

УДК 681.5.015.42

А. С. Писарев

Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Метод автоматического анализа траекторий движения стохастических динамических систем

Разработан метод автоматического анализа траекторий движения стохастических динамических систем, в котором тип движения определяется с помощью модифицированного информационного критерия Акаике. Параметры моделей идентифицируются методом взвешенных наименьших квадратов с модифицированной весовой функцией на основе решения обратной задачи с ограничениями на значения параметров. Разработанный метод реализован в сетевой инфраструктуре с веб-интерфейсом и предоставляет возможность студентам и исследователям выполнять анализ экспериментальных данных в автоматическом режиме.

Идентификация, метод средних квадратов смещений, анализ траекторий

Целью работы является разработка метода автоматического анализа траекторий движения стохастических динамических наносистем.

Исследования в области стохастической динамики широко освещены в литературе, включая фундаментальные труды – А. Н. Колмогорова, В. С. Пугачева, Я. Г. Синая, Р. Л. Стратоновича, Н. Винера, К. Ито, Р. Калмана, П. Леви, Б. Мандельброта, которые, в свою очередь, опирались на работы А. Эйнштейна, М. Смолуховского и П. Ланжевена в исследованиях броуновского движения и диффузионных процессов.

Проблема количественной оценки типов и параметров движения сложных стохастических динамических систем, например дробного броуновского движения (ДБД, fractional Brownian motion – fBm), рассматривается во многих областях применения: телекоммуникации, физике, химии, биологии и других.

Эта проблема осложняется тем, что точность количественных оценок зависит от многих неучтенных факторов неопределенности.

В работе представлен разработанный метод компьютерного анализа параметров движения частиц вдоль траекторий на основе применения предложенного А. А. Вавиловым понятия ранга неопределенности и принципа последовательного раскрытия структурной, параметрической и сигнальной неопределенности [1], [2].

В настоящее время особый интерес представляют программы автоматизированного SPT- (single particle tracking) анализа большого объема

экспериментальных данных, получаемых в лабораториях [9]. Применяются различные подходы и методы на этапе SPT-анализа траекторий – идентификации типов и параметров моделей движения частиц [3]–[10].

В данной работе разработан метод взвешенных средних квадратов смещения (weighted mean squared displacement – WMSD) с модифицированными весовой функцией и критерием Акаике для анализа траекторий частиц [11] на основе единого подхода к идентификации четырех типов моделей движения: броуновская диффузия (D), ограниченная диффузия (DR), аномальная субдиффузия (DA), направленное движение с диффузией (DV) [8]. Ошибки локализации позиций частиц в траекториях учитываются добавлением к перечисленным моделям слагаемого $k\sigma^2$, где k – размерность (1D, 2D, 3D), σ – стандартное отклонение, что условно обозначено символом «E» (DE, DVE, DAE, DRE) [8], [9].

Идентификация параметров моделей движения осуществляется на основе решения обратной задачи с ограничениями на допустимые значения параметров (положительные значения коэффициентов диффузии, квадратов ошибок локализации, показателя субдиффузии, радиуса ограниченного движения) и метода взвешенных наименьших квадратов (weighted least squares method – WLS-метод) [12], [13].

Для корректного выбора структуры модели движения используется модифицированный информационный критерий Акаике [14], [15], в ко-

тором при сравнительной оценке качества моделей учитывается их сложность и точность описания экспериментальных данных.

Разработанный метод реализован в сетевом программном комплексе, доступ к которому осуществляется с использованием веб-интерфейсов.

Модели движений. Приблизительная оценка значений коэффициентов диффузии и скоростей движения может производиться в соответствии с техникой идентификации параметров моделей по экспериментальным данным [3]. Среднеквадратические ошибки локализации частиц могут быть учтены в правой части уравнений [8]. Аппроксимация MSD-моделями показывает хорошие результаты для траекторий движения частиц с большими значениями коэффициентов диффузии при малых значениях ошибки локализации, в других случаях рекомендуется применение моделей, учитывающих ошибки локализации [8], [9].

Модели $M = \{M_1, \dots, M_m\}$ содержат в левой части уравнений средний квадрат смещения (mean square displacement – MSD), а в правой части – зависимость от временного лага $l \Delta t$ MSD и параметров: коэффициента диффузии (D), скорости (V) и др. [5], [8], [11].

Модель свободной диффузии (D) описывает случайное движение частицы (Free 2-D diffusion):

$$\langle r^2 \rangle = kD l \Delta t, \quad (1)$$

где D – коэффициент диффузии, $\text{мкм}^2/\text{с}$; $l \Delta t$ – временной лаг; $k = 2 \dim$, $\dim = 1, 2, 3$ – размерность пространства (1 – одномерное, 2 – двумерное, 3 – трехмерное).

Модель направленного движения (V) описывает направленное движение частиц (Directed motion):

$$\langle r^2 \rangle = (v l \Delta t)^2, \quad (2)$$

где v – скорость перемещения, $\text{мкм}/\text{с}$.

Модель направленного движения с диффузией (DV – Continuous directed motion with free 2-D diffusion) характеризует сложное движение, в котором присутствуют как случайные блуждания частицы в соответствии с моделью D (1), так и направленные движения с моделью V (2):

$$\langle r^2 \rangle = kD l \Delta t + (v l \Delta t)^2. \quad (3)$$

Модель аномальной диффузии (DA – Anomalous diffusion) описывает движение частицы под воздействием разнообразных биофизических процессов, в

том числе препятствий на пути движения и других связывающих событий, которые могут иметь сложное механистическое объяснение:

$$\langle r^2 \rangle = kD (l \Delta t)^\alpha, \quad 0 < \alpha < 1. \quad (3)$$

Модель ограниченной диффузии (DR – Confined 2-D diffusion) характеризует движение частицы, ограниченное в пространстве:

$$\langle r^2 \rangle = R_C^2 \left(1 - e^{-kD l \Delta t / R_C^2} \right),$$

где R_C – радиус ограниченного движения частицы, мкм .

Модель свободной диффузии с ошибкой (DE – Free 2-D diffusion plus error) характеризует диффузионное движение при наличии ошибок локализации позиции частицы со стандартным отклонением σ :

$$\langle r^2 \rangle = kD l \Delta t + k\sigma^2.$$

Модель направленного движения (VE) характеризует направленное движение при наличии ошибок локализации позиции частицы со стандартным отклонением σ :

$$\langle r^2 \rangle = (v l \Delta t)^2 + k\sigma^2.$$

Модель направленного движения с диффузией и ошибкой (DVE – Continuous directed motion with free 2-D diffusion plus error) характеризует направленное движение с диффузией при наличии ошибок локализации позиции частицы со стандартным отклонением σ :

$$\langle r^2 \rangle = kD l \Delta t + (v l \Delta t)^2 + k\sigma^2.$$

Модель аномальной диффузии (DAE – Anomalous diffusion plus error) характеризует движение с аномальной диффузией при наличии ошибок локализации позиции частицы со стандартным отклонением σ :

$$\langle r^2 \rangle = kD (l \Delta t)^\alpha + k\sigma^2, \quad 0 \leq \alpha \leq 1.$$

Модель ограниченной диффузии (DRE – Confined 2-D diffusion plus error, Corralled motion) характеризует движение частицы, ограниченное в пространстве при наличии ошибок локализации позиции частицы со стандартным отклонением σ :

$$\langle r^2 \rangle = R_C^2 \left(1 - e^{-kD l \Delta t / R_C^2} \right) + k\sigma^2.$$

В формулах используется обозначение коэффициента $k = 4$ для случая 2-мерной (2D) диффузии.

Качественный предварительный анализ позволяет по виду зависимости MSD от временного лага оценить тип движения частицы.

Зависимость линейного вида может свидетельствовать о наличии свободной диффузии, в то время как параболическая зависимость – о направленном движении.

Уравнение аномальной диффузии (3) при $0 < \alpha < 1$ характеризует субдиффузию.

Метод средних квадратов смещений (MSD). Средний квадрат смещения (mean square displacement – MSD) для временного лага $l \Delta t$ MSD в микрометрах в квадрате определяется по формуле

$$\langle r^2 \rangle_{l\Delta t} = \frac{\sum_{i=1}^{N-l+1} r_{i,l\Delta t}^2}{N-l+1}, \quad l = 1, 2, \dots, N.$$

Дисперсия $\sigma_{l\Delta t}^2$ (также в микрометрах в квадрате):

$$\sigma_{l\Delta t}^2 = \frac{\sum_{i=1}^{N-l+1} \left(r_{i,l\Delta t}^2 - \langle r^2 \rangle_{l\Delta t} \right)^2}{(N-l)}.$$

Весовой коэффициент среднего квадрата смещения MSD для l -го временного лага

$$w_l = \frac{1}{\sigma_{l\Delta t}^2}.$$

Нормализованный весовой коэффициент среднего квадрата смещения MSD

$$w'_l = w_l \Big/ \sum_{l=1}^{N-1} w_l.$$

Нормализованные весовые коэффициенты применяются в методе взвешенных наименьших квадратов при идентификации модели движения по экспериментальным траекториям.

К особенностям зависимости средних квадратов смещения от временного лага относятся непостоянство среднеквадратических ошибок и снижение достоверности при больших значениях лага. В связи с этим была применена модификация весовой функции, в которой $\sigma_{l\Delta t}^2$ скорректирована с учетом значения доверительного интервала, зависящего от числа усредненных точек.

Метод средних по ансамблю квадратов смещений (ETA-MSD). Средний по ансамблю траекторий квадрат смещения (ensemble time

averaged mean square displacement – ETA-MSD) определяется усреднением MSD индивидуальных траекторий ансамбля по следующей формуле [8]:

$$\begin{aligned} \bar{\rho}_n^{(ens)} &= \frac{1}{N_T(N-n)} \sum_{j=1}^{N_T} \sum_{i=1}^{N-n} \left(\mathbf{r}_{i+n}^{(j)} - \mathbf{r}_i^{(j)} \right)^2 = \\ &= \frac{1}{N_T} \sum_{j=1}^{N_T} \bar{\rho}_n^{(j)}, \quad n = 1, \dots, N-1, \end{aligned}$$

где n – номер позиции в траектории; N – длина наибольшей траектории ансамбля, равная числу позиций в траектории; $\mathbf{r}_i^{(j)}$ – i -я позиция в j -й траектории; $\bar{\rho}_n^{(j)}$ – значение MSD в позиции n , вычисленное для j -й индивидуальной траектории.

Дисперсия σ_n^2 определяется по формуле

$$\sigma_n^2 = \frac{\sum_{j=1}^{N_n} \left(\bar{\rho}_n^{(j)} - \bar{\rho}_n^{(ens)} \right)^2}{(N_n - 1)},$$

где N_n – число траекторий, длина которых более или равна $(n+1)$.

При идентификации модели движения по ETA-MSD с использованием метода взвешенных наименьших квадратов применяются нормализованные весовые коэффициенты:

$$w'_n = w_n \Big/ \sum_{n=1}^{N_{\max}-1} w_n,$$

где $w_n = \frac{1}{\sigma_n^2}$ – весовой коэффициент.

При идентификации параметров модели движения частиц с использованием функции ETA-MSD (ensemble time averaged mean square displacement) используется модификация критерия взвешенных сумм квадратов остатков (WRSS), учитывающая дополнительно вариабельность средних квадратов смещений всех индивидуальных траекторий с номером n позиции в траектории:

$$WRSS = \sum_{i=1}^M \frac{\bar{w}_i^{(ens)}}{\sum_{n=1}^{N_T} \bar{w}_i^{(ens)}} \bar{w}_i \left(y_i^{\text{model}} - y_i^{\text{data}} \right)^2,$$

где \bar{w}_i – нормализованный коэффициент веса i -го измерения:

$$\overline{w}_i = \overline{w}_i / \sum_{j=1}^M \overline{w}_j, \quad \overline{w}_i = 1 / \sum_{n=1}^{N_T} \sigma_n^2.$$

Метод структурной идентификации. Для корректного выбора модели типа движения частицы используется модифицированный информационный критерий Akaike (AICc) [14], [15]. В этом критерии при сравнительной оценке качества моделей учитывается их сложность и точность описания экспериментальных данных.

Для структурной и параметрической идентификации моделей с ошибками локализации применен метод обратных задач в условиях неопределенности [12], [13] с модифицированными информационными критериями BIC, AICc, R2adj [14], [15].

В модифицированном информационном критерии Akaike (AIC_{cw}) при оценке качества моделей учитывается сложность (число параметров) и точность описания экспериментальных данных:

$$AIC_{cw} = 2k + n \ln \frac{WRSS}{n} + \frac{2k(k+1)}{n-k-1},$$

где k – число параметров модели; n – размер выборки [8].

Критерий взвешенной суммы квадратов остатков (WRSS):

$$WRSS = \sum_{i=1}^M w'_i \left(y_i^{\text{model}} - y_i^{\text{data}} \right)^2,$$

где w'_i – нормализованная весовая функция; y_i^{model} ; y_i^{data} – модельные и экспериментальные данные в моменты времени t_i .

Результаты. Метод прошел апробацию на синтетических траекториях, соответствующих вышеуказанным моделям. Результаты эксперимента показали, что метод позволяет правильно определить структуру и значения параметров модели движения частиц в условиях неопределенности.

Метод взвешенных наименьших квадратов с модифицированными весовыми функциями позволил произвести более чем в 10 раз точные оценки значений параметров моделей по сравнению с обычным методом взвешенных наименьших квадратов.

На рис. 1 показан пример траектории с ограниченной диффузией. Погрешность оценки коэффициента диффузии и радиуса ограничивающей сферы составляет менее 1 %.

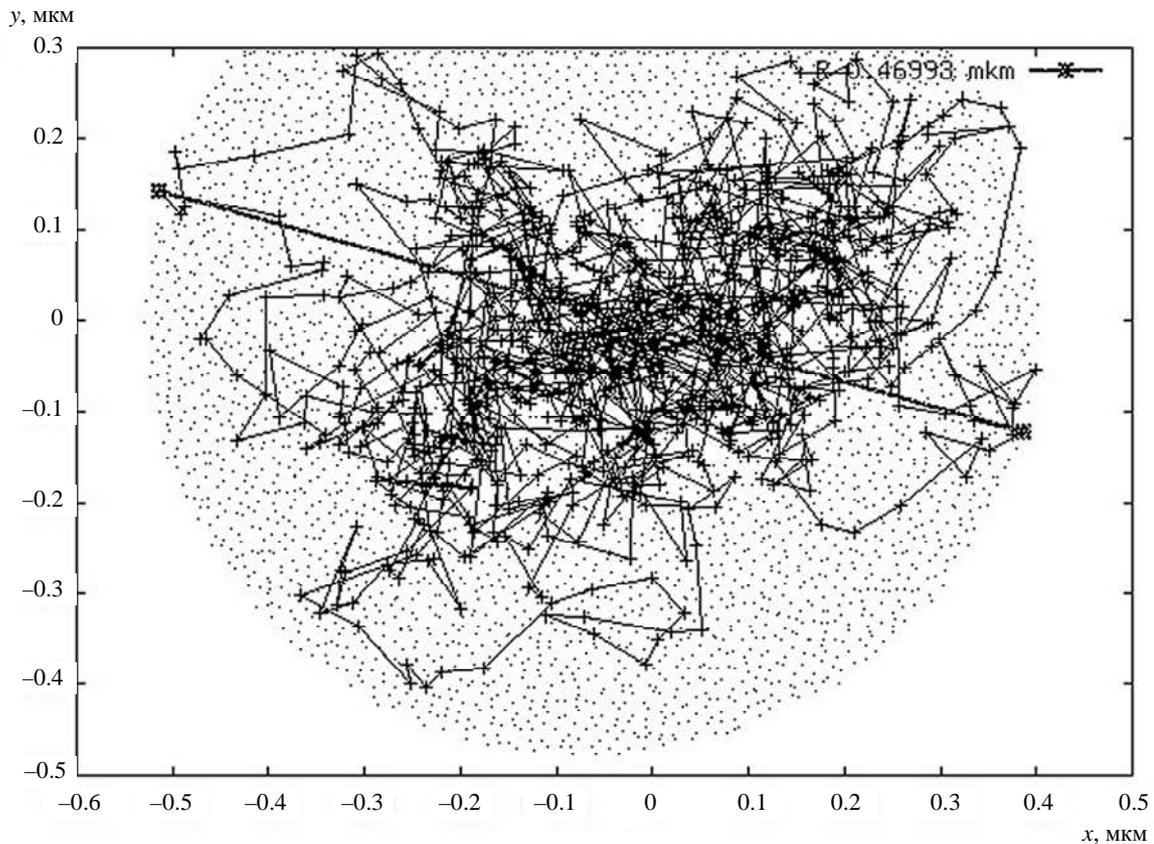


Рис. 1

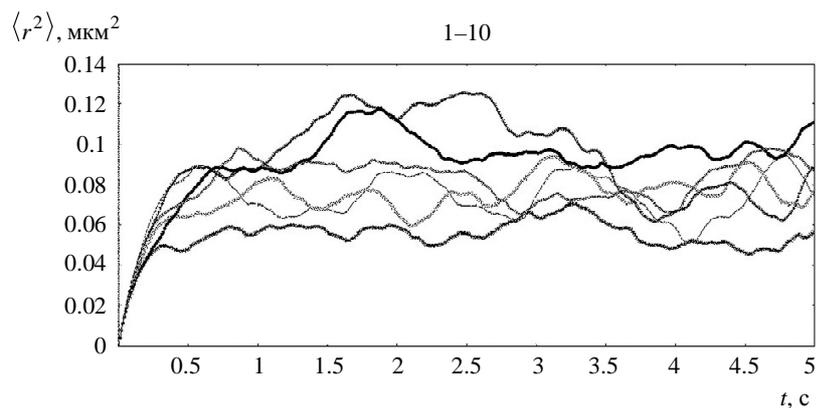


Рис. 2

При определении параметров модели и типа аномальной субдиффузии (рис. 2 – MSD ансамбля траекторий с аномальной диффузией (DA) с заданным параметром $\alpha = 0.7$) были получены результаты с относительной погрешностью менее 1%. Более точные оценки параметров моделей были получены с использованием ETA-MSD.

Сетевой программный комплекс. Метод автоматического анализа траекторий движения реализован в сетевом программном комплексе с веб-интерфейсом и включает следующие компоненты:

- интерфейс для взаимодействия с пользователями с использованием стандартных веб-браузеров;

- веб-сервер (Apache);
- база данных (MySQL);
- сервер приложений (Tomcat);
- программные агенты, выполняющие задачи анализа данных (Java).

Пользователь с использованием веб-браузера загружает траектории частиц на сервер. Программные агенты выполняют анализ данных и сохраняют результаты в базе данных.

Незарегистрированный пользователь (гость) может либо ожидать результаты анализа в периодически обновляемой странице веб-браузера, либо получить уведомление по электронной почте. Если пользователь указывает адрес электронной почты, то ему отправляется сообщение со ссылкой на результаты анализа в виде pdf-файла.

Входные данные для WMSD содержатся в загружаемом на сервер файле ансамбля траекторий, представленном в zip- или xml-форматах. Индивидуальные траектории 1D, 2D и 3D могут быть представлены в виде текстового файла (plain text) со значениями отметок времени и координат частицы (t, x, y, z) или в формате представления результатов трекинга (ISBI Challenge format).

WMSD выполняет автоматический анализ траекторий и формирует отчет анализа в виде html-страницы и файла в формате pdf, которые содержат:

- результаты анализа усредненных по времени MSD индивидуальных траекторий (time averaged mean square displacement – TA-MSD);
- результаты анализа усредненного по ансамблю траекторий MSD (ETA-MSD);
- средние, минимальные и максимальные значения скоростей частиц в индивидуальных траекториях;

- изображение траекторий частиц.

Результаты ETA-MSD анализа включают:

- тип модели движения;
- значения критериев структурной идентификации AICsw, BICw, R2adjw;
- параметры модели движения (DC – коэффициент диффузии, v – скорость, α – показатель экспоненты аномальной субдиффузии, R_c – размер области ограниченного движения, σ – стандартное отклонение среднеквадратической ошибки локализации частиц);

- график реконструированной зависимости ETA-MSD от временного лага.

Результаты TA-MSD анализа включают:

- для каждой индивидуальной траектории: идентифицированный тип модели движения; параметры модели; значения критериев структурной идентификации AICsw, BICw, R2adjw;

- для ансамбля траекторий: доминирующий тип движения в соответствии с моделями движения; процентный состав траекторий по каждому идентифицированному типу движения и статистические характеристики значений параметров моделей.

Разработан метод взвешенных средних квадратов смещений для анализа стохастических динамических систем по траекториям движения.

Метод реализован в сетевом программном комплексе и прошел апробацию на синтетических траекториях, полученных при моделировании различных типов движений, соответствующих моделям броуновской диффузии, аномальной

диффузии и др. Результаты эксперимента показали, что WMSD позволяет на основе единого подхода корректно определять структуру и параметры стохастических моделей по траекториям движения в условиях неопределенности [11].

СПИСОК ЛИТЕРАТУРЫ

1. Вавилов А. А. Структурный и параметрический синтез сложных систем управления / ЛЭТИ. Л., 1979. 94 с.
2. Вавилов А. А., Имаев Д. Х., Фомин Б. Ф. Системное моделирование, анализ и эволюционный синтез сложных систем управления // Имитационное моделирование производственных систем / под общ. ред. чл.-кор. АН СССР А. А. Вавилова. М.: Машиностроение; Берлин: Техника, 1983. С. 5–100.
3. Qian H., Sheetz M. P., Elson E. L. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems // Biophys. j. 1991. Vol. 60, № 4. P. 910.
4. Saxton M. J. Single-particle tracking: the distribution of diffusion coefficients // Biophys. j. 1997. Vol. 72, № 4. P. 1744.
5. Saxton M. J., Jacobson K. Single-particle tracking: applications to membrane dynamics // An. rev. of biophys. and biomolecular structure. 1997. Vol. 26, № 1. P. 373–399.
6. Saxton M. J. A particle tracking meet // Nature methods. 2014. Vol. 11, № 3. P. 247–248.
7. Fractional Brownian motion in crowded fluids / D. Ernst, M. Hellmann, J. Köhler, M. Weiss // Soft Matter. 2012. Vol. 8, № 18. P. 4886–4889.
8. Michalet X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium // Phys. rev. E. 2010. Vol. 82, № 4. P. 1–26.
9. Bayesian approach to MSD-based analysis of particle motion in live cells / N. Monnier, S. M. Guo, M. Mori, J. He, P. Lénárt, M. Bathe // Biophys. j. 2012. Vol. 103, № 3. P. 616–626.
10. Tuerkcan S., Alexandrou A., Masson J. B. A Bayesian inference scheme to extract diffusivity and potential fields from confined single-molecule trajectories // Biophys. j. 2012. Vol. 102, № 10. P. 2288–2298.
11. Писарев А. С. Метод оценки параметров движения микрочастиц по траекториям в условиях неопределенности // Тр. XV Междунар. конф. по мягким вычислениям и измерениям (SCM'2015). Санкт-Петербург. 2015. СПб., 2015. С. 29–33.
12. Писарев А. С. Применение многоагентных технологий и эвристических методов в задачах нечеткой многокритериальной оптимизации // Изв. СПбГЭТУ «ЛЭТИ». 2013. № 5. С. 72–79.
13. Писарев А. С., Самсонова М. Г. Метод решения многокритериальных обратных задач в условиях неопределенности // Биофизика. 2013. Т. 58, № 2. С. 157–166.
14. Akaike H. A new look at the statistical model identification // Selected Papers of Hirotugu Akaike. New York: Springer, 1998. P. 215–222.
15. Burnham K. P., Anderson D. R. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media, 2002. 515 p.

A. S. Pisarev

Saint Petersburg Electrotechnical University «LETI»

AUTOMATIC ANALYSIS METHOD OF STOCHASTIC DYNAMICAL SYSTEMS MOTION TRAJECTORIES

A method for automatic analysis of the trajectories of stochastic dynamical systems, in which the type of movement is determined by the modified information criteria Akaike, and the identification of the model parameters is carried out by the method of weighted least squares with modified weight function on the basis of the inverse problem solution with constraints on the parameter values. The developed method is implemented in the network infrastructure with Web-based interface and provides an opportunity for students and researchers to carry out the analysis of experimental data in the automatic mode.

Identification, the method of the mean square displacement, the analysis of the trajectories