

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/364700763>

Bayesian Multi-level Mixed-effects Model for Influenza Dynamics

Article in *Journal of the Royal Statistical Society Series C Applied Statistics* · October 2022

DOI: 10.1111/rssc.12603

CITATIONS

0

READS

64

1 author:



[Hanwen Huang](#)

University of Georgia

55 PUBLICATIONS 775 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



dynamic model [View project](#)

Bayesian multilevel mixed-effects model for influenza dynamics

Hanwen Huang

Department of Epidemiology and Biostatistics, University of Georgia, Athens, GA 30602

E-mail: huanghw@uga.edu

Summary. Influenza A viruses (IAV) are the only influenza viruses known to cause flu pandemics. Understanding the evolution of different sub-types of IAV on their natural hosts is important for preventing and controlling the virus. We propose a mechanism-based Bayesian multilevel mixed-effects model for characterizing influenza viral dynamics, described by a set of ordinary differential equations (ODE). Both strain-specific and subject-specific random effects are included for the ODE parameters. Our models can characterize the common features in the population while taking into account the variations among individuals. The random effects selection is conducted at strain level through reparameterizing the covariance parameters of the corresponding random effect distribution. Our method does not need to solve ODE directly. We demonstrate that the posterior computation can proceed via a simple and efficient Markov chain Monte Carlo algorithm. The methods are illustrated using simulated data and a real data from a study relating virus load estimates from influenza infections in ducks.

Keywords: Fixed effect, ODE model, Random effect, Reparametrization, Smoothing spline

1. Introduction

Influenza A viruses (IAV) can spread beyond species barriers and adapt rapidly to new hosts and environmental conditions and thus represent a major threat to human and veterinary health. IAV are divided into many sub-types and a pandemic can occur when a new and very different virus emerges that both infects people and has the ability to spread efficiently among people. Wild aquatic birds, particularly certain wild ducks, are the natural hosts for most IAV. Understanding the evolution of different sub-types of IAV in their avian hosts is an important prerequisite to better surveillance and preparedness for pandemic threats.

Our study is motivated by the virus load data from influenza infections in ducks for seven different IAV strains (Figure 1). The seven virus isolates were obtained from an ongoing long-term surveillance of IAV circulation in wild birds in Minnesota, USA (Keeler et al., 2013; Wilcox et al., 2011). Five viruses were isolated from dabbling ducks. Two additional viruses were isolated from surface lake water. For six of the influenza strains, measurements for 5 ducks were collected, for one strain only (H6N2) data for 4 animals is available. There are total 34 ducks. For each duck the virus loads at different time points

after inoculation are measured. Virus load was quantified based on the Matrix gene copy number in cloacal and oropharyngeal samples estimated by real-time PCR. Further details related to the experimental design and methodology are available from the previously published studies (Lebarbenchon et al., 2011, 2012; Handel et al., 2014).

The genetic diversity of influenza A viruses naturally arose from genetic re-assortments and the seven virus strains displayed in Figure 1 were identified on the basis of full-genome sequencing. We aim to investigate the effects of IAV genotype on the within host transmission dynamics in wild duck populations. The mechanism-based ordinary differential equations (ODE) models have been widely used to describe influenza viral dynamics due to their transparent structures. Parameters in ODE often have biologically interpretations and can be used to investigate the effects of possible interventions, such as treatment strategies. When the dynamical process are measured for multiple subjects as shown in Figure 1, the hierarchical mixed-effects ODE models appear to be reasonable for modeling the infection dynamics at population level. The mixed-effects models include random-effects to account for within-subjects correlation and allow for the viral dynamic processes to share certain similar patterns between subjects while still having distinct individual characteristics. For the study described in Figure 1, the random effects can come from both strain and individual levels.

Estimating mixed-effects ODE models is an important but challenging problem because most of the ODEs do not have closed form solutions due to their nonlinear features. Several methods and software have been developed to estimate mixed-effects ODE models by solving ODE numerically. Both the commercial software package, MONMEM (Beal and Sheiner, 1980) and the free popular software, MONOLIX (Lixoft, 2012) use EM-type algorithm such as stochastic approximation EM algorithm (Kuhn and Lavielle, 2005) to estimate nonlinear mixed effect models. Huang et al. (2006) and Huang and Wu (2006) use a Bayesian MCMC approach to estimate mixed-effects for a HIV infection ODE model.

The problem with all these methods is that they need to numerically solve the ODE equation for every updating of the parameters and thus are very intensive in computation. For estimating ODE parameters from a single subject, the two-step based approaches have been proposed in the literature to reduce this computational complexity. Examples include Campbell and Steele (2012); Dondelinger et al. (2013); Macdonald et al. (2016) among many others. For multiple subjects, based on the profiling approach proposed in Ramsay et al. (2007), Wang et al. (2014) developed a semi-parametric method to estimate mixed-effects ODE models. They use spline basis functions to approximate the dynamic process and estimate the spline coefficients, the ODE random effects and fixed effects in three nested levels of optimization. As mentioned in Wu et al. (2012), this spline smoothing-based parameter cascade approaches may not be flexible enough to deal with the complicated local features of the data. In addition, they need more efficient optimization techniques and complicated iterative computation algorithms to obtain an estimator.

Huang et al. (2020) propose a Bayesian model that directly combine the smoothness, system observations, and ODE together. In this model, the ODE constraints are replaced with a probability expression and combined with non-parametric regression into a joint likelihood function. The benefit of this Bayesian approach to parameter estimation in ODEs is that, for some special ODE formulations, the closed form conditional distribution for ODE variables can be obtained which substantially facilitates the convergence of

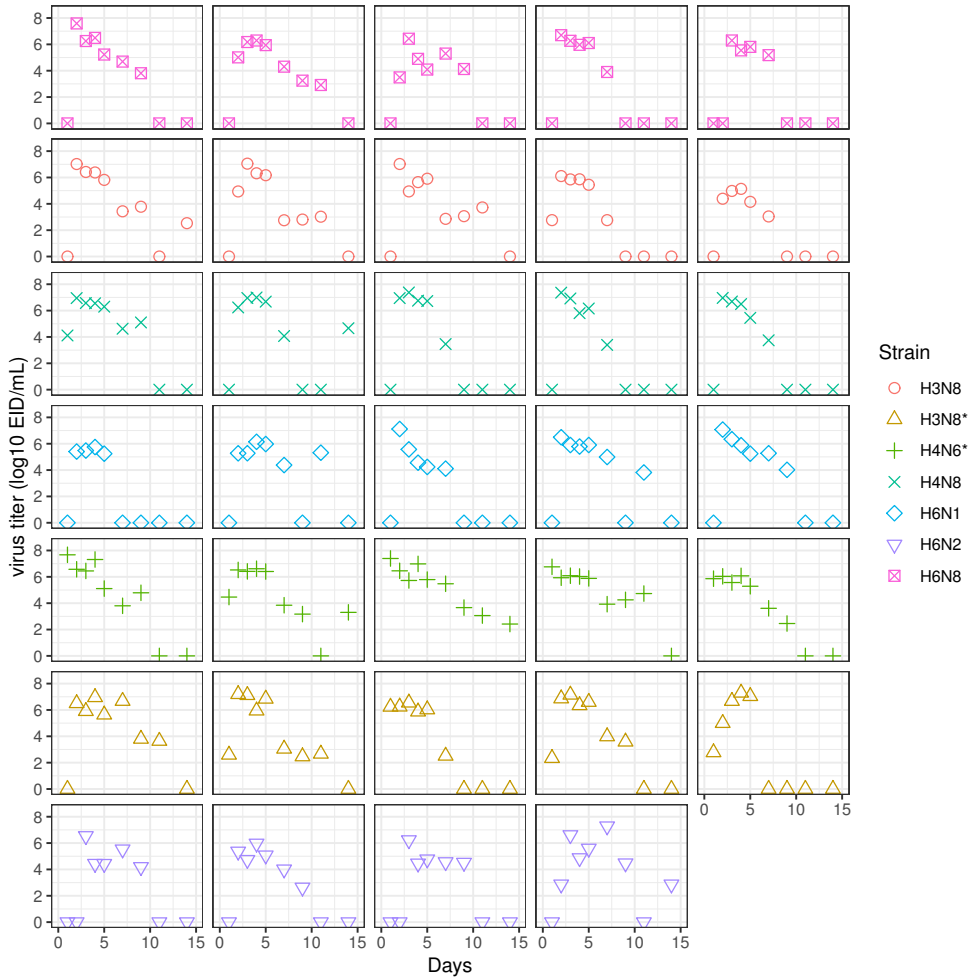


Fig. 1. Virus load data for each infected duck. Five viruses were isolated from dabbling ducks (H3N8; H4N8; H6N1; H6N2; H6N8). Two additional viruses, marked with a \star were isolated from surface lake water (H4N6; H3N8).

MCMC process. Moreover, this approach can be applied to situations where only partial components of the state variables are observed. In this study, we incorporate the hierarchical mixed-effect model into this framework to investigate variability among different levels. We combine all subjects together by including random effects to ODE parameters to account for the variations among different strains and individuals.

An important practical problem in multilevel mixed-effects model is to select the random effect components and decide which coefficients vary at different levels with nonzero variance. Assume there are d potential ODE parameters which can include random effects, one can use standard model selection measures, such as Akaike's information criterion (AIC) or Bayesian information criteria (BIC), to compare all 2^d models for random effect variable selection. However, when 2^d is large, it is very computational intensive because estimating parameters in each ODE model is more complicated than dealing with regular statistical models. Miao et al. (2009) use traditional subset selection methods AIC and BIC to select an ODE model. Zhang et al. (2015) proposed a new ODE model selection method which employs statistical estimation of the full model followed by a combination of a least squares approximation and the adaptive lasso. On Bayesian model selection, a great amount of work exists for fixed effects; however, very little exists on selection of random effects. The main difficulty is that it is not clear what penalty for model complexity is appropriate for comparing models with different numbers of random effects.

Chen and Dunson (2003) proposed an approach to parameterize the mixed model so that functions of the covariance parameters of the random effect distribution are incorporated as regression coefficients on standard normal latent variables. They allow random effects to effectively drop out of the model by choosing mixture priors with point mass at zero for the random effects variances. We incorporate this approach into our multilevel Bayesian ODE framework and perform selection on the random effects component in order to identify ODE parameters that vary across different virus sub-types.

Most closely related to the current paper are results by Liu et al. (2019) that proposed a Bayesian mixed-effects ODE model which also estimated the ODE solution with a linear combination of basis functions. However, the major difference is that they approximated the ODE solution by minimizing a penalty term, while we replace the ODE constraint with a probability expression and implement a fully Bayesian estimation. Moreover, their method was only evaluated by a simple one variable ODE example without considering random effects variable selection.

The rest of the paper is organized as follows. Section 2 describes ODE models for population level influenza dynamics. Section 3 outlines the Bayesian algorithm for ODE parameter estimation. Section 4 presents results from a simulation study. Section 5 illustrates the methodology via an application to the viral load data for influenza infections in ducks illustrated in Figure 1, and Section 6 summarizes the results.

2. ODE model for influenza dynamics at population level

2.1. ODE Model for within-host influenza dynamics

The infection dynamics in avian hosts has been recognized and a mechanistic target-cell limitation model can be used to describe virus kinetics. The model tracks uninfected cells U , infected cells I , and free infectious virus V . They are all functions of the time t . Assume

that cells become infected at rate a , infected cells produce virus at rate p and die at rate b , and free virus is cleared at rate c . The dynamics of the system can be described by the following differential equations

$$\begin{aligned}\dot{U}(t) &= \frac{dU(t)}{dt} = -aU(t)V(t), \\ \dot{I}(t) &= \frac{dI(t)}{dt} = aU(t)V(t) - bI(t), \\ \dot{V}(t) &= \frac{dV(t)}{dt} = pI(t) - cV(t).\end{aligned}\tag{1}$$

In this model, the state variables are $U(t)$, $I(t)$ and $V(t)$, and the unknown parameters are a , b , p , and c . These types of models have been used in several analyses of influenza A virus within-host infection dynamics (see e.e. Nowak and May (2000); Smith and Perelson (2011) for reviews). This simple model can describe most data for influenza virus infections rather well, see e.g. a study in Baccam et al. (2006). After an initial rise in virus load, uninfected target cells become depleted, leading to a subsequent virus decline and resolution of the infection. This so-called target cell limited model is basically equivalent to a simple epidemic model, which produces a single infectious disease outbreak in a susceptible population. This is also a typical inverse problem for ODE and our goal is to estimate the parameters using the measurements of the state variables. The observed state variables are not exactly the ODE solution because they are measured with noise.

In the experiment of influenza infections in ducks, only the virus load was observed among the three components involved in the dynamics. It is necessary to study parameter identifiability of ODE models with only V measured before developing or applying statistical methods to estimate model parameters; otherwise, including unidentifiable unknown parameters in a model is likely to result in biased estimates and thus misleading conclusions. Starting with Model (1) and recalling that only V is experimentally measured, we can eliminate U and I from the model to obtain

$$\begin{aligned}& V\{(\ddot{V} + c\dot{V}) + b(\dot{V} + cV)\} - \dot{V}\{(\ddot{V} + c\dot{V}) + b(\dot{V} + cV)\} \\ & + aV^2\{(\ddot{V} + c\dot{V}) + b(\dot{V} + cV)\} = 0,\end{aligned}$$

which is independent of p . Thus, we can let $p = 1$ and the unknown parameters for ODE model only include a , b , and c .

2.2. Multilevel mixed-effects ODE model for influenza infections

Model (1) describes the infection dynamics at individual level. In order to combine all subjects together and determine how infection kinetics differ among strains and subjects, we consider mixed-effects model which includes both fixed effect components and random effect components. Motivated by the examples of influenza infections in ducks in Figure 1 which show strong variability of infections among both individuals and strains, we consider a two level random effect model in which the first level comes from the strain and the second level comes from the individual.

Denote the number of strains by S and the number of the subjects in the s th strain by n_s . The total number of subjects is $n = \sum_{s=1}^S n_s$. Define a vector of ODE parameters $\boldsymbol{\theta} = (a, b, c)$. Let $\boldsymbol{\theta}_{sj} = (a_{sj}, b_{sj}, c_{sj})$ represent the parameters for the j th subject in the s th strain. For the j th subject in the s th strain, m_{sj} virus load observations $y_{sj1}, \dots, y_{sjm_{sj}}$ are obtained according to independent additive noise model

$$y_{sji} = V_{\boldsymbol{\theta}_{sj}}(t_i) + \epsilon_{sji}, \quad (2)$$

where the noise is Gaussian $\epsilon_{sji} \sim N(0, \sigma_{sj}^2)$ and $V_{\boldsymbol{\theta}_{sj}}(t_i)$ is the solution of virus load to ODE equation (1) at time t_i based on coefficients $\boldsymbol{\theta}_{sj}$. In multilevel mixed-effect model, we include random effects at both strain and individual levels, and have the following formulation

$$\boldsymbol{\theta}_{sj} = \boldsymbol{\mu} + \boldsymbol{\alpha}_s + \boldsymbol{\beta}_{sj}, \quad \boldsymbol{\alpha}_s \sim N(0, \boldsymbol{\Sigma}_S), \quad \boldsymbol{\beta}_{sj} \sim N(0, \boldsymbol{\Sigma}_I). \quad (3)$$

Here $\boldsymbol{\mu}$ represents the fixed effect of the dynamical coefficients, $\boldsymbol{\alpha}_s$ represents the random deviation for the s th strain, and $\boldsymbol{\beta}_{sj}$ represents the random deviation for the j th subject in the s th strain. $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_I$ are d dimensional covariance matrices, where $d = 3$. In hierarchical mixed-effect ODE model (3), it is of primary interest to estimate the fixed effects $\boldsymbol{\mu}$ and random effects $\boldsymbol{\Sigma}_I$ and $\boldsymbol{\Sigma}_S$.

3. Parameter estimation

We estimate the multilevel mixed-effects ODE model by incorporating the hierarchical random effects into the Bayesian ODE framework developed in Huang et al. (2020). Section 3.1 gives a brief description of this Bayesian framework. Then in Section 3.2 we build a multilevel ODE model by incorporating (3) into this framework. A hierarchical Bayesian method for selecting the random effect components in ODE coefficients is also developed in this section. Section 3.3 summarizes the full conditional posterior distributions for all the variables.

3.1. Bayesian ODE parameter estimation approach at individual level

In a general continuous time dynamical system, the evolution of K states $\mathbf{x}(t) \equiv \{x_1(t), \dots, x_K(t)\}$ is described by a set of K ODEs

$$\dot{\mathbf{x}}(t) \equiv \frac{d}{dt} \mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}, t), \quad (4)$$

where $\boldsymbol{\theta} \in \mathbb{R}^d$ is the parameter vector of ODE and $\mathbf{f}(\cdot) = \{f_1(\cdot), \dots, f_K(\cdot)\}$ is a vector of known appropriately smoothing functions. Without loss of generality, assume that the first K_0 states are measured at time points t_1, \dots, t_m . A common situation in practice is that only part of the system is measured, i.e. $K_0 < K$. Model (1) can be fit into this general framework with $d = 3$, $K = 3$, $K_0 = 1$, $x_1 = V$, $x_2 = U$, and $x_3 = I$.

In Huang et al. (2020), \mathbf{x}_k and its derivative were expressed in terms of a basis function expansion

$$x_k(t) = \sum_{j=1}^q c_{kj} \phi_j(t) = \mathbf{c}_k^T \boldsymbol{\phi}(t), \quad \dot{x}_k(t) = \mathbf{c}_k^T \dot{\boldsymbol{\phi}}(t), \quad (5)$$

where $\phi(t) \equiv (\phi_1(t), \dots, \phi_q(t))^T$ is a vector of basis functions such as polynomial spline bases or B-spline bases. Denote $q \times K$ matrix $\mathbf{C} \equiv [\mathbf{c}_1, \dots, \mathbf{c}_K]$ and the observation matrix $\mathbf{Y} \equiv [\mathbf{y}(t_1), \dots, \mathbf{y}(t_m)]$, where $\mathbf{y}(t) \equiv (y_1(t), \dots, y_{K_0}(t))^T$. The likelihood function of the model is

$$P_{OBS}(\mathbf{Y}|\mathbf{C}, \sigma^2) = \prod_{k=1}^{K_0} \prod_{i=1}^m N(y_k(t_i) | \mathbf{c}_k^T \phi(t_i), \sigma_k^2), \quad (6)$$

where $\sigma^2 \equiv \{\sigma_1^2, \dots, \sigma_{K_0}^2\}$. Here we assume that the noise variability σ_k^2 does not depend on the observed state variable and is a constant across different time points. This homoscedastic assumption has been widely used in the literature of ODE parameter estimation. The consistency of the two-stage least square estimator has been established in Liang and Wu (2008) in cases where the true variability in the noise is not necessarily constant. Moreover, for ODE problem (1), if the virus load at a time point is below the threshold, it cannot be precisely measured and usually is assigned a tiny number close to 0. It will cause numerical problem in the analysis if we assume that the variability of noise is proportional to the observed virus load. Therefore, the homoscedasticity assumption we made can also facilitate numerical stability in the analysis. For each σ_k^2 , we employ an inverse Gamma prior

$$P(\sigma_k^2) = \text{Inverse Gamma}(\alpha_\sigma, \beta_\sigma), \quad (7)$$

for some shape and rate hyper-parameters $(\alpha_\sigma, \beta_\sigma)$.

The exact ODE constraint (4) was replaced with the following probability expression which models a normal distribution for the discrepancy between $\dot{\mathbf{x}}(t)$ and $\mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}, t)$.

$$P_{ODE}(\mathbf{C}|\boldsymbol{\theta}, \delta^2) \sim \prod_{k=1}^K \prod_{i=1}^m \frac{1}{\delta_k} \exp \left[-\frac{\{\mathbf{c}_k^T \dot{\phi}(t_i) - f_k(\mathbf{C}^T \phi(t_i), \boldsymbol{\theta}, t_i)\}^2}{2\delta_k^2} \right], \quad (8)$$

where $\delta^2 \equiv \{\delta_1^2, \dots, \delta_K^2\}$. The advantage of the probability (8) for ODE is that in situations where $\mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}, t)$ only depends on \mathbf{x}_k up to the first order of power, the full conditional posterior distribution for \mathbf{c}_k becomes multivariate normal. For each δ_k^2 , we employ an inverse Gamma prior

$$P(\delta_k^2) = \text{Inverse Gamma}(\alpha_\delta, \beta_\delta), \quad (9)$$

for some shape and rate hyper-parameters $(\alpha_\delta, \beta_\delta)$.

3.2. Random effects selection in hierarchical mixed-effect ODE model

Combining all subjects and strains together, the observation model (6) and ODE requirement (8) lead to the following joint distribution of the whole observations over state parameters \mathbf{C}_{sj} , ODE parameters $\boldsymbol{\theta}_{sj}$, observations \mathbf{Y}_{sj} , and the remaining parameters

$$\prod_{s=1}^S \prod_{j=1}^{n_s} P_{OBS}(\mathbf{Y}_{sj} | \mathbf{C}_{sj}, \sigma_{sj}^2) P_{ODE}(\mathbf{C}_{sj} | \boldsymbol{\theta}_{sj}, \delta_{sj}^2) P(\boldsymbol{\theta}_{sj}) P(\delta_{sj}^2) P(\sigma_{sj}^2), \quad (10)$$

where $P(\boldsymbol{\theta})$ is determined by (3), and $P(\boldsymbol{\sigma}^2)$ and $P(\boldsymbol{\delta}^2)$ are properly chosen priors for parameters $\boldsymbol{\sigma}^2$ and $\boldsymbol{\delta}^2$ which are give in (7) and (9) respectively. For those components of \mathbf{C} which are not observed, they are not included in $P_{OBS}(\mathbf{Y}|\mathbf{C}, \boldsymbol{\sigma}^2)$ but can still be estimated because they appear in $P_{ODE}(\mathbf{C}|\boldsymbol{\theta}, \boldsymbol{\delta}^2)$.

To select strain level random effects in (10), we follow Chen and Dunson (2003) and use a modified Cholesky decomposition of $\boldsymbol{\Sigma}_S$: $\boldsymbol{\Sigma}_S = \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T\boldsymbol{\Lambda}$. $\boldsymbol{\Lambda}$ is a positive diagonal matrix with diagonal elements $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ proportional to the random effects standard deviations, so that setting $\lambda_l = 0$ is equivalent to dropping the l th random effect from the model. $\boldsymbol{\Gamma}$ is a lower triangular matrix with diagonal elements equal to 1 and off-diagonal elements describing the random effects correlations. In the case of independent random effects, $\boldsymbol{\Gamma}$ is simply the identity matrix \mathbf{I}_d and the diagonal elements of $\boldsymbol{\Lambda}$ equal to the random effects standard deviations. Applying the covariance decomposition to (3) we have

$$\boldsymbol{\theta}_{sj} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{b}_s + \boldsymbol{\beta}_{sj}, \quad \mathbf{b}_s \sim N(0, \mathbf{I}_d), \quad \boldsymbol{\beta}_{sj} \sim N(0, \boldsymbol{\Sigma}_I), \quad (11)$$

which leads to

$$P(\boldsymbol{\theta}_{sj}) = P(\boldsymbol{\theta}_{sj}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \mathbf{b}_s, \boldsymbol{\Sigma}_I)P(\mathbf{b}_s)P(\boldsymbol{\mu})P(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})P(\boldsymbol{\Sigma}_I), \quad (12)$$

where

$$\begin{aligned} P(\boldsymbol{\theta}_{sj}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \mathbf{b}_s, \boldsymbol{\Sigma}_I) &= N(\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{b}_s, \boldsymbol{\Sigma}_I), \\ P(\mathbf{b}_s) &= N(0, \mathbf{I}_d), \end{aligned}$$

and $P(\boldsymbol{\mu})$, $P(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$, and $P(\boldsymbol{\Sigma}_I)$ are priors for $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Sigma}_I$ respectively.

Motivated by practical considerations, we want to choose priors that facilitate posterior computation. For this reason, prior distributions that are conditionally conjugate are desirable. For $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_I$, we choose the following conjugate priors

$$\boldsymbol{\mu} \sim N(\boldsymbol{\eta}_0, \boldsymbol{\Lambda}_0), \quad \boldsymbol{\Sigma}_I^{-1} \sim \text{Wishart}(\boldsymbol{\Omega}_0, \nu_0), \quad (13)$$

where $\nu_0 > 0$, $\boldsymbol{\eta}_0$ is a d -dimensional vector, $\boldsymbol{\Lambda}_0, \boldsymbol{\Omega}_0$ are $d \times d$ positive definite matrices. Here $\boldsymbol{\eta}_0, \boldsymbol{\Lambda}_0, \boldsymbol{\Omega}_0, \nu_0$ are hyper-parameters and are known.

In choosing priors for $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$, we assume

$$P(\boldsymbol{\Lambda}, \boldsymbol{\Gamma}) = P(\boldsymbol{\lambda}, \boldsymbol{\gamma}) = P(\boldsymbol{\gamma}|\boldsymbol{\lambda})P(\boldsymbol{\lambda}) \propto N(\boldsymbol{\gamma}|\boldsymbol{\gamma}_0, \mathbf{R}_0)I(\boldsymbol{\gamma} \in \mathcal{R}_{\boldsymbol{\lambda}})P(\boldsymbol{\lambda}),$$

where $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_{d(d-1)/2}\}$ are elements of $\boldsymbol{\Gamma}$. Here, the indicator function $I(\cdot)$ sets zero for the elements of $\boldsymbol{\gamma}$ that correspond to zero $\boldsymbol{\lambda}$. Further specify

$$P(\boldsymbol{\lambda}) = \prod_{l=1}^d P(\lambda_l), \quad P(\lambda_l) \sim ZI-N^+(p_{l0}, m_{l0}, s_{l0}^2),$$

where $ZI-N^+(p_{l0}, m_{l0}, s_{l0}^2)$ denotes the density of a zero inflated half normal distribution consisting of a point mass at zero (with probability p_{l0}) and a $N(m_{l0}, s_{l0}^2)$ density truncated below by zero.

3.3. The full conditional posterior distributions

From (10) and (12), the full conditional distributions for all parameters can be obtained as

$$\begin{aligned}
 \mathbf{C}_{sj} &\sim P_{OBS}(\mathbf{Y}_{sj}|\mathbf{C}_{sj}, \boldsymbol{\sigma}_{sj}^2)P_{ODE}(\mathbf{C}_{sj}|\boldsymbol{\theta}_{sj}, \boldsymbol{\delta}_{sj}^2), \\
 \boldsymbol{\theta}_{sj} &\sim P_{ODE}(\mathbf{C}_{sj}|\boldsymbol{\theta}_{sj}, \boldsymbol{\delta}_{sj}^2)P(\boldsymbol{\theta}_{sj}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \mathbf{b}_s, \boldsymbol{\Sigma}_I), \\
 \mathbf{b}_s &\sim \prod_{j=1}^{n_s} P(\boldsymbol{\theta}_{sj}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \mathbf{b}_s, \boldsymbol{\Sigma}_I)P(\mathbf{b}_s), \\
 \boldsymbol{\mu} &\sim \prod_{s=1}^S \prod_{j=1}^{n_s} P(\boldsymbol{\theta}_{sj}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \mathbf{b}_s, \boldsymbol{\Sigma}_I)P(\boldsymbol{\mu}), \\
 \boldsymbol{\Gamma} &\sim \prod_{s=1}^S \prod_{j=1}^{n_s} P(\boldsymbol{\theta}_{sj}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \mathbf{b}_s, \boldsymbol{\Sigma}_I)P(\boldsymbol{\gamma}|\boldsymbol{\lambda}), \\
 \boldsymbol{\Lambda} &\sim \prod_{s=1}^S \prod_{j=1}^{n_s} P(\boldsymbol{\theta}_{sj}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \mathbf{b}_s, \boldsymbol{\Sigma}_I)P(\boldsymbol{\lambda}), \\
 \boldsymbol{\Sigma}_I &\sim \prod_{s=1}^S \prod_{j=1}^{n_s} P(\boldsymbol{\theta}_{sj}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \mathbf{b}_s, \boldsymbol{\Sigma}_I)P(\boldsymbol{\Sigma}_I), \\
 \boldsymbol{\sigma}_{sj}^2 &\sim P_{OBS}(\mathbf{Y}_{sj}|\mathbf{C}_{sj}, \boldsymbol{\sigma}_{sj}^2)P(\boldsymbol{\sigma}_{sj}^2), \\
 \boldsymbol{\delta}_{sj}^2 &\sim P_{ODE}(\mathbf{C}_{sj}|\boldsymbol{\theta}_{sj}, \boldsymbol{\delta}_{sj}^2)P(\boldsymbol{\delta}_{sj}^2).
 \end{aligned}$$

Using priors (13), the full conditional distributions of \mathbf{b}_s and $\boldsymbol{\mu}$ are normal, $\mathbf{b}_s \sim N(\bar{\boldsymbol{\mu}}_s, \bar{\boldsymbol{\Sigma}}_s)$, $\boldsymbol{\mu} \sim N(\bar{\boldsymbol{\mu}}_\mu, \bar{\boldsymbol{\Sigma}}_\mu)$, where

$$\begin{aligned}
 \bar{\boldsymbol{\mu}}_s &= (n_s \boldsymbol{\Gamma} \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}_I^{-1} \boldsymbol{\Lambda} \boldsymbol{\Gamma} + \mathbf{I}_d)^{-1} \left\{ \boldsymbol{\Sigma}_I^{-1} \sum_{j=1}^{n_s} (\boldsymbol{\theta}_{sj} - \boldsymbol{\mu}) \right\}, \\
 \bar{\boldsymbol{\Sigma}}_s &= (n_s \boldsymbol{\Gamma} \boldsymbol{\Lambda}^T \boldsymbol{\Sigma}_I^{-1} \boldsymbol{\Lambda} \boldsymbol{\Gamma} + \mathbf{I}_d)^{-1}, \\
 \bar{\boldsymbol{\mu}}_\mu &= (n \boldsymbol{\Sigma}_I^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1} \left\{ \boldsymbol{\Sigma}_I^{-1} \sum_{s=1}^S \sum_{j=1}^{n_s} (\boldsymbol{\theta}_{sj} - \boldsymbol{\Lambda} \boldsymbol{\Gamma} \mathbf{b}_s) + \boldsymbol{\Lambda}_0^{-1} \boldsymbol{\eta}_0 \right\}, \\
 \bar{\boldsymbol{\Sigma}}_\mu &= (n \boldsymbol{\Sigma}_I^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1}.
 \end{aligned}$$

The full conditional distribution of $\boldsymbol{\Sigma}_I^{-1}$ is Wisharts, $\boldsymbol{\Sigma}_I^{-1} \sim \text{Wishart}(\boldsymbol{\Omega}_I^{-1}, n + \nu_0)$, where

$$\boldsymbol{\Omega}_I = \boldsymbol{\Omega}_0^{-1} + \sum_{s=1}^S \sum_{j=1}^{n_s} (\boldsymbol{\theta}_{sj} - \boldsymbol{\mu} - \boldsymbol{\Lambda} \boldsymbol{\Gamma} \mathbf{b}_s)(\boldsymbol{\theta}_{sj} - \boldsymbol{\mu} - \boldsymbol{\Lambda} \boldsymbol{\Gamma} \mathbf{b}_s)^T.$$

The full conditional distributions of $\boldsymbol{\sigma}_{sj}^2$ and $\boldsymbol{\gamma}_{sj}^2$ are inverse Gammas. The full conditional distribution of $\boldsymbol{\gamma}$ given $\boldsymbol{\Lambda}$ and \mathbf{b}_s has a multivariate normal distribution. The full conditional distribution of $\boldsymbol{\lambda}_l$ is zero inflated half normal. The full conditional distributions of \mathbf{X} and

θ depend on the explicit form of function $\mathbf{f}(\cdot)$. For situations where $\mathbf{f}(\mathbf{x}(t), \theta, t)$ only depends on each argument up to the first order of power, we can derive a closed form full conditional posterior distribution for θ and each component of \mathbf{X} . For model (1), $\mathbf{f}(\cdot)$ is a linear function of both θ and all the components of \mathbf{X} , the full conditional distributions for θ_{sj} and the components of \mathbf{C}_{sj} are all multivariate normal. Thus we can use closed form Gibbs sampling scheme in the MCMC process.

4. Simulation Study

In this section, we analyze data from simulated studies to illustrate our approach. The design of simulations is similar to the motivation experiment of influenza infections in ducks described in Section 1. In particular, we consider 6 strains and generate data for 36 subjects, 6 for each strain. For each subject, we assume that measurements of viral load are taken at every 0.5 time unit on the interval $[1, 14]$, i.e. we have 27 virus load measurements for each subject. We let the population-level dynamic parameters $\mu = (9, 2/3, 2/3)$. The individual dynamic parameters θ_{sj} for the j th subject in the s th strain are generated by (11) with the individual level covariance matrix $\Sigma_I = \text{diag}(0.9, 0.1, 0.1)$ and the strain level covariance matrix Σ_s being the product of Λ and Γ , where Λ takes

$$\Lambda_1 = \begin{pmatrix} 9/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Lambda_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}, \quad \Lambda_3 = \begin{pmatrix} 9/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{pmatrix},$$

and Γ takes

$$\Gamma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \Gamma_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.5 & 0.5 & 1 \end{pmatrix}.$$

Therefore, we have 6 different settings for Σ_s . The random variation at strain level occurs on parameter a for Λ_1 , on parameters b and c for Λ_2 , and on all parameters a , b , and c for Λ_3 . The parameters a , b , and c are generated independently for Γ_1 , but for Γ_2 , the correlation is considered.

We use the initial condition $U_0 = 11$, $I_0 = 0$, and $V_0 = 0.1$. The trajectories of $U(t)$, $I(t)$, and $V(t)$ based on $\mu = (9, 2/3, 2/3)$ and this initial condition is shown in Figure 2. Then, based on generated true parameters θ_{sj} , the observations y_{sjt} are generated by perturbing the solution of the differential equation (1) with a within-subject measurement error, i.e., $y_{sjt} = V_{sj}(t_i) + \epsilon_{sjt}$, where $V_{sj}(t_i)$ is the numerical solution of viral load to the differential equation (1) for the j th subject in the s th strain at time t_i . It is assumed that the within-subject measurement error ϵ_{sjt} is normally distributed with $N(0, 0.1^2)$. For each setting, we replicate the simulation 100 times. Thus our conclusion is based on the summary over analyses of 100 simulated data sets.

We applied the Gibbs sampling algorithm described in Section 3.3 to all the simulated data and test the performance by comparing the estimated dynamic parameters with the true dynamic parameters. We choose hyper-parameters $\eta_0 = 0$, $\Lambda_0 = \Omega_0 = 10^5 \mathbf{I}_3$, $\alpha_\sigma = \alpha_\delta = \beta_\sigma = \beta_\delta = \nu_0 = 10^{-5}$ to reflect the non-informative priors for μ , Σ_I , σ^2 ,

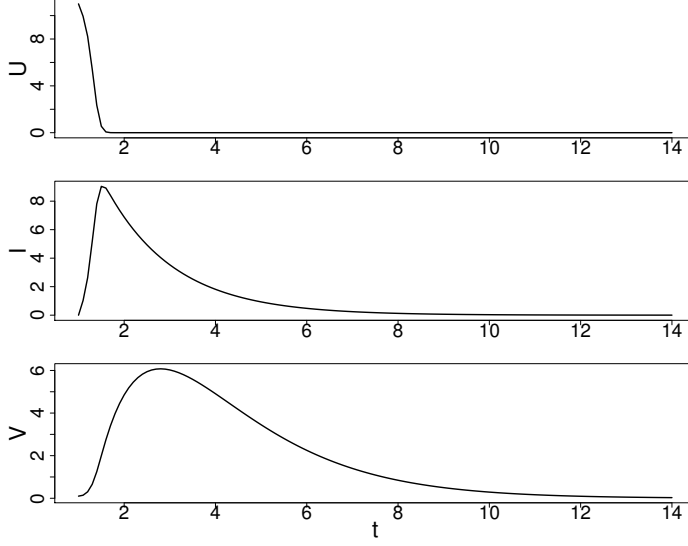


Fig. 2. The trajectories of $U(t)$, $I(t)$, and $V(t)$ as a solution to ODE (1) using parameters $a = 9$, $b = 2/3$, $c = 2/3$, and the initial condition $U_0 = 11$, $I_0 = 0$, and $V_0 = 0.1$.

and δ^2 . The hyper-parameters for λ_l are chosen to be $m_{l0} = 0$ and $s_{l0}^2 = 100$. The hyper-parameters for the elements of Γ are chosen to be $\gamma_0 = 0$ and $\mathbf{R}_0 = \mathbf{I}_3$. To study the effect of the prior probability that $\lambda_l = 0$, we use $p_{l0} = 0.2, 0.5$, and 0.8 , respectively.

To fit the non-parametric model, Cubic B-splines with 25 equally-spaced knots in $[1, 14]$ are used to approximate the ODE solution. To choose appropriate initial values for the elements of θ_{sj} and \mathbf{C}_{sj} , we solve the least square problem for $\text{RSS}_{sj} = \sum_{i=1}^{m_{sj}} (y_{sji} - V_{sj}(t_i))^2$ which are function of the ODE parameters θ_{sj} and initial condition U_0, I_0, V_0 . To ensure the identifiability of this optimization problem, we fix $I_0 = 0$, $V_0 = y_{sj1}$ and obtain U_0, θ_{sj} by minimizing RSS_{sj} . Then the estimated values for θ_{sj} are taken to be the initial values for θ_{sj} and the estimated values for $U(t_i), I(t_i), V(t_i)$ are used to derive the initial values for \mathbf{C}_{sj} in our MCMC procedure.

For the analysis of each data set, the MCMC sampler was run for a total of 25,000 cycles. The first 5000 cycles were discarded as burn-in, and the remainder of the chain was thinned by keeping one out of every ten samples, resulting in a total of 2000 samples for post-MCMC analysis. To compute the posterior probabilities of each of the 2^3 models for strain level random effects, we simply add up the number of occurrences of each model and divide by the total number of iterations.

Table 4 shows the summary of posterior probabilities over 100 simulated data for 8 different strain level random effect models. There is overwhelming evidence that the true model is selected with the highest posterior probabilities. The only exception is for setting $\Sigma_s = \Lambda_2 \Gamma_2$, where the model (a, b, c) has the largest posterior probability while the true specification, i.e. the ODE mixed model with correlated strain level random effects on b and c , has the second largest posterior probability. This conclusion holds regardless of the choice for the prior probability of $\lambda_l = 0$. One possible explanation is that the correlation between a and (b, c) causes the inclusion of parameter a in the model and thus the higher

Table 1. Summary of posterior probabilities over 100 simulation studies for 8 different strain level random effect models. The subset of ODE coefficients that have the strain level random effects are listed. The bold number denotes the largest posterior probability and the underline denotes the true specification.

p_{l0}	Σ_S	null	(a)	(b)	(c)	(a,b)	(a,c)	(b,c)	(a,b,c)
0.2	$\Lambda_1\Gamma_1$	0.008	<u>0.713</u>	0.003	0.007	0.099	0.116	0.006	0.048
	$\Lambda_1\Gamma_2$	0.008	<u>0.703</u>	0.003	0.012	0.102	0.110	0.009	0.054
	$\Lambda_2\Gamma_1$	0.005	0.004	0.050	0.116	0.050	0.076	<u>0.450</u>	0.249
	$\Lambda_2\Gamma_2$	0.008	0.041	0.014	0.044	0.161	0.082	<u>0.208</u>	0.443
	$\Lambda_3\Gamma_1$	0.004	0.024	0.005	0.014	0.032	0.162	0.026	<u>0.735</u>
	$\Lambda_3\Gamma_2$	0.004	0.056	0.012	0.007	0.099	0.145	0.032	<u>0.647</u>
0.5	$\Lambda_1\Gamma_1$	0.025	<u>0.799</u>	0.003	0.011	0.067	0.077	0.004	0.015
	$\Lambda_1\Gamma_2$	0.029	<u>0.793</u>	0.004	0.014	0.065	0.075	0.006	0.015
	$\Lambda_2\Gamma_1$	0.018	0.032	0.077	0.188	0.050	0.057	<u>0.451</u>	0.127
	$\Lambda_2\Gamma_2$	0.010	0.092	0.025	0.092	0.177	0.115	<u>0.217</u>	0.272
	$\Lambda_3\Gamma_1$	0.014	0.046	0.014	0.027	0.074	0.205	0.037	<u>0.584</u>
	$\Lambda_3\Gamma_2$	0.016	0.117	0.025	0.019	0.109	0.187	0.023	<u>0.504</u>
0.8	$\Lambda_1\Gamma_1$	0.054	<u>0.847</u>	0.005	0.015	0.034	0.034	0.006	0.004
	$\Lambda_1\Gamma_2$	0.044	<u>0.865</u>	0.003	0.012	0.035	0.033	0.004	0.004
	$\Lambda_2\Gamma_1$	0.045	0.036	0.173	0.246	0.061	0.022	<u>0.359</u>	0.058
	$\Lambda_2\Gamma_2$	0.051	0.133	0.052	0.110	0.157	0.040	<u>0.192</u>	0.262
	$\Lambda_3\Gamma_1$	0.037	0.064	0.021	0.042	0.115	0.208	0.028	<u>0.485</u>
	$\Lambda_3\Gamma_2$	0.028	0.161	0.049	0.030	0.098	0.150	0.050	<u>0.435</u>

Table 2. Posterior estimates of population-level ODE dynamic parameters based on 100 simulated data sets for 6 different choices of strain level random effect covariance matrix. For each parameter, the first, second, and third rows denote its mean, standard error, and ARE, respectively. Here $p_{l0} = 0.5$ and the true parameters $a = 9$, $b = 2/3$, $c = 2/3$.

	$\Lambda_1\Gamma_1$	$\Lambda_1\Gamma_2$	$\Lambda_2\Gamma_1$	$\Lambda_2\Gamma_2$	$\Lambda_3\Gamma_1$	$\Lambda_3\Gamma_2$
a	7.69	7.75	7.62	6.80	7.29	6.85
	1.43	1.50	0.47	0.92	1.47	1.59
	0.18	0.18	0.15	0.24	0.21	0.26
b	0.73	0.73	0.73	0.76	0.78	0.74
	0.05	0.05	0.13	0.21	0.23	0.17
	0.09	0.09	0.17	0.22	0.23	0.21
c	0.66	0.68	0.72	0.60	0.84	0.68
	0.11	0.16	0.18	0.22	0.29	0.20
	0.07	0.08	0.21	0.25	0.33	0.21

probability for model (a, b, c) . We further studied the marginal probabilities from Table 4 for the true parameters to be included in the model and obtained the consistent results. For example, under $p_{l0} = 0.5$, the probabilities for including random effects of (a) , (b, c) , and (a, b, c) in the model are 95.7%, 57.8%, and 58.4% respectively for settings $\Lambda_1\Gamma_1$, $\Lambda_2\Gamma_1$, and $\Lambda_3\Gamma_1$. This shows that the proposed method obtains good estimates for strain level random effects.

To evaluate the performance of our method in different scenarios, we define the average relative estimation error (ARE) of a parameter θ as

$$\text{ARE} = \frac{1}{N} \sum_{i=1}^N \frac{|\hat{\theta}_i - \theta|}{|\theta|},$$

where $\hat{\theta}_i$ is the estimate of θ in the i -th simulation and N is the number of simulation runs (here $N = 100$). Table 4 presents the posterior summaries for the fixed effect ODE parameters over 100 simulated samples including mean, standard error, and ARE, which are properties associated with the population level. Here we choose non-informative prior $p_{l0} = 0.5$. It can be seen that the estimated values are quite close to the true values, suggesting good performance of the algorithm. Parameter a is slightly underestimated. The reason is that a is mainly determined by the decay rate of uninfected cell U . As shown in the upper panel of Figure 2, U decays very fast and quickly becomes flat. Therefore, only the measurements at very beginning time are useful in the estimation of a . We can increase the accuracy of parameter estimation for a by including more measurements and choosing more knots at the beginning. But in practice, only virus load V is observed, we can only choose knots according to the patterns of virus load changes as described in Section 5.

We have also conducted simulations to study how the performance of our algorithm is influenced by the number of subjects per strain, i.e. n_s . We found that increasing n_s can reduce the bias in the estimation of population level ODE parameters. For example, under setting $\Lambda_1\Gamma_1$ with $p_{l0} = 0.5$, if the number of subjects per strain $n_s = 20$, we obtain that the AREs are 0.13, 0.08, and 0.05 for the fixed effect coefficients a , b , and c respectively, which are smaller than the corresponding values of 0.18, 0.09, and 0.07 obtained based on $n_s = 6$. We also studied the effect of the number of virus load measurements m for each subject and found that the conclusion is pretty similar, i.e. the bias for the estimation of the fixed effect coefficients decreases with m .

5. Application to the experiment of influenza infections in ducks

In this section, we apply the proposed methodology to the data set of viral load dynamics following influenza infections in ducks as shown in Figure 1. Cubic B-splines with 10 equally-spaced knots between day 1 and day 4 and 10 equally-spaced knots between day 4 and day 14 are used to approximate the ODE solution. The knots are chosen in this way in order to capture the patterns of virus load changes which increase sharply from day 1 to day 3 followed by a slow and continuous decrease until day 14. We have tried different number of knots in the analysis and obtained very similar results. For example, if we choose 7 knots between day 1 and day 4 and 5 knots between day 4 and day 14,

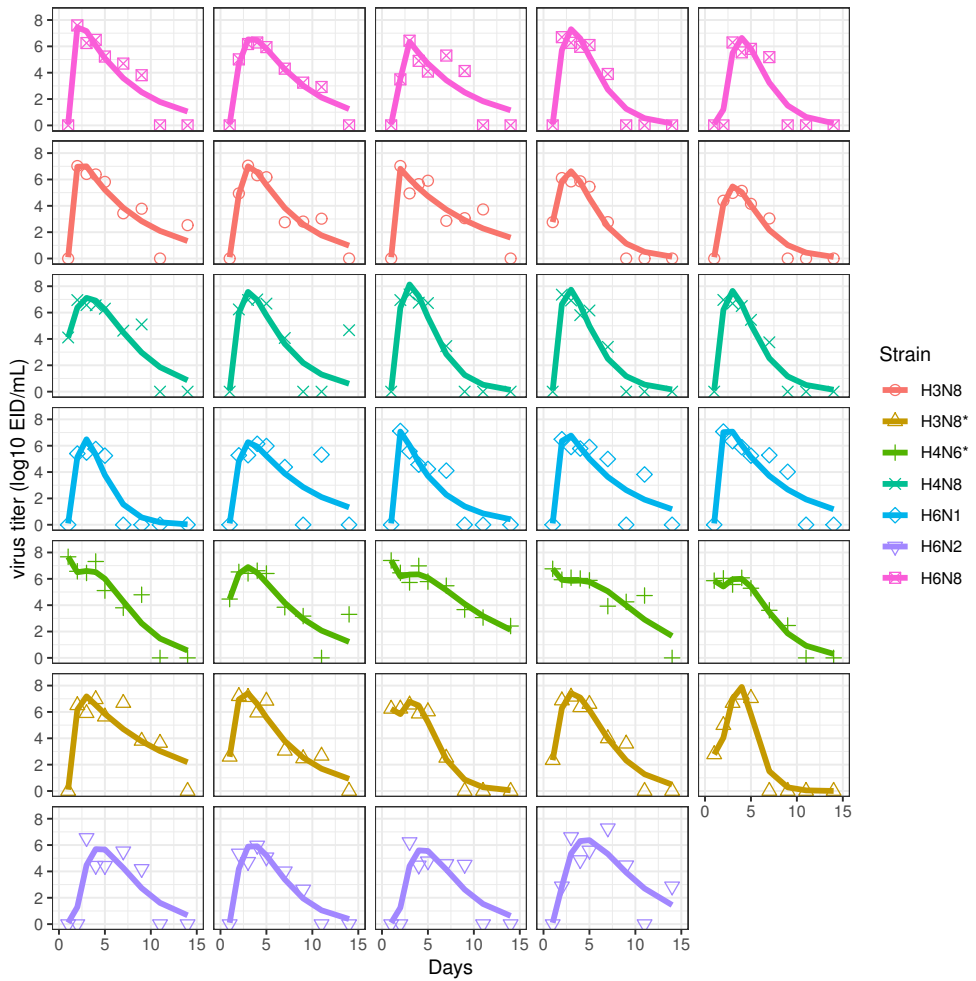


Fig. 3. Fits of the ODE models to the virus load data for each infected duck. The solid lines represent the ODE trajectories based on estimated coefficients for each individual using the Bayesian hierarchical mixed-effect model.

Table 3. Posterior probabilities of 8 different strain level random effect models in the duck influenza applications. The subset of ODE coefficients that have nonzero strain level random effects are listed.

Model	$p_{l0} = 0.2$	$p_{l0} = 0.5$	$p_{l0} = 0.8$
No random effects	0.090	0.300	0.079
(a)	0.451	0.531	0.420
(b)	0.037	0.043	0.050
(c)	0.029	0.032	0.033
(a, b)	0.18	0.042	0.19
(b, c)	0.016	0.006	0.018
(a, c)	0.152	0.045	0.170
(a, b, c)	0.052	0.005	0.045

the conclusion is pretty similar for both visualization and parameter estimation. Since no previous information is available about the strain level variation in duck flu, we choose non-informative prior $p_{l0} = 0.5$ on the set $\lambda_l = 0$ for $l = 1, 2, 3$ to reflect a balance among outcomes. For all the other hyper-parameters, we use the same values as in the simulation study to induce relatively diffuse priors on those parameters. We also follow the similar way as in the simulation study for choosing the initial values of θ_{sj} and C_{sj} . We discard the first 5000 iterations as burn-in. Posterior probabilities are calculated based on 20,000 iterations collected thereafter. To compute posterior summaries of the parameters, we thin the chain by a factor of 10.

Figure 3 displays the individual fitted curves based on the solutions of the viral load V to the ODE equations. The virus load of the j th subject in the s th strain is estimated from ODE using parameters $\hat{\theta}_{sj}$ which is taken as the average of the 2000 posterior samples for θ_{sj} . We found that the model provided a good fit to the observed data for all of the subjects.

To assess sensitivity to prior model probability, the analysis was repeated for $p_{l0} = 0.2$ and $p_{l0} = 0.8$, respectively. Table 5 lists the posterior probabilities of the 8 sub-models. There is evidence that the ODE mixed model which includes between-strain random variations for the parameter a alone is optimal. This conclusion holds for three choices of the prior probability for $\lambda_l = 0$. The marginal probabilities of including the strain level random effects for the parameter a are 83%, 62%, and 82%. In contrast, the corresponding marginal probabilities are 27.9%, 9.4%, 29.85% for including the parameter b and 24.75%, 8.6%, 26.6% for including the parameter c . This indicates that the cells are infected at different rates for different strains, but both the death rates of the infected cell and the clearance rates of the free virus show strong homogeneity across different strains.

In addition to investigating differences among strains, investigators are interested in assessing the overall population-level effect of the dynamic parameters. The population posterior means and the corresponding standard error as well as the 95% highest density interval (HDI) for the three parameters are summarized in Table 5. This result is also robust to the various prior choices. Table 5 displays the posterior estimation of the dynamic parameters for each strain, i.e. $\mu + \alpha_s$ in (3). It can be seen that, for parameters b and c , there are no distinct between-strain differences. But for parameter a , the between-strain differences are quite clear especially for sub-types H4N8 and H4N6* whose values are

Table 4. Posterior estimates of population-level dynamic parameters including mean, standard error (SE), and 95% highest density interval (HDI).

	a	b	c
Mean	8.01	1.11	0.93
SE	2.11	0.26	0.34
HDI	(3.47,12.01)	(0.65,1.62)	(0.30,1.64)

Table 5. Posterior mean and standard error (in parenthesis) of ODE parameters for each individual strain including fixed effect and strain-specific random effect.

Strain	a	b	c
H6N8	8.90 (2.39)	1.15 (0.30)	0.91 (0.38)
H3N8	7.89 (2.30)	1.11 (0.27)	1.01 (0.46)
H4N8	10.41 (3.21)	1.10 (0.28)	0.90 (0.38)
H6N1	7.94 (2.22)	1.12 (0.27)	0.95 (0.37)
H4N6*	4.83 (3.77)	1.06 (0.33)	0.89 (0.40)
H3N8*	7.44 (2.40)	1.16 (0.30)	0.95 (0.37)
H6N2	8.56 (2.52)	1.08 (0.31)	0.90 (0.39)

much larger and smaller respectively than the values of the remaining sub-types. This result is consistent with the posterior probability pattern shown in Table 5.

Since the seven sub-types are identified according to the genomic characterization, our random effect model selection analysis shows that the genetic differences do have some effects on the transmission dynamics of IAV through changing the infection rate parameter a . As our analysis was based on a limited number of tested viruses, such a finding should be reinforced with additional studies focusing on other virus genotypes, at other locations and in different avian hosts. This result nevertheless provides some insight to help biologist for further investigation.

6. Discussion

In this paper, we propose a Bayesian multilevel mix-effects ODE model to describe the dynamics of influenza virus infection. Both the population and individual dynamic parameters can be estimated. In comparison with fixed-effect modeling, mixed-effects modeling has the advantages of obtaining more precise parameter estimation by pooling all data together. By relaxing the ODE constraint with a probability expression, our method does not need to solve ODE directly. The variations at different levels are simultaneously considered by incorporating both strain-specific and subject-specific random effects for each ODE coefficient. One advantage of the proposed method is that the closed form conditional posterior distributions for all variables and parameters can be obtained which substantially facilitate the convergence process. To identify the variation at strain level, we perform random-effects model selection based on a decomposition of the covariance matrix of the strain-specific random-effects distribution. The decomposition enables us to specify a prior distribution which can exclude one or more random effects by setting their variances to

zero. This reparameterization of the covariance matrix results in straightforward and efficient posterior computation via a Gibbs sampler.

We have presented large scale simulation examples and an actual long-term surveillance study of IAV circulation in wild birds to illustrate how the Bayesian procedures can be applied to influenza virus dynamics investigation. For the simulation studies, it was seen that the models provided fairly good fits to both ODE random effects and ODE fixed effects. The estimates for both population and individual dynamic parameters are also very close to the true values. For the real data, the proposed model fitted the observed viral load reasonably well for all subjects in our study. Using random effects variable selection, we identify that the cell infection rate a varies among different strains while the other two ODE parameters are quite homogeneous across the strains.

We use a simplified model to explain the observed patterns and characterize the biological mechanisms of IAV infection with the main goals of retaining crucial features of influenza dynamics. We plan to generalize it to more complicated models with consideration of more infected cell and virus compartments in order to provide more accurate descriptions for long-term IAV dynamics. One challenge is the identifiability problem of model parameters due to the complexity of the models and the factor that only viral loads are observed among ODE variables. We need to consider the trade-off between the complexity and applicability of influenza dynamic models. It has been shown in Lebarbenchon et al. (2011) that the environmental factors, particularly temperature and humidity, appear to be important determinant of IAV persistence and therefore within-host fitness. We will consider modeling ODE parameters as function of environmental factors in order to integrate data and models at both the infection dynamics and the persistence of the virus in the environment within a combined ecological and within-host framework.

Acknowledgment

The author thanks the editor, associate editor, and two referees for many helpful comments and suggestions which led to a much improved presentation. The author also thanks Andreas Handel for sharing the dataset and Xiao Song for helpful discussion. This research is supported by Division of Mathematical Sciences (National Science Foundation) grant DMS-1916411.

References

- Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G., and Perelson, A. S. (2006). Kinetics of influenza a virus infection in humans. *Journal of virology* **80**, 7590–7599.
- Beal, S. and Sheiner, L. (1980). The NONMEM system. *The American Statistician* **34**, 118–119.
- Campbell, D. and Steele, R. J. (2012). Smooth functional tempering for nonlinear differential equation models. *Statistics and Computing* **22**, 429–443.
- Chen, Z. and Dunson, D. B. (2003). Random Effects Selection in Linear Mixed Models. *Biometrics* **59**, 762–769.

- Dondelinger, F., Filippone, M., Rogers, S., and Husmeier, D. (2013). ODE parameter inference using adaptive gradient matching with gaussian processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pages 216–228.
- Handel, A., Lebarbenchon, C., Stallknecht, D., and Rohani, P. (2014). Trade-offs between and within scales: environmental persistence and within-host fitness of avian influenza viruses. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20133051.
- Huang, H., Handel, A., and Song, X. (2020). A new Bayesian approach to estimate parameters of ordinary differential equation. *Computational Statistics* **35**, 1481–1499.
- Huang, Y., Liu, D., and Wu, H. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal hiv dynamic system. *Biometrics* **62**, 413–423.
- Huang, Y. and Wu, H. (2006). A Bayesian approach for estimating antiviral efficacy in hiv dynamic models. *Journal of Applied Statistics* **33**, 155–174.
- Keeler, S. P., Lebarbenchon, C., and Stallknecht, D. E. (2013). Strain-related variation in the persistence of influenza A virus in three types of water: distilled water, filtered surface water, and intact surface water. *Virology Journal* **10**, 13.
- Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Stat. Data Anal.* **49**, 1020–1038.
- Lebarbenchon, C., Sreevatsan, S., Lefvre, T., Yang, M., Ramakrishnan, M. A., Brown, J. D., and Stallknecht, D. E. (2012). Reassortant influenza A viruses in wild duck populations: effects on viral shedding and persistence in water. *Proc Biol Sci* **279**(1744), 3967–3975.
- Lebarbenchon, C., Yang, M., Keeler, S. P., Ramakrishnan, Muthannan A. Brown, J. D., Stallknecht, D. E., and Sreevatsan, S. (2011). Viral replication, persistence in water and genetic characterization of two influenza A viruses isolated from surface lake water. *PLoS One* **6**(10), e26566.
- Lebarbenchon, C., Yang, M., Keeler, S. P., Ramakrishnan, M. A., Brown, J. D., Stallknecht, D. E., and Sreevatsan, S. (2011). Viral replication, persistence in water and genetic characterization of two influenza A viruses isolated from surface lake water. *PLOS ONE* **6**, 1–8.
- Liang, H. and Wu, H. (2008). Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association* **103**, 1570–1583.
- Liu, B., Wang, L., Nie, Y., and Cao, J. (2019). Bayesian inference of mixed-effects ordinary differential equations models using heavy-tailed distributions. *Computational Statistics & Data Analysis* **137**, 233–246.
- Lixoft (2012). Monolix 4.2. <http://www.lixoft.eu/products/monolix/product-monolix-overview>.

- Macdonald, B., Niu, M., Rogers, S., Filippone, M., and Husmeier, D. (2016). Approximate parameter inference in systems biology using gradient matching: a comparative evaluation. *BioMedical Engineering OnLine* **15**, 80.
- Miao, H., Dykes, C., Demeter, L. M., and Wu, H. (2009). Differential equation modeling of hiv viral fitness experiments: Model identification, model selection, and multimodel inference. *Biometrics* **65**, 292–300.
- Nowak, M. and May, R. M. (2000). *Virus dynamics: mathematical principles of immunology and virology: mathematical principles of immunology and virology*. Oxford University Press, UK.
- Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 741–796.
- Smith, A. M. and Perelson, A. S. (2011). Influenza a virus infection kinetics: quantitative data and models. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **3**.
- Wang, L., Cao, J., Ramsay, J., Burger, D., Laporte, C., and Rockstroh, J. (2014). Estimating mixed-effects differential equation models. *Statistics and Computing* **24**, 111–121.
- Wilcox, B. R., Knutsen, G. A., Berdeen, J., Goekjian, V., Poulson, R., Goyal, S., Sreevatsan, S., Cardona, C., Berghaus, R. D., Swayne, D. E., Yabsley, M. J., and Stallknecht, D. E. (2011). Inuenza-a viruses in ducks in northwestern minnesota: ne scale spatial and temporal variation in prevalence and subtype diversity. *PLoS One* **6(9)**, e24010.
- Wu, H., Xue, H., and Kumar, A. (2012). Numerical discretization-based estimation methods for ordinary differential equation models via penalized spline smoothing with applications in biomedical research. *Biometrics* **68**, 344–352.
- Zhang, X., Cao, J., and Carroll, R. J. (2015). On the selection of ordinary differential equation models with application to predator-prey dynamical models. *Biometrics* **71**, 131–138.