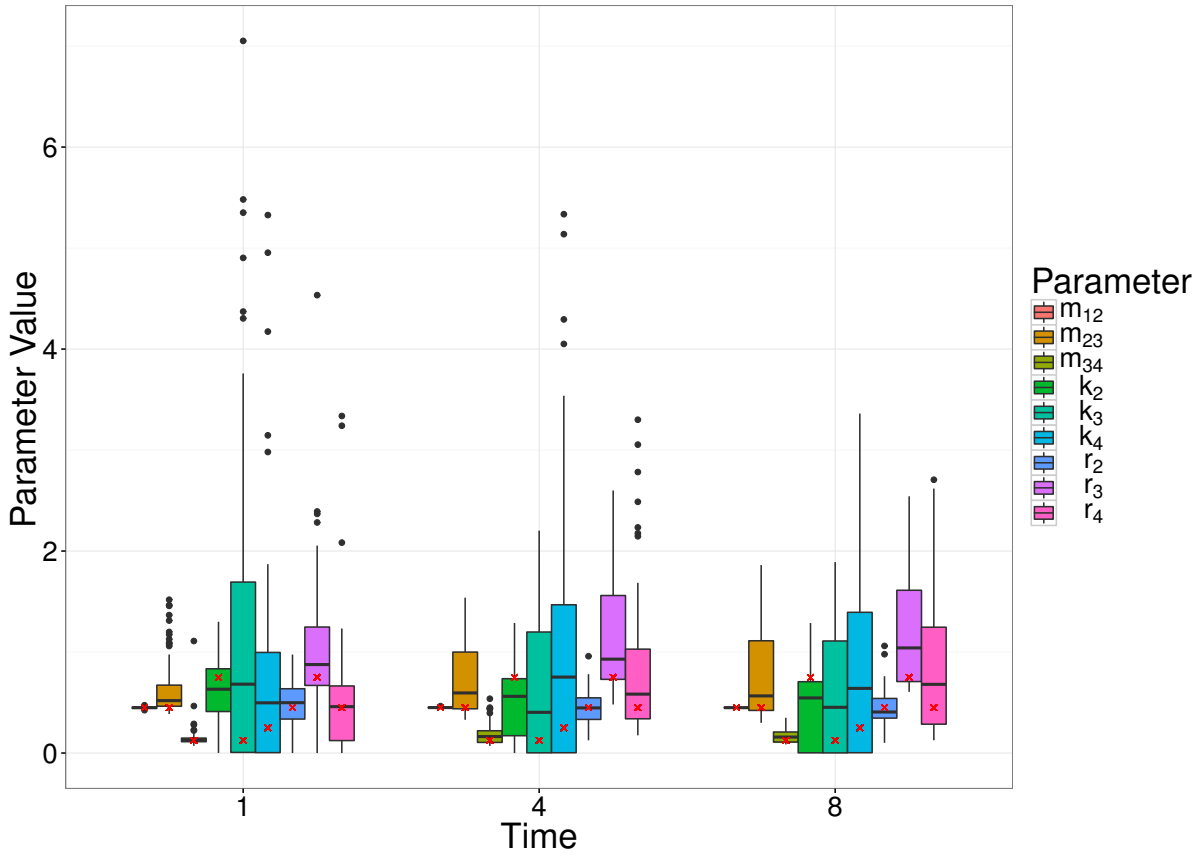


Figure 8: Box plots of MDE's corresponding to simulated data for a random set of parameter values (Scenario 4), at a number of different observation times (1, 4 and 8) of the linear model. The box plots represent 100 parameter estimates corresponding to the 100 simulations, and the red crosses denote the true parameter value.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

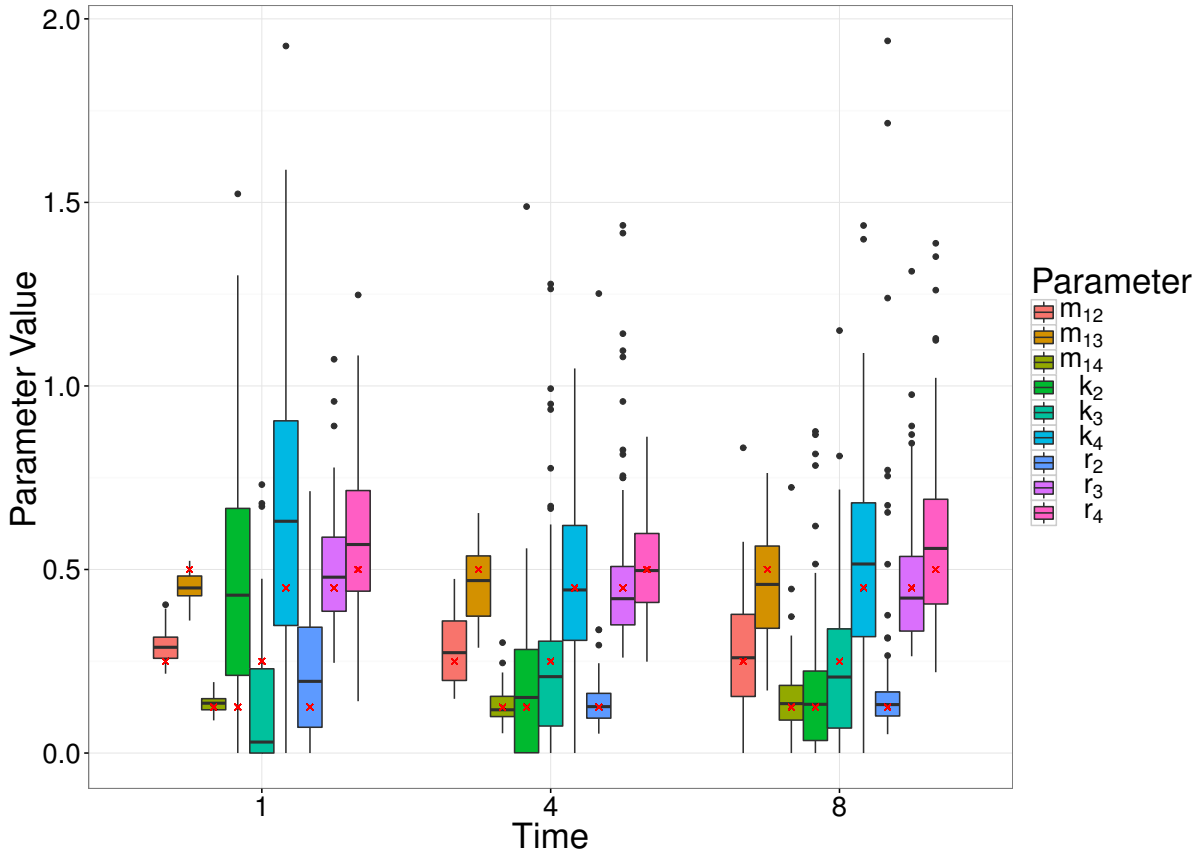


Figure 9: Box plots of MDE's corresponding to simulated data for a random set of parameter values (Scenario 4), at a number of different observation times (1, 4 and 8) of the radial model. The box plots represent 100 parameter estimates corresponding to the 100 simulations, and the red crosses denote the true parameter value.

3.3 Analysis of experimental data

3.3.1 Description of the system

Finally, we performed a re-analysis of experimental data from [11], in which groups of mice received an intravenous dose of *Salmonella enterica* Typhimurium, composed of an even mixture of 8 wildtype isogenic tagged strains (WITS). Bacterial loads and WITS composition were measured in the blood, liver and spleen of 10 mice at each observation time. During the first phase of the WITS experiments (represented by the first two time points at 0.5 and 6 hours post inoculation), the only biologically relevant processes considered are: migration of bacteria from blood to liver or spleen, replication and death

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

inside the liver and the spleen. That is, on state space $\mathcal{S} = \{(n_B, n_L, n_S) \mid n_B, n_L, n_S \geq 0\}$, we have the following transition rates:

$$q_{(n_B,n_L,n_S),(n_B-1,n_L+1,n_S)} = c_L n_L, \qquad q_{(n_B,n_L,n_S),(n_B-1,n_L,n_S+1)} = c_S n_S,$$

$$q_{(n_B,n_L,n_S),(n_B,n_L+1,n_S)} = r_L n_L, \qquad q_{(n_B,n_L,n_S),(n_B,n_L,n_S+1)} = r_S n_S,$$

$$q_{(n_B,n_L,n_S),(n_B,n_L-1,n_S)} = k_L n_L, \qquad q_{(n_B,n_L,n_S),(n_B,n_L,n_S-1)} = k_S n_S,$$

where c_L, c_S are the clearance (=migration) rates from the blood into the liver and spleen respectively, r_L, r_S are the replication rates in the liver and spleen respectively, and k_L, k_S are the killing rates in the liver and spleen, respectively.

One purpose of the study in [11] was to compare the effects of two vaccines on the dynamics of infection. Here, we re-analyse two of the experimental groups: an untreated group (naive) who received no vaccine, and a group who received a live-attenuated vaccine (LV). In the naive group, in addition to the clearance, replication and killing rates, we estimated the effective inoculum size $i_0 = N_B(0)$ to account for the possibility that some of the inoculated bacteria failed to initiate an infection process [11]. In the vaccinated group, no bacteria were recovered from the blood at either of the two observation times, hence we were unable to estimate the corresponding migration rates. Instead we assumed that the transfer of bacteria to the organs was effectively instantaneous, and sought to estimate the effective initial loads in the liver and spleen (i.e., $i_L = N_L(0)$ and $i_S = N_S(0)$), as well as the replication and killing rates. In each case, the inoculum size is assumed to be Poisson distributed.

3.3.2 Parameter estimation with observation noise

We applied the MDE method to estimate the parameters corresponding to these systems, and compared the results to the estimates obtained via the MLE approach, presented in [11]. Having obtained the MDE parameter estimates, we used the bootstrap approach to calculate uncertainty intervals on the estimates, and assess the model goodness-of-fit.

As described in [11], the recorded data were subject to observation noise. In particular, only a fraction of some organs were sampled (e.g., as we are not able to fully recover the total amount of blood from an individual mouse), and there was noise introduced via the qPCR. The observed moments were adjusted using the simulation-

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

based, pre-processing procedure described in Section 2.7: we assumed that the form of the sampling was binomial, and the qPCR noise has been previously modelled using a log-Normal distribution [11]. For example, for the naive group, we sampled 10,000 parameters $(c_L, c_S, k_L, k_S, r_L, r_S)$ each independently from $U[0, 3]$, and initial inoculum $i_0 \sim \text{Poisson}(30.33)$ (the mean WITS population calculated from plating experiments [11]). "Perfect moments" were calculated from simulations of the same size as the experimental data. The corresponding observed moments were then calculated by applying the observation noise model – binomial sampling to represent the fractional sampling of organs within each mouse, and log-Normal noise to represent the qPCR noise – to each simulation, and calculating the moments. The variances were log-transformed for the purpose of the regression models to predict the corresponding perfect moments from those observed from the experiment.

Figure 10 provides a comparison of the MLE and MDE estimation procedures for the naive and LV experimental groups. The box plots represent the bootstrapped parameter estimates, with the MDE parameter estimate (blue dots), and MLE parameter estimates (red crosses).

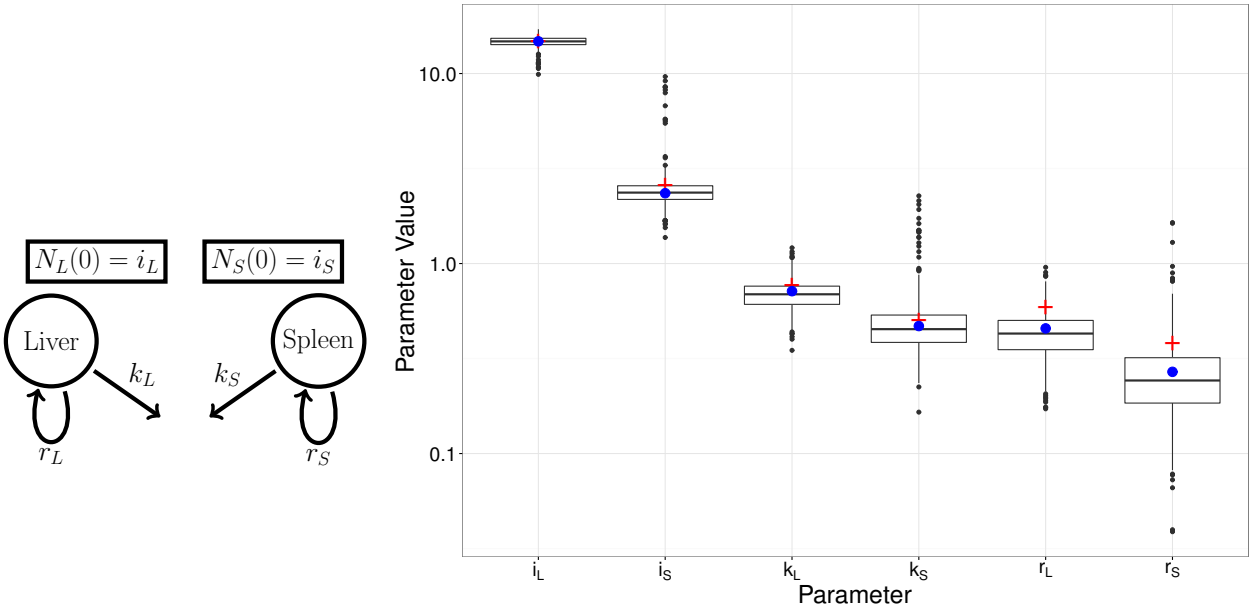
As explained in the previous section, the strong positive correlation between the replication and killing rates within each organ, clearly visible on Figure 11, is characteristic of this system.

3.3.3 Model goodness-of-fit

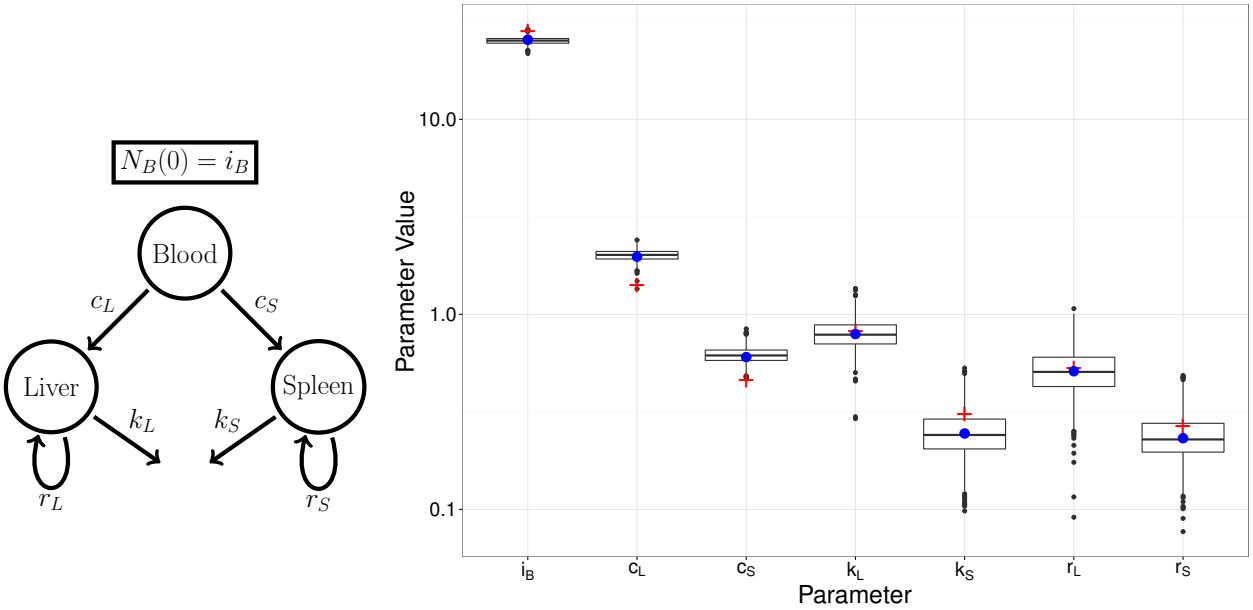
In calculating the MDE parameter estimates for the bootstrap samples, we evaluated divergence measures for datasets simulated from the model. Thus, the collection of these divergences provide a suitable representation of the null distribution of divergences under this model. We compared the divergence for our observed data set to this null distribution in order to assess the model goodness-of-fit. Figure 12 demonstrates the goodness-of-fit measure for the models fit to the two experimental groups.

The approximate p-values associated with the goodness-of-fit tests for the naive and LV experimental groups are 0.002 and 0.882, respectively. This suggests that the model provides a suitable fit to the vaccinated experimental group; however, there is a non-

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



(a) Live-Vaccinated group



(b) Naive group

Figure 10: Left: Diagram of each model illustrating the relevant compartments, initial conditions, and rates of interest. Right: MDE parameter estimate (blue dot), MLE parameter estimate (red cross), and box plots of the bootstrapped parameter estimates.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

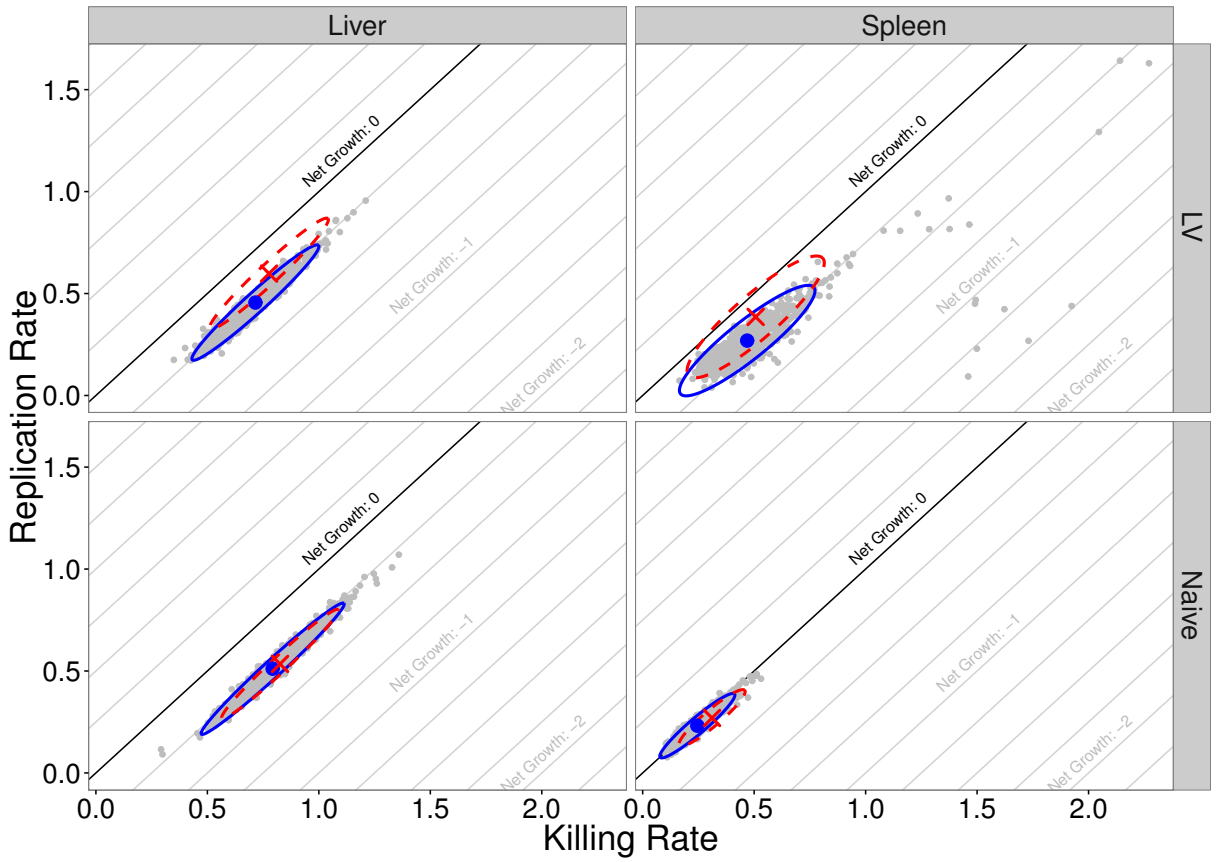


Figure 11: Bivariate distributions of replication and killing rates in liver and spleen, for both naive and live-vaccinated groups. Blue circles are the MDE values, with the blue (solid) ellipses representing the 95% confidence ellipses calculated using the 1000 bootstrap samples (grey points). The red crosses are the MLE values, with red (dashed) ellipse calculated using the hessian evaluated at the MLE.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

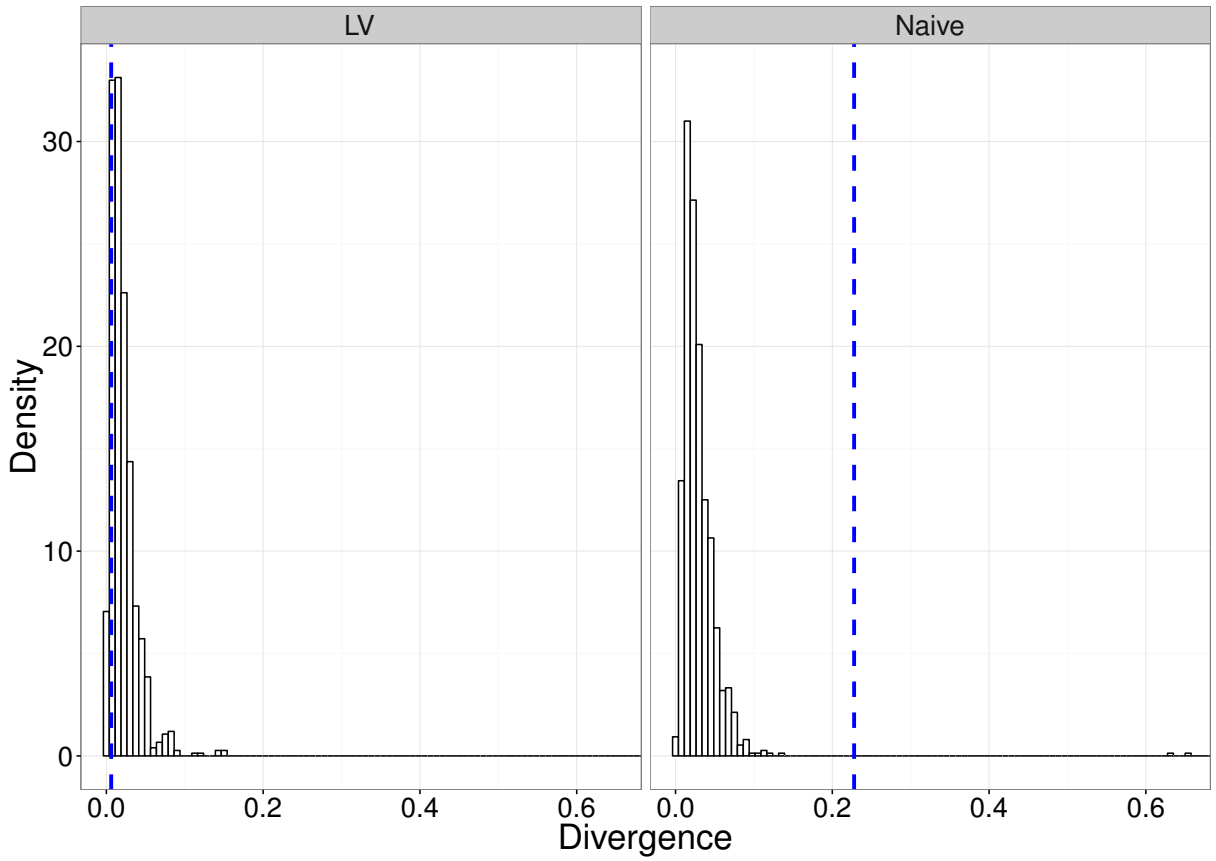


Figure 12: Figures demonstrating the goodness-of-fit of the two models to the respective data sets. The histogram bars are the bootstrapped estimate of the null distribution of divergences under the model at the estimated parameter values for the respective model. The (blue) vertical dashed-line is the divergence corresponding to the observed data set.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

negligible discrepancy between the model output for the naive group fit, and the observed data. This suggests that some of the assumptions in the original dynamic model may be erroneous; however, revisiting them goes beyond the scope of the present study.

As a further demonstration of the model goodness-of-fit, we compare the model output at each of the estimated parameter sets, to the observed data. Figure 13 shows these simulations from each of the MDE and MLE approaches, at each of the 0.5h and 6h observation times, for both the naive and vaccinated experimental groups. Observational noise consistent with that in the experimental data was added to the simulated data (i.e., binomial sampling for observed counts, and log-normal noise consistent with qPCR). The plots indicate that the main source of discrepancy between data and model in the naive-mice group is the distribution of bacterial loads in the blood after 6 hours.

4 Discussion

We have described a functional and flexible moments-based method to estimate parameters of stochastic metapopulation models, and applied it to experimental data on within-host bacterial infection dynamics. Compared to simulation-based methods, our technique delivers accurate estimation orders of magnitude faster. Although simulation-based inference has grown in popularity in computational biology—either within likelihood-free [17] or approximate-likelihood [16] approaches— even with the increasing availability of fast, multi-core computers, algorithms can still take days to converge in multi-parameter and multi-variable problems. While this is not a major issue when fitting a single model to a single dataset, it is a hindrance for more ambitious applications: indeed there is a growing demand in biology for model selection and model-based optimal experimental design, which are much more computationally intensive tasks (e.g., [34, 35, 36, 37]).

In its current form, our method can incorporate any linear, multivariate, continuous-time Markovian model, and fit it to experimental data that include multiple replicates and an arbitrary number of independent time points. These last two characteristics are typical of experimental biology, yet little attention has been given to systems of this form in recent statistical developments. Indeed, many inference methods for stochastic models

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

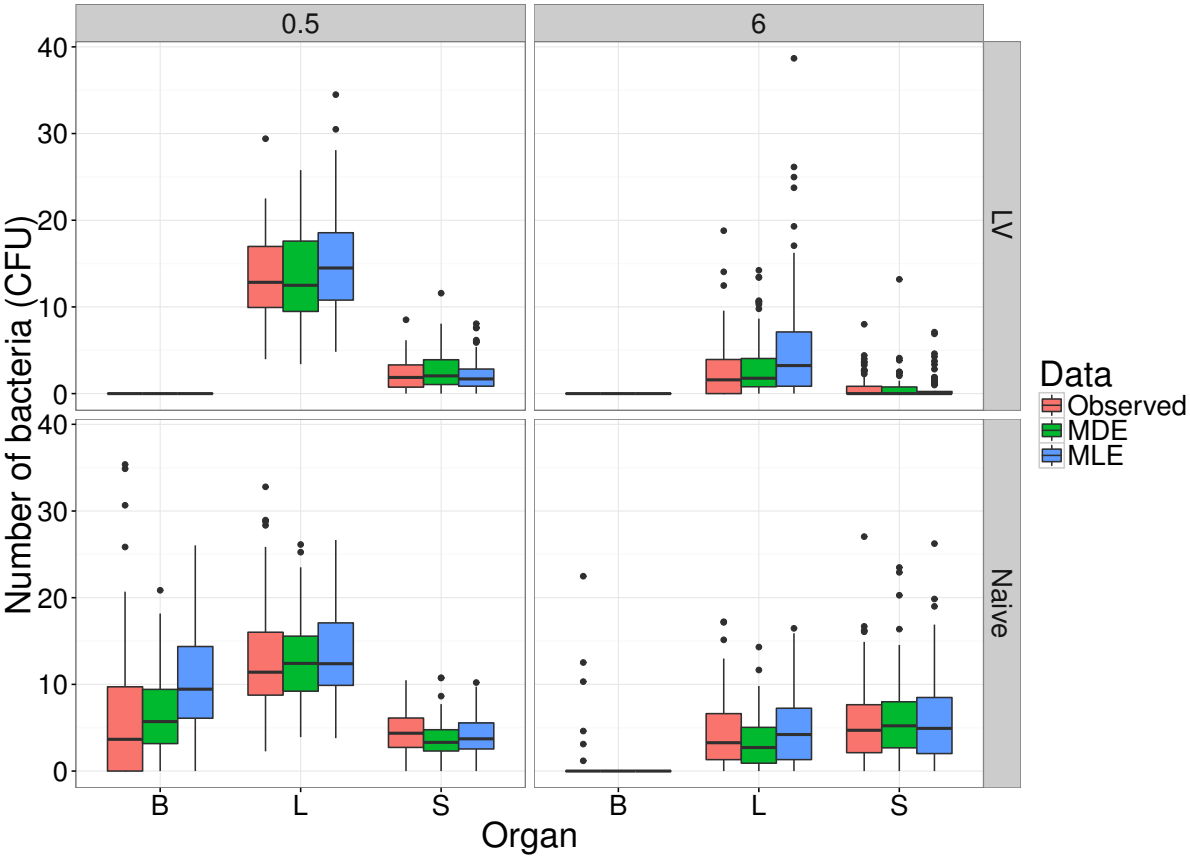


Figure 13: Observed experimental data for both naive and vaccinated groups, at both observation times, compared to simulations from the model at parameters estimated by both the MLE and MDE methods, in each of the blood (B), liver (L) and spleen (S). Numbers of bacteria are the sums of eight identical WITS.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(for example, driven by applications to epidemiology), target inference from a single time-series [38, 16].

In addition, we have provided a worked example for how to correct for observation noise, applicable to any given experimental system. While sampling error is typically taken into account in likelihood-based inference, many inference studies either ignore experimental noise (e.g. [24]) or choose arbitrary distributions (e.g. [39]). Even if experimental error has limited impact on the mean of the observations, it will affect the variance, with implications for the precision and reliability of statistical inference. The method we demonstrated here is based on pre-processing experimental data to effect an empirical correction of the observation noise. Crucially, it is not limited by mathematical tractability when combining multiple error sources, and it does not slow down the inference computation (i.e., it only needs to be performed once before fitting the model, and the same simulations are also used to adjust the moments of the noisy-simulated data).

The choice of a metric to minimise for likelihood-free inference is a common issue in computational statistics, especially in the context of increasingly popular ABC methods [40]. When the data and the model output are summarised by multiple statistics, the default option is to use either euclidian distance or a chi-squared type variant. The latter was chosen, for example, in a recent study using moment-based inference to solve a systems-biology problem [24]. However, because the summary statistics of interest are the moments of statistical distributions, we hypothesised it would be more informative to use a divergence metric instead. While many divergence measures exist for the purpose of comparing mathematically defined distributions [41], we are not aware of standard methods to compute divergence between multivariate distributions generated by complex stochastic processes. Instead, we took a pragmatic approach and used mathematical expressions for the Hellinger distance or the Kullback-Liebler divergence between multivariate normal distributions, because these expressions depend on the first- and second-order moments of the distributions only. Even though the resulting divergence measures are not the actual Hellinger distance or Kullback-Liebler divergence between our observed and predicted distributions, they still outperform the chi-squared metric based on parameter inference accuracy. It is worth noting that, across the three model structures we tested, there

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

was not a single metric that was consistently better than any other, and the differences in accuracy were often relatively small (Fig. 5). Depending on the degree of accuracy sought, it may be worth testing these and other metrics against simulations before applying this method to a different experimental system. It is our intention to provide a flexible blueprint that can be tailored to other problems, rather than a one-size-fits-all black box which may prove unreliable as soon as the circumstances change.

The particular molecular technology (isogenic tagging) that motivated the development of this inference method, has become pervasive in the study of within-host dynamics of bacterial infection in the last 10 years (see reviews by [42, 2, 43]). Yet, to our knowledge, this is the first attempt to provide a general modelling and inference framework that could be applied to any of these experimental systems. Indeed, previous efforts have been tailored to specific case studies [43, 7, 9, 13, 44, 8] despite asking fundamentally similar questions: how fast are bacteria replicating and dying? How much migration is taking place among organs or tissues? As soon as any two of these dynamic processes are co-occurring, it is not possible to evaluate them based solely on average bacterial loads: it is necessary to obtain reliable estimates of the variance, preferably within a single animal, by quantifying a set of independent and isogenic tags. The first known example goes back 60 years, using two naturally occurring mutants of *S. enterica* that could be distinguished by selective growth medium to investigate colonisation dynamics in mice. Although a wide range of bacterial tagging methods (including antibiotic markers and fluorophores) have been used since, non-coding DNA barcodes have opened new prospects as arbitrarily large numbers of truly isogenic tags can be generated and quantified by sequencing [14]. It is our hope that the inference framework presented here will contribute to the field's extension by providing much needed analytical support to analyse and design microbiology experiments.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgements

We thank the many microbiologists whose collaboration has helped initiate and shape this study, including Andrew Grant, Eric Harvill, Emma Slack, Chris Coward and Duncan Maskell. We are also grateful to fellow modellers Roland Regoes, Simon Frost, Julia Gog and Joshua Ross for their technical advice.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

[1] Meynell G. Use of superinfecting phage for estimating the division rate of lysogenic bacteria in infected animals. Microbiology. 1959;21(2):421–437.

[2] Crimmins GT, Isberg RR. Analyzing microbial disease at high resolution: following the fate of the bacterium during infection. Current opinion in microbiology. 2012;15(1):23–27.

[3] Barnes PD, Bergman MA, Mecsas J, Isberg RR. Yersinia pseudotuberculosis disseminates directly from a replicating bacterial pool in the intestine. Journal of Experimental Medicine. 2006;203:1591–1601. doi:10.1084/jem.20060905.

[4] Schwartz DJ, Chen SL, Hultgren SJ, Seed PC. Population dynamics and niche distribution of uropathogenic Escherichia coli during acute and chronic urinary tract infection. Infection and immunity. 2011;79(10):4250–4259.

[5] Walters MS, Lane MC, Vigil PD, Smith SN, Walk ST, Mobley HLT. Kinetics of uropathogenic Escherichia coli metapopulation movement during urinary tract infection. mBio. 2012;3(1):1–11. doi:10.1128/mBio.00303-11.

[6] Lowe DE, Ernst SMC, Zito C, Ya J, Glomski IJ. Bacillus anthracis has two independent bottlenecks that are dependent on the portal of entry in an intranasal model of inhalational infection. Infection and Immunity. 2013;81(12):4408–4420. doi:10.1128/IAI.00484-13.

[7] Grant AJ, Restif O, McKinley TJ, Sheppard M, Maskell DJ, Mastroeni P. Modelling within-host spatiotemporal dynamics of invasive bacterial disease. PLoS Biology. 2008;6:e74. doi:10.1371/journal.pbio.0060074.

[8] Melton-Witt JA, Rafelski SM, Portnoy DA, Bakardjiev AI. Oral Infection with Signature-Tagged Listeria monocytogenes Reveals Organ-Specific Growth and Dissemination Routes in Guinea Pigs. Infection and Immunity. 2012;80(2):720–732. doi:10.1128/iai.05958-11.

[9] Kaiser P, Slack E, Grant AJ, Hardt WD, Regoes RR. Lymph node colonization dynamics after oral salmonella typhimurium infection in mice. PLoS pathogens. 2013;9(9):e1003532. doi:10.1371/journal.ppat.1003532.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[10] Kaiser P, Regoes RR, Dolowschiak T, Wotzka SY, Lengefeld J, Slack E, et al. Cecum Lymph Node Dendritic Cells Harbor Slow-Growing Bacteria Phenotypically Tolerant to Antibiotic Treatment. PLoS Biology. 2014;12(2). doi:10.1371/journal.pbio.1001793.

[11] Coward C, Restif O, Dybowski R, Grant AJ, Maskell DJ, Mastoeni P. The Effects of Vaccination and Immunity on Bacterial Infection Dynamics *In Vivo*. PLOS Pathogens. 2014;10.

[12] Dybowski R, Restif O, Goupy A, Maskell DJ, Mastroeni P, Grant AJ. Single passage in mouse organs enhances the survival and spread of Salmonella enterica. Journal of the Royal Society, Interface / the Royal Society. 2015;12(113):20150702–. doi:10.1098/rsif.2015.0702.

[13] Li Y, Thompson CM, Trzciński K, Lipsitch M. Within-host selection is limited by an effective population of Streptococcus pneumoniae during nasopharyngeal colonization. Infection and immunity. 2013;81(12):4534–4543. doi:10.1128/IAI.00527-13.

[14] Abel S, Abel zur Wiesch P, Chang HH, Davis BM, Lipsitch M, Waldor MK. Sequence tag-based analysis of microbial population dynamics. Nat Methods. 2015;12(3):223–6, 3 p following 226. doi:10.1038/nmeth.3253.

[15] Crawford FW, Suchard Ma. Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution. Journal of mathematical biology. 2012;65(3):553–80. doi:10.1007/s00285-011-0471-z.

[16] King AA, Nguyen D, Ionides EL. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. Journal of Statistical Software. 2015;59(10). doi:10.18637/jss.v069.i12.

[17] Liepe J, Kirk P, Filippi S, Toni T, Barnes CP, Stumpf MPH. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. Nature protocols. 2014;9(2):439–56. doi:10.1038/nprot.2014.025.

[18] Ryan CM, Drovandi CC, Pettitt AN. Optimal Bayesian Experimental Design for Models with Intractable Likelihoods Using Indirect Inference Applied to Biological Process Models. Bayesian Analysis. 2016;11(3):857–883. doi:10.1214/15-BA977.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[19] Gillespie CS. Moment-closure approximations for mass-action models. IET Systems Biology. 2009;3(1):52–58. doi:10.1049/iet-syb:20070031.

[20] Sotiropoulos V, Kaznessis YN. Analytical derivation of moment equations in stochastic chemical kinetics. Chemical Engineering Science. 2011;66(3):268–277. doi:10.1016/j.ces.2010.10.024.

[21] Singh A, Hespanha JP. Approximate Moment Dynamics for Chemically Reacting Systems. IEEE Transactions on Automatic Control. 2011;56(2):414–418. doi:10.1109/TAC.2010.2088631.

[22] Smadbeck P, Kaznessis YN. A closure scheme for chemical master equations. Proceedings of the National Academy of Sciences. 2013;110(35):14261–14265. doi:10.1073/pnas.1306481110.

[23] Lakatos E, Ale A, Kirk PD, Stumpf MP. Multivariate moment closure techniques for stochastic kinetic models. The Journal of chemical physics. 2015;143(9):094107.

[24] Buchholz VR, Flossdorf M, Hensel I, Kretschmer L, Weissbrich B, Gräf P, et al. Disparate individual fates compose robust CD8+ T cell immunity. Science. 2013;340(6132):630–635.

[25] Claudi B, Spröte P, Chirkova A, Personnic N, Zankl J, Schürmann N, et al. Phenotypic variation of Salmonella in host tissues delays eradication by antimicrobial chemotherapy. Cell. 2014;158(4):722–733.

[26] Singh A, Hespanha JP. Moment Closure Techniques for Stochastic Models in Population Biology. Proceedings of the 2006 American Control Conference. 2006;.

[27] Van Loan CF. Computing Integrals Involving the Matrix Exponential. IEEE Transactions on Automatic Control. 1978;23.

[28] R Core Team. R: A Language and Environment for Statistical Computing; 2016. Available from: `https://www.R-project.org/`.

[29] Powell MJD. UOBYQA: unconstrained optimization by quadratic approximation. University of Cambridge; 2000.

[30] Efron B, Tibshirani R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Statistical Science. 1986;1(1):54–75.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[31] Hall P. On the Bootstrap and Likelihood-Based Confidence Regions. Biometrika. 1987;74(3):481–493.

[32] Murdoch D, Chow ED. ellipse: Functions for drawing ellipses and ellipse-like confidence regions; 2013. Available from: `https://CRAN.R-project.org/package=ellipse`.

[33] Casella G, Berger RL. Statistical Inference. 2nd ed. Duxbury; 2002.

[34] Busetto AG, Hauser A, Krummenacher G, Sunnåker M, Dimopoulos S, Ong CS, et al. Near-optimal experimental design for model selection in systems biology. Bioinformatics. 2013;29(20):2625. doi:10.1093/bioinformatics/btt436.

[35] Price DJ, Bean NG, Ross JV, Tuke J. On the efficient determination of optimal Bayesian experimental designs using ABC: A case study in optimal observation of epidemics. Journal of Statistical Planning and Inference. 2016;172:1 – 15. doi:http://dx.doi.org/10.1016/j.jspi.2015.12.008.

[36] Overstall AM, Woods DC. Bayesian Design of Experiments using Approximate Coordinate Exchange. Technometrics. 2016;doi:10.1080/00401706.2016.1251495.

[37] Ryan EG, Drovandi CC, McGree JM, Pettitt AN. A Review of Modern Computational Algorithms for Bayesian Optimal Design. International Statistics Review. 2016;84(1):128–154. doi:10.1111/insr.12107.

[38] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of The Royal Society Interface. 2009;6(31):187–202. doi:10.1098/rsif.2008.0172.

[39] House T, Ford A, Lan S, Bilson S, Buckingham-Jeffery E, Girolami M. Bayesian uncertainty quantification for transmissibility of influenza, norovirus and Ebola using information geometry. Journal of The Royal Society Interface. 2016;13(121):20160279. doi:10.1098/rsif.2016.0279.

[40] Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian Computation. PLOS Computational Biology. 2013;9(1):1–10. doi:10.1371/journal.pcbi.1002803.

[41] Pardo L. Statistical inference based on divergence measures. CRC Press; 2005.

bioRxiv preprint doi: https://doi.org/10.1101/116319; this version posted April 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[42] Mastroeni P, Grant AJ, Restif O, Maskell DJ. A dynamic view of the spread and intracellular distribution of Salmonella enterica. Nature Reviews Microbiology. 2009;7:73–80. doi:10.1038/nrmicro2034.

[43] Abel S, Abel zur Wiesch P, Davis BM, Waldor MK. Analysis of Bottlenecks in Experimental Models of Infection. PLoS Pathog. 2015;11(6):e1004823. doi:10.1371/journal.ppat.1004823.

[44] Lim CH, Voedisch S, Wahl B, Rouf SF, Geffers R, Rhen M, et al. Independent Bottlenecks Characterize Colonization of Systemic Compartments and Gut Lymphoid Tissue by Salmonella. PLoS Pathog. 2014;10(7):e1004270. doi:10.1371/journal.ppat.1004270.