

## Метод спектральной мультиэкспоненциальной аппроксимации для обработки больших массивов экспериментальных данных

Хрущев С.С., Плюснина Т.Ю.

Московский государственный университет имени М.В.Ломоносова,  
биологический факультет, кафедра биофизики

[styx@biophys.msu.ru](mailto:styx@biophys.msu.ru)

Дается подробное описание метода спектральной мультиэкспоненциальной аппроксимации и его применения для анализа кривых индукции флуоресценции хлорофилла *a*. Метод формализует и значительно ускоряет процедуру определения характерных времён и относительного вклада отдельных процессов в общую кинетику наблюдаемого сигнала. В основе метода лежит использование линейного разложения сигнала по базису из показательных функций с характерными временами, взятыми на фиксированной сетке. Устойчивость разложения достигается за счёт введения условия неотрицательности коэффициентов (амплитуд). Предложен удобный способ визуализации амплитуд и характерных времён для большого количества индукционных кривых в виде тепловых карт. На примере данных непрерывного мониторинга роста культуры микроводорослей *Chlamydomonas reinhardtii* в условиях голодания по сере показана эффективность метода для раннего выявления ответа фотосинтетического аппарата на действие стрессовых факторов. Метод может найти применение в анализе различных типов экспериментальных данных при исследовании переходных процессов разной природы.

*Ключевые слова:* спектральная мультиэкспоненциальная аппроксимация, СМЭА, анализ экспериментальных данных, фотосинтез, флуоресценция, хлорофилл, фотосистема 2, экологический мониторинг, ответ растений на стресс.

## Processing Huge Sets of Experimental Data with Spectral Multi-Exponential Approximation

Khruschev S.S., Plyusnina T.Yu.

*Lomonosov Moscow State University, School of Biology, Biophysics Department*

A detailed description of the Spectral Multi-Exponential Approximation method and its application for the analysis of chlorophyll *a* fluorescence transients is given. The method formalizes and significantly accelerates the procedure for estimation of characteristic times and relative contributions of individual processes to the general kinetics of the observed signal. The method is based on linear decomposition of the signal over a basis of exponential functions with characteristic times taken on a fixed grid. Numerical stability of the decomposition is achieved by constraining the coefficients (amplitudes) to be non-negative. A convenient way of visualizing amplitudes and characteristic times for a large number of induction curves in the form of heat maps is proposed. The effectiveness of the method for early detection of the photosynthetic apparatus response to the action of stress factors is shown by the example of continuous monitoring of microalgae *Chlamydomonas reinhardtii* growth on sulfur-deprived media. The method can find wild application in the analysis of various types of experimental data in the study of transient processes of different nature.

*Key words:* Spectral Multi-Exponential Approximation, SMEA, analysis of experimental data, photosynthesis, fluorescence, chlorophyll, Photosystem 2, ecological monitoring, plant response to stress.

### 1. Введение

Основой современной науки является количественный анализ экспериментальных данных с помощью различных математических методов. Такой анализ подразумевает создание математической модели явления и идентификацию

численных значений параметров модели по измеренным в экспериментах наблюдаемым параметрам. До последнего времени измерения проводились вручную, поэтому объёмы анализируемых данных были сравнительно невелики, особенно при исследовании динамических систем – то есть в тех экспериментах, в которых изучается изменение величин

наблюдаемых параметров во времени. При исследовании таких систем число оцениваемых параметров, то есть число степеней свободы математической модели явления, зачастую оказывалось лишь незначительно меньше числа экспериментально измеренных значений. Соответственно, стандартные статистические методы анализа ориентированы именно на такую ситуацию.

Положение кардинальным образом изменилось с появлением быстродействующих электронных измерительных приборов, позволяющих в автоматическом режиме собирать большие массивы экспериментальных данных. Широкое применение автоматических средств измерения в научных исследованиях, на производстве и в быту привело к тому, что в настоящее время общий объём хранимых и обрабатываемых человеком данных измеряется в зеттабайтах ( $2^{70}$  байт). Для обработки столь больших объёмов данных требуются новые статистические методы, ориентированные на полностью автоматический анализ информации. В связи с этим большое развитие получили технологии машинного обучения и искусственного интеллекта, позволяющие идентифицировать параметры модели по наблюдаемым величинам без явного установления характера зависимости между ними. Тем не менее, появление новых методов не принижает важность традиционных методов статистической обработки, основанных на регрессионном анализе. В данной работе мы развиваем предложенный нами в [1] метод спектральной мультиэкспоненциальной аппроксимации (СМЭА), позволяющий в автоматическом режиме оценивать характерные времена отдельных процессов и их вклад в общую кинетику наблюдаемого явления, и рассматриваем его применение на примере обработки кривых индукции флуоресценции хлорофилла *a*.

## 2. Флуорометрический метод оценки функционального состояния растений

Возможность оперативного контроля физиологического состояния живых существ имеет важное значение, как для фундаментальной биологической науки, так и для сельского хозяйства, биотехнологии, охраны окружающей среды. В настоящее время широкое распространение получили флуорометрические методы определения функционального состояния фотосинтезирующих организмов, таких как растения и микроводоросли, основанные на явлении индукции флуоресценции, впервые обнаруженном в 1931 году [2]. Существует большое число экспериментальных протоколов, позволяющих получать информацию о состоянии отдельных растений, частей растения, отдельных клеток и даже отдельных хлоропластов, как в лабораторных, так и в полевых условиях [3, 4]. Размещение измерительной аппаратуры на искусственных

спутниках Земли позволяет дистанционно получать информацию о состоянии наземных растений или морского фитопланктона сразу на большой площади [5]. Разработаны флуориметры, предназначенные для непрерывного мониторинга состояния растений и микроводорослей в природных условиях или при культивировании [6].

Показано, что параметры кривой индукции флуоресценции чувствительны к действию стрессовых факторов разной природы: освещённости, температуры, влажности, засоленности, нехватки минерального питания, причём изменение параметров индивидуально для различных факторов. В работе [7] предложена нейросетевая модель, позволяющая по параметрам индукционной кривой выявить дефицит отдельных элементов минерального питания растений. Однако возможность практического применения подобных моделей пока весьма ограничена из-за очень большого разнообразия форм индукционных кривых в зависимости от вида растений и условий их выращивания.

### 2.1. Анализ кривых индукции флуоресценции

Одним из наиболее распространённых экспериментальных протоколов является так называемый ЛР-тест, при котором адаптированные к темноте опытные образцы освещают светом с интенсивностью  $3000 \text{ мкмоль} \cdot \text{м}^{-2} \cdot \text{с}^{-1}$  в течение одной секунды. За это время интенсивность флуоресценции у высших растений и большинства микроводорослей (за исключением цианобактерий) достигает максимума и начинает постепенно снижаться. Нарастающий участок кривой индукции хлорофилла имеет сложный многофазный характер (рис. 1). В качестве первичных параметров в ЛР-тесте берутся интенсивности флуоресценции на временах 20 (или 50) мкс, 100 мкс, 300 мкс, 2 мс и 30 мс, максимальная интенсивность флуоресценции, время достижения максимума и «площадь над индукционной кривой». Основываясь на теории энергетических потоков [8], из этих данных можно оценить эффективность отдельных стадий переноса электрона в электрон-транспортной цепи (ЭТЦ) фотосинтеза, а также определить общий «индекс производительности» фотосинтеза.

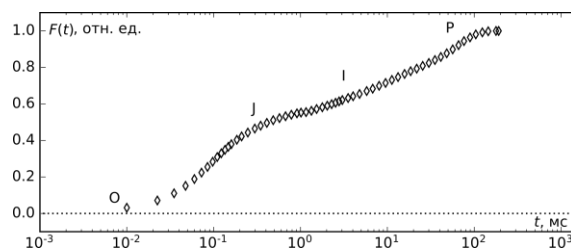


Рис. 1. Нарастающий участок кривой индукции флуоресценции хлорофилла *a* микроводоросли *Chlamydomonas reinhardtii*. Буквами показаны фазы нарастания флуоресценции.

При таком способе анализа существенная часть информации о форме индукционных кривых никак

не используется. Преодолеть этот недостаток призван анализ индукционных кривых с помощью детальных моделей фотосинтетической электрон-транспортной цепи [9–11], основанных на современных представлениях о происхождении наблюдаемой флуоресценции и молекулярных механизмах преобразования энергии в фотосинтезе. Достаточно подробные математические модели, представляющие собой системы из десятков и даже сотен нелинейных дифференциальных уравнений, позволяют с высокой точностью воспроизвести экспериментально полученные кривые за счёт большого количества свободных параметров. Однако вопрос об однозначной идентификации и физическом смысле этих параметров обычно остаётся открытым, так как одна и та же индукционная кривая может в равной степени хорошо описываться различными физическими моделями с разными наборами параметров.

## 2.2. Представление кривых индукции флуоресценции в виде суммы показательных функций

Для того чтобы наиболее полно использовать информацию о форме индукционных кривых, и при этом абстрагироваться от деталей физических процессов, обуславливающих эту форму, рядом авторов было предложено рассматривать такие кривые как сумму нескольких (обычно двух или трёх) показательных функций [12, 13]. В качестве обоснования такого подхода можно рассматривать представление о том, что пигмент-белковый комплекс фотосистемы 2 (ФС2), являющийся основным источником флуоресценции хлорофилла в растениях, в каждый момент времени находится в одном из дискретных состояний. Система не обладает памятью, то есть переход между её состояниями может быть описан с помощью формализма Марковской цепи (рис. 2).

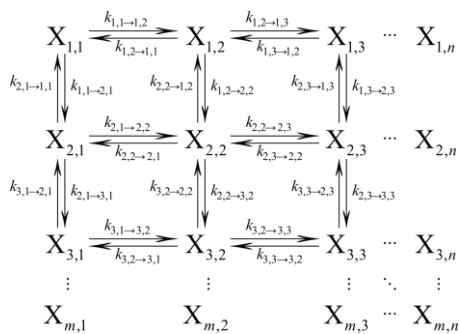


Рис. 2. Обобщённая схема переходов между состояниями фотосистемы 2 в Марковской модели.

Эволюция распределения большого числа комплексов ФС2 по состояниям может быть охарактеризована линейным дифференциальным управляющим уравнением:

$$\frac{dX_i}{dt} = \sum_j k_{i,j} X_j - \sum_j k_{j,i} X_i,$$

где  $X_i$  и  $X_j$  – доли комплексов, находящихся в  $i$ -ом и  $j$ -ом состоянии,  $k_{i,j}$  – константа скорости перехода из  $j$ -го состояния в  $i$ -ое, знак суммы подразумевает суммирование по всем состояниям. Решение управляющего уравнения может быть записано в виде суммы показательных функций:

$$X_i(t) = \sum_j a_{i,j} e^{-t/\tau_j} + c_i,$$

где  $a_{i,j}$  характеризует вклад компоненты с характерным временем  $\tau_j$  в кинетику изменения доли  $i$ -го состояния во времени,  $c_i$  – равновесное значение  $X_i(\infty)$ . Заметим, что  $\tau_j$  выражается действительным числом. Каждое состояние характеризуется определённым квантовым выходом флуоресценции  $k_i^f$ , и изменение интенсивности флуоресценции во времени обусловлено изменением соотношения долей ФС2, находящихся в состояниях с различным квантовым выходом флуоресценции:

$$F(t) = \sum_i k_i^f X_i(t) = \sum_i k_i^f \left( \sum_j a_{i,j} e^{-t/\tau_j} + c_i \right).$$

Это соотношение может быть переписано в виде

$$F(t) = F_0 + \sum_j^n a_j \left( 1 - e^{-t/\tau_j} \right), \quad (1)$$

где  $F_0$  – интенсивность флуоресценции хлорофилла начальный момент времени, а  $n$  – число показательных функций. Уравнение (1) используется для представления экспериментально полученной индукционной кривой в виде суммы показательных функций, при этом обычно исследователь сначала выбирает количество показательных функций  $n$  и начальные приближения для  $F_0$ , характерных времён  $\tau_j$  и амплитуд  $a_j$ , после чего с помощью того или иного метода нелинейной регрессии получает окончательные оценки этих параметров.

Следует отметить, что рассматриваемая задача идентификации параметров  $F_0$ ,  $n$ ,  $\tau_j$  и  $a_j$  является некорректно поставленной (так как имеет бесконечно много решений) и плохо обусловленной. Описанная выше процедура для её решения обладает целым рядом существенных недостатков. Во-первых, задание числа показательных функций  $n$  обычно производится интуитивно, так как существующие методы для оценки  $n$  (описание одного из таких методов можно найти, например, в [14]) неудобны для практического применения. Получаемые оценки параметров сильно зависят от выбора начальных приближений, а многократное повторение процедуры с различными начальными приближениями для поиска наилучшего решения требует больших вычислительных ресурсов. Методы нелинейной регрессии, хорошо «справляющиеся» с данными размерностью в единицы или десятки экспериментальных точек на одну кривую, оказываются неспособны за разумное время обработать массив данных, в которых каждая

индукционная кривая представлена тысячами или даже десятками тысяч экспериментальных точек.

### 3. Метод спектральной мультиэкспоненциальной аппроксимации

В [1] нами был предложен метод спектральной мультиэкспоненциальной аппроксимации, позволяющий (с определёнными ограничениями) решить описанные выше проблемы. Этот метод предназначен для обработки экспериментальных сигналов, оцифрованных с высокой степенью подробности, то есть таких, где число полученных в эксперименте отсчётов  $F(t_i)$  значительно превышает предполагаемое число показательных функций  $n$  в уравнении (1). В таком случае мы можем свести задачу нелинейной регрессии к линейной, выбирая набор из  $N$  базисных функций  $f_k(t) = 1 - e^{-t/\tau_k}$  на фиксированной сетке характерных времён  $\tau_k = \tau_0 \cdot \text{step}^k$ , где  $\text{step}$  – шаг сетки. Минимальное характерное время  $\tau_0$ , шаг сетки и количество функций выбираются, исходя из предполагаемого диапазона искомым характерных времён  $\tau_j$ , или, если этот диапазон неизвестен, из минимального и максимального значения  $t_j$  в экспериментальных данных. При этом  $N$  должно быть не больше числа экспериментальных отсчётов. Если искомые характерные времена  $\tau_j$  лежат в диапазоне от  $\tau_0$  до  $\tau_{N-1}$ , то для каждого  $\tau_j$  в базисном наборе обязательно найдётся соответствующее значение  $\tau_k$ , отличающееся от  $\tau_j$  не более чем в  $\sqrt{\text{step}}$  раз. На практике в большинстве случаев имеет смысл использовать значения шага порядка 1.05...1.1.

После выбора базисных функций уравнение (1) становится линейным, так как неизвестными являются только начальная интенсивность флуоресценции  $F_0$  и амплитуды  $a_k$ . Однако применение методов линейной алгебры не даёт устойчивого решения, так как получаемые оценки амплитуд будут в большей степени определяться присутствующими в экспериментальных данных ошибками измерения и шумами, а не собственно сигналом. Получаемый набор амплитуд  $a_k$  оказывается состоящим из чередующихся очень больших положительных и отрицательных значений. Однако в том случае, если мы знаем, что искомые амплитуды  $a_j$  неотрицательны, мы можем наложить на решение соответствующее дополнительное ограничение и получить численно устойчивое решение с помощью алгоритма NNLS, что оказывается существенно быстрее получения решения с помощью нелинейной регрессии и не требует начальных приближений для оценок параметров.

В общем случае обоснованность ограничения  $a_j \geq 0$  оказывается под вопросом, так как оно не следует из рассматриваемой модели Марковской цепи. Тем не менее, несмотря на то, что методы нелинейной регрессии во многих случаях обеспечивают лучшую аппроксимацию

индукционных кривых по сравнению со СМЭА, большая численная стабильность СМЭА и меньшая требовательность к вычислительным ресурсам делают этот метод удобным для оперативного анализа больших массивов экспериментальных данных.

#### 3.1. Визуализация данных спектральной мультиэкспоненциальной аппроксимации

Ещё одним плюсом СМЭА является возможность удобной визуализации полученных результатов. Метод СМЭА позволяет выявлять на индукционных кривых визуально неразличимые фазы нарастания флуоресценции. Под фазой в данном случае мы подразумеваем группу соседних ненулевых амплитуд  $a_k$ , отделённую от других групп нулевыми значениями. Для визуализации полученных методом СМЭА спектров может быть построена линейчатая диаграмма (рис. 3). На рисунке видно, что, хотя визуально на кривой индукции флуоресценции (рис. 1) можно выделить не более трёх фаз нарастания флуоресценции, фактически имеется ещё одна скрытая фаза.

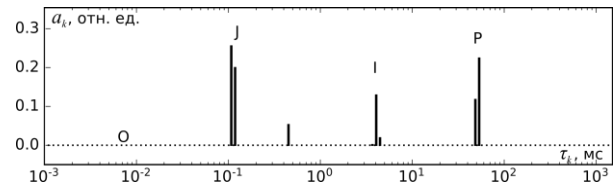


Рис. 3. Амплитуды  $a_k$  спектральной мультиэкспоненциальной аппроксимации для кривой индукции флуоресценции, приведённой на рисунке 1. Буквами показаны соответствующие группам амплитуд фазы нарастания флуоресценции.

Линейчатые диаграммы удобны для определения характерных времён фаз нарастания флуоресценции, однако по ним сложно оценить вклад каждой фазы в общую кинетику процесса. Поэтому более информативными являются графики частичных сумм амплитуд  $\sum_i^k a_i$ . Пример такого графика приведён на рисунке 4.

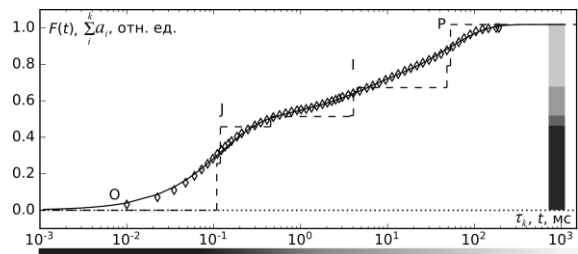
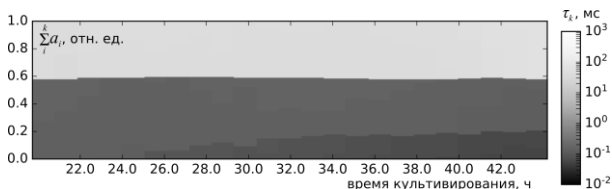


Рис. 4. Экспериментально полученные значения флуоресценции ( $\diamond$ ), их аппроксимация методом СМЭА (—) и частичные суммы амплитуд СМЭА (---) для кривой индукции флуоресценции, приведённой на рисунке 1. Справа показано представление результатов СМЭА в виде столбчатой диаграммы.

Для более компактного представления данных частичные суммы могут отображаться в виде столбчатой диаграммы, на которой высота столбика определяется суммарной амплитудой соответствующей фазы, а значение её характерного времени обозначается цветом (рис. 4, справа). Такой вариант отображения позволяет перейти к визуализации большого массива экспериментальных данных в виде тепловой карты (рис. 5). Анализ данных непрерывного мониторинга фотосинтетической активности микроводорослей *Chlamydomonas reinhardtii* показывает, что через сутки фотогетеротрофного культивирования водорослей на среде без серы на кривой индукции флуоресценции появляется и постепенно увеличивает амплитуду короткая (50–60 мкс) фаза, не характерная для микроводорослей, культивируемых на полной среде [6]. Мы связываем появление этой фазы с повреждением белков реакционного центра ФС2, которые не могут быть репарированы в связи с нарушением синтеза белков из-за дефицита серы. В [15] показано, что аналогичная быстрая фаза появляется у микроводорослей *Chlorella sp.* и высших растений при азотном голодании, которое также приводит к нарушению синтеза белков. Однако при дефиците азота появлению этой фазы предшествует появление более длинной фазы с характерным временем порядка 5 мс, которое может объясняться адаптацией фотосинтетического аппарата к условиям стресса за счёт изменения соотношения числа различных типов белковых комплексов в тилакоидной мембране и размера пластохинонового пула, приходящегося на один реакционный центр ФС2.



**Рис. 5.** Изменение соотношения фаз кривой индукции флуоресценции микроводорослей *Chlamydomonas reinhardtii* в процессе развития стресса, вызванного дефицитом серы в среде (по данным [6]).

#### 4. Заключение

Рассмотрено применение метода спектральной мультиэкспоненциальной аппроксимации для анализа кривых индукции флуоресценции хлорофилла *a*. Данный метод позволил существенно снизить затраты времени на обработку экспериментальных данных, формализовал и автоматизировал процедуру определения характерных времён и относительного вклада отдельных процессов, лимитирующих электронный транспорт в фотосинтетической ЭТЦ, в общую кинетику и сделал возможной обработку и визуализацию большого количества индукционных

кривых. Метод СМЭА реализован в свободно распространяемом программном обеспечении ruPhotoSyn, доступном по запросу. Разработанный метод может найти применение в анализе различных типов экспериментальных данных при исследовании переходных процессов разной природы.

#### 5. Благодарности

Работа выполнена при частичной поддержке гранта РФФИ №17-04-00676.

#### 6. Список литературы

1. Плюснина Т.Ю., Хрущев С.С., Ризниченко Г.Ю., Рубин А.Б. Анализ кинетики индукции флуоресценции хлорофилла с помощью спектральной мультиэкспоненциальной аппроксимации. *Биофизика*. 2015. Т. 60. № 3. С. 487–495.
2. Kautsky H. Energie-Umwandlungen an Grenzflächen, IV. Mittel.: H. Kautsky und A. Hirsch: Wechselwirkung zwischen angeregten Farbstoff-Molekülen und Sauerstoff. *Berichte der deutschen chemischen Gesellschaft (A and B Series)*. 1931. V. 64. № 10. P. 2677–2683. doi: [10.1002/cber.19310641017](https://doi.org/10.1002/cber.19310641017).
3. Гольцев В.Н., Каладжи Х.М., Кузманова М.А., Аллаhverдиев С.И. *Переменная и замедленная флуоресценция хлорофилла a – теоретические основы и практическое приложение в исследовании растений*. Москва-Ижевск: Ижевский институт компьютерных исследований, 2014. 220 с.
4. Kalaji H.M., Schansker G., Ladle R.J., Goltsev V., Bosa K., Allakhverdiev S.I., Brestic M., Bussotti F., Calatayud A., Dąbrowski P., Elsheery N.I., Ferroni L., Guidi L., Hogewoning S.W., Jajoo A., Misra A.N., Nebauer S.G., Pancaldi S., Penella C., Poli D., Pollastrini M., Romanowska-Duda Z.B., Rutkowska B., Serôdio J., Suresh K., Szulc W., Tambussi E., Yannicari M., Zivcak M. Frequently asked questions about in vivo chlorophyll fluorescence: practical issues. *Photosynthesis Research*. 2014. V. 122. № 2. P. 121–158. doi: [10.1007/s11120-014-0024-6](https://doi.org/10.1007/s11120-014-0024-6).
5. Sun Y., Frankenberg C., Wood J.D., Schimel D.S., Jung M., Guanter L., Drewry D.T., Verma M., Porcar-Castell A., Griffis T.J., Gu L., Magney T.S., Köhler P., Evans B., Yuen K. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science* 2017. V. 358. № 6360. doi: [10.1126/science.aam5747](https://doi.org/10.1126/science.aam5747).
6. Antal T., Konyukhov I., Volgusheva A., Plyusnina T., Khruschev S., Kukarskikh G., Goryachev S., Rubin A. Chlorophyll fluorescence induction and relaxation system for the continuous monitoring of photosynthetic capacity in

- photobioreactors. *Physiologia Plantarum*. 2018. doi: [10.1111/ppl.12693](https://doi.org/10.1111/ppl.12693).
7. Goltsev V.N., Aleksandrov V., Paunov M., Velikova V.B., Tsonev T.D., Kalaji H.M. Identification of nutrients deficiencies in growth medium of *Phaseolus vulgaris* by chlorophyll fluorescence methods: application of artificial neural networks as a tool for rapid recognition. In: *7th International Conference Photosynthesis Research for Sustainability, June 19–25 2016*. Pushchino, 2016. 141 p.
  8. Strasser R.J., Tsimilli-Michael M., Srivastava A. Analysis of the chlorophyll a fluorescence transient. In: *Chlorophyll a Fluorescence: A Signature of Photosynthesis*. Ed. Papageorgiou G.C. Dordrecht: Springer, 2004. P. 321–362. doi: [10.1007/978-1-4020-3218-9\\_12](https://doi.org/10.1007/978-1-4020-3218-9_12).
  9. Stirbet A., Riznichenko G.Y., Rubín A.B., Govindjee Modeling chlorophyll a fluorescence transient: relation to photosynthesis. *Biochemistry (Moscow)*. 2014. V. 79. № 4. P. 291–323. doi: [10.1134/S0006297914040014](https://doi.org/10.1134/S0006297914040014).
  10. Lázár D. Modelling of light-induced chlorophyll a fluorescence rise (O-J-I-P transient) and changes in 820 nm-transmittance signal of photosynthesis. *Photosynthetica*. 2009. V. 47. № 4. P. 483–498. doi: [10.1007/s11099-009-0074-8](https://doi.org/10.1007/s11099-009-0074-8).
  11. Маслаков А.С., Антал Т.К., Ризниченко Г.Ю., Рубин А.Б. Моделирование первичных процессов фотосинтеза с помощью кинетического метода Монте-Карло. *Биофизика*. 2016. Т. 61. № 3. С. 467–477.
  12. Boisvert S., Joly D., Carpentier R. Quantitative analysis of the experimental O-J-I-P chlorophyll fluorescence induction kinetics. Apparent activation energy and origin of each kinetic step. *The FEBS journal*. 2006. V. 273. № 20. P. 4770–4777. doi: [10.1111/j.1742-4658.2006.05475.x](https://doi.org/10.1111/j.1742-4658.2006.05475.x).
  13. Pospíšil P., Dau H. Chlorophyll fluorescence transients of photosystem II membrane particles as a tool for studying photosynthetic oxygen evolution. *Photosynthesis Research*. 2000. V. 65. № 1. P. 41–52. doi: [10.1023/A:1006469809812](https://doi.org/10.1023/A:1006469809812).
  14. Варфоломеев С.Д., Гуревич К.Г. *Биокинетика: Практический курс*. М.: ФАИР-ПРЕСС, 1999. 720 с.
  15. Фролов А.Е., Хрущев С.С., Плюснина Т.Ю., Конюхов И.В., Погосян С.И., Ризниченко Г.Ю. Анализ изменений кривой индукции флуоресценции клеток микроводорослей *Chlorella sp.* в процессе роста методом мультиэкспоненциальной аппроксимации. В: *Экологические и биологические системы*. Москва: MAX Press, 2015.