

# **Применение критерия минимальной длины описания для оценки сложности моделей в математической имmunологии**

Д.С. Гребенников<sup>1,2</sup>

<sup>1</sup>ИВМ РАН

<sup>2</sup>Сеченовский Университет

XIV конференция «Математические модели и численные методы в биологии и медицине»  
Москва, ИВМ РАН  
2 ноября 2022

## Статьи

- C.T.H. Baker, G.A. Bocharov, J.M. Ford, P.M. Lumb, S.J. Norton, C.A.H. Paul, T. Junt, P. Krebs, B. Ludewig. Computational approaches to parameter estimation and model selection in immunology. *Journal of Computational and Applied Mathematics* **2005**, 184, 50–76, doi:10.1016/j.cam.2005.02.003.
- A. Ly, M. Marsman, J. Verhagen, R. P. P. P. Grasman and E.-J. Wagenmakers. A Tutorial on Fisher information. *Journal of Mathematical Psychology* **2017**, 80, 40–55, doi:10.1016/j.jmp.2017.05.006.
- D.S. Grebennikov, V.V. Zheltkova, G.A. Bocharov. Application of minimum description length criterion for assessing the complexity of models in mathematical immunology. *Russian Journal of Numerical Analysis and Mathematical Modelling* **2022**, 37(5), 1-9, doi:10.1515/rnam-2022-00...

## Вид рассматриваемых моделей

Рассматриваемые модели (системы ОДУ и ДУЗА) имеют следующий вид:

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

где  $y = [y^1, \dots, y^M] \in \mathbb{R}^M$  – решение модели на интервале  $(0, T)$ ,  
 $p = [p_1, \dots, p_{L-1}, \tau] \in \mathbb{R}^L$  – вектор параметров модели,  
включающий одну задержку  $\tau \geq 0$ ,  
 $h_0$  и  $y_0$  – постоянная функция предыстории и начальное значение.

## Идентификация параметров моделей

Given a set of experimental data  $\{\hat{y}_j^i\}$  for model variables  $i = 1, \dots, M$  at time points  $t_j \in (0, T)$ ,  $j = 1, \dots, N$ , 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:

$$\Phi_{OLS}(p) = \Phi_{OLS}(p, \hat{y}) = \sum_{j=1}^N ||y(t_j, p) - \hat{y}_j||^2$$

## Метод максимального правдоподобия

The OLS estimates and MLE are equivalent under the following assumptions:

- (a1) the errors between the true model solution and experimental data are independent at successive time points  $t_j$ ,
- (a2) the errors in the components of  $\hat{y}_j$  are independent,
- (a3) the errors have the Gaussian distribution with the constant variance  $\sigma^2$ :

$$\hat{y}_j \sim N(y(t_j, p), \Sigma_j), \text{ where } \Sigma_j = \text{diag}\{\sigma, \dots, \sigma\}.$$

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

$$L(p) = L(\hat{y}|p) = \prod_{j=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{1}{2\sigma^2} ||y(t_j, p) - \hat{y}_j||^2} \quad (1)$$

## Метод максимального правдоподобия

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

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

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

The variance  $\sigma^2$  can be estimated using the maximum likelihood method as well. The estimate is given by

$$\hat{\sigma}^2 = \frac{1}{n_{obs}} \Phi_{OLS}(\hat{p}),$$

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

## Информационно-теоретические критерии сравнения моделей

Расстояние Кулльбака-Лейблера от распределения  $Q$  до распределения  $P$ :

$$D_{KL}(P||Q) = \int_{-\infty}^{+\infty} p(x) \log \left( \frac{p(x)}{q(x)} \right) dx$$

Информационно-теоретические критерии (AIC, BIC):

$$\mu_{AIC}(M_k) = -2 \log L_{M_k}(\hat{y}|\hat{p}_k) + 2n_{df_k},$$

$$\mu_{cAIC}(M_k) = -2 \log L_{M_k}(\hat{y}|\hat{p}_k) + 2n_{df_k} + \frac{2n_{df_k}(n_{df_k} + 1)}{n_{obs} - n_{df_k} - 1},$$

$$\mu_{BIC}(M_k) = -2 \log L_{M_k}(\hat{y}|\hat{p}_k) + n_{df_k} \log(n_{obs}).$$

AIC и BIC состоят из степени согласия модели с данными и штрафа на сложность модели, пропорционального размерности задачи.

## Критерий минимальной длины описания

Критерий минимальной длины описания основывается на понятиях информационной сложности и структурных функциях по Колмогорову.

В нашем случае, длину описания  $l_k$  модели  $M_k$  можно определить через нормализованную функцию максимального правдоподобия (NML):

$$l_k = -\log p_{NML}(\hat{y}|M_k) = -\log L_{M_k}(\hat{y}|\hat{p}_k(\hat{y})) + \log \int_{\Omega} L_{M_j}(\tilde{y}|\hat{p}_k(\tilde{y}))d\tilde{y}$$

NML-функцию можно понимать как наблюдаемое согласие модели (с экспериментальными данными  $\hat{y}$ ) нормированное на общее согласие для всех возможных реализациях случайной величины  $\tilde{y} \in \Omega$  вектора состояний модели.

## Критерий минимальной длины описания (аппроксимация на основе информации Фишера)

Аппроксимация MDL через информационную матрицу Фишера:

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

где  $I_{M_k}(p)$  – матрица Фишера модели  $M_k$  с параметрами  $p_k$ ,  
 $\Theta$  – гиперкуб, задающий диапазоны допустимых значений параметров моделей.

Дополнительное слагаемое учитывает геометрическую сложность модели.

## Информационная матрица Фишера

Fisher information, which needs to be computed to determine  $\mu_{FIA}$ , is defined as the variance of the sensitivity to 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:

$$I(p) = E_{y \sim Y} \left( \left( \frac{\partial}{\partial p} \log L(y|p) \right)^2 \right) = \int_{\Omega} \left( \frac{\partial}{\partial p} \log L(\tilde{y}|p) \right)^2 L(\tilde{y}|p) d\tilde{y}, \quad (3)$$

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  $\tilde{y}$ , it is called the observed Fisher matrix,

$$I_{obs}(p) = I_{obs}(p, \tilde{y}) = -\frac{\partial^2}{\partial p^2} \log L(\tilde{y}|p) = \frac{1}{2\sigma^2} \frac{\partial^2}{\partial p^2} \Phi_{OLS}(p, \tilde{y}) \in \mathbb{R}^{L \times L}. \quad (4)$$

## Вычисление матрицы Фишера

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

$$g(\tilde{y}, p) = \left( \frac{\partial}{\partial p} \log L(\tilde{y}|p) \right) = \frac{1}{\sigma^2} \sum_{j=1}^N S(t_j)^T (y(t_j) - \hat{y}_j), \quad (5)$$

where

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

$$\begin{aligned} \frac{d}{dt} S(t) &= \left[ \frac{\partial f}{\partial y} \right] S(t) + \left[ \frac{\partial f}{\partial y_\tau} \right] S(t - \tau) + \left[ \frac{\partial f}{\partial p} \right] - \\ &\quad - \left[ \frac{\partial f}{\partial y_\tau} \right] [0, \dots, y'(t - \tau)], \end{aligned} \quad (6)$$

where, for the equation on the sensitivity to the delay  $\tau$ ,  $S_\tau(t) \equiv S_\tau(t)$  we need to compute the term  $y'(t - \tau)$  in (6), which is nonzero for  $t \geq \tau$  and can be computed as

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

## Вычисление геометрической сложности модели

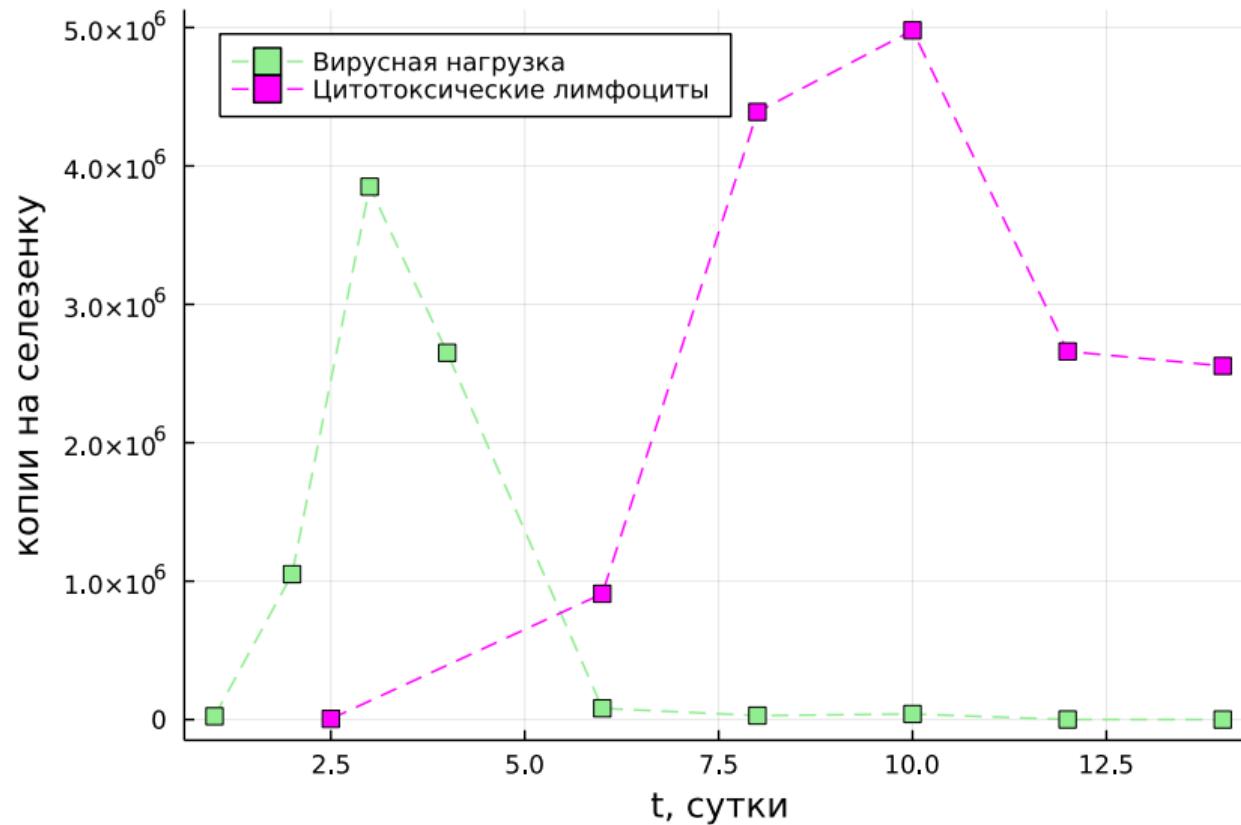
To compute Fisher information  $I(p)$ , we need to estimate the expectation of  $g(\tilde{y}, p) \cdot g(\tilde{y}, p)^T$  over the state space (3). We use the properties of the expectation: by sampling the random numbers  $y_k$  distributed as  $\pi(y) = L(y|p)$  (for fixed  $p$ ), the expectation can 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  $\{s_k\}_1^n, \{z_k\}_1^n$ ,  $s_k \in [0, 1]^{n_{obs}^V}, z_k \in [0, 1]^{n_{obs}^E}$ , where  $n_{obs}^V$  and  $n_{obs}^E$  are the number of observation points for variables  $V$  and  $E$ . The sample points  $y_k = [V_k, E_k]^T$  are obtained through the quantile functions of the independent Gaussian distributions associated with state vector components:  $V_k = F_V^{-1}(s_k), E_k = F_E^{-1}(z_k)$ .

To compute the geometrical complexity term in (9), we need to take the integral of  $\sqrt{\det I(p)}$  over the parameter space  $p \in \Theta$ . Note that the parameter space  $\Theta$  is a part of model definition, and can be different for each model.

# Данные динамики экспериментальной вирусной инфекции (ВЛХМ)



# Конкурирующие модели динамики инфекции ВЛХМ

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

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

- Model 3 ( $M_3$ ):  
 $M_2 +$  accounting for CTL division delay:

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

- Model 2 ( $M_2$ ):  
 $M_1 +$  saturation of CTL proliferation rate:

$$\begin{aligned}\frac{d}{dt}V(t) &= \beta V(t) \left(1 - \frac{V(t)}{K}\right) - \gamma V(t)E(t), \\ \frac{d}{dt}E(t) &= b_2 \frac{V(t)E(t)}{\theta_{Sat} + V(t)} - \alpha_E E(t).\end{aligned}$$

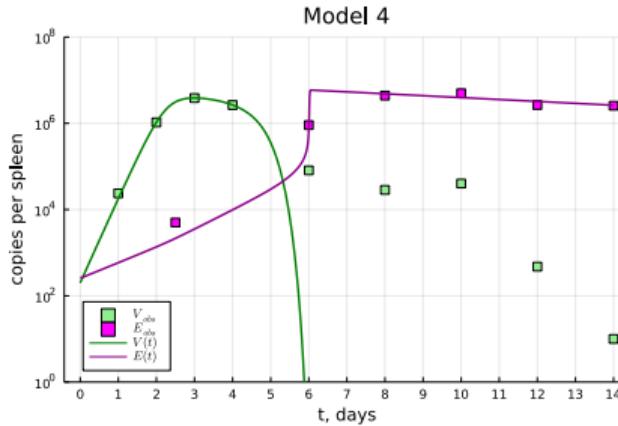
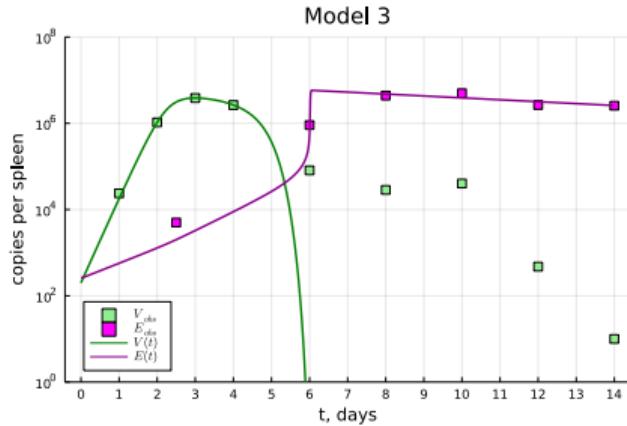
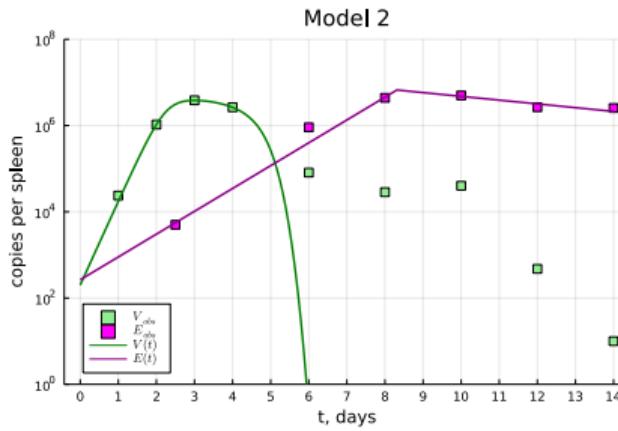
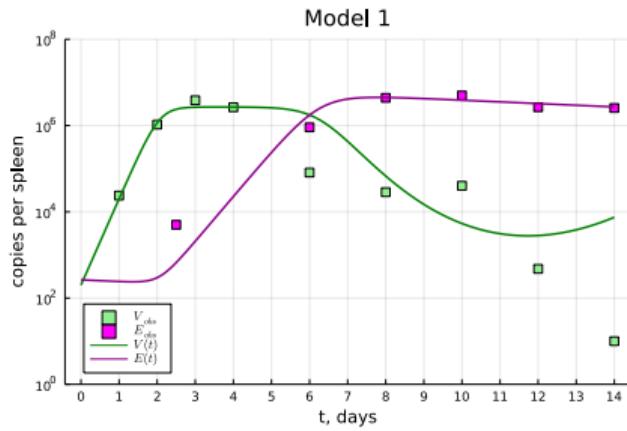
- Model 4 ( $M_4$ ):  
 $M_3 +$  accounting for CTL homeostasis:

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

Initial conditions:

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

# Конкурирующие модели динамики инфекции ВЛХМ (идентификация параметров ММП)



# Конкурирующие модели динамики инфекции ВЛХМ (допустимые диапазоны значений параметров)

	Parameter meaning, units	Range
$\beta$	Virus exponential growth rate, day <sup>-1</sup>	(3, 5)
$K$	Carrying capacity for the virus, virions/spleen	( $10^6$ , $10^8$ )
$\gamma$	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 <sup>-1</sup>	(0.1, 10)
$\tau$	CTL division time, days	( $10^{-3}$ , 0.1)
$\theta_{Sat}$	Viral load for half-maximal CTL stimulation, virions/spleen	(0, 10)
$\alpha_E$	CTL death rate, day <sup>-1</sup>	(0.001, 0.5)
$T^*$	Homeostatic influx of specific CTLs into spleen, cells/spleen/day	(0, 50)

## Ранжирование моделей на основе критериев Акаике и МДО

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

Parameters	$M_1$	$M_2$	$M_3$	$M_4$
$\beta$	4.61	4.51	4.62	4.61
$K$	$2.7 \times 10^6$	$4.69 \times 10^6$	$5.01 \times 10^6$	$4.98 \times 10^6$
$\gamma$	$1.39 \times 10^{-6}$	$8.04 \times 10^{-5}$	$3.29 \times 10^{-4}$	$2.96 \times 10^{-4}$
$b_i$	$9.22 \times 10^{-7}$	1.42	1.14	1.16
$\theta_{Sat}$	—	$0 (3.23 \times 10^{-176})$	$8.79 \times 10^{-6}$	$4.59 \times 10^{-6}$
$\tau$	—	—	$4.38 \times 10^{-2}$	$4.15 \times 10^{-2}$
$\alpha_E$	$9.29 \times 10^{-2}$	$2.01 \times 10^{-1}$	$1.02 \times 10^{-1}$	$1.02 \times 10^{-1}$
$T^*$	—	—	—	1.09
Indicators	$M_1$	$M_2$	$M_3$	$M_4$
$\Phi_{OLS}$	$6.54 \times 10^{12}$	$8 \times 10^{11}$	$1.71 \times 10^{12}$	$1.65 \times 10^{12}$
$\hat{\sigma}$	$6.6 \times 10^5$	$2.3 \times 10^5$	$3.4 \times 10^5$	$3.3 \times 10^5$
$n_{df}$	6	7	8	9
$\mu_{cAIC}$	467	443	464	478
$\mu_{FIA}$	247	—	252	270

## Заключение

- Какой из критериев использовать (AIC, BIC, MDL, FIA)?
- Множественная кросс-валидация по значениям ошибки на валидационных выборках (обобщаемость модели на новые данные).
- Ускорение вычислений MDL/FIA.
- Байесовский подход: BMS + Jeffreys priors

$$\mu_{BMS}(M_k) = -\log \int_{\Theta} L(\hat{y}|p_k) \pi(p_k) dp_k,$$

$$\pi(p_k) = \frac{1}{V_k} \sqrt{\det I_{M_k}(p_k)}, \quad V_k = \int_{\Theta} \sqrt{\det I_{M_k}(p_k)} dp_k$$

**Спасибо за внимание!**

