

STRUCTURAL AND PRACTICAL IDENTIFIABILITY ISSUES OF IMMUNO-EPIDEMIOLOGICAL VECTOR-HOST MODELS WITH APPLICATION TO RIFT VALLEY FEVER

NECIBE TUNCER*, HAYRIYE GULBUDAK, VINCENT L. CANNATARO,
AND MAIA MARTCHEVA

ABSTRACT. In this article, we discuss the structural and practical identifiability of a nested immuno-epidemiological model of arbovirus diseases, where host-vector transmission rate, host recovery, and disease induced death rates are governed by the within-host immune system. We incorporate the newest ideas and the most up-to-date features of numerical methods to fit multi-scale models to multi-scale data. For an immunological model we use Rift Valley Fever Virus (RVFV) time-series data obtained from livestock under laboratory experiments and for an epidemiological model we incorporate a human compartment to the nested model and use the number of human RVFV cases reported by the CDC during the 2006 – 2007 Kenya outbreak. We show that the immunological model is not structurally identifiable for the measurements of time-series viremia concentrations in the host. Thus, we study the non-dimensionalized and scaled versions of the immunological model and prove that both are structurally globally identifiable. After fixing estimated parameter values for the immunological model derived from the scaled model, we develop a numerical method to fit observable RVFV epidemiological data to the nested model for the remaining parameter values of the multi-scale system. For the given (CDC) data set, Monte Carlo simulations indicate that only three parameters of the epidemiological model are practically identifiable when the immune model parameters are fixed. Alternatively, we fit the multi-scale data to the multi-scale model simultaneously. Monte Carlo simulations for the simultaneous fitting suggest that the parameters of the immunological model and the parameters of the immuno-epidemiological model are practically identifiable. We suggest that analytic approaches for studying the structural identifiability of nested models are a necessity, so that identifiable parameter combinations can be derived to reparameterize the nested model to obtain an identifiable one. This is a crucial step in developing multi-scale models which explain multi-scale data.

KEYWORDS: Immuno-epidemiological modeling, Rift Valley Fever, Structural and practical identifiability analysis, parameter estimation, arbovirus-diseases, immune dynamics

AMS SUBJECT CLASSIFICATION: 92D30, 92D40

1. INTRODUCTION

Differential equations are powerful tools for modeling biological systems with broader applications to the fields of biomedical research and infectious disease modeling. In particular, ODEs have been a useful tool for determining the unknown parameters of theoretical models confronting experimental data. Identifiability analysis is a common

Date: June 29, 2016.

*author for correspondence.

methodology to determine unknown parameters in ODE models [24, 10, 6, 26]. Even though identifiability analysis has been used for over the last two decades for ODE models, including structural, practical, and sensitivity-based identifiability analysis, the identifiability analyses of age-structured PDE models have not been studied (but see [26, 27]). Here we study structural identifiability issues for an immuno-epidemiological nested vector-host model with application to Rift Valley Fever disease.

Rift Valley Fever virus (Family, Bunyaviridae; genus, *Phlebovirus*) was originally characterized in 1931 in Kenya and has since been isolated throughout the African continent and in the Arabian Peninsula [5]. The pathogen is transmitted mainly through mosquitoes and may infect several different mammal species, causing illness and death in livestock species and humans and resulting in millions of dollars of economic loss. RVFV persists between epidemic periods, with the leading hypotheses claiming that the virus either remains dormant through the transovarial transmission of the virus within the vector population or through continual amplification events within wildlife host populations [5, 37]. After being first identified in Kenya in 1931, RVFV has largely spread throughout the African continent and has a high risk of spreading to new regions, including European Union (EU) and United States (US), through either an accidental introduction or through the use of RVFV as a biological weapon [45].

Previous researchers have used a range of methods to model RVFV, from individual based models to compartmental ordinary differential equations at the population level, to investigate the impact of the intervention strategies and mechanisms that lead to the persistence of RVFV outbreaks. Gaff et al. [13, 14] modeled RVFV disease in an ODE model with vector-host compartments to assess effectiveness of RVFV interventions. Niu et al. [41] introduced space effects with patch-to-patch movement of species to introduce a methodology for studying the likelihood of pathogen establishment after RVFV introduction to an infection-free area. Fischer et al. [44] investigated the effects of a temperate climate, such as the Netherlands, where both hosts and potential vectors are present. Mpeshe et al. [43] formulated an ODE deterministic model of RVFV with a human compartment, in addition to vector-host compartments, and studied epidemiological threshold properties of reproduction number, \mathcal{R}_0 , and also explored a sensitivity analysis of the model parameters. Xiao et al. [40] recently studied the effect of both seasonality and socioeconomic status in a multi-patch model. André et al. [42] discussed the impact of within-host pathogen and immune dynamics on the pathogen dispersal between hosts. We model RVFV within-host immune dynamics and couple it with an age-since-infection structured vector-host disease model by linking vector-host disease transmission, host recovery rate, and disease induced death rates to the host immune dynamics [31]. In this paper, we are interested in the structural and practical identifiability analysis of the nested model with application to RVFV.

The nested model pertains particularly well to Rift Valley Fever. The immuno-epidemiological model captures the heterogeneity that exists in the ‘macro’ scale among infected individuals in a RVFV epidemic by incorporating within-host pathogen and immune response dynamics. The immunological model captures qualitative within-host immune dynamics in the ‘micro’ scale by modeling IgM and IgG immune response antibodies, which are commonly measured in the laboratory [5, 16]. Although immuno-epidemiological models have garnered attention from scientists and mathematicians hoping to solve problems regarding host-pathogen coevolution, early diagnostics, and

transmission-virulence trade-offs [15, 38, 39], the structural and practical identifiability analysis of multi-scale nested models has not been studied. We use RVFV immunological data obtained under laboratory experiments and epidemiological human data from the 2006 – 2007 outbreak in Kenya [7, 25] to estimate the unknown parameter values of the nested system.

Many researchers have focused on developing methods for solving the ill-posed inverse problems generated by age-structured population models, but there has been less study on the structural and practical identifiability analysis of the age-structured models. Perasso et al. [26] performed identifiability analysis of a nonlinear transport reaction system for a SI model, representing the spread of scrapie in a sheep flock, with realistic assumptions regarding the disease characteristics such as long and variable incubation period and asymptomatic infection. They establish an input-output (IO) relation for the model, which is based on the implicit solution of the PDE model. In another study, Perasso and Razafison [27] investigated the identifiability analysis of the age-dependent mortality rate of the McKendrick-VonFoerster model, where the observation is derived from a given age group of the population. They show how the non-local loopback boundary condition affects the uniqueness of the model parameters. In this study, we are interested in developing a numerical approach for a structural and practical identifiability analysis of a nested immuno-epidemiological model, which couples an ODE immunological model to an age-since-infection structured epidemiological PDE model. To our best knowledge, the identifiability analysis of a nested model, such as the model presented here, has never been studied (but for hierarchical statistical models see [18]), although there is a need for reliable methods for parameter estimation of immuno-epidemiological models to have comparison with experimental data and to provide accurate predictions.

This paper is organized as follows. In the next section we present an immuno-epidemiological model, with an immunological model for IgM and IgG immune response antibodies and an epidemiological model of RVFV with an age-since-infection structure. In section 3, we study the structural identifiability analysis of the immunological model by using the data from [16]. In section 4, we develop a finite difference scheme to approximate the solutions of the immuno-epidemiological model. Then, we estimate the epidemiological parameters from the observable RVFV epidemiological data, by first fixing the immunological parameters. Next, we fit the multi-scale RVFV data to the multi-scale model introduced in this paper. In the last section, we summarize our results and give the conclusion.

2. IMMUNO-EPIDEMIOLOGICAL MODEL

In this section, we first describe a within-host pathogen and immune model, with a logistic pathogen growth rate. Then we introduce an epidemiological model along with linking parameters. The immuno-epidemiological model links the within-host model to the epidemiological model by using the nested model framework described in [15]. The between-host model is structured with time-since-infection parameter τ . Thus the dynamics of pathogen and host immune response determines the transmission, recovery and disease induced mortality properties of the epidemiological model.

2.1. The Immunological Model. The immunological model describes the interaction between pathogen and IgM, IgG immune response antibodies. $M(\tau)$ and $G(\tau)$ denote the within-host IgM and IgG antibody concentrations at time-since-infection τ , respectively.

TABLE 1. Definition of the variables in the within host model

Variable	Meaning
$P(\tau)$	Pathogen concentration τ time-since-infection
$M(\tau)$	IgM concentration τ time-since-infection
$G(\tau)$	IgG concentration τ time-since-infection

We assume that the pathogen, $P(\tau)$, exhibits logistic growth with carrying capacity K and growth rate r . IgM and IgG immune response antibodies contribute to the elimination of the pathogen at rates ϵ and δ , respectively. The IgM immune response antibodies proliferate at a rate proportional to the viremia level in the host, represented by the activation rate a , and decays at rate c . The IgM immune response antibodies are mainly responsible for rapid destruction of invading pathogen. B cells switch production of IgM immune response antibodies to production of a longer lasting class, IgG, at rate q . IgG immune response antibodies activate at a per-capita rate b .

$$(2.1) \quad \text{Within-host Model:} \quad \begin{cases} \frac{dP}{d\tau} &= \left(r \left(1 - \frac{P(\tau)}{K} \right) - \epsilon M(\tau) - \delta G(\tau) \right) P(\tau), \\ \frac{dM}{d\tau} &= aP(\tau)M(\tau) - (q + c)M(\tau), \\ \frac{dG}{d\tau} &= qM(\tau) + bG(\tau)P(\tau). \end{cases}$$

At the initial time of infection (when $\tau = 0$), there are not yet any IgG immune response antibodies, and very little first response antibodies, IgM, capable of detecting the newly presented pathogen. Over the course of the immune response B cells will undergo immunoglobulin class switching and switch from producing IgM to IgG antibodies. Hence the model is equipped with initial conditions: $P(0) = 1$, $M(0) = 0.0001$, $G(0) = 0$. The definitions of the parameters and the variables of the model are given in Table 1 and Table 2.

Dynamics of the novel immunological model (2.1) exhibits the schematically illustrated dynamics described in [5]. Thus either the host dies due to a drastic increase in pathogen concentration, or the pathogen is controlled by the host immune response. If $P(0) \neq K$ and $M(\tau) > 0$ or $G(\tau) > 0$ for $\tau > 0$, the immune response clears the virus, the IgM antibodies die out and the IgG antibodies reach a non-zero steady-state i.e.,

$$\lim_{\tau \rightarrow \infty} P(\tau) = 0; \quad \lim_{\tau \rightarrow \infty} M(\tau) = 0 \quad \text{and} \quad \lim_{\tau \rightarrow \infty} G(\tau) = G^* \quad \text{where } G^* \neq 0$$

We present a detailed analysis of immunological model (2.1) in [31].

2.2. An Epidemiological Model of the Rift Valley Fever: Rift Valley Fever is a vector borne, zoonotic disease. It can cause a significant economic loss due to deaths and abortions among livestock. It is mostly spread among domestic animals and humans by the bite of an infected mosquito. The sporadic outbreaks of RVFV has followed heavy

TABLE 2. Definition of the parameters in the within host model

Parameter	Meaning
r	Pathogen growth rate
K	Carrying capacity of the pathogen
ϵ	Efficiency of the IgM response resulting in pathogen death
δ	Efficiency of the IgG response resulting in pathogen death
a	IgM activation rate
q	IgM \rightarrow IgG
c	IgM decay rate
b	IgG activation rate

rainfall and flooding in Kenya and other parts of Africa since the 1930s. Up to date, no human-to-human transmission has been seen [8].

The mosquito population is divided into two nonintersecting classes; $S_v(t)$ susceptible vectors at time t and $I_v(t)$ infected vectors at time t . The birth and death rates of the vector population are denoted by Λ_v and μ_v , respectively. Susceptible vectors become infected after a blood meal on an infected host. The density of the infected host population $i_H(\tau, t)$ is structured with both time t and time-since-infection τ . The new incidences in the vector population are modeled by the term

$$S_v \int_0^\infty \beta_H(\tau) i_H(\tau, t) d\tau.$$

Infectious hosts have different infectivity at different times-since-infection. Transmission rate from an infected host to a susceptible vector, $\beta_H(\tau)$, is governed by the pathogen level in the infected host at τ time-since-infection. Thus, the transmission rate is linked to the infected hosts immune response dynamics, via [20]

$$\beta_H(\tau) = \frac{c_H P^\nu(\tau)}{B_H + P^\nu(\tau)},$$

where c_H , B_H and ν are constants. The model in the vector population takes the form:

$$(2.2) \quad \text{Vector Model:} \quad \begin{cases} \frac{dS_v}{dt} = \Lambda_v - S_v \int_0^\infty \beta_H(\tau) i_H(\tau, t) d\tau - \mu_v S_v, \\ \frac{dI_v}{dt} = S_v \int_0^\infty \beta_H(\tau) i_H(\tau, t) d\tau - \mu_v I_v. \end{cases}$$

The epidemiology of the RVFV between-host population is described by an SIR model. At any time, the host population (livestock population) is further divided into susceptible hosts, $S_H(t)$ and recovered hosts, $R_H(t)$ classes. The between host model is given by:

$$(2.3) \quad \text{Between-Host Model:} \quad \begin{cases} \frac{dS_H}{dt} &= \Lambda - \beta_v S_H I_v - \mu_H S_H, \\ \frac{\partial i_H}{\partial t} + \frac{\partial i_H}{\partial \tau} &= -(\alpha(\tau) + \gamma(\tau) + \mu_H) i_H(\tau, t), \\ i_H(0, t) &= \beta_v S_H I_v, \\ \frac{dR_H}{dt} &= \int_0^\infty \gamma(\tau) i_H(\tau, t) d\tau - \mu_H R_H, \end{cases}$$

where Λ is the constant recruitment into the host population, μ_H is the natural death rate, and β_v is the transmission rate from vector to host. We suppose that the disease induced death rate, $\alpha(\tau)$, depends not only on the viral load but also on the first immune response antibodies (IgM) in the host. We assume that antibody levels are proportional to the intensity of an immune response, which is costly to the host, similar to the assumptions of previous immuno-epidemiological modelers [15]. Thus, the disease induced death rate, $\alpha(\tau)$, has the following linear dependence on $P(\tau)$ and $M(\tau)$

$$\alpha(\tau) = \sigma P(\tau) + \xi M(\tau),$$

where σ and ξ are the constants. The IgG antibodies have no effect on $\alpha(\tau)$. On the other hand, the recovery rate is proportional to the IgG levels in the host and indirectly proportional to the pathogen level. Hence,

$$\gamma(\tau) = \frac{\kappa G(\tau)}{P(\tau) + \epsilon_0},$$

where κ and ϵ_0 are constants. Definitions of epidemiological variables, parameters, and their linking relations to the host immune response are summarized in Table 3 and Table 4. A schematic illustration of the “nested model” framework for vector-borne diseases is shown in Figure 1.

We study the dynamics of the epidemiological model (2.2)-(2.3) in another paper [31]. Here, we summarize those results. The immuno-epidemiological reproduction number of model (2.2)-(2.3) is,

$$(2.4) \quad \mathcal{R}_0 = \frac{\beta_v \Lambda}{\mu_v \mu_H} \int_0^\infty \frac{\Lambda_v}{\mu_v} \beta_H(\tau) e^{-\int_0^\tau (\alpha(s) + \gamma(s) + \mu_H) ds} d\tau$$

The disease free equilibrium $\epsilon^0 = (S_V^0, I_V^0, S_H^0, i_H^0(\tau), R_H^0) = \left(\frac{\Lambda_v}{\mu_v}, 0, \frac{\Lambda}{\mu_H}, 0, 0 \right)$ is locally asymptotically stable when $\mathcal{R}_0 < 1$ and unstable when $\mathcal{R}_0 > 1$. When $\mathcal{R}_0 > 1$, there exists a unique locally asymptotically stable endemic equilibrium $\epsilon^* = (S_V^*, I_V^*, S_H^*, i_H^*(\tau), R_H^*)$ where

$$S_H^* = \frac{1}{\beta_v \frac{S_V^*}{\mu_v} \int_0^\infty \beta_H(\tau) \pi(\tau) d\tau}, \quad i_H^*(\tau) = i_H^*(0) \pi(\tau), \quad R_H^* = \frac{i_H^*(0)}{\mu_H} \int_0^\infty \gamma(\tau) \pi(\tau) d\tau,$$

$$S_V^* = \frac{\Lambda_v}{i_H^*(0) \int_0^\infty \beta_H(\tau) \pi(\tau) d\tau + \mu_v}, \quad I_V^* = \frac{S_V^*}{\mu_v} \int_0^\infty \beta_H(\tau) i_H^*(\tau) d\tau,$$

TABLE 3. Definition of the variables in the betweenhost model

Variable	Meaning	Variable	Meaning
$S_H(t)$	Susceptible hosts	$S_v(t)$	Susceptible vectors
$i_H(\tau, t)$	Infected hosts	$I_v(t)$	Infected vectors
$R_H(t)$	Recovered hosts		

TABLE 4. Definition of the parameters in the between-host model and their linking-relation to immune response dynamics

Parameter	Meaning	Linking Relation
Λ_v/Λ	Recruitment rate into vector/host population	
μ_v/μ_H	Natural death rate of vector/host	
β_v	Transmission rate from vector to host	
$\beta_H(\tau)$	Transmission rate from host to vector	$\beta_H(\tau) = \frac{c_H P^\nu(\tau)}{B_H + P^\nu(\tau)}$
$\alpha(\tau)$	Disease induced death rate of an infected host	$\alpha(\tau) = \sigma P(\tau) + \xi M(\tau)$
$\gamma(\tau)$	Recovery rate of an infected host	$\gamma(\tau) = \frac{\kappa G(\tau)}{P(\tau) + \epsilon_0}$

and $\pi(\tau)$ is the probability of still being infectious τ time units after becoming infected, which is given by

$$(2.5) \quad \pi(\tau) = e^{-\int_0^\tau (\alpha(s) + \gamma(s) + \mu_H) ds}.$$

3. PARAMETER ESTIMATION OF THE IMMUNOLOGICAL MODEL

We fit the immunological model to the data in [16], which gives times series RVFV viremia, IgM, and IgG immune response antibody concentrations in sheep inoculated subcutaneously with RVFV ([16], Fig 7). Virus concentrations were reported as median tissue culture infectious dose (TCID₅₀), or the viral dilution concentration necessary to elicit cytopathic effects in 50% of cell cultures. Antibody concentrations were reported as ELISA PP values, which are the relative percentage optical density measurements of test specimen in comparison to a mean high-positive control antigens. We extracted the data points (log₁₀ of virmia levels) from the plot using MATLAB code `grabit.m` [36], and present the data in Table 5.

Let $\mathbf{x} = [P, M, G]$ be the vector of state variables, and let $\mathbf{p} = [r, K, \epsilon, \delta, a, q, c, b]$ be the vector of model parameters, then the within host model (2.1) is equivalent to

$$(3.1) \quad \begin{aligned} \mathbf{x}' &= f(\mathbf{x}(\tau), \mathbf{p}), \\ \mathbf{x}(0) &= \mathbf{x}_0. \end{aligned}$$

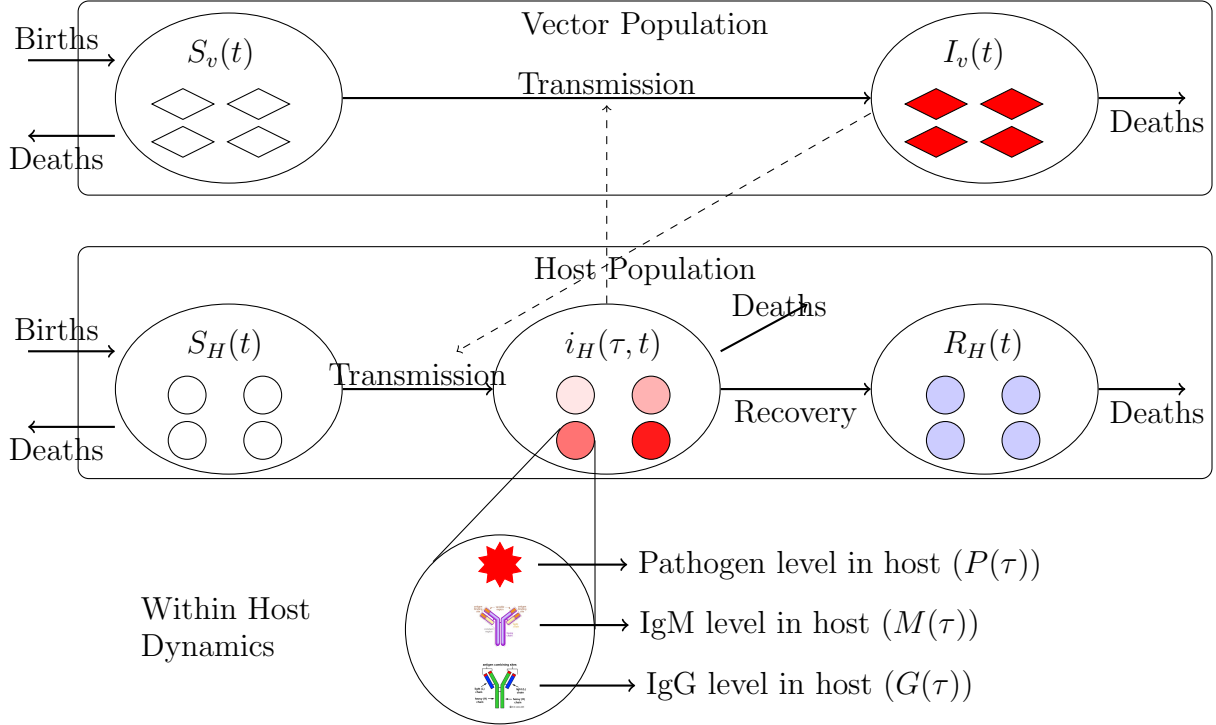


FIGURE 1. Schematic illustration of the “nested model” framework for vector-borne diseases

TABLE 5. Data extracted from Figure 7 in [16]. Data are given as the \log_{10} of the viremia level. Viremia load is given in units (TCID₅₀).

Day:	0	1	2	3	4	5	6	7	8	9	10
Viral Load:	0	0.0179	4.788	7.754	7.024	5.776	5.51	4.53	2.515	0.0347	0

The parameters \mathbf{p} of the model are estimated based on the n observation $\{y_i\}_{i=1}^n$ made at times $\tau_1, \tau_2, \dots, \tau_n$. The data points in Table 5 are the \log_{10} of the viremia levels in blood at time τ_i $i = 1, \dots, n$ ($n = 11$, see Table 5). For the parameter estimation problem, we use the following statistical model as introduced in [3]

$$(3.2) \quad y_i = g(\mathbf{x}(\tau_i), \hat{\mathbf{p}}) + \varepsilon_i$$

where $g(\mathbf{x}(\tau_i), \hat{\mathbf{p}}) = P(\tau_i, \hat{\mathbf{p}})$ are the viremia levels, and $\hat{\mathbf{p}}$ are the “true” parameters we are looking for. The random variables ε_i represents the measurement errors which cause the observations $g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})$ to drift away from the smooth exact path $g(\mathbf{x}(\tau), \hat{\mathbf{p}})$. In this study, we assume that there is no modeling error, hence the mean of the measurement errors is zero, $\mathbb{E}(\varepsilon_i) = 0$. In a more general setting, the errors satisfy the following form, which allows fairly wide range of error models,

$$(3.3) \quad \varepsilon_i = g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})^f \epsilon_i$$

where $f \geq 0$. The ϵ_i are assumed to be independent, identically distributed random variables with mean zero and finite variance σ_0^2 . The mean and the variance of the random variables y_i satisfy, $\mathbb{E}(y_i) = g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})$, $Var(y_i) = g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})^{2f} \sigma_0^2$, respectively. If $f = 0$,

then (3.3) becomes $\varepsilon_i = \epsilon_i$, that is the error in the measurements are independent from the observations $g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})$, which is called *absolute error*. In counting populations, it is reasonable to assume that the measurement errors depends on the population size, hence in that case we set $f = 1$ in (3.3), which is called *relative error*. In some cases, depending on the problem, it might be appropriate to set $f = 1/2$ in (3.3) [3]. When $f = 1/2$, the error model (3.3) was referred to as *Poisson noise* in [6].

3.1. Structural Identifiability. A first step in estimating the parameters of a model is to determine whether the problem is well-posed for a given model and data. A well-posed parameter estimation problem means that the parameters of the model can be recovered uniquely from the given data set. Even in an ideal situation where the data are noise-free and there is no modeling error the quality of parameter estimation results depends on the predictive capability of the mathematical model. If the mathematical problem is not structurally identifiable, then the parameters estimated by a numerical optimization problem might not be unique and thus be unreliable. On the other hand a mathematical model which is structurally identifiable may not be practically identifiable [10, 11]. That is, when noisy data are considered, the parameters may not be uniquely recovered. Moreover, the data might be precise, but if collected at a time interval where dynamics of the model are not captured, then the parameters might still be unidentified. Eisenberg et al. showed that a SIWR (Susceptible-Infected-Water-Recovered) model of Cholera is structurally identifiable, but it is practically unidentifiable when the environmental transmission of the cholera through the water component is very fast [12].

Structural identifiability is concerned with the possibility of estimating the model parameters assuming perfect experimental data. The definition of the structural identifiability in [24] is given as

Definition 3.1. *A parameter set \mathbf{p} is called structurally (or uniquely) identifiable if for every \mathbf{q} in the parameter space, the equation*

$$g(\mathbf{x}(\tau), \mathbf{p}) = g(\mathbf{x}(\tau), \mathbf{q}) \implies \mathbf{p} = \mathbf{q}.$$

That is, any unequal parameter set yields different observations and hence the corresponding noise-free data are distinct.

It is crucial to check the structural identifiability of the model, and several methods have been proposed to perform this task. This can be done by suitable mathematical methods directly on the model without the need of any experimental data. These include Taylor series methods [24, 29], differential algebra-based methods [4, 12, 19, 24], and other methods such as the generating series approach, the direct test, and a method based on the implicit function theorem that have been reviewed in [10, 24].

We will use the differential algebra approach [4, 24] to test the model (2.1). The differential algebra approach assumes that f and g are rational polynomial functions. One of the strengths of this approach is that, if the model is unidentifiable, the identifiable parameter combinations can be obtained, which then can be used to reparametrize the model to get a structurally identifiable model [23]. The differential algebra approach builds upon deriving the input-output equation, which contains all the structural identifiability information of the model. The input-output equations are determined from the characteristic sets which are derived by Ritt's algorithm and depend on the ranking of the state variables and the output. We refer the reader to [23, 24] for the details on

the differential algebra approach. We briefly summarize the method with the following model (3.4), since our structural identifiability analysis is extensively based on this approach.

Suppose, we want to estimate the parameters of a simple immunological model, adapted from [15] using the viremia levels in the blood as the data.

$$(3.4) \quad \begin{aligned} \frac{dP}{d\tau} &= (r - \epsilon M(\tau)) P(\tau), \\ \frac{dM}{d\tau} &= aP(\tau)M(\tau). \end{aligned}$$

The question of structural identifiability is to determine whether the model is structured to estimate the parameters uniquely given noise free data. Within the differential algebra approach, we adapt the ranking $P < M$ of the state variables, then the model (3.4) yields the following input-output equation derived from the characteristic set,

$$(3.5) \quad \epsilon P''P - \epsilon(P')^2 + \epsilon arP^3 - \epsilon aP'P^2 = 0.$$

The characteristic set in general is not unique, but the coefficients of the normalized input-output equation contain all the identifiability information of the dynamical model [12]. The input-output equation (3.5) can be normalized to make it a monic differential polynomial by dividing by ϵ :

$$(3.6) \quad P''P - (P')^2 + arP^3 - aP'P^2 = 0.$$

The structural identifiability can be determined from the injectivity of the coefficients $c(\mathbf{p})$ of the normalized input-output equation [12].

Thus, within the differential algebra approach, the definition of the structural identifiability becomes,

Definition 3.2. *Let $c(\mathbf{p})$ be the coefficients of the normalized input-output equation, then the dynamical model is structurally identifiable if and only if*

$$c(\mathbf{p}) = c(\mathbf{q}) \implies \mathbf{p} = \mathbf{q}$$

Suppose to the contrary that another set of the parameters $[q_1, q_2, q_3]$ give the same output. Then we would have:

$$q_1q_3 = ar \quad \text{and} \quad q_3 = a.$$

From which we would get the parameters a and r uniquely but ϵ can not be identified. So, the model (3.4) is unidentifiable. Thus, we have proved the following result.

Proposition 3.1. *The model (3.4) is structurally unidentifiable from the observations of viremia levels in host.*

The immunological model (2.1) is a generalization of the model (3.4). Following a similar analysis we prove the following result for the model (2.1).

Proposition 3.2. *The immunological model (2.1) is not structured to identify the parameters $\mathbf{p} = [r, K, \epsilon, \delta, a, b, c, q]$ from the pathogen level observations in the host. Only the parameters r, K, a, b , and the combinations of the parameters $q + c$, and $\frac{q\delta}{\epsilon}$ can be identified.*

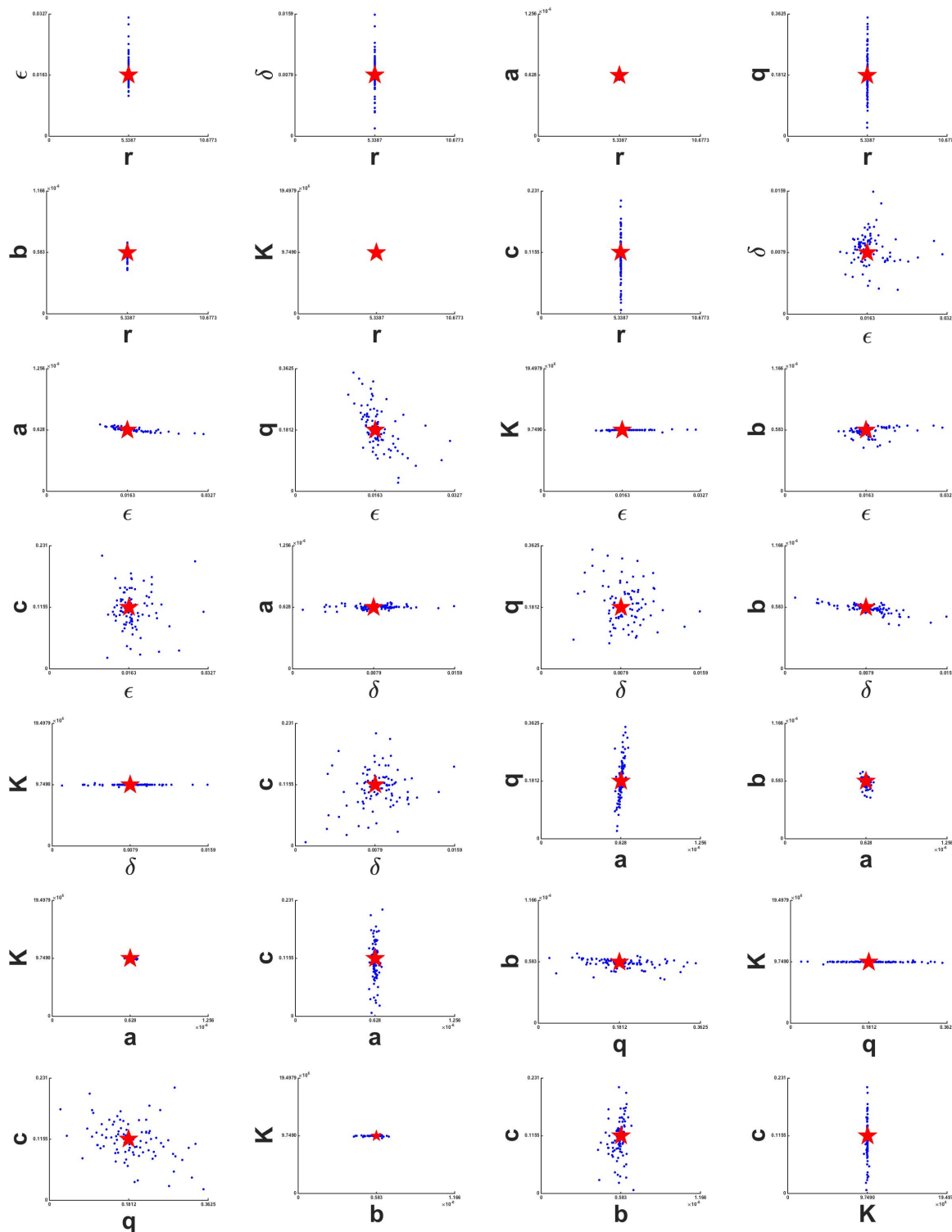


FIGURE 2. The parameter estimates of the model (2.1) for 100 synthetic data generated by Poisson noise. True parameters are indicated by red stars. True parameters are $r = 5.33$, $K = 9.75 \times 10^6$, $\epsilon = 0.0163$, $\delta = 0.0079$, $a = 6.28 \times 10^{-7}$, $q = 0.1812$, $c = 0.1155$, $b = 5.83 \times 10^{-7}$. Note that only the parameters r , K , a and b are identified. The parameters q , c , ϵ and δ can not be identified.

Proof. To derive an input-output equation for the model (2.1) is more complicated. Using the Differential Algebra for Identifiability of Systems (DAISY) software, introduced in [4], we obtain the following normalized input-output equation of the model (2.1).

$$\begin{aligned}
(3.7) \quad & P'''P^3 + P''''P^2 \frac{K(-\epsilon q - \epsilon c + \delta q)}{\epsilon K(a-b)} + 4P''P'P^2 \frac{\epsilon K(-a+b)}{\epsilon K(a-b)} + 3P''P'P \frac{K(\epsilon q + \epsilon c - \delta q)}{\epsilon K(a-b)} + \\
& P''P^4 \frac{\epsilon(ra - rb - a^2K + b^2K)}{\epsilon K(a-b)} + P''P^3 \frac{(-r\epsilon q - r\epsilon c + r\delta q + 2\epsilon aqK + 2\epsilon acK - \delta aqK - \delta qbK)}{\epsilon K(a-b)} + \\
& P''P^2 \frac{K(-\epsilon q^2 - 2\epsilon qc - \epsilon c^2 + \delta q^2 + \delta qc)}{\epsilon K(a-b)} + 3P'^3P + 2P'^3 \frac{K(-\epsilon q - \epsilon c + \delta q)}{\epsilon K(a-b)} + \\
& P'^2P^3 \frac{\epsilon(-ra + rb + a^2K - b^2K)}{\epsilon K(a-b)} + P'^2P^2 \frac{K(-2\epsilon aq - 2\epsilon ac + \epsilon qb + \epsilon cb + \delta aq)}{\epsilon K(a-b)} + \\
& P'^2P \frac{K(\epsilon q^2 + 2\epsilon qc + \epsilon c^2 - \delta q^2 - \delta qc)}{\epsilon K(a-b)} + P'P^5 \frac{\epsilon(-ra^2 + rb^2 + a^2bK - ab^2K)}{\epsilon K(a-b)} + \\
& P'P^4 \frac{(2r\epsilon aq + 2r\epsilon ac + r\epsilon qb + r\epsilon cb - r\delta aq - 2r\delta qb - 2\epsilon aqbK - 2\epsilon acbK + \epsilon qb^2K + \epsilon cb^2K + \delta aqbK)}{\epsilon K(a-b)} + \\
& P'P^3 \frac{(-r\epsilon q^2 - 2r\epsilon qc - r\epsilon qbK - r\epsilon c^2 - r\epsilon cbK + r\delta q^2 + r\delta qc + r\delta qbK + \epsilon q^2bK + 2\epsilon qcbK + \epsilon c^2bK - \delta q^2bK - \delta qcbK)}{\epsilon K(a-b)} + \\
& P^7 \frac{r\epsilon ab(a-b)}{\epsilon K(a-b)} + P^6 \frac{rb(-\epsilon a^2K - 2\epsilon aq - 2\epsilon ac + \epsilon abK + \epsilon qb + \epsilon cb + \delta aq)}{\epsilon K(a-b)} + \\
& P^5 \frac{rb(2\epsilon aqK + 2\epsilon acK + \epsilon q^2 + 2\epsilon qc - \epsilon qbK + \epsilon c^2 - \epsilon cbK - \delta aqK - \delta q^2 - \delta qc)}{\epsilon K(a-b)} + \\
& P^4 \frac{rbK(-\epsilon q^2 - 2\epsilon qc - \epsilon c^2 + \delta q^2 + \delta qc)}{\epsilon K(a-b)} = 0
\end{aligned}$$

Suppose another set of parameters, $\mathbf{q} = [q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8]$ produced the same output P . Since the viremia levels are equal, we compare the coefficients of the input-output equation and obtain a nonlinear system of 11 equations, and derive the Groebner basis using Mathematica.

Solving the Groebner basis we arrive at the following results,

$$r = q_1, \quad K = q_2, \quad a = q_5, \quad b = q_6, \quad q + c = q_8 + q_7 \quad \frac{q\delta}{\epsilon} = \frac{q_8q_4}{q_3}.$$

Thus, the model (2.1) is unidentifiable. Only the parameters r, K, a, b , and the parameter combinations $q + c$ and $\frac{q\delta}{\epsilon}$ can be identified. \square

The Proposition 3.2 states that the parameters of the model (2.1) can not be identified from the observations of viremia levels only. We would like to point out that if all the state variables were observable, in addition to viremia levels, and we were to observe IgM and IgG levels, then all parameters of the model (2.1) would be identified. The data given for IgM and IgG concentrations in [16] are relative percentage optical density measurements from the test specimen in comparison to a mean high-positive control antigens. That is, it is not possible to use the data provided for IgM and IgG in the parameter estimations.

We perform a numerical experiment to show the theoretical results we obtained in Proposition 3.2. The viremia data in [16] are given for \log_{10} of the pathogen levels. But, for the numerical experiment we would like to test the structural identifiability results

obtained by differential algebra approach, which requires g to be a rational function. That is, we would like to estimate the parameters of the immunological model (2.1) from the observations $g(\mathbf{x}(\tau_i), \hat{\mathbf{p}}) = P(\tau_i, \hat{\mathbf{p}})$. (Note that for the practical identifiability, we take $g(\mathbf{x}(\tau_i), \hat{\mathbf{p}}) = \log_{10}(P(\tau_i, \hat{\mathbf{p}}))$ in the next section). The parameters of the model (2.1) are estimated using the Maximum Likelihood approach from 100 synthetic data generated by adding Poisson noise to the model output $g(\mathbf{x}(\tau_i), \hat{\mathbf{p}}) = P(\tau_i, \hat{\mathbf{p}})$, at the data time points $\tau_i, i = 1, 2, \dots, n$ as in Table 5. That is we generate $M = 100$ synthetic data whose mean is $\mathbb{E}(y_i) = g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})$ and $Var(y_i) = g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})$. The true parameter values, $\hat{\mathbf{p}}$, are shown as red stars in Figure 2. We estimate the parameters by solving the following minimization problem for each data set

$$(3.8) \quad \mathbf{p}_j = \min_{\hat{\mathbf{p}}} \left(\sum_{i=1}^n g(\mathbf{x}(\tau_i), \hat{\mathbf{p}}) - \sum_{i=1}^n y_i \ln(g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})) \right), \quad j = 1, 2, \dots, 100.$$

The Nelder-Mead iterative algorithm is used to approximate the minimum of the cost function (3.8). The initial parameter values for the iterative solver are chosen randomly from a normal distribution whose mean is the true parameter values and the standard deviation is 10% of the mean. The structural identifiability conclusion of Proposition 3.2 is consistent with the numerical experiment we performed for estimating the parameters of the model (2.1) (see Figure 2). The scatter plots supports the theoretical results in Proposition 3.2; only the parameters, r, K, a can be identified and parameters q, c, ϵ and δ can not be identified. The scatter plots are inconclusive for the parameter b . The correlations between the unidentifiable parameters can not be deduced from the scatter plots, but Proposition 3.2 gives the correlations among the unidentified parameters.

Model (2.1) is not structured to identify its parameters from viremia observations. In attempt to get an identifiable model, we first non-dimensionalize (2.1) by setting $u_1 = \frac{P}{K}, u_2 = \frac{\epsilon}{r}M, u_3 = \frac{\delta}{r}G$, and $t = \frac{\tau}{r}$.

$$(3.9) \quad \begin{aligned} \frac{du_1}{dt} &= ((1 - u_1) - u_2 - u_3) u_1, \\ \frac{du_2}{dt} &= \alpha u_1 u_2 - \gamma u_2 \\ \frac{du_3}{dt} &= \beta u_2 + \omega u_3 u_1 \end{aligned}$$

where $\alpha = \frac{aK}{r}, \gamma = \frac{q+c}{r}, \beta = \frac{q\delta}{\epsilon r}, \omega = \frac{bK}{r}$.

Proposition 3.3. *The non-dimensionalized model (3.9) is structurally identifiable from viremia observations.*

Proof. Using DAISY, we get the following input-output equation of the model (3.9).

$$\begin{aligned}
(3.10) \quad & P'''P^4 + \frac{(\beta - \gamma)}{(\alpha - \omega)} P'''P^3 - 4P''P'P^3 + 3\frac{(\gamma - \beta)}{(\alpha - \omega)} P''P'P^2 + \frac{(-\alpha^2 + \alpha + \omega^2 - \omega)}{(\alpha - \omega)} P''P^5 + \\
& \frac{(2\alpha\gamma - \alpha\beta - \gamma - \beta\omega + \beta)}{(\alpha - \omega)} P''P^4 + \frac{\gamma(-\gamma + \beta)}{(\alpha - \omega)} P''P^3 + 3P'^3P^2 + 2\frac{(-\gamma + \beta)}{(\alpha - \omega)} P'^3P + \\
& \frac{(\alpha^2 - \alpha - \omega^2 + \omega)}{(\alpha - \omega)} P'^2P^4 + \frac{(-2\alpha\gamma + \alpha\beta + \gamma\omega)}{(\alpha - \omega)} P'^2P^3 + \frac{\gamma(\gamma - \beta)}{(\alpha - \omega)} P'^2P^2 + \\
& \frac{(\alpha^2\omega - \alpha^2 - \alpha\omega^2 + \omega^2)}{(\alpha - \omega)} P'P^6 + \frac{(-2\alpha\gamma\omega + 2\alpha\gamma + \alpha\beta\omega - \alpha\beta + \gamma\omega^2 + \gamma\omega - 2\beta\omega)}{(\alpha - \omega)} P'P^5 + \\
& \frac{(\gamma^2\omega - \gamma^2 - \gamma\beta\omega + \gamma\beta - \gamma\omega + \beta\omega)}{(\alpha - \omega)} P'P^4 + \frac{\alpha\omega(\alpha - \omega)}{(\alpha - \omega)} P^8 + \\
& \frac{\omega(-\alpha^2 - 2\alpha\gamma + \alpha\beta + \alpha\omega + \gamma\omega)}{(\alpha - \omega)} P^7 + \frac{\omega(2\alpha\gamma - \alpha\beta + \gamma^2 - \gamma\beta - \gamma\omega)}{(\alpha - \omega)} P^6 + \frac{\gamma\omega(-\gamma + \beta)}{(\alpha - \omega)} P^5 = 0
\end{aligned}$$

Solving $c(\mathbf{p}) = c(\mathbf{q})$ for $\mathbf{p} = [\alpha, \gamma, \beta, \omega]$ and $\mathbf{q} = [q_1, q_2, q_3, q_4]$ using the Groebner basis in Mathematica we obtain,

$$\alpha = q_1, \quad \gamma = q_2, \quad \beta = q_3, \quad \omega = q_4.$$

Thus, the model (3.9) is structurally identifiable. \square

We note that even though model (3.9) is structurally identifiable, in practice it is not useful. It is not clear how to normalize the data since the carrying capacity, K , of the viremia is not known. Furthermore the parameters $r, \epsilon, \delta, a, q, c$, and b can not be derived from estimated values of α, γ, β , and ω . We search for another model which is structurally identifiable.

We scale the model (2.1) by setting $u_2 = \epsilon M, u_3 = \delta G$ and obtain;

$$(3.11) \quad \text{Scaled Model:} \quad \begin{cases} \frac{dP}{d\tau} = \left(r \left(1 - \frac{P(\tau)}{K} \right) - u_2(\tau) - u_3(\tau) \right) P(\tau), \\ \frac{du_2}{d\tau} = aP(\tau)u_2(\tau) - (q + c)u_2(\tau), \\ \frac{du_3}{d\tau} = wu_2(\tau) + bu_3(\tau)P(\tau), \end{cases}$$

where $w = \frac{q\delta}{\epsilon}$. By Proposition 3.2, we see that the combination $q + c$ is identifiable, but not the parameters q and c . The parameter c denotes the exponential decay rate of the IgM immune response antibody concentration. It is suggested that the biological half-life of the IgM immune response antibodies in lambs is 6 days [35]. Hence, we fix $c = \frac{\ln 2}{6}$.

Proposition 3.4. *When c is fixed, the scaled model (3.11) is structurally identifiable from the viremia observations.*

Proof. Using DAISY, we get the following input-output equation of the model (3.11).
(3.12)

$$\begin{aligned}
& P'''P^4 + P'''P^3 \frac{K(-q+w-c)}{K(a-b)} - 4P''P'P^3 + 3P''P'P^2 \frac{K(q-w+c)}{K(a-b)} + \\
& P''P^5 \frac{(ra-rb-a^2K+b^2K)}{K(a-b)} + P''P^4 \frac{(-rq+rw-rc+2aqK-awK+2aKc-wbK)}{K(a-b)} + \\
& P''P^3 \frac{K(-q^2+qw-2qc+wc-c^2)}{K(a-b)} + 3P'^3P^2 + 2P'^3P \frac{K(-q+w-c)}{K(a-b)} + \\
& P'^2P^4 \frac{(-ra+rb+a^2K-b^2K)}{K(a-b)} + P'^2P^3 \frac{K(-2aq+aw-2ac+qb+bc)}{K(a-b)} + \\
& P'^2P^2 \frac{K(q^2-qw+2qc-wc+c^2)}{K(a-b)} + P'P^6 \frac{(-ra^2+rb^2+a^2bK-ab^2K)}{K(a-b)} + \\
& P'P^5 \frac{(2raq-raw+2rac+rqb-2rwb+rbc-2aqbK+awbK-2abKc+qb^2K+b^2Kc)}{K(a-b)} + \\
& P'P^4 \frac{(-rq^2+rqw-rqbK-2rqc+rwbK+rwc-rbKc-rc^2+q^2bK-qwbK+2qbKc-wbKc+bKc^2)}{K(a-b)} + \\
& P^8 \frac{rab(a-b)}{K(a-b)} + P^7 \frac{rb(-a^2K-2aq+aw+abK-2ac+qb+bc)}{K(a-b)} + \\
& P^6 \frac{rb(2aqK-awK+2aKc+q^2-qw-qbK+2qc-wc-bKc+c^2)}{K(a-b)} + \\
& P^5 \frac{rbK(-q^2+qw-2qc+wc-c^2)}{K(a-b)} = 0
\end{aligned}$$

Solving $c(\mathbf{p}) = c(\mathbf{q})$ for $\mathbf{p} = [r, K, a, b, q, c, w]$ and $\mathbf{q} = [q_1, q_2, q_3, q_4, q_5, q_6, q_7]$ where $\mathbf{p} > 0$ and $\mathbf{q} > 0$ we obtain,

$$r = q_1, \quad K = q_2, \quad a = q_3, \quad b = q_4, \quad q + c = q_5 + q_6, \quad w = q_7.$$

Since c is fixed the scaled model (3.11) is globally identifiable. \square

Each parameter value of the immunological model (2.1) can be uniquely reconstructed from the unique parameter estimates of the scaled model (3.11) except the killing efficacy parameters ϵ and δ . Regarding these two parameters, the parameter estimate of the scaled model only provides their fraction. However, the parameters ϵ and δ are less of a concern to us, since biologically these parameters represent the average effects of multiple complex interactions occurring between antibody-pathogen recognition and the response of immune cells and cannot be easily interpreted as a single parameter.

We estimate the parameters $\mathbf{p} = [r, K, a, b, q, w]$ of model (3.11) to the viremia levels given in Table 5 using Nelder-Mead algorithm implemented in MATLAB. We estimate the parameters $\mathbf{p} = [r, K, a, q, w, b]$, of model (3.11) by minimizing the following cost functional

$$(3.13) \quad J(\mathbf{p}) = \sum_{i=1}^n \omega_i (y_i - \log_{10}(P(\tau_i, \mathbf{p})))^2.$$

where ω_i are the weights. Initial values are fixed at $P(0) = y_1$, $M(0) = 0.0001$ and $G(0) = 0$. We present the parameter estimates which were obtained by weighted least squares in Table 6 together with the uncertainties in the estimates. Uncertainties in parameter estimates are represented in terms of the coefficient of variation (CV) which is calculated by constructing a Fisher Information Matrix using sensitivity functions.

For this analysis, we first compute the sensitivity matrix $\mathbf{S}(\tau_i, \mathbf{p}) = \frac{\partial g}{\partial \mathbf{p}}(\mathbf{x}(\tau_i), \mathbf{p})$, which is the Jacobian of the output with respect to the model parameters. Clearly, we do not have an explicit representation of the pathogen levels $P(\tau_i, \mathbf{p})$, so we compute the sensitivity matrix \mathbf{S} by coupling the system (3.1) with the following system of sensitivity functions [6, 9];

$$(3.14) \quad \begin{aligned} \frac{d}{dt} \frac{\partial \mathbf{x}}{\partial \mathbf{p}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{p}} + \frac{\partial \mathbf{f}}{\partial \mathbf{p}}, \\ \frac{\partial \mathbf{x}}{\partial \mathbf{p}}(\mathbf{0}) &= \mathbf{0}. \end{aligned}$$

Here $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ is the Jacobian matrix of the system (3.1). Solving the initial value problem (3.14) simultaneously with (3.1), we obtain $P(\tau_i, \mathbf{p})$ and $\frac{\partial P}{\partial \mathbf{p}}(\tau_i, \mathbf{p})$ and then compute the sensitivity matrix \mathbf{S} . If the noise in the data has independent and identical normal distribution with zero mean and unity variance, then

$$\mathbf{S}^T \mathbf{S} = F(\mathbf{p})$$

is the Fisher Information Matrix. The inverse of the Fisher Information Matrix provides a lower bound on the variance of the parameter estimates, which is known as Cramer-Rao lower bound [28, 34]. Namely,

$$F^{-1}(\mathbf{p}) \leq Cov(\mathbf{p}).$$

The standard deviations of the parameter estimation can be obtained by taking the square roots of the diagonal elements of the $F^{-1}(\mathbf{p})$ matrix. We will use the standard deviations to quantify uncertainty in the parameter estimation using $F^{-1}(\mathbf{p})$ in the following way. Let $p \in \mathbf{p}$ be the i^{th} parameter, and $\sigma(p)$ be the standard deviation of the estimated parameter p , then

$$\sigma(p) = \sqrt{F^{-1}(\mathbf{p})_{i,i}},$$

where $F^{-1}(\mathbf{p})_{i,j}$ denotes (i, j) component of the matrix $F^{-1}(\mathbf{p})$. The coefficient variation $CV(p)$ of the parameter can be computed by;

$$CV(p) = \sigma(p)/p.$$

Practical Identifiability. Structural identifiability is a property of the dynamical model and is independent from the accuracy of the experimental data. A parameter which is structurally identifiable may still be non-identifiable in practice. For a practically non-identifiable parameter, it might be very difficult or impossible to detect the reasons behind the non-identifiability. On the other hand, knowing that the parameter is structurally identifiable assures that the cause of the practical non-identifiability is either lack of information captured from the experimental data (too few or too noisy data) or the inability of the numerical optimization algorithm to locate the minimum of $J(\mathbf{p})$.

The parameters of the scaled model (3.11) are structurally identifiable, next we test whether the parameters are identifiable in practice. For practical identifiability, we carry out *Monte Carlo Simulations*. Monte Carlo simulations have been widely used for practical identifiability of ODE models [24]. We perform Monte Carlo simulations

TABLE 6. Parameter estimates of model (3.11) fitted to \log_{10} of viremia levels given in Table 5. The estimated values are precise numbers given by MATLAB, including the values for a and b .

Parameter	Estimate	Units	CV
r	6.57701933	$(\text{TCID}_{50} \times \text{days})^{-1}$	6.7×10^{-7}
K	6.46580327×10^7	TCID_{50}	33.31
a	2.5×10^{-7}	$(\text{ELISA PP} \times \text{days})^{-1}$	1.5×10^{-4}
q	0.423269696	days^{-1}	2.38×10^{-4}
w	0.126327973	days^{-1}	1.64×10^{-4}
b	9×10^{-8}	$(\text{ELISA PP} \times \text{days})^{-1}$	3.81×10^{-5}

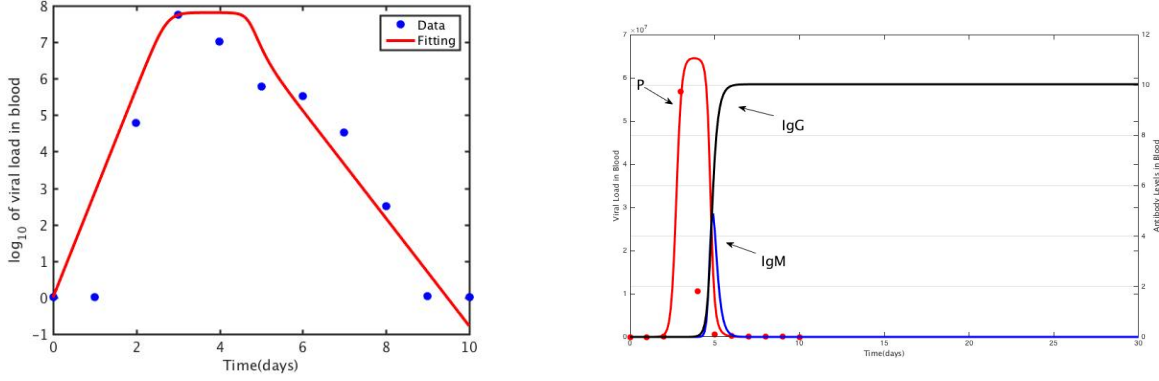


FIGURE 3. Fitting results. We plot the solutions of the model (3.11) for fitted parameters in Table 6. The left plot in Figure 3 is showing the \log_{10} of the pathogen levels for the time-since infection period $\tau \in [0, 10]$ together with the data presented in Table 5. The right plot in Figure 3 is presenting the dynamics of the model with fitted parameters for $\tau \in [0, 30]$ and the data (red dots).

first by taking the estimated parameters presented in Table 6 as the true parameter set $\hat{\mathbf{p}}$. Then, we generate $M = 1000$ synthetic data by evaluating the viremia observations at the true parameter set $\hat{\mathbf{p}}$ and adding noise at increasing levels. The Monte Carlo Simulations we performed are outlined in the following steps.

- (1.) Solve the scaled immunological model (3.11) numerically with the true parameters $\hat{\mathbf{p}}$ and obtain the output vector $\mathbf{g}(\mathbf{x}(\tau_i), \hat{\mathbf{p}}) = P(\tau_i, \hat{\mathbf{p}})$ at the discrete data time points $\{\tau_i\}_{i=1}^n$.
- (2.) Generate $M = 1000$ simulated data from the statistical model (3.2)-(3.3) with a given measurement error structure. Data sets are drawn from a normal distribution whose mean is the output vector obtained in step (1.) and standard deviation is the $\sigma_0\%$ of the mean. That is, in the statistical model (3.2) -(3.3),

we set $f = 1$ and obtain the synthetic data points by

$$y_i = \log_{10} (g(\mathbf{x}(\tau_i), \hat{\mathbf{p}}) + g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})\epsilon_i),$$

where $Var(\epsilon_i) = \sigma_0^2$.

- (3.) Fit the scaled immunological model (3.11) to each of the M simulated data sets to estimate the parameter set \mathbf{p}_j for $j = 1, 2, \dots, M$. That is

$$\mathbf{p}_j = \min_{\mathbf{p}} \sum_{i=1}^n \omega_i (y_i - \log_{10}(g(\mathbf{x}(\tau_i), \mathbf{p})))^2, \quad j = 1, 2, \dots, M.$$

- (4.) Calculate the average relative estimation error (ARE) for each parameter in the set \mathbf{p} as introduced in [24],

$$ARE(p^{(k)}) = 100\% \frac{1}{M} \sum_{j=1}^M \frac{|\hat{p}^{(k)} - p_j^{(k)}|}{\hat{p}^{(k)}}$$

where $p^{(k)}$ is the k^{th} parameter in the set \mathbf{p} , $\hat{p}^{(k)}$ is the k^{th} parameter in the true parameter set $\hat{\mathbf{p}}$, and $p_j^{(k)}$ is the k^{th} parameter in the estimated parameter set \mathbf{p}_j .

- (5.) Repeat steps 1 through 4 with increasing level of noise, that is take $\sigma_0 = 0, 1, 5, 10, 20\%$.

The computed AREs give an insight about the identifiability of the parameters of the immunological model. Since the scaled immunological model (3.11) is structurally identifiable, when $\sigma_0 = 0$, the AREs are 0 (see Table 7). That is Monte Carlo simulations agree with the structural identifiability analysis via differential algebra approach when there is no noise in the data. As expected, increasing the noise level in the data raises the AREs. We see from Table 7 that the AREs of the parameters q and w are relatively high, showing that these two parameters are sensitive to the noise in the data. A parameter is not practically identifiable, if the ARE of that parameter is significantly high even for a reasonable level of measurement error. We say that a parameter is practically identifiable if the ARE of that parameter is less than the measurement error level. The AREs of the parameters r, K , and b are consistently below the measurement errors, so we claim that these parameters are practically identifiable. Based on the Monte Carlo Simulation results presented in Table 7, we conclude that the parameters q, w , and a are not practically identifiable.

Practical identifiability of the parameters of an ODE model might be improved by increasing the data points. Next, we perform the Monte Carlo simulations with increased data time points by inserting additional time points at the midpoint between the two original time points. That is, we take the new increased data time points to be $\tau_i = \{0, 0.5, 1, 1.5, 2.2.5, \dots, 10.5, 11\}$. The computed AREs are presented in Table 8. We observe that increasing the data points did not improve the practical identifiability of the parameters. As was shown in [18], if the parameters are non-identifiable, the accuracy does not increase with the number of data points. The parameters a, q , and w are not practically identifiable. On the other hand, increasing the data points lowered the computed AREs of the identifiable parameters K and r .

TABLE 7. Practical identifiability analysis of parameters of the scaled immunological model (3.11) by Monte Carlo simulations for measurement error levels $\sigma_0 = 0\%$, 1% , 5% , 10% and 20% .

Parameter	ARE $\sigma_0 = 0\%$	ARE $\sigma_0 = 1\%$	ARE $\sigma_0 = 5\%$	ARE $\sigma_0 = 10\%$	ARE $\sigma_0 = 20\%$
r	0.00%	0.08%	0.32%	0.64%	1.56%
K	0.00%	0.77%	4.28%	8.57%	17.30%
a	0.00%	0.94%	5.86%	12.51%	26.18%
q	0.00%	1.26%	7.06%	14.19%	29.73%
w	0.00%	1.88%	12.37%	22.53%	33.89%
b	0.00%	0.77%	4.34%	8.87%	16.93%

TABLE 8. Monte Carlo simulations of the scaled immunological model with increased data points (3.11). The AREs are computed for measurement error levels $\sigma_0 = 0\%$, 1% , 5% , 10% and 20% .

Parameter	ARE $\sigma_0 = 0\%$	ARE $\sigma_0 = 1\%$	ARE $\sigma_0 = 5\%$	ARE $\sigma_0 = 10\%$	ARE $\sigma_0 = 20\%$
r	0.00%	0.04%	0.23%	0.50%	1.3%
K	0.00%	0.84%	3.9%	7.8%	11.40%
a	0.00%	2.03%	6.72%	10.07%	15.36%
q	0.00%	2.55%	8.37%	12.38%	20.55%
w	0.00%	5.18%	15.52%	21.15%	27.82%
b	0.00%	1.01%	4.32%	7.14%	12.27%

4. PARAMETER ESTIMATION OF IMMUNO-EPIDEMIOLOGICAL MODEL

We develop a numerical method to fit multi-scale models to multi-scale data. For epidemiological data, we use the number of human RVFV cases reported by the CDC during the 2006-2007 outbreak in Kenya [7] (given in Table 9). RVF is transmitted to humans mainly through direct and indirect contact with body fluids and aerosols of infected animals during slaughtering, butchering, or assisting animal births and necropsy or laboratory procedures [5]. Another route of transmission for humans are bites of infected vectors. No human-to-human cases have been seen so far. Since the majority of human infections results from handling infected animals, we model only transmission from infected host to human and augment the following human outbreak model to the system (2.2) - (2.3)

$$(4.1) \quad \text{RVFV Human Model:} \quad \begin{cases} \frac{dS}{dt} = -S \int_0^\infty \beta(\tau) i_H(\tau, t) d\tau \\ \frac{dI}{dt} = S \int_0^\infty \beta(\tau) i_H(\tau, t) d\tau, \end{cases}$$

TABLE 9. Number of human RVF cases as reported by CDC during the 2006-2007 Kenya outbreak [7].

Day:	0	1	4	6	7	8	10	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Human cases:	1	1	3	1	5	1	5	2	4	2	10	3	9	5	8	16	23	16	13	24	20
Day:	26	27	28	29	30	31	32	33	34	35	36	37	38	39	41	42	43	44	45	46	
Human cases:	19	17	12	14	5	5	10	6	16	8	8	9	3	6	4	1	2	5	2	1	

where $S(t)$ denotes the susceptible human, and $I(t)$ denotes the infected human at time t . The transmission rate from an infected host to human varies according to the infectivity of the host via,

$$\beta(\tau) = \frac{c_B P^v(\tau)}{B + P^v(\tau)},$$

where c_B , B , and v are constants. Because data are the number of new cases, we fit the new human incidences $\hat{g}(t, \hat{\mathbf{p}})$ at time t to the given data where

$$\hat{g}(t, \hat{\mathbf{p}}) = S(t) \int_0^\infty \beta(\tau) i_H(\tau, t) d\tau,$$

and $\hat{\mathbf{p}}$ is the vector of epidemiological parameters to be fitted. We develop a finite difference scheme and combine it with a MATLAB ODE solver for the immunological model to approximate the solutions of the age-structured PDE model.

4.1. Finite Difference Scheme for the Immuno-Epidemiological Model. We develop a finite difference method that discretizes both the time-since-infection variable τ and the time variable t . Let T be the final time of interest and A be the final time-since-infection. Hence, $D = \{(\tau, t) : 0 \leq \tau \leq A, 0 \leq t \leq T\}$ is the domain of the system (2.2) - (2.3). First, we discretize the domain D . Let N and M be nonnegative integers such that

$$\Delta t = \frac{T}{N} \quad \text{and} \quad \Delta \tau = \frac{A}{M} \quad \text{and} \quad \Delta t = \Delta \tau.$$

We generate a square mesh and the discrete point (τ_k, t_n) of the rectangular domain D is given by $\tau_k = k\Delta\tau$ and $t_n = n\Delta t$, $k = 1, \dots, M$, $n = 1, \dots, N$. The approximate values at any point (τ_k, t_n) of the discretized domain are denoted by $S(t_n) \approx S^n$, $i_H(\tau_k, t_n) \approx i_H^{k,n}$, $\beta_H(\tau_k) \approx \beta_H^k$. We approximate the time derivatives by backward Euler difference quotient and obtain the following implicit method for the system (2.2) - (2.3).

(4.2) **Finite Difference Algorithm:**

$$\left\{ \begin{array}{l} S_v^{n+1} = \frac{S_v^n + \Delta t \Lambda_v}{1 + \Delta t \sum_{k=1}^M \beta_H^k i_H^{k,n} \Delta t + \mu_v \Delta t}, \\ I_v^{n+1} = \frac{I_v^n + \Delta t S_v^{n+1} \sum_{k=1}^M \beta_H^k i_H^{k,n} \Delta t}{1 + \mu_v \Delta t}, \\ S_H^{n+1} = \frac{S_H^n + \Delta t \Lambda}{1 + \beta_v I_v^n \Delta t + \mu_H \Delta t}, \\ i_H^{k+1,n+1} = \frac{i_H^{k,n}}{1 + (\alpha^k + \gamma^k + \mu_H) \Delta t}, \\ i_H^{0,n+1} = \beta_v S_H^{n+1} I_v^{n+1}, \\ R_H^{n+1} = \frac{R_H^n + \Delta t \sum_{k=1}^M \gamma^k i_H^{k,n} \Delta t}{1 + \mu_H \Delta t}, \end{array} \right.$$

To evaluate the linked parameters $\beta_H(\tau_k)$, $\alpha(\tau_k)$, $\gamma(\tau_k)$, the immunological model (3.11) is solved using the MATLAB ODE solver `ode45`. Not all MATLAB ODE solvers can be set to preserve the positivity of solutions. Since it is crucial that the numerical method also preserves the positivity of the solutions, we set the function `ode45` to preserve the positivity of the approximate solutions for all the numerical experiments in this paper. Integrals $\int_0^\infty \beta_H(\tau) i_H(\tau, t) d\tau$ and $\int_0^\infty \gamma(\tau) i_H(\tau, t) d\tau$ are approximated using the right-end point rule. The nonlinear terms are linearized using a single Picard Iteration at each time level. For instance the nonlinear term, $\beta_v S_H^{n+1} I_v^{n+1}$ is linearized by $\beta_v S_H^{l+1} I_v^l$ where l denotes the Picard Iteration counter at time $t = t_{n+1}$. The approximate solutions obtained by the implicit method (4.2) converge to the solutions of the system (2.2) - (2.3) with single Picard Iteration, that is with $l = 1$ (results are not shown). The order of convergence is $O(\Delta t)$. The numerical method (4.2) preserves the positivity of the solutions. More details about finite difference schemes for age structured models can be found in [20].

Structural identifiability of age structured PDE models have not been studied as extensively as dynamical ODE systems with few exceptions in [26, 27]. In recent years, researchers have focused on developing methods for solving the ill-posed inverse problems generated by age-structured population models [1, 2]. However, these studies did not consider the structural identifiability of the age-structured models. Analyzing the structural identifiability of the nested immuno-epidemiological models such as (2.2) - (2.3) is the topic of our current research. In this paper we take a numerical approach and use Monte Carlo simulations for studying identifiability issues of the parameter estimation problem in a nested immuno-epidemiological model.

The parameters of importance to determine from the epidemiological data are the ones related to transmission, recovery and disease-induced death rates of the RVFV. Other epidemiological parameters such as Λ_v , Λ , μ_v , μ_H and β_v are fixed to the values obtained

TABLE 10. Fixed parameter values of model (2.2) - (2.3)

Parameter	Fixed Value	Source	Unit	Parameter	Fixed Value	Source	Unit
Λ_v	1/40	text	$\frac{\text{vector}}{\text{time}}$	$I_v(0)$	0.000005	text	vector
μ_v	1/40	[22]	$\frac{1}{\text{time}}$	$S_H(0)$	0.9999	text	host
Λ	1/(365 × 10)	text	$\frac{\text{host}}{\text{time}}$	$i_H(\tau, 0)$	0.0000001	text	$\frac{\text{host}}{\text{time}}$
μ_H	1/(365 × 10)	[14]	$\frac{1}{\text{time}}$	$R_H(0)$	0	text	host
β_v	0.2	text	$\frac{1}{\text{host} \times \text{time}}$	$S(0)$	2300000	text	human
$S_v(0)$	0.999995	text	vector	$I(0)$	1	text	human

from the literature and presented in Table 10. Studies find the median adult longevity of *Aedes aegypti*, a mosquito known to transmit RVFV, in the lab to be approximately 61 days [22]. We took the average life span of the vector population to be $\mu_v = \frac{1}{40}$ days⁻¹, intermediate between laboratory measured lifespan and the lifespan previously used to model wild populations of mosquitoes ([13, 14]). The total vector and host populations are normalized to unity, and hence the recruitment rate of the vector population is set to $\Lambda_v = \frac{1}{40}$ vector × days⁻¹. Average life span of livestock is 10 years [14]. Hence $\mu_H = \frac{1}{365 \times 10}$ days⁻¹ and $\Lambda = \frac{1}{365 \times 10}$ host × days⁻¹. We suppose that initially 0.005% of the vector population is infected and initial distribution of the infected host is 0.0001% of the population. 85% of the reported cases for the 2006-2007 Kenya outbreak came from only 4 districts (out of 69) in Kenya [25]; Garissa (623060), Ijara (62571), Baringo (555561), Kilifi (41109735). Initial susceptible human population is estimated to $S(0) = 2300000$ (an approximate total population of above 4 districts in Kenya). We take $\nu = 1$ and $\upsilon = 2$.

4.2. Fitting epidemiological data when the immune model parameters are fixed. To estimate the parameters of the immuno-epidemiological model, we first fix the parameters of the immunological model at the fitted values given in Table 6. The remaining parameters are fitted using a MATLAB code combining an ODE solver for the immunological model, finite difference scheme (4.2) for the PDE model and an optimization routine to minimize the least squares error. Hence, we estimate the parameters $\hat{\boldsymbol{p}} = [\sigma, \xi, \kappa, \epsilon_0, c_H, B_H, c_B, B]$ of the epidemiological model (2.2) - (2.3), (4.1) by solving the following optimization problem using the Nelder-Mead algorithm

$$(4.3) \quad \min_{\hat{\boldsymbol{p}}} \hat{J}(\hat{\boldsymbol{p}}) = \min_{\hat{\boldsymbol{p}}} \sum_{j=1}^{\hat{n}} (\hat{g}(t_j, \hat{\boldsymbol{p}}) - \hat{y}_j)^2,$$

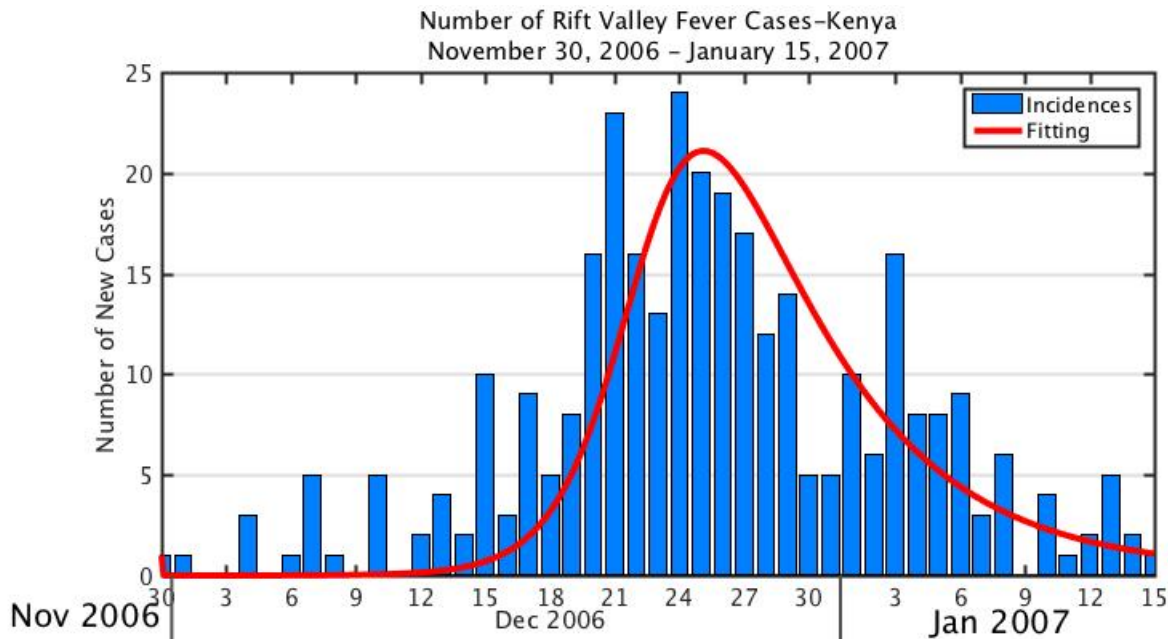


FIGURE 4. Fitting results. The blue bars are the new incidences reported by CDC[7] as given in Table 9. The red curve is the computed new incidences with the fitted parameters presented in Table 11. The immune parameters are fixed at the estimated values given in Table 6.

TABLE 11. Parameter estimates of the epidemiological model (2.2) - (2.3) fitted to human incidences as reported by CDC. The immune parameters are fixed at the estimated values given in Table 6. The immuno-epidemiological model is approximated by the finite difference scheme (4.2) with $\Delta t = 0.01$.

Parameter	Estimate	Unit	Parameter	Estimate	Unit
σ	4.9×10^{-9}	$\frac{1}{\text{pathogen} \times \text{time}}$	κ	0.0694	$\frac{\text{pathogen}}{\text{antibody} \times \text{time}}$
ξ	2.9281×10^{-5}	$\frac{1}{\text{antibody} \times \text{time}}$	ϵ_0	9.3011×10^{-5}	pathogen
c_H	0.6123	$\frac{1}{\text{host} \times \text{time}}$	c_B	1.1873×10^{-5}	$\frac{1}{\text{host} \times \text{time}}$
B_H	5.9504×10^3	pathogen	B	2.7443×10^{-8}	pathogen

where $\hat{n} = 47$, \hat{y}_j are the number of new RVFV human cases at time t_j as reported by CDC [7]. The fitted parameter values are presented in Table 11. In Figure 4 the new incidences computed with the fitted values are plotted with the epidemiological data (blue bars) provided by CDC.

Currently, there are no analytical approaches for studying structural identifiability of nested models such as the epidemiological model we study in this paper (2.2) - (2.3).

We are working on developing analytical tools to study the structural identifiability of nested models. But, within this paper we analyze the identifiability of the parameters of the nested immuno-epidemiological model (2.2) - (2.3) using Monte Carlo simulations. Monte Carlo simulations have been widely used for practical identifiability of ODE models [24]. We perform Monte Carlo simulations first by taking the estimated parameters presented in Table 11 as the true parameter set $\hat{\mathbf{p}}$ and then performed the steps (1) through (4) as explained in details in Section 3 with the following modifications. In step (1), $g(\mathbf{x}(\tau_i), \hat{\mathbf{p}})$ is replaced with $\hat{g}(t_i, \hat{\mathbf{p}})$, in step (2) we set $M = 1000$, and

$$y_j = \hat{g}(t_j, \hat{\mathbf{p}}) + \hat{g}(t_j, \hat{\mathbf{p}})\epsilon_j \quad j = 1, 2, \dots, \hat{n}$$

in step (3) we evaluated (4.3). The immuno-epidemiological model is approximated by the finite difference scheme (4.2) with $\Delta t = 0.1$ (see Table 12) and $\Delta t = 0.01$ (see Table 13). Monte Carlo simulations for the immuno-epidemiological PDE model are computationally expensive compared to the immunological ODE model. Computation time increases as the time step (Δt) in the finite difference scheme is refined. We performed the Monte Carlo simulations for $M = 1000$ synthetic data sets for the time step $\Delta t = 0.1$ (Table 12) and for $M = 100$ data sets for the time step $\Delta t = 0.01$ (Table 13). The parameters of the scaled immunological model (3.11) are fixed at the estimated values presented in Table 6. The computed AREs are presented in Table 12 and Table 13 for measurement errors $\sigma_0 = 0\%, 1\%, 5\%, 10\%, 30\%$. The AREs of the parameters c_H , B_H , and c_B increase gradually as the measurement error level increases, but remain below the measurement error level. Hence, we claim that the parameters c_H , B_H , and c_B are practically identifiable. We also observe that refining the time step, Δt , in the finite difference scheme (4.2) did not affect the ARE values.

On the other hand, the ARE values of the parameters σ , ξ , κ , and ϵ_0 reveal an unusual relation to the measurement error levels; the AREs do not increase as the measurement error levels increase (see Table 12 and Table 13). The parameters σ , ξ , κ , and ϵ_0 appear in the formulation of $\pi(\tau)$ as given in (2.5) which is the probability of still being infectious τ time units after becoming infected. Note that the infected host population, $i_H(\tau, t)$ satisfies the following relation which is derived by the method of characteristics,

$$(4.4) \quad i_H(\tau, t) = \begin{cases} \beta_v S_H(t - \tau) I_v(t - \tau) \pi(\tau), & \tau < t, \\ i_H(\tau - t, 0) \frac{\pi(\tau)}{\pi(\tau - t)}, & \tau > t. \end{cases}$$

Then new incidences are given as

$$(4.5) \quad \hat{g}(t, \hat{\mathbf{p}}) = \begin{cases} \beta_v S(t) \int_0^\infty \beta(\tau) S_H(t - \tau) I_v(t - \tau) e^{-\int_0^\tau (\alpha(s) + \gamma(s) + \mu_H) ds} d\tau & \tau < t, \\ S(t) \int_0^\infty \beta(\tau) i_H(\tau - t, 0) \frac{e^{-\int_0^\tau (\alpha(s) + \gamma(s) + \mu_H) ds}}{e^{-\int_0^{\tau-t} (\alpha(s) + \gamma(s) + \mu_H) ds}} d\tau & \tau > t, \end{cases}$$

We hypothesize that the strange behavior of the AREs of the parameters σ , ξ , κ , and ϵ_0 is due to the fact that the measurement level error does not affect the disease induced

TABLE 12. Practical identifiability analysis of the epidemiological model (2.2) - (2.3) by Monte Carlo simulations. The immune parameters are fixed at the estimated values given in Table 6. The immuno-epidemiological model is approximated by the finite difference scheme (4.2) with $\Delta t = 0.1$. AREs of the parameters are presented for each error level. $M = 1000$.

Error Level	σ	ξ	c_H	B_H	c_B	B	κ	ϵ_0
0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1%	3.48	2.55	0.05	0.54	0.20	3.11	1.20	1.94
5%	3.83	2.54	0.2	1.68	1.17	4.37	1.35	2.37
10%	3.91	2.50	0.45	2.76	2.27	4.71	1.42	2.67
30%	3.78	2.24	1.51	4.46	7.15	5.62	1.68	2.40

death rate, $\alpha(s)$, and the recovery rate, $\gamma(s)$, as much as it affects the transmission rate from an infected host to human, $\beta(\tau)$, and the transmission rate from an infected host to a vector, $\beta_H(\tau)$, since they are related through the integral of a negative exponential term. We test our hypothesis with the following simplified version of (4.5). Assume that data are given by the following output function,

$$(4.6) \quad h(t, \hat{\mathbf{p}}) = S(0)e^{-\left(\int_0^\infty \beta(\tau)\pi(\tau)d\tau\right)t} \int_0^\infty \beta(\tau)\pi(\tau)d\tau$$

The immunological parameters are fixed at the estimated values given in Table 6, and the parameters of the transmission rate $\beta(\tau)$, c_H , and B_H are fixed at the estimated values given in Table 11. We perform Monte Carlo simulations to compute the AREs of the parameters σ , ξ , κ , and ϵ_0 and set the true parameters as the fitted values given in Table 11. We present the results in Table 14. We observe that the AREs of σ increase gradually as the measurement error level increases, but the AREs of the parameters ξ , κ , and ϵ_0 preserve the unusual behavior. Clearly, fixing the parameters c_H and B_H improved the identifiability of the σ . We conclude that fitting the immuno-epidemiological model (2.2) - (2.3) when the immunological parameters are fixed, only the parameters c_H , B_H , and c_B are practically identifiable.

4.3. Fitting multi-scale data to multi-scale model. Our initial approach in fitting multi-scale data to a multi-scale model was to fit the immunological model first (3.13), then use those fitted parameters in estimating the parameters of the epidemiological model (4.3). Next, we fit the multi-scale data to the multi-scale model simultaneously. Thus, the vector of parameters to be fitted are $\mathbf{p}^* = [\mathbf{p}, \hat{\mathbf{p}}] = [r, K, a, b, q, w, \sigma, \xi, \kappa, \epsilon_0, c_H, B_H, c_B, B]$, then we estimate \mathbf{p}^* by solving the following minimization problem,

$$(4.7) \quad \min_{\mathbf{p}^*} J^*(\mathbf{p}^*) = \min_{\mathbf{p}, \hat{\mathbf{p}}} \left(J(\mathbf{p}) + \hat{J}(\hat{\mathbf{p}}) \right) = \min_{\mathbf{p}, \hat{\mathbf{p}}} \left(\sum_{i=1}^n \omega_i (\log_{10}(g(\mathbf{x}(\tau_i), \mathbf{p})) - y_i)^2 + \sum_{j=1}^{\hat{n}} (\hat{g}(t_j, \hat{\mathbf{p}}) - \hat{y}_j)^2 \right)$$

where y_i represents the immunological data, and \hat{y}_j represents the epidemiological data. To solve the optimization problem (4.7), we use the Nelder-Mead algorithm implemented

TABLE 13. Practical identifiability analysis of the epidemiological model (2.2) - (2.3) by Monte Carlo simulations. The immune parameters are fixed at the estimated values given in Table 6. The immuno-epidemiological model is approximated by the finite difference scheme (4.2) with $\Delta t = 0.01$. AREs of the parameters are presented for each error level. $M = 100$.

Error Level	σ	ξ	c_H	B_H	c_B	B	κ	ϵ_0
0%	0.02	0.02	0.00	0.00	0.00	0.00	0.02	0.02
1%	3.52	2.47	0.04	0.63	0.21	3.54	1.07	2.16
5%	3.87	2.36	0.19	1.57	1.09	4.19	1.40	2.13
10%	3.97	2.44	0.36	2.27	2.24	4.75	1.54	2.69
30%	3.66	2.21	1.36	4.48	6.56	5.62	1.31	2.09

TABLE 14. Monte Carlo simulations for the output function $h(t, \hat{\mathbf{p}})$ given in (4.6). The immune parameters are fixed at the estimated values given in Table 6. AREs of the parameters σ , ξ , κ , and ϵ_0 are presented for each error level. $M = 1000$.

Error Level	σ	ξ	κ	ϵ_0
0%	0.00	0.00	0.00	0.00
1%	0.25	3.19	2.21	4.01
5%	1.51	3.46	4.64	5.45
10%	3.42	3.60	4.38	6.89
30%	9.09	4.25	4.71	8.82

in MATLAB which is one of the most widely used derivative-free algorithms for unconstrained optimization problems such as (4.7). The estimated values are presented in Table 15. The model output at the estimated values together with the data are plotted in the Figure 5. Next, we perform Monte Carlo simulations to test the practical identifiability of the nested models (3.11) and (2.2) - (2.3) when both immunological and epidemiological data are available for the parameter estimation. The computed AREs of the parameters are presented in Table 16. We observe that the practical identifiability of all the parameters have improved. All the parameters of the immuno-epidemiological model are practically identifiable for multi-scale data. It is very well known that in dynamical ODE models, the identifiability of the model parameters improves as the more state variables are observed. We conclude that same phenomena holds true for the nested models (3.11) and (2.2) - (2.3).

5. DISCUSSION

Models must be properly parameterized if conclusions drawn from their predictions are to be trusted. Structural identifiability issues for parameter estimation have been studied for many biological systems. From a modeling and biological perspective, it is clear that within-host dynamics have an effect on the between-host disease transmission. Hence,

TABLE 15. Parameter estimates of the nested immuno-epidemiological model (2.2) - (2.3) fitted to human incidences as reported by CDC and the immunological model (3.11) fitted to the viremia levels given in Table 5. The immuno-epidemiological model is approximated by the finite difference scheme (4.2) with $\Delta t = 0.1$ and the immunological model (3.11) is approximated with MATLAB built in function `ode45`. The optimization problem (4.7) is solved by Nelder-Mead algorithm.

Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
σ	6.21×10^{-8}	B	4×10^{-11}	q	7.6944
ξ	8.51×10^{-5}	κ	0.7212	w	1.2914
c_H	0.5365	ϵ_0	7.43×10^{-4}	b	9.58×10^{-8}
B_H	3.03×10^3	r	6.3033	K	7.57×10^7
c_B	7.63×10^{-6}	a	3.93×10^{-7}		

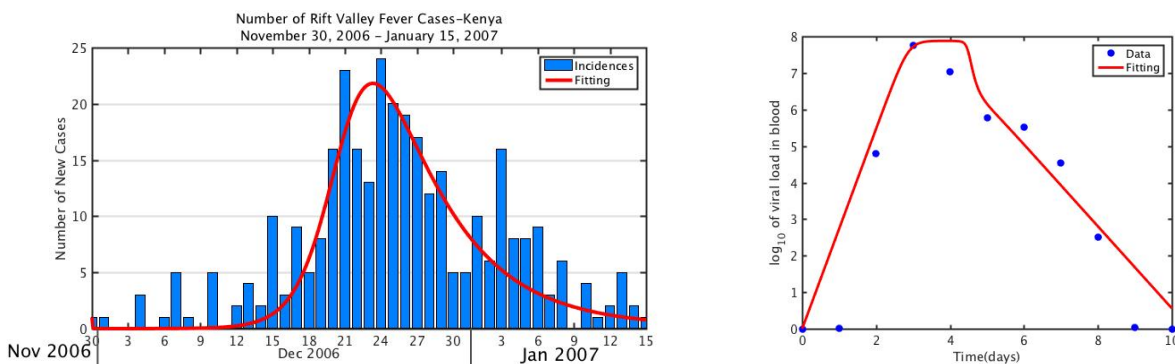


FIGURE 5. Fitting results of the parameter estimate problem (4.7). In the left figure, the blue bars are the new incidences reported by CDC[7]. The red curve is the computed new incidences with the fitted parameters presented in Table 15. In the right figure, the blue dots are the data given in Table 5 and the red curve is the \log_{10} of the viremia levels computed with the estimated values presented in Table 15.

nested models have been developed to study host-pathogen coevolution, early diagnostics, and transmission-virulence trade-offs [15, 38, 39]. In this paper, we are interested in the structural and practical identifiability analysis of nested immuno-epidemiological models, where between-host disease transmission, host recovery and disease-induced death rates are governed by within-host immune dynamics. Although there are well established theories for identifiability of dynamical systems described by ODE models, identifiability issues for structured PDE models have been a recent topic of interest. We extend the structural and practical identifiability analysis of PDE models to a multi-scale coupled model.

TABLE 16. Practical identifiability analysis of nested immuno-epidemiological model by Monte Carlo simulations for measurement error levels $\sigma_0 = 0\%, 1\%, 5\%, 10\%$ and 30% .

Parameter	ARE $\sigma_0 = 0\%$	ARE $\sigma_0 = 1\%$	ARE $\sigma_0 = 5\%$	ARE $\sigma_0 = 10\%$	ARE $\sigma_0 = 30\%$
σ	1.00	1.80	3.73	6.78	13.49
ξ	1.00	1.50	3.14	5.46	15.92
c_H	4×10^{-9}	0.07	0.34	0.82	2.57
B_H	1×10^{-7}	0.49	2.95	6.83	11.70
c_B	4×10^{-8}	0.21	1.08	2.29	6.27
B	2.63×10^{-6}	3.03	5.05	6.50	19.90
κ	1.00	1.66	4.57	7.71	16.91
ϵ_0	1.00	1.80	4.27	6.86	16.41
r	1.86×10^{-8}	0.11	0.61	1.26	3.44
a	1.22×10^{-8}	0.06	0.35	0.87	3.21
q	2.46×10^{-8}	0.10	0.56	1.15	3.59
w	1.25×10^{-8}	0.07	0.51	1.06	3.22
b	8.26×10^{-9}	0.02	0.16	0.35	1.76
K	1.80×10^{-8}	0.06	0.45	0.73	3.44

Regarding constructing an immunological model for arbovirus diseases, although within-host immune dynamics have the same qualitative behavior, the immune response hasn't been modeled with IgM and IgG antibodies which are commonly measured in the laboratory [5, 16]. Here we used a novel immunological model that represents the within-host immune dynamics resulting from infection from an arbovirus disease, to estimate the unknown parameters by using RVFV time-series data obtained under experimental conditions. For the 'macro scale' perspective, we used 2006 – 2007 Kenya outbreak data [7, 25] for RVFV human cases to determine the unknown parameter values for the immuno-epidemiological model.

We develop a finite difference scheme and combine it with a MATLAB ODE solver for the immunological model to approximate the solutions of the age-structured PDE model. In recent years, although many researchers have focused on developing methods for solving the ill-posed inverse problems generated by the age-structured population models, there have been few studies on the structural and practical identifiability of the age-structured models. In this study, we take a numerical approach for the structural and practical identifiability analysis of a nested immuno-epidemiological model, which couples an ODE immunological model to an age-since-infection structured epidemiological PDE model.

For dynamical systems described by ordinary differential equations, the state isomorphism method, the Taylor series expansion method, and the algebra-differential elimination method are mainly used for identifiability analysis. We use the differential algebra method to study the structural identifiability of the immunological model. One of the advantages of this approach is that if the model is not structurally identifiable, the method provides the identifiable parameter combinations. The parameter combinations

are then used to reparametrize the system to obtain a structurally identifiable model [23]. We show that the immunological model is not structurally identifiable for the measurements of time-series viremia concentrations in the host. Thus we study the non-dimensionalized and scaled version of the immunological model and prove that both are structurally globally identifiable. Although the non-dimensionalized system is structurally identifiable, we cannot use the non-dimensionalized system to reconstruct the parameter values for the immune model (2.1) from the uniquely estimated parameter values of non-dimensionalized system. We estimate the unique value of all *key parameters* in the immunological model (2.1) via the scaled model by fixing the biologically estimated parameter c (half life of IgM immune response antibodies). However, in practice, structural identifiability may not imply practical identifiability. Here we also study the practical identifiability of the scaled model by Monte Carlo simulations and show that the parameters K , r , and b are practically identifiable, but the parameters q , a and w are not practically identifiable.

After fixing parameters to the fitted values for the immunological model, we develop numerical methods to fit the observable epidemiological RVFV human data to the multi-scale model to estimate the rest of the immuno-epidemiological model parameter values. For this fitting, we use MATLAB by combining an ODE solver for the immunological model, finite difference scheme for the PDE model, and an optimizing tool to minimize the least square errors. Monte Carlo simulations suggest that only three parameters of the epidemiological model are identifiable when the immunological parameters are fixed at fitted values. Alternatively, we fit the multi-scale data to the multi-scale model simultaneously. To solve the obtained optimization problem, we use a modified Nelder-Mead algorithm `mfminsearch` implemented in MATLAB. Monte Carlo simulations for simultaneous fitting suggest that the parameters of immunological model and the parameters of the immuno-epidemiological model are practically identifiable.

We suggest developing analytical methods to study structural identifiability of nested models and this is our current topic of research. Developing such analytical methods is a necessity especially in deriving multi-scale models that adequately explain multi-scale data. Structural identifiability analysis of unidentifiable nested model can reveal the parameter relationships, and this can further be used to reparameterize the nested (PDE) model to obtain an identifiable model. A technique that has been widely used for ODE models should be developed for nested models.

From a methodological perspective, accurate estimation of the nested model parameter values has more caveats and multi-scale nested models add one more layer complexity to these analyses in comparison to ODE models. We take a numerical approach for practical identifiability analysis of nested models. Estimation for accurate values of model parameters allows a biological system to have a satisfactory mathematical representation, which is necessary for theoretical studies confronting experiments. Structural identifiability analysis allows for clear model predictions to model parameter values.

ACKNOWLEDGMENT

The authors N. Tuncer and M. Martcheva acknowledge support from the National Science Foundation (NSF) under grants DMS-1515661/DMS-1515442. Authors H. Gulbudak and V. Cannataro would also like to acknowledge partial support from IGERT

grant NSF DGE-0801544 in the Quantitative Spatial Ecology, Evolution and Environment Program at the University of Florida. We would like to thank the reviewers for their constructive comments which lead to the improvement of the paper.

REFERENCES

- [1] A. ACKLEH, H.T. BANKS, K. DENG, S. HU, Parameter Estimation in a Coupled System of Nonlinear Size-Structured Populations, *Mathematical Biosciences and Engineering*, **2(2)**, (2005), p. 289-315.
- [2] A. ACKLEH, Parameter Identification in Size-Structured Population Models With Nonlinear Individual Rates, *Mathematical and Computer Modeling*, **30**, (1999), p. 81-92.
- [3] H. T. BANKS, S. HU, AND W.C. THOMPSON, Modeling and inverse problems in the presence of uncertainty, CRC Press, 2014.
- [4] G. BELLU, M.P. SACCOMANI, S. AUDOLY, L. D'ANGIO, DAISY: a new software tool to test global identifiability of biological and physiological systems, *Comput. Methods Programs Biomed.*, **88**,1, (2007) p. 52-61.
- [5] B.H. BIRD, T. G. KSIAZEK, S. T. NICHOL, J. MACLACHLAN, Rift Valley fever virus, *JAVMA* **234**(7), (2009), p. 883-893.
- [6] A.CAPALDI, S. BEHREND, J. SMITH, B. BERMAN, J. WRIGHT, A. L. LLOYD, Parameter estimation and uncertainty quantification for an epidemic model, *Mathematical Biosciences*, **9(3)**,(2012), p. 553-576.
- [7] CDC, Rift Valley Fever Outbreak, Kenya, November 2006-January 2007, Morbidity and Mortality Weekly Report (MMWR) 56 (04) (2007), p. 73-76.
- [8] CDC, Rift Valley Fever Transmission, www.cdc.gov/vhf/rvf/transmission/index.html.
- [9] A. CINTRN-ARIAS, H.T. BANKS, A.CAPALDI, A. L. LLOYD, A sensitivity matrix based methodology for inverse problem formulation, *Journal of Inverse and Ill-Posed Problems* **17** (6), (2009), p. 545-564.
- [10] O.T. CHRIS, J.R. BANGA, E. BALSACANTO, Structural identifiability of systems biology models: A critical comparison of methods, *PloS One*, (2011) 6(11): e27755.
- [11] C. COBELLI, J.J. DiSTEFANO, Parameter and structural identifiability concepts and ambiguities: a critical review and analysis, *Am J Physiol.* **239(1)**,(1980), p. R7-24.
- [12] M. EISENBERG, S. ROBERTSON AND J. TIEN, Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease, *JTB* **324**, (2013), p. 84-102.
- [13] GAFF, H., HARTLEY, D., LEAHY, N. A mathematical Model of Rift Valley Fever. *Electronic Journal of Differential Equations (EJDE)*, **2007(115)**, (2007),p. 1-12.
- [14] GAFF, H., BURGESS, C., JACKSON, J., NIU, T., PAPELIS, Y., HARTLEY, D. Mathematical Model to Assess the Relative Effectiveness of Rift Valley Fever Countermeasures. *International Journal of Artificial Life Research (IJALR)*, **2(2)**, (2011),p. 1-18.
- [15] M.A. GILCHRIST, A. SASAKI, Modeling host-parasite coevolution: a nested approach based on mechanistic models, *J. Theor. Biol.* 218(3), (2002), p. 289-308.
- [16] P. JANSEN VAN VUREN, J.T. PAWESKA, Laboratory safe detection of nucleocapsid protein of Rift Valley fever virus in human and animal specimens by a sandwich ELISA, *Journal of Virological Methods*, **157**, (2009), p. 15-24.
- [17] LAGARIAS, J.C., J. A. REEDS, M. H. WRIGHT, AND P. E. WRIGHT, Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions, *SIAM Journal of Optimization*, **9**,1,(1998) p. 112-147.
- [18] S. R. LELE, K. NADEEM, AND B. SCHMULAND m Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning, *J. Am. Stat. Assoc.* (2010) 105, p.16171625. (doi:10.1198/jasa.2010.tm09757)
- [19] L. LJUNG, T. GLAD, On the global identifiability of arbitrary model parametrizations. *Automatica*, **30**, (1994), p. 265-276.
- [20] M. MARTCHEVA, An introduction to mathematical epidemiology, Springer, New York, 2015.
- [21] MCKINNON, K.I.M., Convergence of the Nelder-Mead simplex method to a nonstationary point, *SIAM J. Optim* **9**, (1998) p.148?158.

- [22] McMENIMAN CJ, LANE RV, CASS BN, FONG AW, SIDHU M, WANG YF, O'NEILL SL., Stable introduction of a life-shortening Wolbachia infection into the mosquito *Aedes aegypti*, *Science*, **323** (2009) p. 141-144. doi: 10.1126/science.1165326.
- [23] N. MESHKAT, M. EISENBERG, J. DiSTEFANO III An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Grbner Bases, *Mathematical Biosciences*, **222**, (2009), p. 61-72.
- [24] H. MIAO, X. XIA, A. S. PERELSON, H. WU On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM review*. **53**, (1), (2011), p. 3-39.
- [25] P. MUNYUA, R.M. MURITHI, S. WAINWRIGHT, J. GITHINJI, A. HIGHTOWER, D. MUTONGA, J. MACHARIA, P.M. ITHONDEKA, J. MUSAA, R.F. BREIMAN, P. BLOLAND, M.K. NJENGA Rift Valley fever outbreak in livestock in Kenya, 2006-2007. *Am J Trop Med Hyg*, (2010), p. 58-64, doi: 10.4269/ajtmh.2010.09-0292.
- [26] A. PERASSO, B. LAROCHE, Y. CHITOUR, S. TOUZEAU Identifiability analysis of an epidemiological model in a structured population. *Journal of Mathematical Analysis and applications* **374**(1) (2011), p.154-165.
- [27] A. PERASSO, U. RAZAFISON, Identifiability problem for recovering the mortality rate in an age-structured population dynamics model, (2014), hal-00995676
- [28] B. PETERSEN, K. GERNAEY AND P.A. VANROLLEGHEM, Practical identifiability of model parameters by combined respirometric-titrimetric measurements *Water Science and Technology*, **43**, No 7 (2001) p. 347-355.
- [29] H. POHJANPALO, System identifiability based on power-series expansion of solution, *Math. Biosci.*, **41**, (1978), p. 21-33.
- [30] A. PUGLIESE, The role of host population heterogeneity in the evolution of virulence, *JBD*, **00**, (2010), p. 1-16.
- [31] H. GULBUDAK, V. CANNATARO, N. TUNCER, M.MARTCHEVA, From Ecology to Evolution of Host and Vector-Borne Pathogen in a Structured Immuno-epidemiological Model, (2015). Manuscript in preparation.
- [32] RAUE A, KREUTZ C, MAIWALD T, BACHMANN J, SCHILLING M, KLINGMULLER U, TIMMER J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, *Bioinformatics* **25**(15), (2009) p,1923-1929.
- [33] J.G. REID, Structural identifiability in linear time-invariant systems, *IEEE Trans. Automat. Control*, **22**, (1977), p. 242-246.
- [34] A.W. VAN DER VAART, Asymptotic Statistics. Cambridge University Press, 1998.
- [35] D. WATSON, Biological half-life of ovine antibody in neonatal lambs and adult sheep following passive immunization, *Veterinary immunology and immunopathology*, **30**, (1992), p. 221-32.
- [36] <http://extractdata.blogspot.com>
- [37] C. MANORE, B. R. BEECHLER Inter-Epidemic and Between-Season Persistence of Rift Valley Fever: Vertical Transmission or Cryptic Cycling? *Transboundary and emerging diseases* (2013) p. 1-11
- [38] M. MOHTASHEMI, R. LEVINS Transient dynamics and early diagnostics in infectious disease. *Journal of Mathematical Biology* **470** (2001) p.446470.
- [39] S. ALIZON, M. VAN BAALEN Emergence of a convex trade-off between transmission and virulence. *The American naturalist*. **165** (2005), p. E155-167.
- [40] Y. XIAO, J. C. BEIER, R. S. CANTRELL, C. COSNER, D. L. DEANGELIS, S. RUAN Modelling the Effects of Seasonality and Socioeconomic Impact on the Transmission of Rift Valley Fever Virus. *PLoS neglected tropical diseases* **9**(1) (2015), p.e3388.
- [41] T. NIU, H. D. GAFF, Y. E. PAPELIS, D. M. HARTLEY An epidemiological model of Rift Valley fever with spatial dynamics. *Computational and mathematical methods in medicine* **2012** (2012).
- [42] J.B. ANDRÉ, J.B. FERDY, B. GODELLE Within-host parasite dynamics, emerging trade-off, and evolution of virulence with immune system. *Evolution* **57** (2003) p. 1489-1497
- [43] S. C. MPESHE, H. HAARIO, J. M. TCHUENCHE A mathematical model of Rift Valley fever with human host. *Acta biotheoretica* **59**(3-4) (2011), p.231-250.

- [44] E. A. J. FISCHER, G. J. BOENDER, G. NODELIJK, A. A. DE KOEIJER, H. J. VAN ROERMUND, The transmission potential of Rift Valley fever virus among livestock in the Netherlands: a modelling study. *Vet. Res.* **44(1)** (2013):58.
- [45] A.I. ROLIN, L. BERRANG-FORD, M. A. KULKARNI The risk of Rift Valley fever virus introduction and establishment in the United States and European Union. *Emerging Microbes & Infections* **2(12)** (2013), p.e81.

DEPARTMENT OF MATHEMATICAL SCIENCES, FLORIDA ATLANTIC UNIVERSITY, SCIENCE BUILDING, ROOM 234 777 GLADES ROAD BOCA RATON, FL 33431

E-mail address: `ntuncer@fau.edu`

SCHOOL OF BIOLOGY AND SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, 310 FERST DR., ATLANTA, GA 30332

E-mail address: `hgulbudak3@gatech.edu`

DEPARTMENT OF BIOLOGY, UNIVERSITY OF FLORIDA, 220 BARTRAM HALL, PO Box 118525, GAINESVILLE, FL 32611-8525

E-mail address: `vcannataro@ufl.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF FLORIDA, 358 LITTLE HALL, PO Box 118105, GAINESVILLE, FL 32611-8105

E-mail address: `maia@math.ufl.edu`