pp. 1–12 (2022)

Application of minimum description length criterion to assess the complexity of models in mathematical immunology

D. S. Grebennikov[*†‡] V. V. Zheltkova [*†§] G. A. Bocharov [*†§]

Abstract — Mathematical models in immunology differ enormously in the dimensionality of the state space, the number of parameters and the parameterizations used to describe the immune processes. The ongoing diversification of the models needs to be complemented by rigorous ways to evaluate their complexity and select the parsimonious ones in relation to the data available/used for their calibration. A broadly applied metrics for ranking the models in mathematical immunology with respect to their complexity/parsimony is provided by the Akaike information criterion. In the present study, a computational framework is elaborated to characterize the complexity of mathematical models in immunology using a more general approach, namely, the Minimum Description Length criterion. It balances the model goodness-of-fit with the dimensionality and geometrical complexity of the model. Four representative models of the immune response to acute viral infection formulated with either ordinary or delay differential equations are studied. Essential numerical details enabling the assessment and ranking of the viral infection models include: (1) the optimization of the likelihood function, (2) the computation of the model sensitivity functions, (3) the evaluation of the Fisher information matrix and (4) the estimation of multidimensional integrals over the model parameter space.

Keywords: Model selection, information criteria, minimum description length, geometric complexity, maximum likelihood estimates, delay differential equations

MSC 2010: 92-08

Mathematical modelling of infectious diseases in humans and experimental animal systems is a rapidly expanding area of applied mathematics [7]. The recently observed strong boost in infection modelling is driven by the SARS-CoV-2 pandemic which is a medical problem for the human population worldwide. Mathematical models are used to link various phenotypes of infection dynamics with the parameters of virus spreading and immune responses as well as to analyze plausible ther-

[*]Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow 119333

[†]Moscow Center for Fundamental and Applied Mathematics at INM RAS

[‡]World-Class Research Center ‘Digital Biodesign and Personalized Healthcare’, Sechenov First Moscow State Medical University, Moscow 119991, Russia

[§]Institute of Computer Science and Mathematical Modelling, Sechenov First Moscow State Medical University, Moscow 119991, Russia

E-mails: dmitry.ew@gmail.com, valeryaaziattseva@ya.ru, g.bocharov@inm.ras.ru

This work was financed by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers ‘Digital biodesign and personalized healthcare’ No. 075-15-2022-304.

apies ahead of their clinical applications [17]. A comprehensive set of references and discussion can be found in [11].

The collection of the so far developed mathematical models consists of a broad set of systems including systems of ODEs, reaction-diffusion-type PDEs and hybrid models. The models differ enormously in the specific parameterizations used to describe specific processes, the dimensionality of the state space and the number of parameters. These features determine a trait known as the model complexity. The ongoing model expansion needs to be complemented by rigorous ways to evaluate their complexity and select the parsimonious ones in relation to the data available/used for their calibration. These days, a broadly applied metrics for ranging the models in mathematical immunology with respect to their complexity is provided by Akaike information criterion [3, 5]. A more general framework, i.e., the Minimum Description Length (MDL) principle [20], has not yet become a universal tool for model evaluation and selection. Some applications of the MDL principle to ranking simple models of immune cell viability can be found in [8].

The Fisher information (FI) is a key element in the analysis of the model complexity and the estimated parameter uncertainty. We discuss in detail the numerical aspects of computing the FI and the MDL in Section 2. A first systematic application of the MDL principle to the analysis of four representative models of the immune response to acute viral infection formulated with ordinary- and delay differential equations (respectively, ODEs and DDEs) is presented in Section 3. Finally, the challenges and future direction of the computational assessment of the model complexity using the MDL principle are discussed in Section 4.

1. Methods

The models of the experimental virus infection analyzed in this paper are based on either ordinary differential equations (ODEs) or delay differential equations (DDEs), and have the following general structure:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}y(t,p) = f(y(t,p),y(t-\tau,p),p) \equiv f(y,y_\tau,p), \quad t \in (0,T) \\ &\quad y(t,p) \equiv h_0, \quad t \in [-\tau,0), \quad y(0,p) \equiv y_0 \end{aligned} \tag{1.1}$$

where state vector $y = [y^1,\dots,y^M] \in \mathbb{R}^M$ is a model solution which we are interested to obtain during the time course of infection $(0,T)$, $p = [p_1,\dots,p_{L-1},\tau] \in \mathbb{R}^L$ is a vector of model parameters, which includes a single delay $\tau \geqslant 0$ ($\tau = 0$ corresponds to the ODE model), h_0 and y_0 are a constant history function and an initial value, respectively. More generally, we could consider any piecewise-continuous initial function $h_0(t,p)$ and a finite number of constant delays τ_k, but for simplicity we will stick with form (1.1).

1.1. Estimation of model parameters

Given a set of experimental data $\{\widehat{y}_j^i\}$ for model variables $i = 1,\dots,M$ at time moments $t_j \in (0,T)$, $j = 1,\dots,N$, one can formulate an inverse problem to estimate the model parameters. This typically involves minimization in the parameter space of some objective function $\Phi(p)$ which measures the discrepancy between the experimental data and the model solution. A classical example of the objective function is the one used in the ordinary least squares (OLS) method:

$$\Phi_{\mathrm{OLS}}(p) = \Phi_{\mathrm{OLS}}(p,\widehat{y}) = \sum_{j=1}^{N} \|y(t_j,p) - \widehat{y}_j\|^2. \tag{1.2}$$

More general framework for obtaining the point estimates for model parameters is the maximum likelihood approach. Instead of minimizing the discrepancy, maximum likelihood estimation (MLE) involves maximizing the likelihood function, i.e., the function which defines the probability to observe the experimental data with a given model (a model solution with given parameters). The formulation of the likelihood function depends on the assumptions about the distribution of the noise in experimental data. Note that the OLS estimates and MLE are equivalent under the following assumptions [3]:

(a1) the errors between the true model solution and experimental data are independent at successive time moments t_j;

(a2) the errors in the components of $\widehat{y}_j$ are independent;

(a3) the errors have the Gaussian distribution with the constant variance σ^2:

$$\widehat{y}_j \sim N(y(t_j,p),\Sigma_j), \quad \Sigma_j = \mathrm{diag}\{\sigma,\dots,\sigma\}.$$

Under assumptions (a1)–(a3), the likelihood function is defined as

$$L(p) = L(\widehat{y}|p) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}||y(t_j,p) - \widehat{y}_j||^2\right). \tag{1.3}$$

The logarithm of likelihood function (log-likelihood function) is linked with the objective function in the following way:

$$\log L(p) = -\frac{1}{2} n_{\mathrm{obs}} \log(2\pi) - \frac{1}{2} n_{\mathrm{obs}} \log(\sigma^2) - \frac{1}{2\sigma^2} \Phi_{\mathrm{OLS}}(p) \tag{1.4}$$

where n_{obs} is the total number of experimental data points ($n_{\mathrm{obs}} = MN$ for complete data sets).

The variance σ^2 can be estimated using the maximum likelihood method as well. The estimate [3] is given by

$$\widehat{\sigma}^2 = \frac{1}{n_{\mathrm{obs}}} \Phi_{\mathrm{OLS}}(\widehat{p}) \tag{1.5}$$

where $\widehat{p} = \arg\min_p \Phi_{\mathrm{OLS}}(p)$ is the vector of identified parameters. We will use this estimate in the likelihood function (1.3), which gives the following number of degrees of freedom: $n_{\mathrm{df}} = L+1$.

We will use parameter values already identified using MLE in [3]. These estimates are provided in Table 2 for each model. Other types of objective functions and likelihood functions can be considered, as done in [3]. We will consider only the basic ones ((1.2) and (1.3)) in this research.

1.2. Information criteria for model selection

Information criteria (IC) are used to rank the models based on the balance between their goodness-of-fit and complexity to favour more generalizable models and to prevent overfitting [18]. To select an optimal (parsimonious) model among a set of competing models, one needs to choose a model with minimal criterion value (indicator). There are different IC which provide a compact formula for the indicators. They are derived under certain asymptotic approximations of the general notion of Kullback-Leibler distance between the true probabilistic model (which may not be present among the competing models) and the model under consideration (defined by its likelihood function). In practice, the most commonly used criterion is the Akaike IC (AIC), as well as its correction for small sample sizes AICc, which are defined for model M_k (see [3]) as

$$\begin{aligned} \mu_{\mathrm{AIC}}(M_k) &= -2\log L_{M_k}(\widehat{y}|\widehat{p}_k) + 2n_{\mathrm{df}} \\ \mu_{\mathrm{cAIC}}(M_k) &= -2\log L_{M_k}(\widehat{y}|\widehat{p}_k) + 2n_{\mathrm{df}} + \frac{2n_{\mathrm{df}}(n_{\mathrm{df}}+1)}{n_{\mathrm{obs}}-n_{\mathrm{df}}-1}. \end{aligned} \tag{1.6}$$

One can see that AIC provides the parsimony by penalizing the goodness-of-fit (proportional to the negative log-likelihood function) with the number of estimated model parameters, i.e., the complexity of the model. Other similar IC can be applied, e.g., the Bayesian IC (BIC), which includes the penalty term $n_{\mathrm{df}}\log(n_{\mathrm{obs}})$ instead of $2n_{\mathrm{df}}$. In both cases, the model complexity is related to the dimensionality of the problem.

In this work, we aim to compute more general indicator, which not only accounts for the goodness-of-fit and the dimensionality of the problem, but also for the so called geometric complexity of the model. The theory of minimum description length (MDL) criterion [12] is based on the concepts of Kolmogorov complexity and Kolmogorov structure functions [13]. In our context, the description length l_k of model M_k can be defined through the normalized maximum likelihood (NML) function [15] as

$$l_k = -\log p_{\mathrm{NML}}(\widehat{y}|M_k) = -\log L_{M_k}(\widehat{y}|\widehat{p}_k(\widehat{y})) + \log \int_{\Omega} L_{M_j}(\widetilde{y}|\widehat{p}_k(\widetilde{y}))\,\mathrm{d}\widetilde{y} \tag{1.7}$$

where the first term on the right-hand side is the measure of goodness-of-fit, and the second term, the logarithm of the denominator of p_{NML}, is the measure of model

complexity. The NML function can be understood as the observed likelihood (for given experimental data $\widehat{y}$) normalized by the overall likelihood for all possible outcomes for the state vector $\widetilde{y} \in \Omega$ treated as random variable. The second term can be hard to compute, as it requires to obtain the ML estimates $\widehat{p}_k(\widetilde{y})$ for each sample in the state space $\widetilde{y}$. However, it can be approximated through the use of Fisher information approximation (FIA) [20] via the sum of dimensionality and geometrical complexity terms. This gives the following indicator for the minimum description length criterion:

$$\mu_{\mathrm{FIA}}(M_k) = -\log L_{M_k}(\widehat{y}|\widehat{p}_k) + \frac{1}{2} n_{\mathrm{df}} \log \frac{n_{\mathrm{obs}}}{2\pi} + \log \left(\int_{\Theta} \sqrt{\det I_{M_k}(p_k)}\, \mathrm{d}p_k \right) \quad (1.8)$$

where $I_{M_k}(p)$ is the Fisher information for model M_k with parameters p_k, and Θ is the hypercube on the range of admissible parameter values.

The last term in (1.8), which we will denote as μ_{gc}, is called geometric complexity due to its possible interpretation as a logarithm of the volume of model in the space of probabilistic models [15]. The model volume is also used in Bayesian inference as normalizing constant for the Jeffrey's prior distributions on model parameters [15], which have an important property of being invariant to model reparameterizations.

1.3. Computation of the Fisher matrix

Fisher information, which needs to be computed to determine μ_{FIA}, is defined as the variance of the sensitivity of the log-likelihood function with respect to parameters, also called score. As the expected value of the score is zero, Fisher information matrix $I(p) \in \mathbb{R}^{L \times L}$ is given by

$$I(p) = E_{y \sim Y} \left(\left(\frac{\partial}{\partial p} \log L(y|p) \right)^2 \right) = \int_{\Omega} \left(\frac{\partial}{\partial p} L(\widetilde{y}|p) \right)^2 L(\widetilde{y}|p)\, \mathrm{d}\widetilde{y} \quad (1.9)$$

where $(x)^2$ is understood as $x \cdot x^T$. Fisher matrix can be expressed, under certain regularity conditions, as the expectation of the Hessian of negative log-likelihood or objective function. For the given observation data $\widetilde{y}$, it is called the observed Fisher matrix,

$$I_{\mathrm{obs}}(p) = I_{\mathrm{obs}}(p, \widetilde{y}) = -\frac{\partial^2}{\partial p^2} \log L(\widetilde{y}|p) = \frac{1}{2\sigma^2} \frac{\partial^2}{\partial p^2} \Phi_{\mathrm{OLS}}(p, \widetilde{y}) \in \mathbb{R}^{L \times L}. \quad (1.10)$$

Observed Fisher information (1.10) is usually used instead of the Fisher information (1.9), for example, to obtain the confidence intervals for parameter values. To compute Fisher information, i.e., the expectation of the observed Fisher matrix, it can be beneficial to use (1.9) instead of (1.10), as in practice, computed Hessian (1.10) can be non-symmetrical or non-positive definite matrix in some areas of Ω. In addition, it is no longer required to compute the second-order derivatives for

the model solution, but only the first-order sensitivities. The score, i.e., the sensitivity to the log-likelihood, is given by sensitivities $S(t_j)$ of the model solution at time moments t_j as

$$g(\widetilde{y},p)=\left(\frac{\partial}{\partial p}\log L(\widetilde{y}|p)\right)=\frac{1}{\sigma^2}\sum_{j=1}^{N}S(t_j)^T\left(y(t_j)-\widehat{y}_j\right) \tag{1.11}$$

where

$$S(t_j)=\frac{\partial y(t_j,p)}{\partial p}=[S^1(t_j),\ldots,S^M(t_j)]^T\in\mathbb{R}^{M\times L}.$$

By applying the chain rule to differentiate the original system (1.1), we obtain the direct method to compute $S(t)$. In matrix form, the system is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t)=\left[\frac{\partial f}{\partial y}\right]S(t)+\left[\frac{\partial f}{\partial y_\tau}\right]S(t-\tau)+\left[\frac{\partial f}{\partial p}\right]-\left[\frac{\partial f}{\partial y_\tau}\right]\left[0,\ldots,y'(t-\tau)\right] \tag{1.12}$$

where, for the equation on the sensitivity to the delay τ, $S_L(t)\equiv S_\tau(t)$ we need to compute the term $y'(t-\tau)$ in (1.12), which is nonzero for $t\geqslant\tau$ and can be computed as

$$y'(t-\tau)=f(y(t-\tau),y(t-2\tau),p),\quad t\geqslant\tau.$$

There are several alternative ways to obtain $S(t)$. First, one can use the adjoint method to compute sensitivities, which can be beneficial for the models with many parameters (cf. [2, 16, 19]). Alternatively, one can use the automatic differentiation software, such as the package ForwardDiff.jl [10] which we used for verification. These packages track all the elementary operations made to numerically obtain the solution and apply the differentiation rules to get the gradients and Hessians to the solution. Finally, one can use the finite difference methods to obtain the approximation for $S(t)$, or directly for (1.10), although this method can be prone to numerical instabilities.

1.4. Computation of the geometrical complexity and description length of the model

To compute Fisher information $I(p)$, we need to estimate the expectation of $g(\widetilde{y},p)\cdot g(\widetilde{y},p)^T$ over the state space (1.9). We use the properties of the expectation: by sampling the random numbers y_k distributed as $\pi(y)=L(y|p)$ (for fixed p), the expectation can be approximated as the mean value of observable matrices at sampling points:

$$I(p)\approx\frac{1}{n}\sum_{k=1}^{n}g(y_k,p)\cdot g(y_k,p)^T.$$

For a deterministic estimator, we use the inverse transform sampling with Sobol quasi-random low-discrepancy sequences. We generate Sobol sequences $\{s_k\}_1^n$,

$\{z_k\}_1^n$, $s_k \in [0,1]^{n^V_{\text{obs}}}$, $z_k \in [0,1]^{n^E_{\text{obs}}}$ where n^V_{obs} and n^V_{obs} are the number of observation points for variables V and E, respectively. The sample points $y_k = [V_k, E_k]^T$ are obtained through the quantile functions of the independent Gaussian distributions associated with the state vector components: $V_k = F_V^{-1}(s_k)$, $E_k = F_E^{-1}(z_k)$.

To compute the geometrical complexity term in (1.8), we need to take the integral of $\sqrt{\det I(p)}$ over the parameter space $p \in \Theta$. Note that the parameter space Θ is a part of model definition, and can be different for each model. In our case, $\Theta = \prod_{i=1}^{L}[p^i_{\min}, p^i_{\max}]$ is the same for all competing models, and determines the admissible range of values for model parameters due to their biological meaning (defined in Table 1). Once again, we use the quasi-Monte-Carlo integration method with Sobol sequences generated on Θ.

Overall, we have two quasi Monte-Carlo methods, the internal one is used to compute $I(p)$ for particular p, the external one is used to compute the logarithm of the geometrical complexity μ_{gc}. Depending on available computing resources, one can employ different schemes to compute μ_{gc} in parallel. For external Monte-Carlo loop, we distributed sequential streams along several nodes run in parallel, and used batched multithreading for the internal loop. The estimate variances needed to control the convergence of Monte-Carlo methods were combined from parallel streams using the parallel version of the Welford's online algorithm for variances.

2. Results

2.1. Competing mathematical models of the experimental virus infection

We analyze the models which were proposed in [3] to describe a data set of $n_{\text{obs}} = 15$ data points for viral load and number of cytotoxic T lymphocytes (CTLs) in spleen during experimental infection of mice with lymphocytic choriomeningitis viruses (LCMV). Their MLE parameter estimation and ranking via Akaike IC was performed in [3]. We apply the MDL criterion for their ranking by using a Fisher information approximation indicator μ_{FIA}. The models and their characteristics are described below.

- Model 1 (M_1): Basic predator-prey model with logistic virus growth term:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}V(t) &= \beta V(t)\left(1 - \frac{V(t)}{K}\right) - \gamma V(t)E(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}E(t) &= b_1 V(t)E(t) - \alpha_E E(t). \end{aligned} \tag{2.1}$$

- Model 2 (M_2): Model 1 + saturation of CTL proliferation rate:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}V(t) &= \beta V(t)\left(1 - \frac{V(t)}{K}\right) - \gamma V(t)E(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}E(t) &= b_2 \frac{V(t)E(t)}{\vartheta_{\text{Sat}} + V(t)} - \alpha_E E(t). \end{aligned} \tag{2.2}$$

- Model 3 (M_3): Model 2 + accounting for CTL division delay:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V(t) &= \beta V(t)\left(1-\frac{V(t)}{K}\right)-\gamma V(t)E(t)\\ \frac{\mathrm{d}}{\mathrm{d}t}E(t) &= b_3\frac{V(t-\tau)E(t-\tau)}{\vartheta_{\mathrm{Sat}}+V(t)}-\alpha_E E(t).\end{aligned} \tag{2.3}$$

- Model 4 (M_4): Model 3 + accounting for CTL homeostasis:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V(t) &= \beta V(t)\left(1-\frac{V(t)}{K}\right)-\gamma V(t)E(t)\\ \frac{\mathrm{d}}{\mathrm{d}t}E(t) &= b_4\frac{V(t-\tau)E(t-\tau)}{\vartheta_{\mathrm{Sat}}+V(t)}-\alpha_E E(t)+T^*.\end{aligned} \tag{2.4}$$

Models are subject to the following initial conditions:

$$\begin{aligned}&V(0)=200 \text{ virions}, \quad E(0)=265 \text{ cells}\\ &V(t)=0 \text{ virions}, \qquad E(t)=265 \text{ cells}, \quad t<0 \quad (\text{for } M_3, M_4).\end{aligned} \tag{2.5}$$

The MLE parameters for each model are presented in Table 2, and the corresponding numerical solutions (alongside the experimental data points) are shown in Fig. 1. Note that we excluded from the analysis Model 2 due to its singularity with respect to parameter ϑ_{Sat} (the MLE estimate presented in [3], $\vartheta_{\mathrm{Sat}} = 3.23 \cdot 10^{-176}$ was obtained without specifying constraints on the values of model parameters) which makes obtaining the solution prone to numerical instabilities. More appropriate analysis of Model 2 should be made by transforming it to a stochastic model with discrete state space, i.e., based on a Markov chain.

To calculate the geometrical complexity μ_{gc}, we need to specify the ranges of admissible values for parameter space Θ. Note that the parameter space is a part of model definition. The numerical solution of the model should also be stable and biologically reasonable within the specified ranges. The lower and upper bounds for model parameter values are specified in Table 1. They were derived from the range estimates available in LCMV modelling studies [6, 14]. Note that we needed to narrow down the range for the CTL division time lag τ for models M_3, M_4 (see Table 1) for numerical stability of these models at the extreme points of the parameter space. For parameter ϑ_{Sat}, the lower bound was set to machine epsilon 10^{-15} to avoid the singularity in the corresponding model terms. In addition, during the Monte-Carlo sampling process to compute μ_{gc}, at some sampling points the determinants of the Fisher information matrices were slightly negative (deviated from zero less than on machine epsilon), in which cases, these determinants were set to zero.

2.2. Ranking of models based on the Akaike and MDL information criteria

The results of the parameter estimation and evaluation of the Akaike and MDL criteria for the four considered mathematical models are summarized in Table 2.

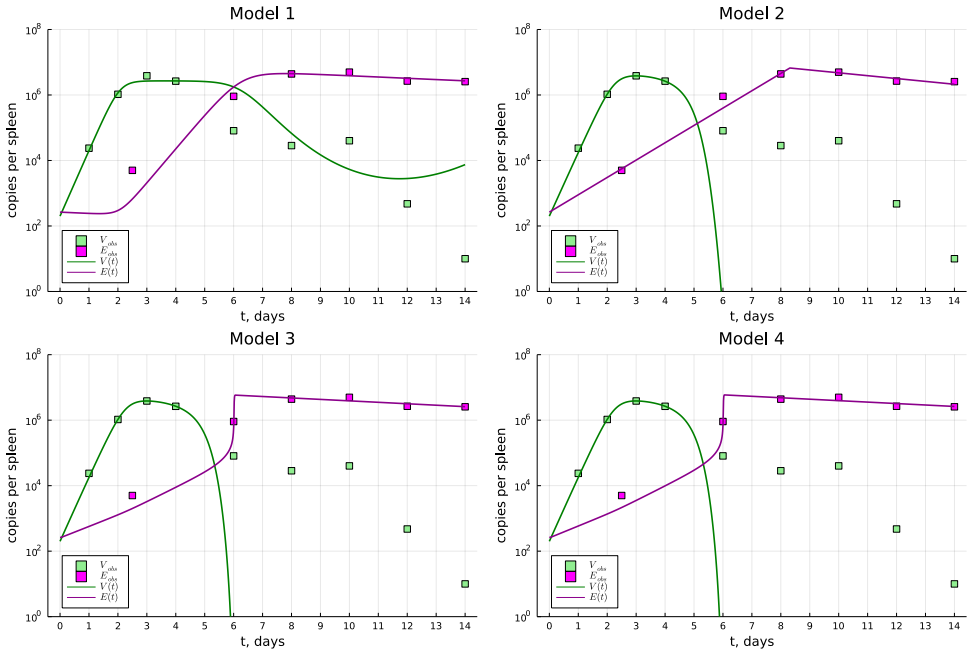

Figure 1. Numerical solutions for the models M_1-M_4 with parameters given in Table 2.

Model 2 appears to describe the activation of the CTL response in a degenerate way as the parameter ϑ_{Sat} is effectively zero. This results in difficulties with computing the respective MDL value. The Akaike criterion ranks the models as follows: M_2, M_3, M_1, M_4, suggesting that the second model (2.2) is the most parsimonious one. In contrast, the description length is smallest for the first one, ranking the models as M_1, M_3, M_4. The values of the indicators of the model complexities are obtained under the assumption that the experimental measurements of the viral load and CTL numbers follow the Gaussian distribution. Visually, the model M_1 provides a better match to the specific data set as compared to other models.

3. Conclusions

In this study, a computational framework has been elaborated to characterize the complexity of mathematical models in immunology. The issue of dimensionality of a model is one of the fundamental importance across many disciplines in life sciences when dealing with processes for which their mathematical descriptions are constructed rather than derived from first principles, in particular in mathematical immunology [1, 7]. So far, for model ranking and selection the most popular approach makes use of the Akaike information criterion [9]. The more general MDL-based approach balances the model goodness-of-fit with the model dimensionality, in terms of the free parameters and the data sample size, and the geometrical complexity of the model, i.e. the volume of distinguishable distributions that the model can describe [4].

Table 1. Definition of model parameters: biological meaning and admissible ranges.

	Parameter meaning, units	Range
β	Virus exponential growth rate, day^{-1}	$(3,5)$
K	Carrying capacity for the virus, virions/spleen	$(10^6, 10^8)$
γ	Virus elimination rate, 1/virions/day	$(10^{-6}, 10^{-3})$
b_1	CTL stimulation rate, 1/virions/day	$(10^{-8}, 10^{-6})$
$\{b_i\}_{i=2}^4$	CTL stimulation rate, day^{-1}	$(0.1, 10)$
τ	CTL division time, days	$(10^{-3}, 1)$, narrowed to $(10^{-3}, 0.1)$
ϑ_{Sat}	Viral load for half-maximal CTL stimulation, virions/spleen	$(0, 10)$
α_E	CTL death rate, day^{-1}	$(0.001, 0.5)$
T^*	Homeostatic influx of specific CTLs into spleen, cells/spleen/day	$(0, 50)$

Table 2. Maximum likelihood estimates of model parameters and the goodness-of-fit and complexity measures of the corresponding models.

Parameters	M_1	M_2	M_3	M_4
β	4.61	4.51	4.62	4.61
K	$2.7 \cdot 10^6$	$4.69 \cdot 10^6$	$5.01 \cdot 10^6$	$4.98 \cdot 10^6$
γ	$1.39 \cdot 10^{-6}$	$8.04 \cdot 10^{-5}$	$3.29 \cdot 10^{-4}$	$2.96 \cdot 10^{-4}$
b_i	$9.22 \cdot 10^{-7}$	1.42	1.14	1.16
ϑ_{Sat}	–	0 $(3.23 \cdot 10^{-176})$	$8.79 \cdot 10^{-6}$	$4.59 \cdot 10^{-6}$
τ	–	–	$4.38 \cdot 10^{-2}$	$4.15 \cdot 10^{-2}$
α_E	$9.29 \cdot 10^{-2}$	$2.01 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$
T^*	–	–	–	1.09
Indicators	M_1	M_2	M_3	M_4
Φ_{OLS}	$6.54 \cdot 10^{12}$	$8 \cdot 10^{11}$	$1.71 \cdot 10^{12}$	$1.65 \cdot 10^{12}$
$\widehat{\sigma}$	$6.6 \cdot 10^5$	$2.3 \cdot 10^5$	$3.4 \cdot 10^5$	$3.3 \cdot 10^5$
n_{df}	6	7	8	9
μ_{cAIC}	467	443	464	478
μ_{FIA}	247	–	252	270

In future work, we would like to explore the Bayesian setting both for the inference of these and other competing models and for their comparison using the Bayesian Model Selection indicator, defined as

$$\mu_{BMS}(M_k) = -\log \int_{\Theta} L(\widehat{y}|p_k)\pi(p_k)\,\mathrm{d}p_k.$$

For the prior parameter distribution $\pi(p_k)$, the Jeffrey's priors can be used which depend on the geometrical complexity of models computed in this paper.

We have presented essential numerical details enabling the assessment and ranking of four representative models of immune response to acute viral infection. These include: (1) optimization of the likelihood function, (2) computation of the sensitivity model functions, (3) evaluation of the Fisher information matrix, and (4) es-

timation of multidimensional integrals over the model parameter space. Our study shows the feasibility of the MDL approach to the analysis of dynamical systems formulated with ODEs and DDEs. We hope that it will promote a broader application of the MDL framework based on the fundamental concept of Fisher information for analysis and selection of optimal models in biology and medicine.

References

1. S. M. Andrew, C. T. H. Baker, and G.y A. Bocharov, Rival approaches to mathematical modelling in immunology. *Journal of Computational and Applied Mathematics* **205** (2007), 669–686.
2. C. T. H. Baker and G. A. Bocharov, Computational aspects of time-lag models of Marchuk type that arise in immunology. *Russian Journal of Numerical Analysis and Mathematical Modelling* **20** (2005), 247–262.
3. C. T. H. Baker, G. A. Bocharov, J. M. Ford, P. M. Lumb, S. J. Norton, C. A. H. Paul, T. Junt, P. Krebs, and B. Ludewig, Computational approaches to parameter estimation and model selection in immunology. *Journal of Computational and Applied Mathematics* **184** (2005), 50–76.
4. V. Balasubramanian, MDL, Bayesian inference, and the geometry of the space of probability distributions. *Advances in Minimum Description Length: Theory and Applications*, MIT Press Cambridge, MA, USA, 2005, pp. 81–98.
5. R. Blanco-Rodrguez, X. Du, and E. Hernndez-Vargas, Computational simulations to dissect the cell immune response dynamics for severe and critical cases of SARS-CoV-2 infection. *Computer Methods and Programs in Biomedicine* **211** (2021), 106412.
6. G. A. Bocharov, Modelling the dynamics of LCMV infection in mice: Conventional and exhaustive CTL responses. *Journal of Theoretical Biology* **192** (1998), 283–308.
7. G. Bocharov, V. Volpert, B. Ludewig, and A. Meyerhans, *Mathematical Immunology of Virus Infections*. Springer International Publishing, Cham, 2018.
8. G. Bocharov, R. Zst, L. Cervantes-Barragan, T. Luzyanina, E. Chiglintsev, V. A. Chereshnev, V. Thiel, and B. Ludewig, A systems immunology approach to plasmacytoid dendritic cell function in cytopathic virus infections. *PLoS Pathogens* **6** (2010), e1001017.
9. K. P. Burnham, D. R. Anderson, and K. P. Burnham, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed ed. Springer, New York, 2002, OCLC: ocm48557578.
10. Forward mode automatic differentiation on Julia. https://juliadiff.org/ForwardDiff.jl/stable/
11. D. Grebennikov, A. Karsonova, M. Loginova, V. Casella, A. Meyerhans, and G. Bocharov, *Predicting the cross-coordinated immune response dynamics in SARS-CoV-2 infection: Implications for disease pathogenesis*. Mathematics & Computer Science, Preprint, July 2022.
12. P. D. Grunwald, In Jae Myung, and M. A. Pitt (Eds.), *Advances in Minimum Description Length: Theory and Applications*, Neural information processing series. MIT Press, Cambridge, Mass, 2005.
13. M. Li and P. Vitnyi, *An Introduction to Kolmogorov Complexity and its Applications*, Texts in Computer Science. Springer International Publishing, Cham, 2019.
14. T. Luzyanina, K. Engelborghs, S. Ehl, P. Klenerman, and G. Bocharov, Low level viral persistence after infection with LCMV: a quantitative insight through numerical bifurcation analysis. *Mathematical Biosciences* **173** (2001), 1–23.
15. A. Ly, M. Marsman, J. Verhagen, R. P. P. P. Grasman, and E.-J. Wagenmakers, A tutorial on Fisher information. *Journal of Mathematical Psychology* **80** (2017), 40–55.

16. G. I. Marchuk, V. Shutyaev, and G. Bocharov, Adjoint equations and analysis of complex systems: Application to virus infection modelling. *Journal of Computational and Applied Mathematics* **184** (2005), 177–204.

17. M. Ostaszewski et al., COVID19 Disease Map, a computational knowledge repository of virus-host interaction mechanisms. *Molecular Systems Biology* **17** (2021), e10387.

18. M. A. Pitt and In Jae Myung, When a good fit can be bad. *Trends in Cognitive Sciences* **6** (2002), 421–425.

19. F. A. Rihan, A. A. Azamov, and H. J. Al-Sakaji, An inverse problem for delay differential equations: Parameter estimation, nonlinearity, sensitivity. *Applied Mathematics & Information Sciences* **12** (2018), 63–74.

20. J. J. Rissanen, Fisher information and stochastic complexity. *IEEE Transactions on Information Theory* **42** (1996), 40–47.