

УДК 57.087+519.254+519.651

© А. Л. Буляница

МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ КРИВОЙ ЛОГИСТИЧЕСКОГО РОСТА.

Ч. 2. ОЦЕНИВАНИЕ И КОМПЕНСАЦИЯ СИСТЕМАТИЧЕСКОЙ ПОГРЕШНОСТИ ИЗМЕРЕНИЯ. СОПОСТАВЛЕНИЕ ДВУХ РАЗЛИЧНЫХ ФОРМ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ЛОГИСТИЧЕСКОГО РОСТА

Логистический рост является одной из традиционных математических моделей, применяемых для описания многих популяционных процессов. Динамика количества ДНК при проведении полимеразной цепной реакции в реальном времени (ПЦР-РВ) адекватно описывается подобной моделью. В ч. 1 данной работы (см. этот выпуск НП) были описаны алгоритмы построения эффективных оценок параметров и приведено описание характеристик случайных погрешностей предложенных оценок. Задачами ч. 2 статьи является выявление и компенсация систематической составляющей кривой логистического роста и сопоставление двух различных форм (с использованием алгебраической (степенной) и показательной функций) математической модели процесса.

Кл. сл.: полимеразная цепная реакция, логистический рост, систематическая погрешность, приближение функций, итерация

ВВЕДЕНИЕ

Одной из принятых форм информативного сигнала в приборах, реализующих полимеразную цепную реакцию в реальном масштабе времени, является кривая логистического роста

$$X_n = \frac{X^*}{1 + (B-1)a_0^{-n}}. \quad (1)$$

Здесь X_n — величина сигнала на n -м шаге измерения ($n = 0, 1, \dots, N$ при $N = 30 \div 40$).

Характеристика зависимости (1), интерпретация роли информативных параметров a_0 , X^* и X_0 , обоснование алгоритмов их оценивания, а также оптимизация выбора отсчетов сигнала для повышения эффективности оценок были даны в первой части статьи [1].

В ч. 2 данной работы описаны алгоритмы выявления и компенсации систематических составляющих сигнала. Кроме того, проанализировано другое достаточно распространенное описание процесса логистического роста (2), включающее степенную зависимость

$$X_n = A_2 + \frac{A_1 - A_2}{1 + (n/n^*)^p}, \quad (2)$$

и установлена связь между параметрами обеих моделей.

КОМПЕНСАЦИЯ НАЧАЛЬНОГО ДРЕЙФА СИГНАЛА ЛОГИСТИЧЕСКОГО РОСТА

Проведение измерений может осуществляться в условиях, когда имеется дрейф базовой линии, интерпретируемый как систематическая погрешность измерения. При этом малому информативному сигналу на начальных отсчетах n , когда количество ДНК еще мало, соответствует значительный фоновый сигнал. Очевидно, что он требует оценивания или/и компенсации.

Однако предварительный анализ кривой (1) позволяет осуществить две простые экспресс-оценки сигнала насыщения X^* .

а) В [1] были сформулированы критерии поиска порогового цикла (точки кривой с наименьшей кривизной). В этой точке информативная составляющая сигнала должна быть равна $0.5X^*$.

б) Размах сигнала даже с учетом возможности недостижения уровня насыщения должен примерно соответствовать X^* .

Таким образом, приближенные экспресс-оценки X^* могут быть выполнены на основе таких простых рассуждений.

Рассматриваются алгоритмы компенсации фонового сигнала, исходя из предположения, что этот сигнал адекватно аппроксимируется полиномом либо нулевого, либо первого порядка. В этом случае возможны два подхода: аппроксимация или/и компенсация соответствующего тренда ма-

лого порядка по Кендаллу—Стьюарту [2], либо оценивание параметров, например, по МНК с последующей экстраполяцией на все отсчеты и компенсацией.

Определение величины фонового сигнала $Z_n = Z = \text{const}(n)$ базируется на гипотезе, что Z аддитивно суммируется со всеми информативными сигналами, т. е. вместо X_n измеряется $X_n + Z$. Экспресс-оценивание параметров X^* и X_0 кривой (1), а также Z возможно на основе следующих предположений.

а) По критерию смены знака второй разности можно определить примерное положение порогового цикла n^* . При указанном $n = n^*$ $X_{n^*} = 0.5X^* + Z$.

б) Полагая в первом приближении $a_0 = 1.90$, можно оценить $B = X^* / X_0 + 1 = a_0^{n^*}$.

в) На начальных отсчетах n , когда $B / a_0^n \gg 1$, соответствующий оцениваемый сигнал близок к Z ; на конечных отсчетах, близких к стадии насыщения, если она достигается, при $B / a_0^n \ll 1$ величина сигнала должна быть близка к $X^* + Z$. Таким образом, можно приближенно оценить величину Z и произвести ее компенсацию.

После компенсации Z можно исследовать информативный сигнал в тех интервалах отсчетов, когда он достаточно слабо меняется по величине, т. е. при $n \ll n^*$ (информативная составляющая мала по величине) или/и $n \gg n^*$ (информативная составляющая практически не меняется из-за достижения насыщения). По нескольким отсчетам строится линейная аппроксимация, оценивается величина параметра положения сигнала и значимость наличия тренда первого порядка (например, по критерию Фишера или аналогичному [3, 4]). Далее линейный тренд экстраполируется на весь диапазон отсчетов n и компенсируется (вычитается) эта составляющая систематической погрешности.

Вышеописанный алгоритм проиллюстрируем простым примером: генерируем кривую (1) с параметрами $a_0 = 1.950$, $X^* = 2000$, $X_0 = 1$, $n = 0, 1, \dots, 30$. Внесем аддитивную случайную центрированную помеху, сгенерированную по равномерному закону в диапазоне $[-10; +10]$, и постоянную систематическую погрешность $Z = +200$. Некоторые значимые отсчеты суммарного сигнала приведены в табл. 1.

Расчет сигнала второй разности по отсчетам 10, 11, 12 дает (-15.9) ; по отсчетам 11, 12, 13 — $(+22.3)$. Можно убедиться, что по крайней мере в 2 последующих и предшествующих отсчетах знак второй разности сохраняется. Тогда на интервале 11–12 отсчетов наблюдается перегиб. Более точная его локализация с помощью линейной интерполяции по величине второй разности дает $n^* = 11.41$ и, следовательно, $X_{n^*} = 1205.1$.

Экспресс-оценка $B = 1.90^{11.41} + 1 \approx 1516$. Отсчетам $n = 0 \dots 3$ соответствуют $t \geq 221$ и, следовательно, сигналы, близкие по величине к Z . Отсчетам $n \geq 24$ соответствуют $t \leq 3 \cdot 10^{-4}$ и сигнал, практически равный $X^* + Z$. С помощью МНК для $n = 0 \dots 3$ и $n = 24 \dots 30$ строим линейные аппроксимации:

а) для первого участка $X_n = 191.08 + 2.90 \cdot n$;

б) для второго участка $X_n = 2208.58 - 1.24 \cdot (n - 24)$ либо с соизмеримой точностью $X_n = 2204.90 = \text{const}(n)$.

Средние значения сигнала на первом и втором участках рассматриваются как Z и $X^* + Z$. (При этом $0.5X^* + Z$ примерно равен 1205.1). Тогда $Z \approx 191.08 + 2.90 \cdot 1.5 = 195.4$ и $X^* + Z \approx 2204.9$. Следовательно, $X^* \approx 2009.5$ (значение сигнала при $n = n^*$ должно быть примерно равно $0.5 \cdot 2009.5 + 195.4 = 1200.2$).

Табл. 1. Отсчеты сигнала на участках малого изменения информативной составляющей

n	X_n	n	X_n	n	X_n
0	191.2	7	295.7	24	2208.6
1	192.0	8	393.4	25	2206.6
2	200.5	9	547.8	26	2208.4
3	198.1	10	764.6	27	2206.2
4	211.5	11	1072.4	28	2199.0
5	222.2	12	1396.1	29	2202.1
6	254.3	13	1697.5	30	2203.2

Оценочное значение $X_0 = X^* / B \approx 2009.5 / 1516 = 1.33$. Тогда на начальных отсчетах логистический рост практически неограничен, носит экспоненциальный характер и осуществляется примерно по закону $X_0 \cdot a_0^n$, т. е. при $n = 0, 1, 2, 3$ получим величины 1.33, 2.53, 4.80 и 9.12 (линейная аппроксимация соответствующего фрагмента сигнала позволяет получить оценку параметра положения, равную 2.56). Таким образом, наклон линейной аппроксимации практически полностью объясняется логистическим ростом, и, следовательно, аддитивной систематической погрешности типа линейного тренда первого порядка *не имеется*.

СРАВНЕНИЕ ДВУХ МОДЕЛЕЙ КРИВЫХ ЛОГИСТИЧЕСКОГО РОСТА

Кривая логистического роста описывается двумя зависимостями (1) и (2), отличающимися различным числом информативных параметров и типом используемых функций. В (1) входит показательная функция, в (2) — степенная. Существенным представляется вопрос о связи параметров моделей: (X^*, X_0, a_0) и (A_1, A_2, n^*, p) . Первые два параметра в обеих моделях имеют один и тот же смысл: начальный сигнал $X_0 = A_1$ и сигнал насыщения $X^* = A_2$. Два оставшихся параметра модели (2) можно связать с параметрами модели (1) на основе сопоставления временного положения и динамики информативного сигнала в области наиболее линейной фазы сигнала, т. е. на цикле $n_0 = \ln(X^* / X_0 - 1) / \ln(a_0)$. Имеем $n_0 \gg 1$ при достаточно большом соотношении X^* / X_0 . Приравняв временное положение порогового цикла на основе моделей (1) и (2), а также dX_n / dn — скорость роста информативного сигнала в указанной временной области, получим соотношения

$$(n^* / n_0)^p = (p - 1) / (p + 1)$$

$$\text{и } X^* \ln(a_0) / 4 = \frac{(X^* - X_0)(p^2 - 1)}{4pn^*}.$$

Т. к.

$$X^* / X_0 \gg 1 \Rightarrow n^* \gg 1 \Rightarrow p \gg 1 \Rightarrow p \approx n^* \ln(a_0),$$

$n^* = n_0 \sqrt[p]{\frac{p-1}{p+1}} \approx n_0 \left(1 - \frac{2}{p(p-1)}\right) \approx n_0$. Т. е. характерный временной масштаб n^* из (2) практически совпадает с величиной порогового цикла n^* . Тем самым в первом приближении возможно связать параметры обеих моделей кривых логистического роста.

Однако, выбрав в качестве критерия минимизацию среднеквадратичного отклонения обеих моделей на интервале отсчетов $n = [0; N]$, получим, что оптимальные оценки параметров A_1, A_2, n^* и p расходятся с названными выше. Это связано, в частности, с тем, что скорости роста степенной и показательной функций различны, и в областях стыковки участков кривой "начальный участок" — "экспоненциальный рост" и "экспоненциальный рост" — "участок насыщения" имеются существенные расхождения обеих кривых.

Имитационное моделирование свидетельствует, что оптимальные значения $n_0 \approx n^*$, а остальные оптимальные значения имеют некоторые расхождения с приближенными оценками:

$$A_1 = X_0; A_2 = X^*; p = n_0 \ln(a_0). \quad (3)$$

Задачу оптимизации предлагается решить следующим образом. Указанные три значения параметров (3) рассматриваются как первое итерационное приближение. Необходимые поправки для второго итерационного приближения определим в соответствии с алгоритмом:

а) выпишем функционал качества

$$\Phi(A_1, A_2, p) = \sum_{n=0}^N \left(\frac{X^*}{1 + (X^* / X_0 - 1) \cdot a_0^n} - \left(A_2 + \frac{A_1 - A_2}{1 + (n / n^*)^p} \right) \right)^2; \quad (4)$$

б) вычисляются частные производные первого порядка $\frac{\partial \Phi}{\partial A_1}; \frac{\partial \Phi}{\partial A_2}; \frac{\partial \Phi}{\partial p}$ при значениях параметров (3);

в) после необходимых итерационных приращений $\Delta A_1, \Delta A_2, \Delta p$ мы должны получить стационарную точку функционала Φ , в которой все частные производные первого порядка должны стать равными нулю (необходимое условие экстремума функции нескольких переменных);

г) приращение частных производных первого порядка в приближении второго дифференциала функции нескольких переменных равны:

$$\begin{aligned} \Delta \frac{\partial \Phi}{\partial A_1} &\approx \frac{\partial^2 \Phi}{\partial A_1^2} \Delta A_1 + \frac{\partial^2 \Phi}{\partial A_1 \partial A_2} \Delta A_2 + \frac{\partial^2 \Phi}{\partial A_1 \partial p} \Delta p; \\ \Delta \frac{\partial \Phi}{\partial A_2} &\approx \frac{\partial^2 \Phi}{\partial A_1 \partial A_2} \Delta A_1 + \frac{\partial^2 \Phi}{\partial A_2^2} \Delta A_2 + \frac{\partial^2 \Phi}{\partial A_2 \partial p} \Delta p; \\ \Delta \frac{\partial \Phi}{\partial p} &\approx \frac{\partial^2 \Phi}{\partial A_1 \partial p} \Delta A_1 + \frac{\partial^2 \Phi}{\partial p \partial A_2} \Delta A_2 + \frac{\partial^2 \Phi}{\partial p^2} \Delta p. \end{aligned} \quad (5)$$

При этом производные первого порядка выражаются, как

$$\begin{aligned} \frac{\partial \Phi}{\partial A_1} &= -2 \sum (X_n^1 - X_n^2) \frac{1}{1 + V_n}; \\ \frac{\partial \Phi}{\partial A_2} &= -2 \sum (X_n^1 - X_n^2) \frac{V_n}{1 + V_n}; \\ \frac{\partial \Phi}{\partial p} &= 2(A_1 - A_2) \sum (X_n^1 - X_n^2) \frac{V_n \ln(n/n^*)}{(1 + V_n)^2}. \end{aligned}$$

Здесь и далее введены обозначения X_n^1 и X_n^2 — отсчеты кривых логистического роста в соответствии с моделями (1) и (2), $V_n = (n/n^*)^p$.

Производные второго порядка определяются выражениями:

$$\begin{aligned} \frac{\partial^2 \Phi}{\partial A_1^2} &= 2 \sum \frac{1}{(1 + V_n)^2}; \quad \frac{\partial^2 \Phi}{\partial A_1 \partial A_2} = 2 \sum \frac{V_n}{(1 + V_n)^2}; \\ \frac{\partial^2 \Phi}{\partial A_1 \partial p} &= 2(A_2 - A_1) \sum \frac{V_n \ln(n/n^*)}{(1 + V_n)^3} + \\ &+ 2 \sum (X_n^1 - X_n^2) \frac{V_n \ln(n/n^*)}{(1 + V_n)^2}; \\ \frac{\partial^2 \Phi}{\partial A_2 \partial p} &= 2(A_2 - A_1) \sum \frac{V_n^2 \ln(n/n^*)}{(1 + V_n)^3} - \\ &- 2 \sum (X_n^1 - X_n^2) \frac{V_n \ln(n/n^*)}{(1 + V_n)^2}; \\ \frac{\partial^2 \Phi}{\partial p^2} &= 2(A_2 - A_1)^2 \sum \frac{V_n^2 \ln^2(n/n^*)}{(1 + V_n)^4} - \\ &- 2(A_2 - A_1) \sum (X_n^1 - X_n^2) \frac{V_n(1 - V_n) \ln^2(n/n^*)}{(1 + V_n)^3}. \end{aligned}$$

Таким образом, на основе линейной системы уравнений (5) вычисляются необходимые приращения параметров ΔA_1 , ΔA_2 и Δp и второе итерационное приближение решения задачи минимизации функционала качества (4).

ПРИМЕР РАСЧЕТА ОПТИМАЛЬНО СОГЛАСОВАННЫХ ПО ПАРАМЕТРАМ МОДЕЛЕЙ (1) И (2)

Рассмотрим кривую (1) с параметрами $X^* = 2000$, $X_0 = 0.2$, $a_0 = 1.94$. Первое итерационное приближение параметров степенной модели (2): $A_1 = 0.2$; $A_2 = 2000$; $n^* = 13.898$; $p = 9.210$ (n^* вычислен [1], как $\ln(X^*/X_0 - 1) / \ln(a_0) = \ln(9999) / \ln(1.94)$). Дисперсию расхождения моделей можно оценить, как $\Phi(A_1, A_2, p) / (N + 1 - 3)$,

где Φ определен согласно (4), поскольку выборка содержит $(N + 1)$ отсчет и 3 оцениваемых параметра. Вычисление частных производных и решение системы (5) для определения приращений параметров позволяет получить новое итерационное приближение: $A_1 = 0.2 + 17.001 = 17.201$; $A_2 = 2000 + 17.369 = 2017.369$; $p = 9.210 + 0.043 = 9.253$. (Построение 10 итераций в соответствии с алгоритмом, встроенным в пакет ORIGIN 6.1, позволяет получить весьма близкие оценки: $A_1 = 17.198$, $A_2 = 2017.290$, $p = 9.256$). Среднеквадратичное отклонение моделей при первой итерации значений параметров равно 25.06; при втором (уточненном) наборе параметров — 17.13. Взяв модель (1) с $X^* = 3000$; $X_0 = 0.05$; $a_0 = 1.95$, т. е. условия более позднего достижения порогового цикла ($n^* = 16.474$) и более эффективной ПЦР, в качестве оптимальных параметров по описанной выше методике получим: $A_1 = 16.691$; $A_2 = 3028.084$; $p = 10.945$. Реализация итерационной процедуры по алгоритму ORIGIN дает весьма близкие оценки $A_1 = 16.601$; $A_2 = 3028.149$; $p = 10.941$. Тем самым подтверждается высокая точность решений задачи оптимизации "пересчета" параметров модели на основе итерационного приближения первого дифференциала и матрицы Гессе (т. е. системы (5)).

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ВЫПОЛНЕНИЯ ЦИКЛА АЛГОРИТМОВ ОБРАБОТКИ КРИВЫХ ЛОГИСТИЧЕСКОГО РОСТА

В качестве модельного сигнала генерируем последовательность отсчетов $n = 0, \dots, 34$ при заданных параметрах: $X^* = 3200$, $X_0 = 0.2$, $a_0 = 1.92$. Далее аддитивно добавляем сгенерированную равномерно распределенную центрированную случайную величину в диапазоне $[-16; +16]$ (оговоренное ранее условие $X^* / \delta \geq 120$ выполнено). В качестве систематической погрешности задается линейный тренд первого порядка $250 + 2.0 \cdot n$. График сгенерированной зашумленной кривой и ее значимых фрагментов представлен далее.

На рис. 1 приведена форма кривой логистического роста с наложенными систематической и случайной погрешностями измерения. На графике схематически выделены участки: 1 — фаза начального роста, соответствующая малому количеству ДНК; 2 — наиболее линейный участок роста в области порогового цикла (точки перегиба); 3 — участок насыщения.

Далее проиллюстрированы основные стадии обработки указанного информативного сигнала. Заметим, что идеальной кривой логистического роста (1) соответствуют $B = 3200 / 0.2 = 16000$ и $n^* = \ln(16000 - 1) / \ln(1.92) = 14.840$.

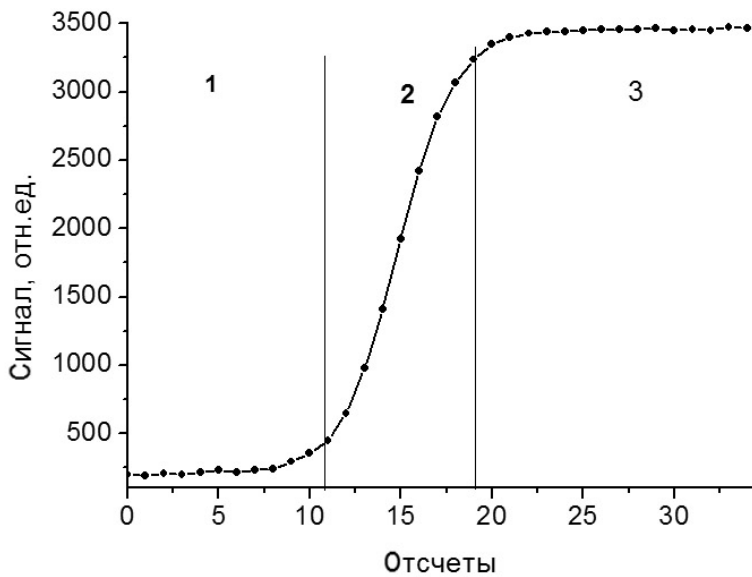


Рис. 1. Схематическое представление информативного сигнала в соответствии с моделью логистического роста. Пояснения см. в тексте

Табл. 2. Величина сигнала на этапе насыщения логистической кривой

n	X_n	n	X_n	n	X_n
23	3440.4	27	3456.1	31	3458.9
24	3438.3	28	3457.0	32	3451.8
25	344.1	29	3464.7	33	3471.4
26	3455.0	30	3447.5	34	3467.2

Табл. 3. Величина сигнала на начальном этапе кривой логистического роста

n	X_n	n	X_n	n	X_n
0	195.6	3	197.7	6	214.8
1	193.3	4	217.2	—	—
2	205.9	5	229.6	—	—

Первый этап обработки предполагает выявление систематической погрешности, ее компенсацию и грубую оценку основных параметров кривой логистического роста.

Анализируем динамику второй разности исходного сигнала: при $n = 12 \div 14$ эти разности положительны (при $n = 14$ равна +86.07), далее при $n = 15 \div 17$ происходит смена знака (при $n = 15$ величина второй разности -17.08). Таким образом, на интервале 14–15 отсчетов находится точка перегиба (пороговый цикл). Методом линейной интерполяции определяется точка перехода второй разности через ноль: $n^* = 14.83$ и $X_{n^*} = 1834.4$.

В предположении формы графика систематической погрешности, как линейной, тренд первого

порядка X_{n^*} примерно соответствует $0.5X^* + \alpha_0 + \alpha_1 n^*$. Эффективность ПЦР априорно полагаем средней, т. е. $\alpha_0 = 1.90$. Т. к. $a_0^8 = 1.9^8 \approx 170$, то отклонение от порогового цикла на +8 тактов соответствует стадии насыщения сигнала (информативная составляющая практически совпадает с X^*); участок, предшествующий пороговому циклу на 8 тактов и более, — начальный участок экспоненциального роста (полезная составляющая сигнала близка к $X_0 \cdot a_0^n$). На основе временного положения порогового цикла получаем примерное отношение $X^*/X_0 = 1 + 1.9^{14.83} = 13610$.

В табл. 2 и 3 представлены значения фрагментов сгенерированного информативного сигнала —

на этапе насыщения (отсчеты 23–34) и на начальном этапе (отсчеты 0–6).

На основе МНК данные таблицы аппроксимированы трендом первого порядка $3394 + 2.13n$, что иллюстрируется на рис. 2. Таким образом, сумма $X^* + \alpha_0 \approx 3394$ при $\alpha_1 \approx 2.13$.

Аппроксимация $X_n = f(n)$ данных табл. 3 по МНК дает зависимость $192.6 + 5.04 \cdot n$. Можно сделать следующие выводы:

а) систематическая погрешность содержит постоянное смещение ($\alpha_0 \approx 190$);

б) параметр положения линейного тренда ($\alpha_1 \approx 2.1$);

в) большие значения параметра $\alpha_1 \approx 5.0$ связаны с информативным слагаемым (экспоненциальным ростом);

г) разность между свободными членами на обоих участках ($3394 - 190 = 3204$) примерно соответствует величине сигнала насыщения X^* .

Для проверки проанализируем предполагаемую точку перегиба (измерение на пороговом цикле): компенсировав из сигнала 1834.4 и систематическую составляющую, равную $\alpha_0 + \alpha_1 \cdot n^* \approx 190 + 2.1 \cdot 14.83 \approx 221$, должны получить около 1613, что примерно соответствует требуемому уровню $0.5X^* = 1602$. Таким образом, в первом приближении можно полагать сигнал насыщения 3204. Следовательно, оценка начального информативного сигнала X_0 составит $3204 / 13610 \approx 0.235$. Информативная составляющая сигнала на начальных отсчетах приближенно соответствует $X_0 \cdot a_0^n \approx 0.235 \cdot 1.9^n$. При аппроксимации этой зависимости линейным трендом по МНК получим $1.57 + 1.63 \cdot n$. Разность вычисленных параметров положения $5.0 - 2.1 = 2.9$ в значительной степени

объясняется "наложенным" информативным экспоненциальным ростом. Кроме того, истинное значение параметра a_0 может превосходить априорно заданный средний уровень 1.90, а наличие случайной помехи большого (по сравнению с изменением сигнала на начальных отсчетах) диапазона также может исказить оцененную величину параметра положения систематической погрешности. В целом можно полагать, что в первом приближении систематическую погрешность можно аппроксимировать зависимостью $190 + 2.1n$. Далее эта систематическая погрешность компенсируется.

После компенсации систематической составляющей сигнала осуществляется более точное оценивание информативных составляющих сигнала — a_0 , X^* и X_0 по методикам, описанным выше. Также следует заметить, что проведенная компенсация линейного тренда первого порядка не повлияла на оценку положения порогового цикла (точки перегиба), т. к. последний определяется вторым разностным сигналом, а вторая разность от линейного тренда первого порядка есть ноль.

Ранее [1] были найдены условия оптимального оценивания параметра a_0 (4-точечная схема, включающая измерения на отсчетах n , $n + 1$, $n + 4$ и $n + 5$, т. к. в большинстве случаев оптимальным является $k = 4$). Было определено, что оптимальный отсчет n должен соответствовать $t = 4.5 \div 5$, т. е. примерно на 3 отсчета предшествовать пороговому циклу. Однако, поскольку предварительно осуществлялась компенсация систематической составляющей, целесообразно использовать отсчеты с большими величинами информативного сигнала, т. к. эффект неточной компенсации систематической составляющей информативного сигнала будет в этом случае менее значим.

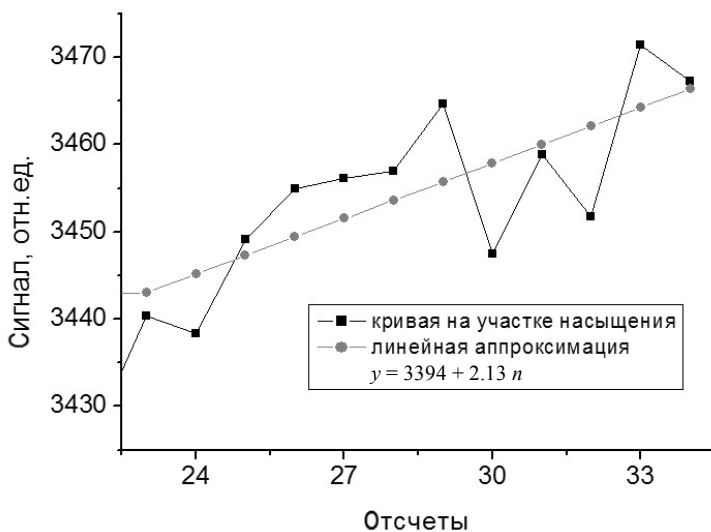


Рис. 2. Выделение систематической погрешности как линейного тренда первого порядка на стадии насыщения

Табл. 4. Величины отсчетов сигнала после компенсации систематической составляющей

n	X_n	n	X_n	n	X_n	n	X_n
27	3209	29	3214	31	3204	33	3212
28	3208	30	3195	32	3195	34	3206

Табл. 5. Выборка оценок параметра F

	n					
	11	10	12	13	14	15
F	9.654	9.589	9.681	9.643	9.656	9.702

Для иллюстрации выберем четыре 4-точечные схемы, соответствующие $n = 12 \div 15$. Расчет параметра a_0 даст следующие оценки: 1.944; 1.948; 1.894 и 1.838. Выбрав медиану этих 4 измерений, получим $(1.944 + 1.894) / 2 \approx 1.919$.

Для сравнения, использование близких 4-точечных схем по отсчетам (13, 14, 16, 17), (13, 14, 18, 19) и (12, 13, 18, 19) дает достаточно близкие оценки: 1.895, 1.889 и 1.922 соответственно.

После оценивания a_0 осуществляется расчет оценок X^* . Как говорилось ранее, более эффективные оценки будут при использовании наибольших по величине измерений (близких к стадии насыщения). Измерения на заключительных отсчетах кривой (от 27 до 34-го) после компенсации аппроксимированной систематической погрешности и округленные до целочисленных значений представлены в табл. 4.

Используя пары значений сигнала на отсчетах 27–31, 28–32, 29–33, 30–34, применили формулу расчета X^* из [1] при $k = 4$ и оцененное значение параметра $a_0 = 1.919$. Тогда оценки X^* практически совпадают с вышеуказанными измерениями, поскольку стадия насыщения практически достигнута: 3203.6; 3194.0; 3211.8 и 3206.9. В этих условиях дисперсия ошибки оценивания должна лишь незначительно превышать дисперсию исходной помехи (стандартное отклонение исходной равномерной помехи с полуразмахом $\delta = 16$ есть 9.24). В соответствии с медианной оценкой получим, что $X^* \approx (3206.9 + 3203.6) / 2 = 3205.3$.

Заключительный оцениваемый параметр F связан с отношением X^* / X_0 . В соответствии с методикой [1] наиболее эффективное оценивание этого параметра осуществляется при отсчетах, примерно соответствующих $t = 12 \div 15$ (требуется решение трансцендентного уравнения). Т. к. $t = 1$ — точка перегиба (пороговый цикл), a_0 близко к 2.0, то требуется выбрать отсчет, предшествующий пороговому циклу на Δn : $\Delta n = \ln(t) / \ln(a_0) = 4 \div 5$. Однако по соображениям, высказанным ранее, следу-

ет несколько сместить отсчеты в сторону больших измерений. В табл. 5 приведены оценки параметра F на основе измерений $n = 10 \div 15$. По медианному алгоритму получено $F = 9.655$. Следовательно, $B = 15600$ и $X_0 = 0.205$. Пороговый цикл вычисляется, как $F / \ln(a_0) = 14.816$. Поскольку оценочные значения параметров получились весьма близкими к истинным, то и достаточно устойчивая оценка положения порогового цикла оказалась весьма близкой к истинному значению.

В заключительной стадии обработки информативного сигнала можно осуществить сопоставление найденных параметров модели (1) с параметрами альтернативной модели (2). Первая итерация параметров: $n_0 = n^* = 14.816$, $A_1 = X_0 = 0.205$, $A_2 = X^* = 3205.3$ и $p = n_0 / \ln(a_0) = 9.655$. Применяя алгоритм уточнения параметров, основанный на решении системы (5), получим необходимые приращения параметров и новые итерационные значения: $\Delta A_1 = 25.995 \Rightarrow A_1 = 26.200$, $\Delta A_2 = 21.71 \Rightarrow A_2 = 3227.01$ и $\Delta p = 0.083 \Rightarrow A_1 = 9.738$. Среднеквадратичное отклонение двух кривых составит: при первой итерации параметров 36.17 отсчетов, при второй итерации — 25.88. На рис. 3 иллюстрируется совпадение кривых, построенных на основе различных моделей логистического роста. Взяты значимые фрагменты, поскольку на начальной стадии и на стадии насыщения обе модели обеспечивают близкие значения (соответственно к величинам X_0 и X^*). Наиболее "расходящейся" стадией являются участки с наибольшей кривизной сигнала, т. е. на границах областей 1—2 и 2—3 (см. рис. 2). Более детализированные укрупненные фрагменты представлены на рис. 4 (а и б).

Разумеется, проиллюстрирована лишь общая схема оценивания информативных параметров кривых логистического роста. В частности, возможно формирование как более представительной статистики оценок параметров (за счет варьирования выбора n и k вблизи оптимальных условий),

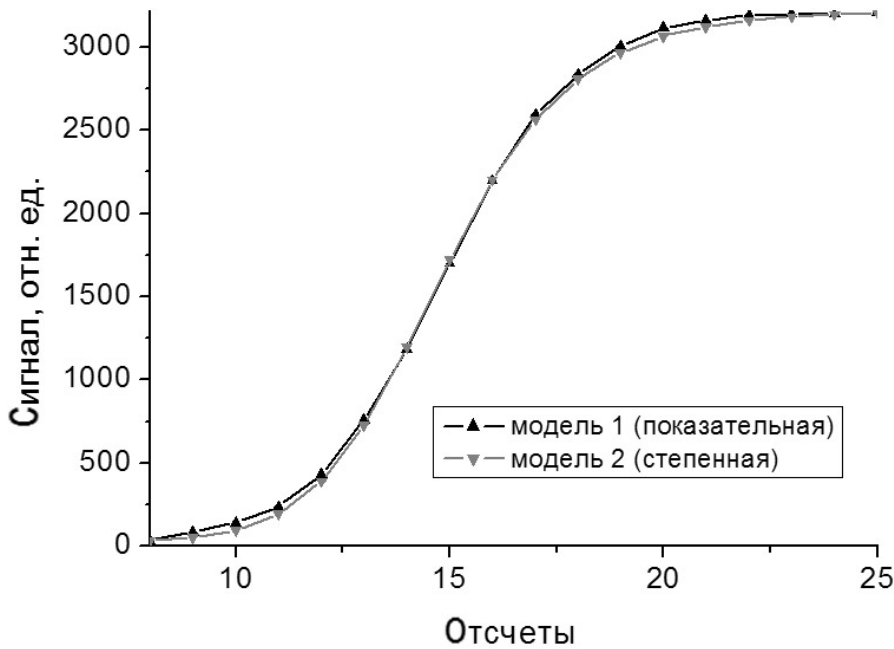


Рис. 3. Сравнение форм кривых логистического роста, сформированных на основе различных моделей

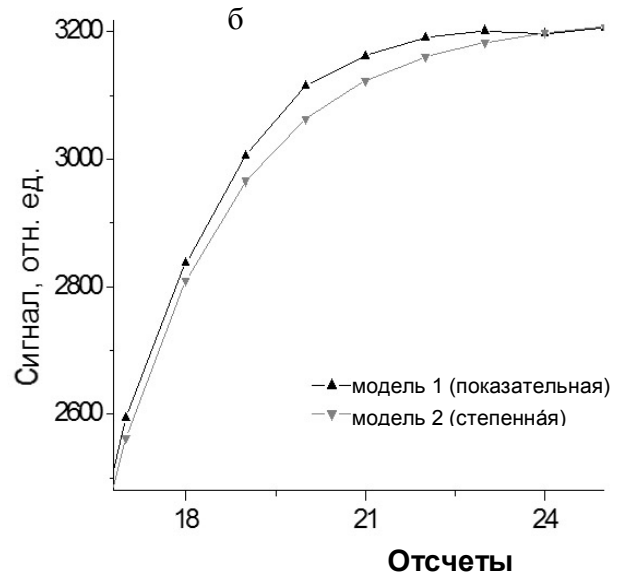
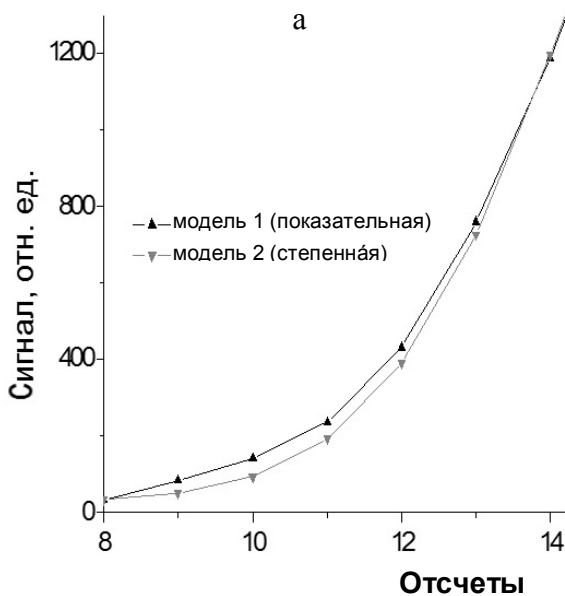


Рис. 4. Различие кривых логистического роста, сформированных в соответствии с различными математическими моделями.

а — фрагмент в области перехода режимов 1—2; б — фрагмент в области перехода режимов 2—3

так и другие принципы отбора результирующей оценки из выборки. Например, в [1] для формирования результирующей оценки параметров рассмотрен алгоритм стохастической аппроксимации.

Другая предложенная оценка в форме медианы по четному числу измерений алгоритмически является усеченным средним, т. е. учитывает различие оценок, обладает необходимой робастностью (ус-

тойчивостью к выбросам) и мало отличается по своим характеристикам от оценки на основе алгоритма стохастической аппроксимации.

ЗАКЛЮЧЕНИЕ

Показаны возможности эффективной компенсации систематической составляющей сигнала, а также надежного оценивания параметров моделей логистического роста. Кроме того, детально описана процедура сопоставления двух наиболее распространенных форм (1) и (2) решения логистического уравнения, использующих показательные и степенные функции. Рассмотренные в ч. 1 и 2 алгоритмы позволяют осуществить обработку соответствующих кривых (информативных сигналов) и в конечном счете способствуют получению необходимой аналитической информации при реализации ПЦР-РВ.

Данная работа выполнялась при финансовой поддержке в рамках программ и проектов: Аналитической ведомственной целевой программы "Развитие научного потенциала высшей школы" — проект "Исследования и диагностика клеточных структур: новые методические подходы и инструментальные решения на основе сканирующей зондовой микроскопии и микрочиповых технологий" (№ 4247); Программы фундаментальных исследований Президиума РАН № 20 "Создание и совершенствование методов химического анализа и исследования структуры веществ и материалов" — проект "Микрочиповые аналитические системы для метода молекулярных колоний"; Программы фундаментальных исследований Президиума РАН на 2009 г. № 27 "Основы фундаментальных исследований нанотехнологий и наноматериалов".

шенствование методов химического анализа и исследования структуры веществ и материалов" — проект "Микрочиповые аналитические системы для метода молекулярных колоний"; Программы фундаментальных исследований Президиума РАН на 2009 г. № 27 "Основы фундаментальных исследований нанотехнологий и наноматериалов".

СПИСОК ЛИТЕРАТУРЫ

1. Буляница А.Л. Методы оценивания параметров кривой логистического роста. Ч. 1. Оптимизация условий оценивания при наличии аддитивной случайной помехи // Здесь. С. 3–11.
2. Kendall M.G. & Stuart A. The Advanced Theory of Statistics. V. 3. Design and analysis, and time-series. 2nd edn. London: Charles Griffin & Company Limited, 1968.
3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. 416 с.
4. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений. М.: Наука, 1968. 288 с.

*Институт аналитического приборостроения РАН,
Санкт-Петербург*

Материал поступил в редакцию 24.04.2009.

METHODS FOR LOGISTIC GROWTH CURVE PARAMETERS ESTIMATION.

PART. 2. ESTIMATION AND COMPENSATION OF THE SYSTEMATIC COMPONENT OF MEASUREMENT. COMPARISON OF TWO DIFFERENT FORMS OF THE MATHEMATICAL MODEL FOR LOGISTIC GROWTH

A. L. Bulyanitsa

Institute for Analytical Instrumentation RAS, Saint-Petersburg

Logistic growth is one of the traditional mathematical models, used for describing many population processes. The dynamics of DNA quantity in conducting polymerase chain reaction in real time (PCR-RT) is adequately described by a similar model, as it was noted earlier [1]. In part. 1 of this paper the algorithms for the development of the efficient parameters estimation were described, and the description of the characteristics of the random errors in the estimations proposed was given. The main aims of part. 2 of this paper are determination and compensation of the systematic component of logistic growth curve, and comparison of two different forms (using algebraic (power) and exponential functions) of the mathematical model of the process.

Keywords: polymerase chain reaction, logistic growth, systematic error compensation, approximation of functions, iterations