

Идентифицируемость математических моделей медицинской биологии

С.И. Кабанихин^{1,2}, Д.А. Воронов^{1,2}, А.А. Гродзь², О.И. Криворотко^{1,2}

¹ Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия
² Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

Анализ биологических данных является важнейшим вопросом в биоинформатике, вычислительной геномике, молекулярном моделировании и системной биологии. Рассматриваемые в статье подходы позволяют сократить затраты на проведение экспериментов по получению биологических данных. В статье рассмотрен вопрос идентифицируемости математических моделей физиологии, фармакокинетики и эпидемиологии. Рассматриваемые процессы моделируются с помощью нелинейных систем обыкновенных дифференциальных уравнений. Математическое моделирование динамических процессов основано на использовании закона сохранения масс. В процессе решения задачи по оценке параметров, характеризующих исследуемый процесс, нередко возникает вопрос неединственности решения. В случае, когда известны результаты эксперимента (данные на выходе) и данные на входе, целесообразно проводить априорный анализ информативности этих данных. В статье рассмотрено понятие идентифицируемости математических моделей. Представлен обзор методов анализа идентифицируемости динамических систем. В работе приведен обзор следующих подходов: метод передаточной функции, применяемый для линейных моделей (удобен для анализа фармакокинетических данных, так как большой класс препаратов характеризуется линейной кинетикой); метод разложения в ряды Тейлора, применяемый для нелинейных моделей; метод, основанный на теории дифференциальной алгебры (структура данного алгоритма допускает его реализацию на ЭВМ); метод, основанный на теории графов (данный метод не только определяет идентифицируемость модели, но и позволяет найти замену переменных специального вида, приводящую исходную модель к идентифицируемой). На конкретных примерах продемонстрирована необходимость проводить априорный анализ идентифицируемости модели перед проведением численных расчетов по определению параметров, характеризующих тот или иной процесс. Рассмотрены примеры анализа идентифицируемости математических моделей медицинской биологии.

Ключевые слова: идентифицируемость; математические модели медицинской биологии; система обыкновенных дифференциальных уравнений; фармакокинетика; эпидемиология; физиология; дифференциальная алгебра.

HOW TO CITE THIS ARTICLE?

Kabanikhin S.I., Voronov D.A., Grodz A.A., Krivorotko O.I. Identifiability of mathematical models in medical biology. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2015;19(6):738-744. Doi 10.18699/VJ15.097

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ?

Кабанихин С.И., Воронов Д.А., Гродзь А.А., Криворотко О.И. Идентифицируемость математических моделей медицинской биологии. Вавиловский журнал генетики и селекции. 2015;19(6):738-744. Doi 10.18699/VJ15.097

Identifiability of mathematical models in medical biology

S.I. Kabanikhin^{1,2}, D.A. Voronov^{1,2}, A.A. Grodz², O.I. Krivorotko^{1,2}

¹ Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
² Novosibirsk State University, Novosibirsk, Russia

Analysis of biological data is a key topic in bioinformatics, computational genomics, molecular modeling and systems biology. The methods covered in this article could reduce the cost of experiments for biological data. The problem of identifiability of mathematical models in physiology, pharmacokinetics and epidemiology is considered. The processes considered are modeled using nonlinear systems of ordinary differential equations. Math modeling of dynamic processes is based on the use of the mass conservation law. While addressing the problem of estimation of the parameters characterizing the process under the study, the question of nonuniqueness arises. When the input and output data are known, it is useful to perform an a priori analysis of the relevance of these data. The definition of identifiability of mathematical models is considered. Methods for analysis of identifiability of dynamic models are reviewed. In this review article, the following approaches are considered: the transfer function method applied to linear models (useful for analysis of pharmacokinetic data, since a large class of drugs is characterized by linear kinetics); the Taylor series expansion method applied to nonlinear models; a method based on differential algebra theory (the structure of this algorithm allows this to be run on a computer); a method based on graph theory (this method allows for analysis of the identifiability of the model as well as finding a proper reparametrization reducing the initial model to an identifiable one). The need to perform a priori identifiability analysis before estimating parameters characterizing any process is demonstrated with several examples. The examples of identifiability analysis of mathematical models in medical biology are presented.

Key words: identifiability; mathematical models in medical biology; system of ordinary differential equations; pharmacokinetic; epidemiology; physiology; differential algebra.

Динамические процессы фармакокинетики, эпидемиологии и физиологии моделируются (Bellman, Astrom, 1970; Brown, 1980; Walter et al., 1984; Tunali, Tarn, 1987; Bellu et al., 2007) с помощью систем обыкновенных дифференциальных уравнений (СОДУ) вида:

$$\begin{cases} \dot{x} = f(x(t), p) + \sum_{i=1}^n h(x(t), p) u_i(t) \\ y(t) = g(u(t), x(t), p) \end{cases}, \quad (1)$$

где $x(t)$ – n -мерная функция состояния (в фармакокинетике – доза препарата, в эпидемиологии – количество больных разного типа), $y(t)$ – k -мерная функция экспериментальных данных (в фармакокинетике – концентрация препарата в крови и/или моче, в эпидемиологии – количество больных по годам), $u(t)$ – функция входных данных (в фармакокинетике – способ введения препарата в организм) (Goodwin, Payne, 1977), g – функция, связывающая модель с измерениями, h – функция, определяющая структуру входных данных, $f(x(t), p)$ – функция, определяющая структуру модели, p – s -мерный вектор параметров, характеризующий рассматриваемый процесс (в фармакокинетике – скорость перехода препарата между органами, в эпидемиологии – смертность, приток индивидов, скорость развития болезни и др.).

При изучении динамических процессов (1) возникает вопрос неединственности набора параметров $p = [p_1, \dots, p_s]$, удовлетворяющего имеющимся экспериментальным данным. С другой стороны, зачастую по измеренным данным невозможно определить набор параметров p . Нередко в физиологических и клинических исследованиях проведение экспериментов по получению данных $y(t)$ является финансово затратной процедурой. Кроме того, получение экспериментальных данных – не всегда приятная процедура для пациента, и избежание «неинформативных» измерений – важный момент и с этической стороны.

На практике у исследователя есть два типа информации: входные данные – функция $u(t)$ и данные на выходе – $y(t)$. Априорный анализ идентифицируемости математической модели отвечает на вопрос: можно ли по информации на входе и на выходе утверждать, что искомый набор параметров определяется единственным образом? Если ответ на данный вопрос положительный, то можно проводить серию измерений для получения экспериментальных данных $y(t)$, не опасаясь иметь, с одной стороны, бесполезные, а с другой – затратные данные.

Определение идентифицируемости (Brown, 1982; Carson, Cobelli, 2001, 2008; Bellu et al., 2007) формулируется следующим образом:

Определение 1. Модель (1) называется априорно идентифицируемой, если ее параметры $p = [p_1, \dots, p_n]$ можно однозначно определить по входным данным $u(t)$ и данным измерений $y(t) = g(t, p)$.

Информация, содержащаяся в измерениях, является функцией от времени и вектора параметров $p = [p_1, \dots, p_n]$ вида:

$$y(t) = g(t, p), \quad (2)$$

где $g(t, p)$ – функция, связывающая модель с измерениями.

На практике *точные* экспериментальные данные $y(t)$ недоступны из-за ошибок измерений (например, настроек прибора по измерению данных). Обычно в наборе дискретных точек по времени $[t_1, \dots, t_N]$ экспериментальные данные имеют вид:

$$z(t_i) = y_i + v_i = g(t_i, p) + v_i, \quad (3)$$

где v_i – погрешность (ошибка) i -го измерения, которую можно рассматривать как случайную величину, имеющую нулевое математическое ожидание (Carson, Cobelli, 2008). Зачастую некоторая статистическая информация о погрешности известна заранее.

На рис. 1 приведен пример фармакокинетических данных: крестиками отмечены значения измеренных концентраций препарата, внутривенно введенного в момент времени $t = 0$. Данные z_1, \dots, z_n могут быть представлены в виде (3) с помощью функции $g(t, p) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}$ следующим образом:

$$z(t_i) = y_i + v_i = g(t_i, p) + v_i = A_1 e^{-\lambda_1 t_i} + A_2 e^{-\lambda_2 t_i} + v_i.$$

Известно (Carson, Cobelli, 2008), что если для точных данных $y(t)$ модель является неидентифицируемой, то набор $p = [p_1, \dots, p_n]$ невозможно определить единственным образом. А при использовании «зашумленных» данных вида (3) возможно получение данных, не имеющих никакого смысла.

В первом разделе статьи рассмотрены подходы к анализу идентифицируемости. Так как в определении идентифицируемости фигурируют функции $y(t)$ и $u(t)$, и отсутствует $x(t)$, то большинство методов основано на «исключении» $x(t)$ и получении соотношений между $y(t)$ и $u(t)$. Во втором разделе рассмотрены примеры анализа идентифицируемости математических моделей.

1. Обзор методов

1.1. Метод передаточной функции

Для динамических процессов, которые могут быть представлены с помощью линейной системы обыкновенных дифференциальных уравнений вида (4), чаще всего применим метод передаточной функции.

$$\begin{cases} \dot{x}(t) = A(p)x(t) + B(p)u(t) \\ y(t, p) = g(x(t, p); p) \end{cases}, \quad (4)$$

где $A(p)$ и $B(p)$ – постоянные матрицы. Функция измерений чаще всего имеет вид: $y(t, p) = C(p)x(t, p)$, где $C(p)$ – постоянная матрица.

Метод передаточной функции основан на использовании преобразования Лапласа и так называемой передаточной функции $H(s) = \frac{Y(s)}{U(s)}$, которая выражает связь между параметрами модели и наблюдаемыми параметрами (Cobelli, DiStefano, 1980; Audoly, D'Angio, 1983; Jacquez, Greif, 1985; Ben-Zvi et al., 2004; Carson, Cobelli, 2008). Здесь $U(s)$ и $Y(s)$ – преобразования Лапласа функций $u(t)$ и $y(t)$ соответственно. Метод заключается в анализе коэффициентов матрицы передаточной функции вида:

$$H(s, p) = [H_{ij}(s, p)] = \frac{Y_i(s, p)}{U_j(s)}, \quad (5)$$

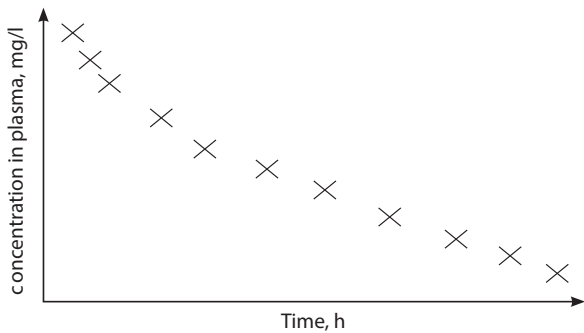


Fig. 1. Drug concentration in plasma.

где каждый элемент H_{ij} матрицы H является преобразованием Лапласа функций выхода $y_i(t, p)$ и входа $u_j(t)$. $H_{ij}(s, p)$ представляют собой рациональные полиномы с коэффициентами: $\beta_1^{ij}(p), \dots, \beta_n^{ij}(p)$ в числителе и $\alpha_1^{ij}(p), \dots, \alpha_n^{ij}(p)$ в знаменателе. Коэффициенты перед мономами представляются в виде комбинаций параметров $p = [p_1, \dots, p_n]$. Приравняв эти комбинации к неопределенным параметрам φ_j^{ij} , которые назовем наблюдаемыми величинами, и составим систему алгебраических уравнений вида:

$$\begin{cases} \beta_1^{11}(p) = \varphi_1^{11} \\ \dots \\ \alpha_n^{11}(p) = \varphi_{2n}^{11} \\ \dots \\ \beta_1^{rm}(p) = \varphi_1^{rm} \\ \dots \\ \alpha_n^{rm}(p) = \varphi_{2n}^{rm} \end{cases} \quad (6)$$

Если система (6) разрешима единственным образом относительно вектора параметров p , значит исходная математическая модель (4) является идентифицируемой. Таким образом, анализ идентифицируемости СОДУ сводится к вопросу разрешимости системы алгебраических уравнений вида (6). Во втором разделе рассмотрен фармакокинетический пример такого анализа.

1.2. Метод, основанный на разложении в ряд Тейлора

При анализе идентифицируемости можно использовать разложение функции выходных данных (2) в ряд Тейлора. Данный метод в отличие от метода передаточной функции применим и к системам нелинейных дифференциальных уравнений. Подробное описание данного метода можно посмотреть в работе (Carson, Cobelli, 2008).

Разложение в ряд Тейлора функции $y_i(t, p)$ в точке $t_0 = 0$ выглядит следующим образом: $y_i(t) = y_i(t_0) + t\dot{y}_i(t_0) + \frac{t^2}{2!}\ddot{y}_i(t_0) + \frac{t^3}{3!}\dddot{y}_i(t_0)$. Обозначим:

$$y_i^{(k)}(t_0, p) = \varphi_k^i, k = 0, 1, 2 \dots \quad (7)$$

Данный метод основан на замене в равенстве (7) левых частей на выражения, получаемые из исходной СОДУ (1). После такой замены получается система нелинейных алгебраических уравнений. Как и в методе передаточной функции, вопрос идентифицируемости математической

модели сводится к анализу разрешимости алгебраической системы. В разделе 2 рассмотрен пример такого анализа.

1.3. Метод, основанный на теории дифференциальной алгебры

Перепишем систему (1) в виде:

$$\begin{cases} \dot{x} - f(x(t), p) - \sum_{i=1}^n h(x(t), p)u_i(t) = 0 \\ y(t) - g(u(t), x(t), p) = 0 \end{cases} \quad (8)$$

Рассмотрим ее как систему алгебраических уравнений в переменных состояния $x(t)$, измерения $y(t)$, входных данных $u(t)$, а также соответствующих производных этих функций. Ниже приведена краткая схема анализа идентифицируемости.

1. Введем формальное упорядочение: будем считать, что $u < \dot{u} < \ddot{u} < \dots < y < \dot{y} < \ddot{y} < \dots < x < \dot{x} < \ddot{x} < \dots$. Если дифференциальные полиномы являются рациональными функциями, то необходимо привести их к общему знаменателю.

2. Проводится упорядочение всех уравнений системы (8) в соответствии со старшим слагаемым в каждом дифференциальном полиноме.

3. Проводится специальная процедура псевдоделения, исключая из рассматриваемых уравнений функции $x, \dot{x}, \ddot{x}, \dots$. Подробно данная процедура изложена в работах (Carson, Cobelli, 2001; Bellu et al., 2007; Meshkat et al., 2012). В результате такого алгоритма остается столько дифференциальных многочленов относительно $y(t), u(t)$, их производных, а также всевозможных комбинаций, сколько в исходной системе (8) выходных данных $y_i(t)$.

4. Составляем нелинейную систему алгебраических уравнений, приравнивая коэффициенты при каждом слагаемом дифференциальных полиномов к наблюдаемым параметрам (аналогично предыдущим методам). Анализ разрешимости системы проводится с использованием теории базисов Гребнера.

5. Вопрос идентифицируемости системы (8) сводится к вопросу разрешимости нелинейной алгебраической системы.

Данный метод привлекателен тем, что структура алгоритма позволяет автоматизировать его и проводить анализ идентифицируемости нелинейных динамических систем с использованием ЭВМ (Bellu et al., 2007; Meshkat et al., 2014). Подробное описание данного подхода можно посмотреть, например, в работах (Carson, Cobelli, 2001; Bellu et al., 2007; Meshkat et al., 2014).

1.4. Метод, основанный на теории графов

Иногда неидентифицируемую модель с помощью специальной замены можно привести к идентифицируемой. Ниже приведен алгоритм поиска так называемой рациональной репараметризации для линейных динамических систем (4). Замена такого типа крайне полезна в фармакокинетических приложениях (Meshkat, 2014). Этот метод позволяет найти такую рациональную замену, которая, во-первых, позволяет сохранить вид функции $x_1(t, p)$, так как чаще измерения доступны в виде $y_1(t) = x_1(t, p)$. Во-вторых, после такой замены переменных в качестве

искомых параметров выступают комбинации исходных параметров, обладающие физическим смыслом.

Метод заключается в сопоставлении линейной СОДУ графа G (Cobelli et al., 1976). По данному графу G с n вершинами и $m \leq 2n - 2$ дугами мы либо указываем конкретную замену, которая сведет нашу модель к идентифицируемой, либо говорим, что такой замены не существует. Пусть пространство параметров $P \subseteq \mathbb{R}^{2n-1}$ состоит из таких матриц, структура которых индуцируется графом G . Отображение $c: P \rightarrow \mathbb{R}^{2n-1}$ сопоставляет матрице $A \subseteq P$ вектор $(c_1(A), \dots, c_n(A), d_1(A), \dots, d_{n-1}(A))$, составленный из коэффициентов при характеристических многочленах. Отображение $c: P \rightarrow \mathbb{R}^{2n-1}$ называется двойным характеристическим полиномиальным отображением. На первом этапе алгоритма мы вычисляем $d = \dim(\text{Im}c)$ – размерность образа так называемого двойного характеристического полиномиального отображения. Если $d \neq m + 1$, то не существует замены переменных в необходимом для нас виде. Иначе:

а) ищем остовное дерево T из графа G с вершинами $j_1 \rightarrow i_1, \dots, j_{n-1} \rightarrow i_{n-1}$;

б) формируем матрицу инцидентности $E(T)$ таким образом, что первые $n-1$ столбцов отвечают ребрам T . Далее вычеркиваем первую строку и получаем матрицу E_1 . Другими словами, получили E_1 – матрицу размера $(n-1) \times (n-1)$, отвечающую ребрам T ;

в) определяем мономиальное масштабирование: $X_i = f_i(A)x_i$. Положим $f_1(A) = 1$.

Пусть также $r_i = (r_{1,i}, \dots, r_{n-1,i})^T$ – i -ый столбец матрицы $C_1 = E_1^{-1}$. Тогда $f_{i+1}(A) = a_{i,j_1}^{r_{1,i}} \dots a_{i,j_{n-1}}^{r_{n-1,i}}$;

г) заменяем элементы a_{ij} матрицы A на элементы $a_{ij}f_i(A)/f_j(A)$.

Сделаем некоторые замечания по описанному алгоритму. Во-первых, d может быть вычислено как с помощью определения ранга матрицы Якоби отображения c в общей точке, так и с помощью поиска обнуляющего идеала образа отображения c , при использовании базисов Грёбнера. Также следует отметить, что в некоторых случаях этот шаг можно пропустить: если G является индуктивно сильно связным графом, то условие $d = m + 1$ автоматически выполняется и нет необходимости вычислять d трудоемкими методами.

В случае существования идентифицируемой масштабируемой замены переменных, новая матрица будет иметь $n-1$ элементов, отвечающих остовному дереву T , равных единице, и оставшиеся внедиагональные элементы (их $m-n+1$) понимаются как новые параметры в репараметризованной системе. Получаемая с помощью такой замены модель является идентифицируемой относительно рациональных комбинаций исходных параметров. Более подробно этот алгоритм изложен в работе (Meshkat, 2014).

2. Примеры анализа идентифицируемости

2.1. анализ идентифицируемости фармакокинетической модели контроля глюкозы-инсулина методом передаточной функции
Рассмотрим модель контроля глюкозы-инсулина (9) (Carson, Cobelli, 2008). Следует отметить, что в рамках

камерного подхода к моделированию фармакокинетических процессов в качестве камер могут выступать не только органы, но и различные вещества внутри камеры (например, глюкозы и инсулина в крови).

$$\begin{cases} \frac{dx_1(t)}{dt} = -q_1x_1(t) - q_2x_2(t) + u(t), \\ \frac{dx_2(t)}{dt} = -q_3x_2(t) + q_4x_1(t), \\ y_1(t) = \frac{x_1(t)}{V_1}, \\ x_1(0) = x_2(0) = 0, \end{cases} \quad (9)$$

где $x_1(t)$, $x_2(t)$ – массы глюкозы (назовем глюкозу – 1-я камера) и инсулина (2-я камера), $u(t) = D \cdot \delta(t)$ – доза. $p = [q_1, q_2, q_3, q_4, V_1]$ – вектор неизвестных параметров. Пусть V_2 – объем второй камеры. Предположим, что в качестве данных измерений нам известна только концентрация глюкозы $y_1(t)$. Следует отметить, что на практике начальные данные системы (9) не могут быть равными нулю. В этом примере такие значения выбраны для удобства и никак не влияют на схему анализа.

Для определения идентифицируемости данной модели воспользуемся методом передаточной функции. Запишем передаточную функцию в виде: $H_1(s) = \frac{Y_1(s)}{U(s)}$.

Применяя преобразования Лапласа для уравнений системы (9), получим:

$$\begin{cases} sX_1(s) - x_1(0) = -q_1X_1(s) - q_2X_2(s) + D \\ sX_2(s) - x_2(0) = -q_3X_2(s) + q_4X_1(s) \\ Y_1(s) = \frac{X_1(s)}{V_1} \end{cases}$$

Прописными буквами обозначены преобразования Лапласа для соответствующих функций. Находим выражение для $Y_1(s)$ путем выражения $X_1(s)$ из первых двух уравнений в последней системе. Получим следующее выражение:

$$Y_1(s) = \frac{X_1(s)}{V_1} = \frac{s^2 + (q_1 + q_3)s + q_1q_3 + q_2q_4}{DV_1(s + q_3)}$$

Теперь подставляем значения $Y_1(s)$ и $U(s)$ в формулу передаточной функции:

$$H_1(s) = \frac{Y_1(s)}{U(s)} = \frac{\frac{s + q_3}{V_1}}{s^2 + (q_1 + q_3)s + q_1q_3 + q_2q_4} = \frac{\varphi_2 s + \varphi_1}{s^2 + \varphi_4 s + \varphi_3}$$

$H_1(s)$ – рациональный многочлен с коэффициентами, зависящими от неизвестных параметров. Записываем систему алгебраических уравнений (10), приравнявая коэффициенты при мономах к наблюдаемым параметрам φ_i :

$$\begin{cases} \frac{q_3}{V_1} = \varphi_1 \\ \frac{1}{V_1} = \varphi_2 \\ q_1q_3 + q_2q_4 = \varphi_3 \\ q_1 + q_3 = \varphi_4 \end{cases} \quad (10)$$

Система (10) не имеет единственного решения, следовательно, модель априорно неидентифицируема. Если рассмотреть вместо параметров q_2 и q_4 – их произведение,

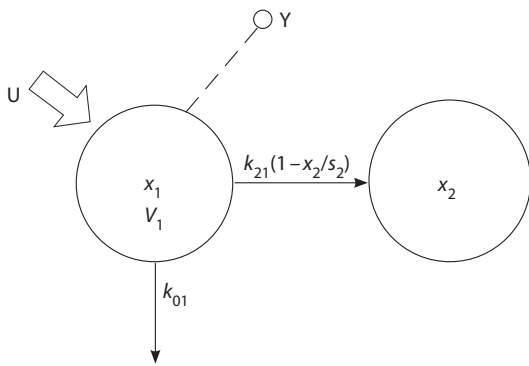


Fig. 2. A nonlinear two-compartment pharmacokinetic model. x_1 , drug amount in blood; x_2 , drug amount in the peripheral compartment; V_1 , volume of the first compartment.

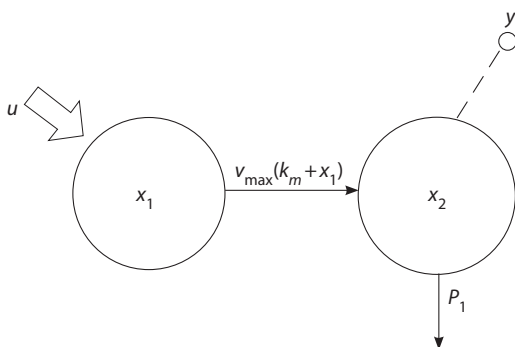


Fig. 3. A nonlinear two-compartment model with Michaelis-Menten kinetics. x_1, x_2 , drug amounts in the blood and tissues; u , the drug loading function; y , measurements in the second compartment.

считая их одним параметром, то система разрешится, и набор параметров, определяемых единственным образом, имеет вид: $p = [q_1, q_3, V_1, q_2q_4]$.

Предположим, что в качестве данных измерений известна также концентрация инсулина, т.е. $y_2(t) = \frac{x_2(t)}{V_2}$. Для таких данных измерений также выпишем передаточную функцию. Применяя преобразование Лапласа для соответствующих уравнений системы, получим выражение для $Y_2(s)$, передаточная функция $H_2(s)$ имеет следующий вид:

$$H_2(s) = \frac{Y_2(s)}{U(s)} = \frac{q_4}{s^2 + (q_1 + q_3)s + q_1q_3 + q_2q_4}$$

К системе алгебраических уравнений (10) добавляется еще одно уравнение, связывающее наблюдаемые данные с данными измерений: $\frac{q_4}{V_2} = \varphi_5$. Тогда набор идентифицируемых параметров имеет вид: $p = [q_1, q_3, V_1, q_2V_2, \frac{q_4}{V_2}]$.

Следует отметить, что если известно значение V_2 (или, например, что $V_1 = V_2$), модель станет идентифицируемой. Таким образом, можно заключить, что добавление инфор-

мации об исследуемом процессе делает вычисление параметров более надежным, что согласуется с интуитивным представлением.

2.2. анализ идентифицируемости двухкамерной фармакокинетической модели методом разложения в ряд Тейлора

Рассмотрим нелинейную двухкамерную фармакокинетическую модель (Carson, Cobelli, 2008) (рис. 2).

Предположим, что препарат вводится в момент времени $t = 0$ ($U(t) = D\delta(t)$) в плазму крови (камера 1). В этой же камере измеряется концентрация $y(t) = \frac{x_1(t)}{V_1}$. Элиминация препарата происходит линейно и характеризуется параметром k_{01} , переход вещества в недоступную для измерений камеру 2 (группа органов с низким кровоснабжением) осуществляется «ленгмюровской кинетикой насыщения» – со скоростью $k_{21}(1 - \frac{x_2(t)}{s_2})$. Математическая модель имеет следующий вид:

$$\begin{cases} \dot{x}_1(t) = -k_{01}x_1(t) - k_{21}\left[1 - \frac{x_2(t)}{s_2}\right]x_1(t) + U(t) \\ \dot{x}_2(t) = k_{21}\left[1 - \frac{x_2(t)}{s_2}\right]x_1(t) \\ y(t) = \frac{x_1(t)}{V_1} \end{cases} \quad (11)$$

Для определения идентифицируемости данной модели применим метод разложения в ряд Тейлора. Для этого нужно определить значения $y(0)$, $\dot{y}(0)$, $\ddot{y}(0)$ и $\ddot{\ddot{y}}(0)$.

Из описания модели имеем $x_1(0) = D$. Для $\dot{x}_1(0)$, учитывая структуру системы (11) и тот факт, что $x_2(0) = 0$, получаем $\dot{x}_1(0) = -(k_{01} + k_{21})x_1(0)$. Аналогично $\ddot{x}_1(0) = -(k_{01} + k_{21})\dot{x}_1(0) + \frac{k_{21}}{s_2}k_{21}\dot{x}_1^2(0)$ и $\ddot{\ddot{x}}_1(0) = -\frac{k_{21}^2}{s_2}\left[\frac{k_{21}}{s_2}x_1^2(0) - 3\dot{x}_1(0)\right]x_1(0) - (k_{01} + k_{21})\ddot{x}_1(0)$.

Далее легко найти $y(0)$, $\dot{y}(0)$, $\ddot{y}(0)$ и $\ddot{\ddot{y}}(0)$ из полученных соотношений, разделив каждое из уравнений на V_1 . Согласно алгоритму выражения для $y(0)$, $\dot{y}(0)$, $\ddot{y}(0)$ и $\ddot{\ddot{y}}(0)$, приравняем к наблюдаемым параметрам φ_j для получения системы алгебраических уравнений относительно искомым параметров:

$$\begin{cases} \frac{D}{V_1} = \varphi_1 \\ \frac{1}{V_1}[-(k_{01} + k_{21})x_1(0)] = \varphi_2 \\ \frac{1}{V_1}\left[-(k_{01} + k_{21})\dot{x}_1(0) + \frac{k_{21}}{s_2}k_{21}\dot{x}_1^2(0)\right] = \varphi_3 \\ \frac{1}{V_1}\left[-\frac{k_{21}^2}{s_2}\left[\frac{k_{21}}{s_2}x_1^2(0) - 3\dot{x}_1(0)\right]x_1(0) - (k_{01} + k_{21})\ddot{x}_1(0)\right] = \varphi_4 \end{cases}$$

Система имеет единственное решение, а значит, рассматриваемая модель является идентифицируемой.

2.3. анализ идентифицируемости модели методом, основанным на теории дифференциальной алгебры

Рассмотрим двухкамерную фармакокинетическую модель, описывающую распространение препарата с нелинейной

кинетикой в крови и тканях (Bellu et al., 2007) (рис. 3). Система (1) для такой модели имеет вид:

$$\begin{cases} \dot{x}_1 = -\frac{v_{\max}}{k_m + x_1} x_1 + u \\ \dot{x}_2 = \frac{v_{\max}}{k_m + x_1} x_1 - p_1 x_2 \\ y = x_2 \end{cases} \quad (12)$$

Данные измерений известны в тканях, т. е. функция выходных данных имеет вид: $y = x_2$, p_1 – константа скорости, v_{\max} и k_m – классические параметры кинетики Михаэлиса-Ментен. Неизвестный набор параметров $p = [p_1, v_{\max}, k_m]$.

Проведем анализ идентифицируемости такой модели методом из раздела 2.3.

1. Формальное упорядочение слагаемых имеет вид: $u < y < \dot{u} < \ddot{u} < \dot{y} < \ddot{y} < x_1 < x_2 < \dot{x}_1 < \dot{x}_2$.

2. Упорядочение полиномов имеет следующий вид:

$$\begin{cases} y - x_2 = 0 \\ \dot{x}_1(k_m + x_1) + v_{\max}x_1 - u(k_m + x_1) = 0 \\ \dot{x}_2(k_m + x_1) - v_{\max}x_1 + p_1x_2(k_m + x_1) = 0 \end{cases}$$

3. После применения процедуры псевдоделения получаем одно соотношение на данные входа и выхода:

$$\begin{aligned} & -\ddot{y}k_m^2v_{\max}^2 - \dot{y}^3k_mv_{\max} + \dot{y}^2(-3p_1k_mv_{\max}y + k_m(uv_{\max} + 2v_{\max}^2)) + \\ & -\dot{y}(-3p_1^2k_mv_{\max}y^2 + p_1(-k_m^2v_{\max}^2 + k_m)(2v_{\max}uy + 4v_{\max}^2y)) + \\ & + k_m(-2v_{\max}^2u - v_{\max}^3) - p_1^3k_mv_{\max}y^3 + p_1^2k_m(uv_{\max}y^2 + 2v_{\max}^2y^2) + \\ & + p_1k_m(-2uv_{\max}y - v_{\max}^3y) + k_mv_{\max}^3u = 0. \end{aligned}$$

4. Нелинейная система алгебраических уравнений имеет вид:

$$\begin{cases} k_mv_{\max} = \gamma\beta \\ -p_1 = -\alpha \\ v_{\max} = \beta \\ -p_1^2 = -\alpha^2 \\ -p_1v_{\max} = -\alpha\beta \\ v_{\max}^2 = \beta^2 \\ -p_1^3 = -\alpha^3 \\ -p_1^2v_{\max} = -\alpha^2\beta \\ -p_1v_{\max}^2 = -\alpha\beta^2 \\ -k_mp_1v_{\max} = -\gamma\alpha\beta \end{cases} \quad (13)$$

5. Система алгебраических уравнений (13) разрешима единственным образом относительно искомого набора параметров $p = [p_1, v_{\max}, k_m]$, а значит, исходная фармакокинетическая модель (12) является идентифицируемой.

2.4. Пример замены переменных, приводящей неидентифицируемую модель к идентифицируемой

Рассмотрим фармакокинетическую трехкамерную модель (Meshkat, 2014) с введением препарата и взятием проб из первой камеры ($y = x_1$) и элиминацией из всех трех камер

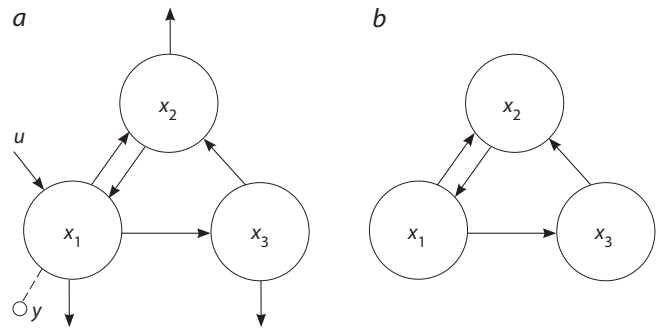


Fig. 4. (a) A three-compartmental pharmacokinetic linear model; u , the drug loading function; y , measurements in the first compartment. (b) A graph with three vertices.

(рис. 4). Соответствующая система дифференциальных уравнений имеет вид:

$$\begin{cases} \dot{x}_1 = a_{11}x_1 + a_{12}x_2 \\ \dot{x}_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ \dot{x}_3 = a_{31}x_1 + a_{33}x_3 \\ y = x_1 \end{cases}$$

Согласно алгоритму из 1.4 первым шагом является построение остовного дерева, в нашем случае остовное дерево соответствует ребрам a_{23} и a_{12} . Второй шаг состоит

в построении матрицы инцидентности $E(T) = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$.

Вычеркиваем 1-ю строку матрицы $E(T)$, получим E_1 и находим $E_1^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Делаем замену согласно формуле

в пункте (в) раздела 1.4. Получаем явный вид рациональной репараметризации для этой модели: $X_1 = x_1, X_2 = a_{12}x_2, X_3 = a_{12}a_{23}x_3$.

Итак, полученная модель является идентифицируемой и имеет следующий вид:

$$\begin{cases} \dot{X}_1 = a_{11}X_1 + X_2 \\ \dot{X}_2 = a_{12}a_{21}X_1 + a_{22}X_2 + X_3 \\ \dot{X}_3 = a_{12}a_{31}a_{23}X_1 + a_{33}X_3 \\ y = X_1 \end{cases}$$

При определенной структуре модели и данных на входе и выходе анализ априорной идентифицируемости позволяет избежать лишних экспериментов, если искомые параметры являются неидентифицируемыми. Проведение такого анализа позволяет определить комбинации параметров рассматриваемой модели, относительно которых модель является идентифицируемой.

В случае неидентифицируемости модели следует предпринять один из следующих шагов:

- 1) добавить данные о характере эксперимента, привлечь дополнительные данные измерений;
- 2) упростить исходную модель, меняя ее структуру. Добавить априорную информацию о некоторых параметрах модели;
- 3) искать замену переменных, приводящую исходную модель к идентифицируемой.

Acknowledgments

This work was supported by the Ministry of Education and Science of the Russian Federation, contract No. 107 “Design of program tools for the study of numerical solution of direct and inverse problems in pharmacokinetics and epidemiology” and by the Ministry of Education and Science of the Kazakhstan Republic, contract No. 1746/GF4 “Theory and numerical methods of the solution of inverse and ill-posed problems in natural science”.

Conflict of interest

The authors declare no conflict of interest.

References

- Audoly S., D’Angio L. On the identifiability of linear compartmental system: a revisited transfer function approach based on topological properties. *Mathematical Biosciences*. 1983;66:201-228.
- Bellman R., Astrom K. On structural identifiability. *Mathematical Biosciences*. 1970;7(3):329-339.
- Bellu G., Saccomani P., Audoly S., D’Angio L. Daisy: a new software tool to test global identifiability of biological and physiological system. *Comput Methods Programs Biomed*. 2007;88(1):52-61. DOI 10.1016/j.cmpb.2007.07.002
- Ben-Zvi A., McLellan P.J., McAuley B.K. Identifiability of linear time-invariant differential-algebraic systems. *Industrial Engineering Chemistry Res*. 2004;43(8):1251-1259.
- Brown R. Compartmental system analysis: state of the Art. *IEEE Transactions on Biomedical Engineering*. 1980;27(1):1-38.
- Brown R.F. Identifiability: role in design of pharmacokinetic experiments. *IEEE Transactions on Biomedical Engineering*. 1982;29:49-54.
- Carson E., Cobelli C. *Modelling Methodology for Physiology and Medicine*. San Diego: Academic Press, 2001.
- Carson E., Cobelli C. *Introduction to Modelling in Physiology and Medicine*. San Diego: Academic Press, 2008.
- Cobelli C., DiStefano III J. Parameter and structural identifiability concepts and ambiguities: a Critical review and analysis. *Amer. J. Physiology-Regulatory, Integrative and Comparative Physiology*. 1980;23(9):7-24.
- Cobelli C., Lepschy A., Jacur G.R. Identifiability of compartmental systems and related structural properties. *Mathematical Biosciences*. 1976;48:1-18.
- Goodwin G.C., Payne R.L. *Dynamic system identification: experiment design and data analysis*. N.Y.: Academic Press, 1977.
- Jacquez J.A., Greif P. Numerical parameter identifiability and estimability: integrating identifiability, estimability and optimal design. *Mathematical Biosciences*. 1985;77:201-227.
- Meshkat N. Identifiable reparametrizations of linear compartment models. *J. Symbolic Computation*. 2014;63:46-67. DOI 10.1016/j.jsc.2013.11.002
- Meshkat N., Anderson C., DiStefano III J. Alternative to Ritt’s Pseudodivision for finding the input-output equations of multi-output models. *Mathemat. Biosciences*. 2012;239(1):117-123. DOI 10.1016/j.mbs.2012.04.008
- Meshkat N., Eisenberg M., DiStefano III J. On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: A novel web implementation. *Plos One*. 2014;9(10). DOI 10.1371/journal.pone.0110261
- Tunali T., Tarn T.J. New Results for Identifiability of Nonlinear Systems. *IEEE Transactions on Automatic Control*. 1987;32(2):146-154.
- Walter E., Lecourtier Y., Happel J., On the structural output distinguishability of parametric models, and its relation with structural identifiability. *IEEE Trans. Aut. Control*. 1984;29:56-57.