

УДК 519.711

## **ВЫЧИСЛИТЕЛЬНЫЕ АСПЕКТЫ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ РАСПРОСТРАНЕНИЯ ТУБЕРКУЛЕЗА**

**В.Н. Леоненко, К.К. Логинов**

В работе рассматриваются подходы, позволяющие повысить быстродействие моделирующей программы для индивидуум-ориентированной модели распространения туберкулеза. Описывается процесс оптимизации вычислительных алгоритмов, а также использование распределенных вычислений на основе системы MONC и параллельных вычислений с применением технологии OpenMP. Приведены результаты вычислительных экспериментов.

**Ключевые слова:** имитационное моделирование, эпидемиология, распределенные вычисления, параллельное программирование.

### **Введение**

Одним из инструментов исследований закономерностей распространения туберкулеза является метод математического моделирования. В настоящее время разработан широкий класс детерминированных моделей, описывающих распространение туберкулеза и контроль за этим заболеванием [1, 2]. Вместе с тем, практически отсутствуют модели, опирающиеся на стохастический подход и индивидуально-ориентированное описание неинфицированных, инфицированных и больных индивидуумов.

Индивидуум-ориентированные модели позволяют детально провести формализацию принципов передачи и развития туберкулеза и исследовать динамику заболеваемости туберкулезом от различных факторов. Одними из первых в этом направлении являются модели, представленные в работах [3, 4]. В работе [3] предложена стохастическая модель в форме высокоразмерной системы рекуррентных уравнений с дискретным временем, учитывающая зависимость развития заболевания от возраста индивидуумов. Модель, построенная в работе [4], имеет следующие отличительные свойства: используется непрерывное время, продолжительность пребывания индивидуумов в различных стадиях болезни задается с помощью случайных величин с произвольными законами распределения, вероятность контактов индивидуумов считается пропорциональной произведению численностей когорт.

В данной работе рассматривается стохастическая модель [5], созданная на основе дифференциальной модели динамики туберкулеза [2]. В индивидуум-ориентированной модификации модели, в отличие

от предшествующих работ [3, 4], учитывается влияние продолжительности болезни на вероятность гибели индивидуумов. Введение индивидуальных параметров привело к тому, что объем обрабатываемых данных повысился и время, затрачиваемое на моделирование динамики популяции, существенно возросло. Основной целью данной работы является рассмотрение подходов, позволяющих повысить быстродействие моделирующих программ путем оптимизации исходного алгоритма и более эффективного применения имеющихся компьютерных мощностей и вычислительных технологий.

### Описание модели

Кратко опишем особенности модификаций модели, рассматриваемых в данной работе (полное формальное описание приведено в [5]).

**Базовая модель.** В популяции выделяется шесть когорт населения: восприимчивые индивиды ( $S$ ), инфицированные индивидуумы ( $L$ ), невыявленные больные индивидуумы без бактериовыделения ( $D$ ), невыявленные больные индивидуумы с бактериовыделением ( $B$ ), выявленные больные индивидуумы без бактериовыделения ( $D_0$ ) и выявленные больные индивидуумы с бактериовыделением ( $B_0$ ). Система уравнений модели имеет следующий вид:

$$\begin{aligned} x_S(t) &= \hat{x}_S(t) - u_{S,L \cup D}(t) + f_1(t), \\ x_L(t) &= \hat{x}_L(t) + u_{S,L}(t) + u_{D,L}(t) + u_{D_0,L}(t) - u_{L,D}(t) + f_2(t), \\ x_D(t) &= \hat{x}_D(t) + u_{S,D}(t) + u_{L,D}(t) + u_{B,D}(t) - u_{D,L}(t) - u_{D,B}(t) - u_{D,D_0}(t) + f_3(t), \\ x_B(t) &= \hat{x}_B(t) + u_{D,B}(t) - u_{B,D}(t) - u_{B,B_0}(t) + f_4(t), \\ x_{D_0}(t) &= \hat{x}_{D_0}(t) + u_{D,D_0}(t) + u_{B_0,D_0}(t) - u_{D_0,B_0}(t) - u_{D_0,L}(t) + f_5(t), \\ x_{B_0}(t) &= \hat{x}_{B_0}(t) + u_{B,B_0}(t) + u_{D_0,B_0}(t) - u_{B_0,D_0}(t) + f_6(t), t = 0, 1, \dots \end{aligned} \quad (1)$$

Система (1) дополняется начальными данными при  $t = 0$ . Здесь  $x_Y(t)$  – численность индивидуумов когорты  $Y$  в момент времени  $t$ ,  $\hat{x}_Y(t)$  – численность индивидуумов когорты  $Y$ , доживших от момента времени  $t-1$  до момента времени  $t$ , случайные величины вида  $u_{Y,Z}(t)$  описывают переход из когорты  $Y$  в когорту  $Z$  в течение промежутка времени  $(t-1, t]$ ,  $Y, Z \in \{S, L, D, B, D_0, B_0\}$ . Слагаемые  $f_i(t)$ ,  $i = \overline{1, 6}$  задают приток населения в соответствующие когорты за промежуток времени  $(t-1, t]$ .

Для описания процессов инфицирования индивидуумов и спонтанной активации болезни используется схема, предложенная в работе [3]. Инфицирование происходит при одновременном посещении больными и здоровыми индивидуумами мест возможного контакта. Считается, что за единицу времени индивидуумы из когорт  $B$ ,  $B_0$  посещают в совокупности  $\xi_t^{(B)}$  и  $\xi_t^{(B_0)}$  мест соответственно, где  $\xi_t^{(B)} = \sum_{i=1}^{x_B(t)} \Psi_{it}^{(B)}$ ,  $\xi_t^{(B_0)} = \sum_{i=1}^{x_{B_0}(t)} \Psi_{it}^{(B_0)}$ ,  $\Psi_{it}^{(B)} \geq 0$ ,  $\Psi_{it}^{(B_0)} \geq 0$  – взаимно независимые, одинаково распределенные случайные величины, не зависящие от  $x_B(t)$ ,  $x_{B_0}(t)$ , причем  $E(\Psi_{it}^{(B)}) = r > 0$ ,  $E(\Psi_{it}^{(B_0)}) = k \cdot r$ ,  $k \in (0; 1)$ . Вероятности того, что индивидуумы из когорт  $S$ ,  $L$  посетят одно из этих мест, обозначим через  $\lambda_S \in (0; 1)$  и  $\lambda_L \in (0; 1)$  соответственно. Вероятности заражения индивидуумов после контакта с больным индивидуумом обозначим через  $\delta_S \in (0; 1)$  и  $\delta_L \in (0; 1)$  соответственно. Тогда вероятности того, что индивидуум из конкретной когорты  $S$ ,  $L$  будет инфицирован в течение суток, при фиксированном значении  $\xi_t$  задаются следующими формулами:  $\mu_t^{(S)} = 1 - (1 - \lambda_S \delta_S)^{\xi_t^{(B)} + \xi_t^{(B_0)}}$ ,  $\mu_t^{(L)} = 1 - (1 - \theta_L)(1 - \lambda_L \delta_L)^{\xi_t^{(B)} + \xi_t^{(B_0)}}$ , где  $\theta_L$  – вероятность спонтанной активации болезни для индивидуумов когорты  $L$ . Данные величины определяют количество инфицированных на интервале  $(t-1, t]$ .

**Индивидуум-ориентированная модель.** В популяцию вводятся когорты  $R$  и  $R_0$ , которые используются для описания индивидуумов, находящихся в состоянии ремиссии. Считается, что в когорту  $R$  поступают самоизлечившиеся индивидуумы из когорты  $D$ , а в когорту  $R_0$  – вылеченные индивидуумы из когорты  $D_0$ . Переходы индивидуумов из  $D$  и  $D_0$  в  $L$  заменяются на переходы из  $D$  в  $R$  и из  $D_0$  в  $R_0$ . Предполагается, что больные индивидуумы после перехода в состояние ремиссии остаются инфицированными и существует риск повторного развития заболевания. Это означает, что возможны переходы индивидуумов из  $R$  в  $D$  и из  $R_0$  в  $D_0$ .

Для индивидуумов когорт  $D$ ,  $B$ ,  $D_0$ ,  $B_0$ ,  $R$  и  $R_0$  вводится параметр  $\Delta_x(t) = \max\{0, \pi_x(t) - \sigma_x(t)\}$ ,  $t = 0, 1, 2, \dots$ , отражающий влияние продолжительности болезни индивидуума на вероятность его дожития от момента времени  $t-1$  до  $t$ . Здесь  $\pi_x(t) = \gamma_D \omega_{D,x}(t) + \gamma_B \omega_{B,x}(t) + \gamma_{D_0} \omega_{D_0,x}(t) + \gamma_{B_0} \omega_{B_0,x}(t)$  – величина,

отражающая негативное влияние заболевания на продолжительность жизни индивидуума  $x$ ,  $\sigma_x(t) = \delta_R \omega_{R,x}(t) + \delta_{R_0} \omega_{R_0,x}(t)$  – величина, отражающая снижение эффекта воздействия заболевания на продолжительность жизни индивидуума  $x$  за счет его пребывания в состоянии ремиссии,  $\gamma_D, \gamma_B, \gamma_{D_0}, \gamma_{B_0}, \delta_R, \delta_{R_0}$  – положительные константы,  $\omega_{D,x}(t), \omega_{B,x}(t), \omega_{D_0,x}(t), \omega_{B_0,x}(t), \omega_{R,x}(t), \omega_{R_0,x}(t)$  – продолжительности пребывания индивидуума  $x$  в соответствующих когортах до момента времени  $t$  включительно. Вероятность дожития от момента времени  $t-1$  до  $t$  для индивидуума  $x$  из когорты  $H$  при заданной  $\Delta_x(t)$  равна  $\rho_{H,x}(t) = \rho_H e^{-\Delta_x(t-1)}$ ,  $H \in \{D, B, D_0, B_0, R, R_0\}$  (в отличие от базовой модели, где  $\rho_{H,x}(t) \equiv \rho_H = \text{const}$ ).

### **Повышение быстродействия работы моделирующих программ**

Рассмотрим далее способы ускорения работы моделирующих программ, позволяющих исследовать динамику численностей когорт в зависимости от параметров модели.

**Оптимизация вычислительных алгоритмов.** Рассмотрим индивидуум-ориентированную модификацию модели. Для всех вычислительных алгоритмов, использованных в моделирующей программе, справедливо следующее.

1. Когорты  $S$  и  $L$  полностью характеризуются своими численностями.
2. Динамика когорт  $S$  и  $L$  моделируется розыгрышем общей численности этих когорт на каждом шаге с помощью заданных вероятностей дожития индивидуумов.

Основные сложности возникают при моделировании динамики когорт, члены которых обладают индивидуальными параметрами.

Пусть  $T = \{D, B, D_0, B_0, R, R_0\}$ . Наиболее простым является итеративный вариант моделирования всех взаимодействий, выглядящий следующим образом.

#### **Алгоритм № 1.**

1. Для каждой когорты множества  $T$  создается список, содержащий индивидуальные параметры членов когорты.
2. Дожитие и переходы индивидуумов когорт множества  $T$  **в течение суток** (т.е. в течение одной итерации цикла) разыгрываются для каждого из этих индивидуумов последовательно в процессе прохода по спискам параметров.

Заметим, что вследствие линейности процессов дожития и переходов между когортами для индивидуумов когорт  $D, B, D_0, B_0$  судьбу каждого такого индивидуума можно разыгрывать независимо. Тогда алгоритм моделирования принимает следующий вид.

#### **Алгоритм № 2.**

1. Для когорт  $R$  и  $R_0$  создаются списки, содержащие индивидуальные параметры членов когорт.
2. Процессы дожития и переходов индивидуумов когорт  $R$  и  $R_0$  **в течение суток** разыгрываются для каждого из этих индивидуумов последовательно путем прохода по спискам параметров.
3. Процессы дожития и переходов индивидуумов когорт  $D, B, D_0, B_0$  **вплоть до их гибели или выздоровления** разыгрываются для каждого индивидуума с помощью функции розыгрыша судьбы отдельно взятого индивидуума. В процессе розыгрыша принадлежность индивидуума к заданной когорте в заданный момент времени регистрируется в массиве численностей индивидуумов множества  $T$ . В случае перехода рассматриваемого индивидуума в состояние ремиссии значение его параметра передается в качестве выходного параметра функции и помещается в массив параметров индивидуумов когорт  $R$  и  $R_0$ .

Формальное описание алгоритмов №№ 1, 2 для базовой модели и алгоритма № 1 для индивидуум-ориентированной модели приведено в работе [5].

Функция розыгрыша судьбы отдельно взятого индивидуума реализована в двух вариантах.

- (а) Итеративный вариант. Судьба индивидуума разыгрывается итеративно, с шагом в одни сутки (на каждом шаге с помощью метода Монте-Карло моделируется, погибнет ли индивидуум на данном шаге, и если нет, то перейдет ли он в другую когорту).
- (б) Вариант со скачками. На основании значений вероятностей гибели индивидуума и перехода его в другую когорту разыгрывается момент ближайшей смены статуса рассматриваемого индивидуума, после чего внутри функции происходит «скачок» во времени до этого момента с одновременным пересчетом значения индивидуального параметра.

Для сравнения быстродействия работы моделирующих программ проводился вычислительный эксперимент (табл. 1). Расчеты производились для двух наборов параметров на персональном компьютере Intel Core2 Quad 2,66 ГГц, объем оперативной памяти – 3 Гб. Отрезок моделирования – 5500 суток. Получаем, что моделирующая программа, выполненная на основе алгоритма №2 и функцией розыгрыша

судьбы в варианте (б), производит расчеты с первым набором параметров примерно на 70% быстрее, а со вторым набором – на 55% быстрее по сравнению с исходным вариантом алгоритма.

Заметим, что реализация розыгрыша судьбы индивидуума в виде отдельной функции также облегчает модификацию программного кода при необходимости изменения законов протекания инфекции (в перспективе возможна реализация выбора из нескольких модулей протекания инфекции и схем заражения для разных типов заболеваний).

	Время работы на наборе № 1, с	Время работы на наборе № 2, с
Алгоритм №1	1611,39	2468,01
Алгоритм №2 (а)	970,14	1636,69
Алгоритм №2 (б)	478,43	1111,56

Таблица 1. Быстродействие моделирующей программы с разными алгоритмами

### Распределенные вычисления с использованием системы MONC

Система MONC [6] (сокращение от Monte Carlo), разработанная М.А. Марченко (Институт вычислительной математики и математической геофизики СО РАН, г. Новосибирск), является универсальной высокопроизводительной системой параллельных вычислений для методов Монте-Карло на базе сети персональных компьютеров. Она распределяет независимые копии задания по персональным компьютерам в сети, отдает команды на их исполнение, следит за ходом выполнения заданий и по завершению выполнения заданий выполняет копирование и усреднение файлов с результатами расчетов.

Был проведен вычислительный эксперимент с использованием системы MONC на комплексе из четырех компьютеров с ОС Windows XP. Исполняемым приложением была модификация моделирующей программы для индивидуум-ориентированной модели, доработанная с учетом требований, предъявляемых системой (см. [6]). Производилось нахождение оценок математических ожиданий численностей когорт по ста реализациям для двух различных наборов параметров (отрезок моделирования – 5500 суток). Сначала измерялось время выполнения расчетов на каждом из компьютеров по отдельности, затем общее количество моделируемых реализаций разделялось по компьютерам пропорционально их быстродействию. После окончания расчетов на всех машинах система MONC усредняла полученные данные и сохраняла результат.

Таким образом, в данном случае запуск модели с помощью системы MONC на комплексе из четырех компьютеров позволяет сократить время выполнения задания примерно на 66% для обоих наборов параметров по сравнению с запуском на самом быстром из имеющихся компьютеров (О94-11).

Сетевое имя	Процессор	Оперативная память	Время работы на наборе № 1, с	Время работы на наборе № 2, с
О94-31	Intel Pentium 4, 3,2 ГГц	512 Мб	5156,7	17307,2
О91-01	AMD Athlon 64 x2 Dual Core Processor 5200+, 2,61 ГГц	1 Гб	4958,23	16755,6
О94-09	Intel Pentium 4, 3 ГГц	512 Мб	5826,1	19228,4
О94-11	Intel Core 2 Duo E7200, 2,53ГГц	1 Гб	3679,15	12031,9

Таблица 2. Характеристики компьютеров, задействованных для расчетов с использованием системы MONC

Сетевое имя	Набор № 1		Набор № 2	
	Кол-во реализаций	Время выполнения расчетов, с	Кол-во реализаций	Время выполнения расчетов, с
О94-31	23	1216	23	3885
О91-01	24	1233	24	4073
О94-09	21	1183	20	3850
О94-11	32	1171	33	4019
<b>Итого</b>	<b>100</b>	<b>1233</b>	<b>100</b>	<b>4073</b>

Таблица 3. Результаты вычислительного эксперимента по применению системы MONC

### Параллельные вычисления с использованием технологии OpenMP

Технология OpenMP (Open Multi-Processing) представляет собой набор директив компилятора, функций и переменных окружения, позволяющих создавать на базе последовательных программ многопоточные приложения, предназначенные для выполнения на многопроцессорных системах с единой па-

мятью. Важным достоинством OpenMP является то, что исходная программа не переписывается заново, а модифицируется добавлением в текст соответствующих директив [5].

К моменту написания статьи реализована поддержка технологии OpenMP в моделирующей программе для базовой модели с алгоритмом №1. Проводился вычислительный эксперимент с двумя наборами параметров, расчеты выполнялись на компьютере Intel Core 2 Duo E7400 2.8ГГц с объемом оперативной памяти 2 Гб. Отрезок моделирования – 5500 суток.

Количество потоков	Время работы на наборе № 1, с	Время работы на наборе № 2, с
1	129,968	191,093
2	71,578	99,515

Таблица 4. Результаты испытаний программы с поддержкой технологии OpenMP

Проведение вычислений в одном потоке соответствует выполнению последовательной версии моделирующей программы. Таким образом, применение технологии OpenMP при выполнении программы на двухъядерном процессоре дает экономию времени в 45% для первого набора параметров и 48% для второго набора параметров по сравнению с последовательным вычислением.

### Заключение

В работе были описаны подходы, которые позволили существенно увеличить скорость работы моделирующей программы. Это особенно актуально в связи с прогнозируемым усложнением модели (введением возрастной структуры в популяции и реализацией более детального механизма заражения). Также важной задачей является освоение моделирования на высокопроизводительных ЭВМ с распределенной памятью, в связи с чем планируется адаптировать программу под распределенные вычисления на основе технологии MPI.

Авторы благодарят своего научного руководителя профессора Н. В. Перцева (ОФ ИМ им. С.Л. Соболева СО РАН) за постановку задачи и обсуждение результатов работы.

### Литература

1. Авилов К. К., Романюха А. А. Математические модели распространения и контроля туберкулеза (обзор) // Математическая биология и биоинформатика. – 2007. – Т. 2. – № 2. – С. 188–318.
2. Perelman M.I., Marchuk G.I., Borisov S.E., et. al. Tuberculosis epidemiology in Russia: the mathematical model and data analysis // Russ. J. Numer. Anal. Math. Modelling. – 2004. – V. 19. – № 4. – P. 305–314.
3. Перцев Н.В., Романюха А.А., Касаткина В.С. Нелинейная стохастическая модель распространения туберкулеза // Системы управления и информационные технологии. – 2008. – № 1.2 (31). – С. 246–250.
4. Перцев Н.В., Пичугин Б.Ю. Индивидуум-ориентированная стохастическая модель распространения туберкулеза // Сибирский журнал индустриальной математики. – 2009. – Т. 12. – № 2(38). – С. 97–110.
5. Pertsev N.V., Leonenko V.N. Stochastic individual-based model of spread of tuberculosis // Russ. J. Numer. Anal. Math. Modelling. – 2009. – V. 24. – № 4. – P. 341–360.
6. Марченко М.А. Комплекс программ MONC для распределенных вычислений методом Монте-Карло. [Электронный ресурс]. – Режим доступа: [http://osmf.sssc.ru/~mam/monc\\_rus.htm](http://osmf.sssc.ru/~mam/monc_rus.htm), своб.
7. Антонов А.С. Параллельное программирование с использованием технологии OpenMP. – М.: Изд-во МГУ, 2009.

*Леоненко Василий Николаевич*

– Омский филиал Института математики им. С. Л. Соболева СО РАН, аспирант, VNLeonenko@yandex.ru

*Логинов Константин Константинович*

– Омский филиал Института математики им. С. Л. Соболева СО РАН, аспирант, kloginov85@mail.ru