**У**ДК: 51.76

# Моделирование ВИЧ инфекции: идентификация моделей и исследование глобальной чувствительности Желткова В.В.<sup>\*1</sup>, Желтков Д.А.<sup>2</sup>, Бочаров Г.А.<sup>†2</sup>

<sup>1</sup>Московский государственный университет им. М.В. Ломоносова, Москва, Россия <sup>2</sup>Институт вычислительной математики им. Г.И. Марчука РАН, Москва, Россия

Аннотация. Для решения задач современной иммунологии широко применяется математическое моделирование. В данной работе мы описали задачу моделирования динамики численностей популяций клеток при ВИЧ-инфекции, на примере которой рассмотрели основные этапы построения и идентификации моделей. Были изучены модели возрастающей степени сложности, описывающие различные аспекты заболевания (клеточный и гуморальный иммунный ответ, антигенпрезентирование, динамика регуляторных Т-лимфоцитов). Была представлена модификация модели инфекционного заболевания Марчука–Петрова для описания ВИЧ-инфекции, учитывающая запаздывание, при оценивании параметров которой был проведён сравнительный анализ численных методов оптимизации. Для рассмотренных моделей была сформулирована и решена задача оценивания параметров по клиническим данным с использованием двух моделей ошибок наблюдений, а также проведено исследование глобальной чувствительности. Был проведен сравнительный анализ качества приближения экспериментальных данных по критерию значений функционала невязки и на основе информационно-теоретического подхода.

*Ключевые слова:* mathematical modelling, global sensitivity analysis, parameter estimation, global optimization.

#### **ВВЕДЕНИЕ**

В задачах современной иммунологии широко применяется математическое моделирование. Использование математических моделей позволяет лучше понимать сложные взаимосвязи между компонентами систем, проводить численные эксперименты, формулировать гипотезы о причинах тех или иных явлений, выявлять возможные мишени для разработки препаратов, прогнозировать эффект применения различных режимов терапии. Однако при разработке математической модели иммунологического процесса исследователи сталкиваются с рядом сложностей. Во-первых, параметры описываемых систем не всегда могут быть измерены экспериментально, поэтому возникает необходимость решения задачи оценивания параметров. Рассматриваемая система может обладать достаточно высокой размерностью, что приводит к большим вычислительным затратам при решении задачи оптимизации. Таким образом, важной задачей является выбор наиболее эффективных методов оптимизации. Во-вторых, возникает вопрос изучения влияния неопределённости в значениях параметров модели

<sup>\*</sup>valeryaaziattseva@yandex.ru

<sup>&</sup>lt;sup>†</sup>gbocharov@gmail.com

на получаемые решения, а также выявления тех параметров, изменение которых сильнее всего влияет на решение модели. В-третьих, как правило, даже в рамках описания на одном уровне, например, динамики клеточных популяций, одно и то же явление может быть описано с помощью моделей разной степени детализации. Возникает задача выбора оптимальной модели, такой, чтобы, с одной стороны, она хорошо описывала экспериментальные данные, с другой – была максимально простой. В данной работе обозначенные задачи будут рассмотрены на примере моделирования динамики клеточных популяций при ВИЧ-инфекции. Мы рассмотрим несколько математических моделей различной степени детализации: M1 – простейшую модель ВИЧ-инфекции [1], M2 – модель с учётом клеточного иммунного ответа [2, 3], МЗ – модель с учётом динамики регуляторных Т-лимфоцитов [4], и разработанную нами ранее адаптацию модели Марчука–Петрова (М4.1) [5], для которой будет представлена версия с запаздываниями (М4.2). Для них будет решена задача оценивания параметров по экспериментальным данным. Для решения задачи оценивания параметров наиболее сложных моделей М4.1, М4.2 будет проведён сравнительный анализ эффективности методов глобальной оптимизации. Для моделей M1, M2, M3 будет рассмотрено решение задачи анализа глобальной чувствительности с помощью подхода LHS-PRCC [6], а также проведён сравнительный анализ качества приближения экспериментальных данных на основе анализа функционала невязки (один из вариантов функционала рассматривался нами в [7]), а также информационно-теоретического подхода [8].

#### МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВИЧ ИНФЕКЦИИ

ВИЧ-инфекция – хроническое заболевание, вызываемое вирусом иммунодефицита человека [3], характеризующееся длительным течением и постепенным снижением уровня CD4 Т-лимфоцитов до критического уровня, при котором наступает фаза СПИД. Данное заболевание обладает некоторыми специфическими особенностями. Так, клетками-мишенями при ВИЧ-инфекции являются CD4 Т-лимфоциты и антиген-презентирующие клетки (АПК), которые играют ключевую роль при развитии иммунного ответа. Следует также отметить, что постепенное снижение численности СD4 Т-лимфоцитов и наступление фазы СПИД не имеет однозначного объяснения. Триггерами данного явления могут быть хроническая иммунная активация [9]. пространственные аспекты развития заболевания, связанные с повреждением структуры лимфоидных органов [10], механизмы истощения [11]. Изучение всех аспектов ВИЧ-инфекции, даже в рамках моделирования на уровне популяций клеток, приводит к огромному разнообразию математических моделей [3]. В данном разделе мы рассмотрим математические модели динамики клеточных популяций при ВИЧ-инфекции разной степени детализации: М1 – простейшая модель без учёта иммунного ответа; М2 – расширение M1 путем добавления уравнения для CD8 Т-лимфоцитов; M3 – модель, учитывающая клеточный иммунный ответ и динамику регуляторных Т-лимфоцитов; M4.1 и M4.2 – адаптации модели Марчука–Петрова без учёта и с учётом запаздываний, отражающих длительность деления и дифференцировки клеток.

#### 1. Простейшие модели

Базовая модель ВИЧ-инфекции [1] описывает динамику клеток-мишеней (T), зараженных клеток-мишеней ( $T^*$ ) и вирусных частиц (V). Схема модели М1 представлена на рисунке 1,А, параметры – в таблице 1.

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \lambda - d_T T - kVT, \ \frac{\mathrm{d}T^*}{\mathrm{d}t} = kVT - \delta T^*, \ \frac{\mathrm{d}V}{\mathrm{d}t} = N\delta T^* - cV.$$
(1)

Математическая биология и биоинформатика. 2019. Т. 14.№ 1. doi: 10.17537/2019.14.19



Рис. 1. Схемы моделей: А – модель М1, В – модель М2, С - модель М3.

Расширенная версия модели М2 описана в [2, 3] и включает блок клеточного иммунного ответа в виде динамики эффекторных клеток (*E*). Схема модели представлена на рисунке 1,В, параметры – в таблице 1.

$$\frac{dT}{dt} = \lambda - d_T T - kVT,$$

$$\frac{dT^*}{dt} = kVT - \delta T^* - mET^*,$$

$$\frac{dV}{dt} = N\delta T^* - cV - \rho kVT,$$

$$\frac{dE}{dt} = \lambda_E + \frac{b_E T^*}{T^* + K_b}E - \frac{d_E T^*}{T^* + K_d}E - \delta_E E.$$
(2)

## 2. Модель с регуляторными Т-лимфоцитами

В работе [4] описана модель, учитывающая динамику регуляторных Т-лимфоцитов (R -натуральные (естественные) регуляторные Т-лимфоциты,  $R_{\alpha}$  – адаптивные регуляторные Т-лимфоциты), в которой рассматриваются незрелые  $(E_i)$  и зрелые  $(E_m)$  эффекторные Т-лимфоциты. Регуляторные Т-лимфоциты играют ключевую роль в развитии механизма истощения, и их динамика описывается в ряде моделей онкологических заболеваний и аутоиммунных процессов, например, [12] [13], однако практически никогда не учитываются при разработке математических моделей ВИЧ инфекции. Мы модифицировали модель, предложенную в данной работе, переписав её в виде системы ОДУ, и видоизменили описание процесса подавления эффекторных клеток регуляторными Т-лимфоцитами, с тем, чтобы уравнения модели удовлетворяли условию положительной инвариантности (схема модели представлена на рисунке 1,С, параметры – в таблице 1):

$$\frac{dT}{dt} = \lambda - d_T T - kVT - \alpha_1 RT,$$

$$\frac{dT^*}{dt} = kVT - \delta T^* - mE_m T^*,$$

$$\frac{dE_i}{dt} = p_{E_i} - d_{E_i}E_i - k_E V E_i - \alpha_2 R E_i$$

$$\frac{dE_m}{dt} = k_E V E_i - \delta_E E_m - \alpha_3 R_\alpha E_m$$

$$\frac{dV}{dt} = N\delta T^* - cV,$$

$$\frac{dR}{dt} = p_R - d_r R - \gamma RV,$$

$$\frac{dR_\alpha}{dt} = \gamma RV - \delta_R R_\alpha,$$
(3)

Математическая биология и биоинформатика. 2019. Т. 14.№ 1. doi: 10.17537/2019.14.19

Таблица 1. Параметры моделей М1, М2, М3. Для каждого параметра в верхнем индексе
перечислены номера моделей, в которых он используется

Г

Параметр	Биологическое значение	ед. изм.
$\lambda^{1,2,3}$	Скорость производства клеток-мишеней	кл/(мл∙д.)
$d_T^{1,2,3}$	Скорость гибели клеток-мишеней	1/д.
$k^{1,2,3}$	Скорость заражения клеток-мишеней	мл/(част.·д.)
$\delta^{1,2,3}$	Скорость гибели инфицированных клеток-мишеней	1/д.
$N^{1,2,3}$	Скорость производства вирусных частиц	част./кл
$c^{1,2,3}$	Скорость гибели вирусных частиц	1/д.
$m^{2,3}$	Скор. уничтожения зараженных клеток СD8 Т-лимф.	мл/(кл.•д.)
$\delta_E^{2,3}$	Скорость гибели CD8 Т-лимфоцитов	1/день
$\lambda_E^2$	Скорость производства CD8 Т-лимфоцитов	кл/(мл·д.)
$\rho^2$	К-во вир. частиц, инфицирующих клетку	част./кл
$b_E^2$	Макс. скорость деления CD8 Т-лимфоцитов	1/д.
$K_b^2$	Константа полумакс. скор. для деления CD8 Т-лимфоцитов	кл./мл
$d_E^2$	Макс. скорость гибели CD8 Т-лимфоцитов	1/день
$K_d^2$	Конст. полумакс. скор. для гибели CD8 Т-лимфоцитов	кл./мл
$\alpha_1^3$	Скор. подавления CD4 Т-лимф. натур. рег. Т-лимф.	мл/(кл.•д.)
$p_{E_i}^3$	Скор. производства незрелых CD8 Т-лимф.	кл/(мл∙д.)
$d_{E_i}^3$	Скор. гибели незрелых CD8 Т-лимфоцитов	1/д.
$k_E^3$	Скор. перехода $E_i - > E_m$	мл/(част.•д.)
$\alpha_2^3$	Скор. подавления $E_i$ натуральными рег. Т-лимф.	мл/(кл.•д.)
$\alpha_3^3$	Скор. подавления $E_m$ адаптивными рег. Т-лимф.	мл/(кл.•д.)
$p_R^3$	Скор. производства натуральных рег.Т-лимф.	кл/(мл∙д.)
$d_R^3$	Скор. гибели натуральных рег. Т-лимфоцитов	1/д.
$\gamma_R^3$	Скор. перехода $R - > R_{lpha}$	мл/(част.•д.)
$\delta_R^3$	Скор. гибели адаптивных рег. Т-лимфоцитов	1/д.

## 3. Адаптация модели Марчука-Петрова для описания ВИЧ-инфекции

Модель инфекционного заболевания Марчука–Петрова [14] описывает динамику АПК, гуморальное и клеточное звенья иммунного ответа, а также учитывает поражение органа-мишени. В работе [5] мы адаптировали модель Марчука–Петрова для описания ВИЧ-инфекции, используя следующие предположения: (i) клетками-мишенями являются CD4 Т-лимфоциты и антигенпрезентирующие клетки; (ii) CD4 Т-лимфоциты могут быть заражены как от свободных вирусных частиц, так и при контакте с инфицированными АПК; (iii) вместо поражения органа-мишени рассматривалось общее количество погибших в ходе инфекции клеток. В [5] описана версия модели без учета запаздываний, приведены уравнения, параметры и схема модели. Однако процесс активации и деления клеток при развитии иммунного ответа не является моментальным, поэтому более точным является использование версии модели с запаздываниями. В этом случае уравнения для соответствующих переменных должны быть переписаны в виде:

$$\begin{aligned} \frac{dP}{dt} &= b_P^p \xi \rho_P (D_V (t - \tau_P) + D_V^* (t - \tau_P)) (H_{Bsp} (t - \tau_P) + H_{Bsp}^* (t - \tau_P)) B_{sp} (t - \tau_P) + \\ \alpha_P (\theta P_0 - P); \\ \frac{dH_{Bsp}}{dt} &= \alpha_{H_B} (\xi \theta H_B^0 - H_{Bsp}) - \sigma_{H_B} H_{Bsp} V - \sigma_{H_B}^D H_{Bsp} D_V^* + b_{H_B} (2(D_V (t - \tau_{HB}) + D_V^* (t - \tau_{HB})) H_{Bsp} (t - \tau_{HB}) - (D_V + D_V^*) H_{Bsp}) - b_{H_B}^P (D_V + D_V^*) H_{Bsp} B_{sp}; \\ \frac{dH_{Esp}}{dt} &= \alpha_{H_E} (\xi \theta H_E^0 - H_{Esp}) - \sigma_{H_E} H_{Esp} V - \sigma_{H_E}^D H_{Esp} D_V^* + b_{H_E} (2(D_V (t - \tau_{HE}) + D_V^* (t - \tau_{HE})) H_{Esp} (t - \tau_{HE}) - (D_V + D_V^*) H_{Esp}) - b_{H_E}^P (D_V + D_V^*) H_{Esp} E_{sp}; \\ \frac{dB_{sp}}{dt} &= \alpha_B (\theta B^0 - B_{sp}) \\ &+ b_B^P (2(D_V (t - \tau_B) + D_V^* (t - \tau_B))) (H_{Bsp} (t - \tau_B) + H_{Bsp}^* (t - \tau_B)) B_{sp} (t - \tau_E) + \\ H_{Esp}^* (t - \tau_E)) E_{sp} (t - \tau_E) - b_{ED_V} D_V^* E_{sp} - b_{EH_E} H_E^* E_{sp} - b_{EH_B} H_B^* E_{sp}; \end{aligned}$$
(4)

Для модели с запаздыванием помимо параметров, рассмотренных в [5], требуется оценить также величины запаздываний:  $\tau_P$ ,  $\tau_{HB}$ ,  $\tau_{HE}$ ,  $\tau_B$ ,  $\tau_E$ . Обозначим модифицированную версию модели Марчука–Петрова без запаздываний М4.1, версию с запаздываниями – М4.2.

#### ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛЕЙ

## 4. Статистические подходы. Метод максимального правдоподобия

Пусть математическая модель задана в следующей форме:

$$\begin{cases} \frac{d\mathbf{y}(t)}{dt} &= f(\mathbf{y}(t), \mathbf{p}), t \in [0, T] - \text{ уравнения модели;} \\ \mathbf{y}(t) &= \{y_i(t), \ i = 1, ..., M, y_i(t) \in C^1(0, T)\} - \text{переменные модели;} \\ \mathbf{y}(0) &= \mathbf{y_0} - \text{ начальные условия;} \\ \mathbf{p} &= \{p_i\}, i = 1, ..., L - \text{ вектор параметров.} \end{cases}$$

Методология оценивания параметров моделей, которой мы следуем, подробно описана в [8]. Как правило, для решения задачи применяется Байесовский подход, в основе которого лежит предположение о том, что заранее известна некоторая априорная информация о распределении параметров, полученная, например, в результате предыдущих измерений ( $p(\mathbf{p})$ ). Таким образом, рассматривается апостериорная вероятность:

$$P(\mathbf{p}|D) = \frac{p(D|\mathbf{p}) \cdot p(\mathbf{p})}{\int p(D|\mathbf{p}) \cdot p(\mathbf{p}) d\mathbf{p}}$$

В случае отсутствия априорной информации о распределении параметров можно сделать предположение о равномерном распределении. В этом случае, Байесовский метод редуцируется к методу максимального правдоподобия. Предположим, вектор параметров **p** состоит из двух частей  $\mathbf{p} = [\mathbf{a}, \mathbf{b}]$ , где  $[\mathbf{a}]$  - множество неизвестных параметров, которые мы будем оценивать. Экспериментальные данные представлены в виде набора моментов времени и соответствующих результатов измерений:  $\{\mathbf{y}_j, t_j\}, j = 1, ..., N$ . Будем искать такой вектор  $\mathbf{p}^* = [\mathbf{a}^*, \mathbf{b}]$ , для которого решение модели наилучшим образом приближает экспериментальные данные в точках измерения. В качестве меры близости будем использовать соответствующий функционал  $\Phi(\mathbf{p})$ . Таким образом, задача оценивания параметров может быть переформулирована в виде [15]:

$$\mathbf{p}^* = argmin_{\mathbf{p}\in\Omega\subset\mathbf{R}^l}\Phi(\mathbf{p}). \tag{5}$$

Полученная задача минимизации в общем случае является некорректной по Адамару, так как, с учетом погрешностей измерений, статистически допустимыми решениями будут все те значения параметров, при которых решение модели уклоняется от экспериментальных данных не более, чем на некоторую величину. Для решения задачи приближения модели к экспериментальным данным необходимо выбрать критерий близости, то есть, целевой функционал невязки  $\Phi(\mathbf{p})$ , отражающий характер распределения и статистические свойства погрешностей измерений. Будем использовать следующие предположения: (1) погрешности измерений не зависят от момента времени; (2) погрешности измерений распределены по нормальному закону, то есть  $\mathbf{y}_j = N(y(t_j, \mathbf{p}), \Sigma_j)$ , где  $y(t_j, \mathbf{p}) -$  средние, определяемые моделью,  $\Sigma_j - j$ -я ковариационная матрица; (3) погрешности измерений являются независимыми между собой, то есть ковариационная матрица является диагональной:  $\Sigma_j = \sigma^2 \{ diag[\omega_1^j, ... \omega_M^j] \}$ . Полная функция правдоподобия имеет вид:

$$L(\mathbf{p}) = \prod_{j=1}^{N} H(y_j, \mathbf{p}), \ H(y_j, \mathbf{p}) = \frac{1}{\sqrt{(2\pi)^M det \Sigma_j}} exp\left\{-\frac{1}{2}[y(t_j, \mathbf{p}) - y_j]^T \Sigma_j^{-1}[y(t_j, \mathbf{p}) - y_j]\right\}$$
(6)

Метод максимального правдоподобия заключается в том, чтобы найти такие параметры  $\mathbf{p}^*$ , при которых функция правдоподобия достигает максимума. Функционал взвешенных наименьших квадратов, с учетом предположения (3), можно переписать в виде:

$$\Phi_{WLS}(\mathbf{p}) = \sum_{j=1,\dots,N} [y(t_j, \mathbf{p}) - y_j]^T \Sigma_j^{-1} [y(t_j, \mathbf{p}) - y_j] = \sigma^{-2} \Phi_{\Omega LS}(\mathbf{p}),$$
(7)

где

$$\Phi_{\Omega LS}(\mathbf{p}) = \sum_{j=1,..N} \left\| diag^{-1}[\boldsymbol{\omega}_1^{[j]}, ... \boldsymbol{\omega}_M^{[j]}] \left[ y(t_j, \mathbf{p}) - y_j \right] \right\|^2.$$
(8)

Таким образом, мы видим, что максимизация функции правдоподобия  $L(\mathbf{p})$  эквивалентна минимизации функционала  $\Phi_{\Omega LS}(\mathbf{p})$ . Задача поиска точечных оценок максимального правдоподобия может быть переформулирована в виде задачи минимизации функционала наименьших квадратов. В ходе анализа большого количества экспериментальных данных по динамике численности лимфоцитов было установлено, что наиболее адекватными моделями ошибок являются нормальное и логнормальное распределения [16]. Далее

мы будем предполагать, что для остальных исследуемых величин это наблюдение так же является справедливым. Таким образом, можно рассмотреть следующие варианты функционала невязки:

 Функционал, соответствующий нормальному распределению, с дисперсией, одинаковой по всем переменным и не зависящей от момента времени, то есть, метод наименьших квадратов (Ф<sub>OLS</sub>). При различных дисперсиях используется взвешенный метод наименьших квадратов (Ф<sub>WLS</sub>):

$$\Phi_{OLS}(\mathbf{p}) = \sum_{j=1,\dots N} \sum_{i=1\dots M} [y^i(t_j, \mathbf{p}) - y^i_j]^2; \Phi_{WLS}(\mathbf{p}) = \sigma^{-2} \sum_{j=1,\dots N} \sum_{i=1\dots M} \omega^{[j]}_i [y^i(t_j, \mathbf{p}) - y^i_j]^2.$$

• Функционал, соответствующий логнормальному распределению, с дисперсией, равной по всем переменным и не зависящей от момента времени (аналогично предыдущему случаю, можно использовать данный функционал с весами):

$$\Phi_{LogOLS}(\mathbf{p}) = \sum_{j=1,..N} \sum_{i=1..M} [ln(y^i(t_j, \mathbf{p})) - ln(y^i_j)]^2.$$

Для оценивания параметров моделей М1, М2, МЗ мы использовали набор клинических данных, описывающий типичную динамику ВИЧ-инфекции [17], представляющий собой численность CD4 Т-лимфоцитов, и вирусную нагрузку, взятые в 17 моментах времени в течение острой фазы ВИЧ-инфекции для каждой величины. Для оценивания параметров модели М4.1, М4.2 к данному набору для каждого момента времени были добавлены численности CD8 Т -лимфоцитов [17]. Для моделей М1, М2, МЗ использовалось две формулировки задачи минимизации, соответствующие модели ошибок с нормальным и логнормальным распределением:

$$\Phi_{OLS} = \sum_{i=1}^{17} \left[ CD4^i - CD4^i_{obs} \right]^2 + \sum_{i=1}^{17} \left[ V^i - V^i_{obs} \right]^2$$
(9)

$$\Phi_{LogOLS} = \sum_{i=1}^{17} \left[ \log(CD4^i) - \log(CD4^i_{obs}) \right]^2 + \sum_{i=1}^{17} \left[ \log(V^i) - \log(V^i_{obs}) \right]^2.$$
(10)

Для моделей M4.1 M4.2 рассматривался только логарифмический функционал наименьших квадратов  $\Phi_{LogOLS}$ :

$$\begin{split} \Phi_{LogOLS} &= \sum_{i=1}^{17} \left[ \log(CD4^i) - \log(CD4^i_{obs}) \right]^2 + \sum_{i=1}^{17} \left[ \log(CD8^i) - \log(CD8^i_{obs}) \right]^2 + \\ &\sum_{i=1}^{17} \left[ \log(V^i) - \log(V^i_{obs}) \right]^2. \end{split}$$

Здесь  $CD4_{obs}^i, CD8_{obs}^i, V_{obs}^i$  – наблюдаемые величины в момент времени  $t_i$ ;  $V^i = V(t_i)$ ; для моделей М1, М2, М3:  $CD4^i = T(t_i)$ ; для моделей М4.1 М4.2:  $CD4^i = H_B(t_i) + H_E(t_i) + H_{Bsp}(t_i) + H_{Esp}(t_i) + H_B^*(t_i) + H_B^*(t_i) + H_B^*(t_i) + H_{Bsp}^*(t_i) + H_{Esp}^*(t_i) CD8^i = E(t_i) + E_{sp}(t_i).$ 

Задача оценивания параметров моделей М1, М2, М3, соответствующая модели ошибок с логнормальным распределением, была рассмотрена нами в [7]. Найденные значения параметров моделей представлены в таблице 2, а соответствующие им решения – на рисунке 2.

Параметр	M1, Norm	M1, logN	M2, Norm	M2, logN	M3, Norm	M3, logN
$\lambda^{1,2,3}$	$1.44 \cdot 10^4$	$1.7 \cdot 10^{4}$	$1.52 \cdot 10^{4}$	$1.59\cdot 10^4$	$2.4 \cdot 10^{4}$	$3 \cdot 10^3$
$d_T^{1,2,3}$	0.029	0.027	0.027	0.021	0.037	0.0034
$k^{1,2,3}$	$4.7 \cdot 10^{-9}$	$5.8 \cdot 10^{-9}$	$3.5 \cdot 10^{-9}$	$1.4 \cdot 10^{-8}$	$8.2 \cdot 10^{-9}$	$8.2 \cdot 10^{-9}$
$\delta^{1,2,3}$	2	2.37	1.18	0.6705	0.1551	0.2
$N^{1,2,3}$	653	999	999	314.7393	999	709
$c^{1,2,3}$	1.57	3.34	1.47	1.7600	1.15	1.13
$m^{2,3}$	-	-	$10^{-4}$	$8.4 \cdot 10^{-5}$	$10^{-4}$	$5.7 \cdot 10^{-5}$
$\delta_E^{2,3}$	-	-	0.002	0.0124	$10^{-4}$	0.0052
$\lambda_E^2$	-	-	97.7	82.1372	-	-
$\rho^2$	-	-	2	1.68	-	-
$b_E^2$	-	-	0.5	0.2181	-	-
$K_b^2$	-	-	50	122.3	-	-
$d_E^2$	-	-	0.1	0.26	-	-
$K_d^2$	-	-	600	500	-	-
$\alpha_1^3$	-	-	-	-	$8 \cdot 10^{-8}$	$1.5 \cdot 10^{-7}$
$p_{E_i}^3$	-	-	-	-	$1.9 \cdot 10^{3}$	$2.1 \cdot 10^{3}$
$d_{E_i}^3$	-	-	-	-	0.08	0.003
$k_E^3$	-	-	-	-	10 <sup>-7</sup>	$10^{-7}$
$\alpha_2^3$	-	-	-	-	$7.4 \cdot 10^{-8}$	$5.6 \cdot 10^{-5}$
$\alpha_3^3$	-	-	-	-	$10^{-4}$	$8.4 \cdot 10^{-5}$
$p_R^3$	-	-	-	-	10	10
$d_R^3$	-	-	-	-	0.016	0.04
$\gamma_R^3$	-	-	-	-	$9.6 \cdot 10^{-5}$	$6 \cdot 10^{-5}$
$\delta_R^3$	-	-	-	-	0.094	0.1

**Таблица 2.** Найденные параметры моделей М1, М2, МЗ при использовании модели ошибки с нормальным (Norm) и логнормальным (logN) распределением. Для каждого параметра в верхнем индексе перечислены номера моделей, в которых он используется



**Рис. 2.** Приближение данных моделями M1, M2, M3 при использовании модели ошибок с нормальным (Norm) и логнормальным (LogN) распределением.

#### ИССЛЕДОВАНИЕ ГЛОБАЛЬНОЙ ЧУВСТВИТЕЛЬНОСТИ

Методология исследования неопределённости и глобальной чувствительности моделей описана в [6]. При построении модели биологической системы часто приходится сталкиваться с неопределённостью экспериментальных данных и параметров. Для изучения неопределённости результатов моделирования при изменении параметров, как правило, применяются техники семплирования (такие, как метод Монте–Карло (МС) и латинского гиперкуба (LHS)). Анализ чувствительности помогает понять, изменение каких параметров вносит наибольший вклад в наблюдаемый процесс. В отличие от исследования локальной чувствительности, при котором все параметры, кроме изучаемого, фиксируются, анализ глобальной чувствительности позволяет получить более полное представление о поведении модели во всем пространстве параметров. В [6] описаны несколько алгоритмов оценивания глобальной чувствительности моделей, такие, как анализ коэффициентов корреляции Пирсона, ранговой корреляции Спирмана, ранговой частичной корреляции, тест чувствительности амплитуды Фурье.

## 5. LHS-PRCC

Преимуществом использования метода семплирования на основе Латинского гиперкуба (LHS) является то, что оно позволяет использовать меньший размер выборки для получения оценок той же точности по сравнению с MC. Пусть K – размер сгенерированной выборки параметров  $\mathbf{p}_i = \{p_j, j = 1, ..., L\}^i, i = 1, ..., K$ ,  $\mathbf{y}^i = \mathbf{y}(\mathbf{p}_i) = \{y_q(\mathbf{p}_i), q = 1, ..., M\}$  – изучаемая величина. Тогда коэффициент корреляции Пирсона вычисляется по формуле (где Cov(p, y) – ковариация, Var(p) – вариация p):

$$CC_{p_j y_q} = \frac{Cov(p_j, y_q)}{Var(p_j)Var(y_q)} = \frac{\sum_{i=1}^{K} (p_j^i - \bar{p_j})(y_q^i - \bar{y_q})}{\sqrt{\sum_{i=1}^{N} (p_j^i - \bar{p_j})^2 \sum_{i=1}^{K} (y_q^i - \bar{y_q})^2}}, j = 1, 2, ..., L$$

...

Коэффициент частичной корреляции РСС, характеризует линейную зависимость между  $p_i, y_a$ , вычисляется по формуле [6]:

$$PCC_{p_j y_q} = CC_{(p_j - \hat{p_j}), (y_q - \hat{y_q})}, \ \hat{p_j} = c_0 + \sum_{l=1, l \neq j}^k c_l p_l, \ \hat{y_q} = b_0 + \sum_{l=1, l \neq j}^k b_l x_l.$$

Математическая биология и биоинформатика. 2019. Т. 14.№ 1. doi: 10.17537/2019.14.19

Коэффициенты ранговой корреляции Спирмана RCC и ранговой частичной корреляции PRCC рассчитываются так же, но не для самих величин, а для их рангов. PRCC является устойчивой мерой оценки чувствительности для нелинейных, но монотонных зависимостей между  $p_i, y$ , при условии независимости  $p_i$ .

## 6. Полученные результаты

Нами была изучена глобальная чувствительность численности клеток-мишеней и вирусных частиц в хронической фазе инфекции к изменению параметров для моделей M1, M2, M3. Для проведения расчётов использовались алгоритмы пакета [18]. По алгоритму LHS было сгененировано по 5000 векторов параметров, соответствующих каждой модели, для которых были найдены решения. Далее были найдены величины PRCC, соответствующие значениям переменных T ( $PRCC_T$ ), V ( $PRCC_V$ ) в момент времени, соответствующий последнему наблюдению. Далее параметры моделей были отсортированы в порядке убывания величины  $|PRCC_V| + |PRCC_T|$ , результаты представлены на рисунке 3. По результатам исследования глобальной чувствительности можно сделать следующие выводы:

- Для всех трех рассмотренных моделей в числе значимых параметров оказались *N*, *c*, *k*, λ. Таким образом, для всех трех рассмотренных моделей, изменение параметров, описывающих динамику вирусных частиц, а так же скорости производства клеток-мишеней, ведет к наибольшему изменению вирусной нагрузки и численности клеток-мишеней в хронической фазе.
- Для модели МЗ наибольшее значение  $|PRCC_V| + |PRCC_T|$  соответствует параметру скорости производства регуляторных Т-лимфоцитов  $p_R$ . Таким образом, можно предположить, что при моделировании ВИЧ-инфекции учет динамики регуляторных Т-лимфоцитов может оказать значительное влияние на результаты моделирования.



Рис. 3. Результаты исследования глобальной чувствительности для М1, М2, М3.

## СРАВНЕНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ВИЧ-ИНФЕКЦИИ РАЗНОЙ СТЕПЕНИ ДЕТАЛИЗАЦИИ НА ОСНОВЕ КАЧЕСТВА ВОСПРОИЗВЕДЕНИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ И КРИТЕРИЯ АКАИКЕ

Для сравнения качества приближения экспериментальных данных моделью можно использовать величину функционала невязки. Чем меньше значение функционала, тем лучше приближение данных моделью. Однако, при данном подходе не учитывается сложность модели. Использование информационного критерия Акаике позволяет выбрать модель, обладающую наибольшей простотой, при этом, хорошо описывающую экспериментальные данные [8]. Величина информационного критерия Акаике зависит не только от величины функционала правдоподобия в найденной точке, но и от количества наблюдений (n = NM) и параметров (L) [8]. Величины критерия Акаике, а также скорректированного критерия Акаике, могут быть вычислены по формулам:

$$u_{AIC} = -2lnL(\mathbf{\hat{p}}) + 2(L+1); \ u_{cAIC} = u_{AIC} + \frac{2(L+1)(L+2)}{n-L-2}.$$

Скорректированный критерий Акаике применяется, если количество наблюдений мало (n < 40(L + 1)). При оценивании параметров в рамках метода максимального правдоподобия для сравнительного анализа можно использовать величины:

$$\check{u}_{AIC} = nln(\Phi_{\Omega LS}(\mathbf{\hat{p}})) + 2(L+1); \ \check{u}_{cAIC} = \check{u}_{AIC} + \frac{2(L+1)(L+2)}{n-L-2}.$$

Чем ниже значение критерия Акаике для модели, тем лучше модель приближает данные с точки зрения информационно-теоретического подхода. Полученные результаты для моделей М1, М2, М3 приведены в таблице 3. Мы видим, что сравнение по значению функционала невязки и критерию Акаике дает разные результаты: с точки зрения приближения данных как в рамках модели ошибок с нормальным распределением, так и в рамках модели с логнормальным распределением, наилучший результат показывает модель М3. Однако с точки зрения информационно-теоретического подхода оптимальной является модель М1. Возможно, это связано с тем, что для оценивания параметров использовался относительно небольшой набор данных, включающий наблюдения только по двум переменным модели.

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ НА ПРИМЕРЕ МОДЕЛИ МАРЧУКА–ПЕТРОВА В ВИДЕ СИСТЕМ ОДУ И ДУЗА

Для нахождения оптимального вектора параметров требуется решить задачу минимизации функционала невязки. Для модели с большим количеством оцениваемых параметров решение оптимизационной задачи является вычислительно затратным. Более того, функционал может иметь несколько локальных минимумов. Существует большое количество разработанных и реализованных алгоритмов оптимизации, поэтому возникает вопрос выбора наиболее оптимального алгоритма. Для решения полученных задач минимизации было использовано несколько численных методов из различных классов, реализованных в библиотеке NLopt [19], а так же метод глобальной оптимизации, основанный на применении TT (tensor train)-разложения тензоров [20]. Нами не рассматривались методы, требующие явного задания градиента функционала, например, такие, как StoGO [21]. Рассмотрим результаты расчётов при помощи следующих методов:

- ТТ (метод, основанный на применении ТТ-разложения тензоров);
- CRS2 (метод контролируемого случайного поиска с локальной оптимизацией [22]);
- rMLSL, qrMLSL (многоуровневый метод разбиения области на односвязные кластеры [23] с псевдослучайными (квазислучайными) стартовыми точками);
- rMLSL+SBPLX, qrMLSL+SBPLX (многоуровневый метод разбиения области на односвязные кластеры с псевдослучайными (квазислучайными) стартовыми точками и локальной оптимизацией методом Subplex [24]);
- ISRES (улучшенная эволюционная стратегия стохастического ранжирования [25]);

• ESCH (эволюционные алгоритмы [26]).

Для решения задачи идентификации параметров модели, сформулированной в виде системы обыкновенных дифференциальных уравнений, методом глобальной минимизации, было установлено ограничение в 10<sup>6</sup> вычислений функционала. Рассматриваемая система является жёсткой, поэтому для её интегрирования применялся метод, основанный на формулах дифференцирования назад (BDFs), реализованный [27]. Для решения системы использовались значения в библиотеке Sundials относительной и абсолютной погрешности, равные  $10^{-4}$  и максимальное количество вычислений функционала, равное 10000. При идентификация параметров модели, сформулированной в виде системы дифференциальных уравнений с запаздывающим аргументом производилось как более точное интегрирование системы (  $\epsilon = 10^{-4}$ ), так и интегрирование со сниженной точностью (  $\epsilon = 10^{-2}$ ). Снижение точности позволило ускорить решение задачи Коши для модели при заданных значениях параметров примерно в 10 раз, поэтому при сниженной точности интегрирования было в 10 раз увеличено число вычислений функционала для каждого метода глобальной оптимизации: функционал вычислялся  $10^7$  раз, а для высокой точности интегрирования –  $10^6$  раз. После работы методов глобальной оптимизации полученные оценки параметров были уточнены с помощью алгоритма локальной оптимизации SBPLX. Для интегрирования системы уравнений с запаздыванием применялся метод Гира (формулы дифференцирования назад), реализованный в системе DIFSUBDEL [28]. Результаты приведены в таблице 4.

**Таблица 3.** Значения  $\Phi_{LogOLS}$ ,  $\Phi_{OLS}$  и критерия Акаике в найденных минимумах для моделей М1, М2, МЗ

Модель	К-во параметров	$\Phi_{LogOLS}$	$\Phi_{OLS}$	$\check{u}_{cAIC}$
M1	6	3.27	$6.2 \cdot 10^{11}$	941.5106
M2	14	2.11	$4.1 \cdot 10^{11}$	965.8107
M3	18	1.4	$1.2 \cdot 10^{11}$	959.4878

Метод	ОДУ	ДУЗА, высокая точность	ДУЗА, низкая точность
TT	2.28	2.56	2.15
CRS2	2.47	5.06	2.47
rMLSL	2.50	3.20	2.63
qrMLSL	2.51	2.89	2.72
rMLSL+SBPLX	2.13	2.82	2.70
qrMLSL+SBPLX	2.22	3.00	2.46
ISRES	2.34	2.72	2.16
ESCH	3.49	4.38	2.56

Таблица 4. Минимум функционала, полученный после локальной дооптимизации

Исходя из полученных результатов мы можем отметить, что, в отличие от решения задачи оценивания параметров модели, сформулированной в виде системы ОДУ, для моделей в виде системы ДУЗА наилучшие результаты показали методы TT и ISRES, особенно при интегрировании системы со сниженной точностью.

## ЗАКЛЮЧЕНИЕ

В данной работе мы рассмотрели основные этапы построения и идентификации математический модели сложного биологического процесса на примере моделирования ВИЧ-инфекции. Были рассмотрены модели различной степени сложности, для которых была сформулирована и решена задача оценивания параметров по данным, описывающим типичный сценарий развития заболевания. Были рассмотрены 2 модели ошибки измерений, соответствующие нормальному и логнормальному распределению. Было проведено исследование глобальной чувствительности, которое показало, что в числе наиболее значимых параметров для всех рассмотренных моделей находятся параметры динамики вирусных частиц, а также скорость производства клеток-мишеней. Для модели с блоком, описывающим динамику регуляторных Т-лимфоцитов, наиболее значимым параметром оказалась скорость производства регуляторных Т-лимфоцитов, что может говорить о том, что динамика регуляторных Т-лимфоцитов оказывает существенное влияние на течение заболевания. Также был проведен сравнительный анализ рассмотренных моделей на основе значений функционала невязки и критерия Акаике. С точки зрения приближения экспериментальных данных наиболее хороший результат показала модель с учетом динамики регуляторных Т-лимфоцитов, однако с точки зрения информационно-теоретического подхода наиболее оптимальной для заданного набора данных является самая простая из рассмотренных моделей. Вероятно, для использования более сложных моделей для калибровки и оценивания параметров требуется использовать более полные наборы данных о динамике компонент ВИЧ инфекции. В работе также была представлена модификация модели Марчука–Петрова с запаздываниями для описания ВИЧ-инфекции. Эта модель является сложной динамической системой и обладает большим количеством степеней свободы, поэтому задача оценивания параметров для неё является вычислительно сложной. На примере рассмотренной задачи калибровки модели мы провели провели сравнительный анализ численных методов оптимизации и определили варианты калибровки, для которых рассматриваемые методы являются наиболее эффективными.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-31-00356 и проекта РФФИ № 17-01-00636.

## СПИСОК ЛИТЕРАТУРЫ

- 1. Perelson A.S. Modelling viral and immune system dynamics. *Nature Reviews Immunology*. 2002. V. 2. № 1. P. 28. doi: 10.1038/nri700
- 2. Adams B.M., Banks H.T., Davidian M., Kwon H.D., Tran H.T., Wynne S.N., Rosenberg E.S. HIV dynamics: modeling, data analysis, and optimal treatment protocols. *Journal of Computational and Applied Mathematics*. 2005. V. 184. № 1. P. 10–49. doi: 10.1016/j.cam.2005.02.004
- 3. Bocharov G., Chereshnev V., Gainova I., Bazhan S., Bachmetyev B., Argilaguet J., Martinez J., Meyerhans A. Human immunodeficiency virus infection: from biological observations to mechanistic mathematical modelling. *Mathematical Modelling of Natural Phenomena*. 2012. V. 7. № 5. P. 78–104. doi: 10.1051/mmnp/20127507
- 4. Simonov M., Rawlings R.A., Comment N., Reed S.E., Shi X., Nelson P.W. Modeling adaptive regulatory T-cell dynamics during early HIV infection. *PloS one*. 2012. V. 7. № 4. P. e33924. doi: 10.1371/journal.pone.0033924
- 5. Zheltkova V.V., Zheltkov D.A., Grossman Z., Bocharov G.A., Tyrtyshnikov E.E. Tensor based approach to the numerical treatment of the parameter estimation problems in mathematical immunology. *Journal of Inverse and Ill-posed Problems*. 2018. V. 26. № 1.

P. 51–66. doi: 10.1515/jiip-2016-0083

- 6. Marino S., Hogue I.B., Ray C.J., Kirschner D.E. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*. 2008. V. 254. № 1. P. 178–196. doi: 10.1016/j.jtbi.2008.04.011
- 7. Желткова В.В., Бочаров Г.А. Математическое моделирование динамики ВИЧ-инфекции с разной степенью детализации В: *Современные проблемы физико-математических наук*. Под ред. Можаровой Т.Н. Орел: ОГУ им. И. С. Тургенева, 2018, С. 347–350.
- 8. Baker C.T., Bocharov G.A., Ford J.M., Lumb P.M., Norton S.J., Paul C.A.H., Junt T., Krebs P., Ludewig B. Computational approaches to parameter estimation and model selection in immunology. *Journal of Computational and Applied Mathematics*. 2005. V. 184. Nº 1. P. 50–76. doi: 10.1016/j.cam.2005.02.003
- 9. Hazenberg M.D., Otto S.A., van Benthem B.H., Roos M.T., Coutinho R.A., Lange J.M., Hamann D., Prins M., Miedema F. Persistent immune activation in HIV-1 infection is associated with progression to AIDS. *Aids*. 2003. V. 17. № 13. P. 1881–1888. doi: 10.1097/00002030-200309050-00006
- 10. Zeng M., Haase A.T., Schacker T.W. Lymphoid tissue structure and HIV-1 infection: life or death for T cells. *Trends in Immunology*. 2012. V. 33. № 6. P. 306–314. doi: 10.1016/j.it.2012.04.002
- Day C.L., Kaufmann D.E., Kiepiela P., Brown J.A., Moodley E.S., Reddy S., Mackey E.W., Miller J.D., Leslie A.J., DePierres C., Mncube Z. PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature*. 2006. V. 443. № 7109. P. 350. doi: 10.1038/nature05115
- 12. Wilson S., Levy D. A mathematical model of the enhancement of tumor vaccine efficacy by immunotherapy. *Bulletin of Mathematical Biology*. 2012. V. 74. № 7. P. 1485–1500. doi: 10.1007/s11538-012-9722-45
- 13. Kim Y., Lee S., Kim Y.S., Lawler S., Gho Y.S., Kim Y.K., Hwang H.J. Regulation of Th1/Th2 cells in asthma development: a mathematical model. *Mathematical Biosciences* & *Engineering*. 2013. V. 10. № 4. P. 1095–1333. doi: 10.3934/mbe.2013.10.1095
- 14. Marchuk G.I. *Mathematical Modelling of Immune Response in Infectious Diseases*. Springer Science+Business Media Dordrecht, 1997. (Mathematics and Its Applications. V. 395). doi: 10.1007/978-94-015-8798-3
- 15. Ashyraliyev M., Fomekong-Nanfack Y., Kaandorp J.A., Blom J.G. Systems biology: parameter estimation for biochemical models. *FEBS Journal*. 2009. V. 276. № 4. P. 886–902. doi: 10.1111/j.1742-4658.2008.06844.x
- 16. Бочаров Г.А., Медведева Н.А. Численные алгоритмы анализа чувствительности и сложности описания в задачах идентификации моделей математической иммунологии. *Численные методы, параллельные вычисления и информационные технологии.* Под ред. Вл.В. Воеводина и Е.Е. Тыртышникова. Изд-во: НИВЦ МГУ, 2008. С. 67–90.
- 17. Munier M.L., Kelleher A.D. Acutely dysregulated, chronically disabled by the enemy within: T-cell responses to HIV-1 infection. *Immunology and cell biology*. 2007. V. 85. № 1. P. 6–15. doi: 10.1038/sj.icb.7100015
- 18. Dr. Denise Kirschner *LHS-PRCC* URL: http://malthus.micro.med.umich.edu/lab/usadata/ (дата обращения: 27.12.2018).
- 19. Johnson S. *The NLopt nonlinear-optimization package*. URL: http://ab-initio.mit.edu/wiki/index.php/NLopt (дата обращения: 27.12.2018).
- 20. Zheltkov D.A., Oferkin I.V., Katkova E.V., Sulimov A.V., Sulimov V.B., Tyrtyshnikov E.E. TTDock: a docking method based on tensor train decompositions. *Vychislitel'nye Metody i*

*Programmirovanie*. 2013. V. 4. № 3. P. 279–291.

- 21. Gudmundsson S. *Parallel Global Optimization*: M.Sc. Thesis. Technical University of Danmarks, 1998.
- 22. Kaelo P., Ali M.M. Some variants of the controlled random search algorithm for global optimization. *Journal of Optimization Theory and Applications*. 2006. V. 130. № 2. P. 253–264. doi: 10.1007/s10957-006-9101-0
- 23. Kan A.R., Timmer G.T. Stochastic global optimization methods. *Mathematical programming*. 1987. V. 39. № 1. P. 27–56. doi: 10.1080/01966324.1984.10737135
- 24. Thomas Harvey Rowan. *Functional stability analysis of numerical algorithms*: The doctoral dissertation. The University of Texas at Austin, 1990.
- 25. Runarsson T.P., Yao X. Search biases in constrained evolutionary optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews).* 2005. V. 35. № 2. P. 233–243.
- 26. Silva-Santos C.H., Goncalves M.S., Hernández-Figueroa H.E. Designing novel photonic devices by bio-inspired computing. *Photonics Technology Letters*, *IEEE*. 2010. V. 22. № 15. doi: 10.1109/LPT.2010.2051222
- 27. Hindmarsh A.C., Brown P.N., Grant K.E., Lee S.L., Serban R., Shumaker D.E., Woodward C.S. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)*. 2005. V. 31. № 3. P. 363–396.
- 28. Бочаров Г.А., Романюха А.А. Численное решение дифференциальных уравнений с запаздывающим аргументом на основе линейных многошаговых методов. Алгоритм и программа: препринт ОВМ АН СССР №117. М., 1986.

Рукопись поступила в редакцию 28.12.2018. Дата опубликования 06.02.2019.