

ПРИМЕНЕНИЕ МЕТОДОВ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ ВАЛИДАЦИИ СИМУЛЯЦИОННОЙ МОДЕЛИ КОЛОНКИ КОРТЕКСА МЫШИ

Зырянова П.А.

Зырянова Полина Александровна – бакалавр,
кафедра интеллектуальной робототехники,
Казанский (Приволжский) федеральный университет, г. Казань

Аннотация: эпилепсия представляет собой хроническое неврологическое заболевание головного мозга, характеризующееся склонностью организма к внезапным судорожным приступам. Для выявления динамики эпилептического приступа была разработана симуляционная модель колонки кортекса мыши, для дальнейшего использования которой необходима её валидация на предмет соответствия биологической модели. В данной статье описывается валидация симуляционной модели с помощью 2 непараметрических критериев математической статистики – критерия однородности Смирнова и критерия ядерной оценки плотности вероятности. В статье обосновывается невозможность применения критерия Смирнова для валидации и приводятся положительные результаты соответствия симуляционной и биологической моделей с помощью критерия ЯОП.

Ключевые слова: валидация, кортикальная колонка, критерий однородности Смирнова, потенциал локального поля, статистическая гипотеза, уровень значимости, ядерная оценка плотности вероятности (ЯОП), kde-тест, p-значение.

APPLICATION OF MATHEMATICAL STATISTICS METHODS TO VALIDATE A SIMULATION MODEL OF A MOUSE CORTICAL COLUMN

Zyryanova P.A.

Zyryanova Polina Alexandrovna – Bachelor,
DEPARTMENT OF INTELLIGENT ROBOTICS,
KAZAN (VOLGA REGION) FEDERAL UNIVERSITY, KAZAN

Abstract: epilepsy is the chronic neurological disease of the brain characterized by the sudden body cramps. To identify the dynamics of epileptic seizure, a simulation model of a mouse cortical column was developed. For further using of this simulation model, it needs to validate with the biological model. This article describes the validation of a simulation model using 2 nonparametric statistics tests – Kolmogorov-Smirnov test (K-S test) and kernel density estimation (kde) test. This article proves the impossibility of Smirnov test application to validate a simulation model and shows positive results of validate simulation and biological models using kde-test.

Keywords: validation, cortical column, Kolmogorov-Smirnov test (K-S test), local field potential (lfp), statistical hypothesis, significance level, kernel density estimation (kde), kde-test, p-value.

Введение

По данным Всемирной организации здравоохранения, около 50 миллионов человек во всём мире страдают от приступов эпилепсии, что делает её одним из наиболее распространённых неврологических заболеваний (от 4 до 10 человек на 1000 населения) [1]. В России данный показатель достигает 3,4 человек на 1000 населения, что сопоставимо с Европейскими странами [2].

До сих пор причина заболеваний не известна примерно в 50% случаев по всему миру [1], чем и обусловлена **актуальность** данного исследования. Причины эпилепсии могут быть инфекционными, структурными, генетическими и т.п. Пороки развития коры головного мозга (так называемые Malformations of cortical development или MCD) представляют собой аномалии развития коры головного мозга, влияющие на нормальные процессы её развития и вовлекающие клетки, которые при нормальных условиях участвовали бы в формировании структуры коры головного мозга. Точная частота MCD неизвестна, однако подсчитано, что 25-40% эпилепсии имеют своей причиной MCD и до 75% пациентов, страдающих пороками развития коры головного мозга, будут также страдать и эпилепсией [3, с. 57; 4, с. 48].

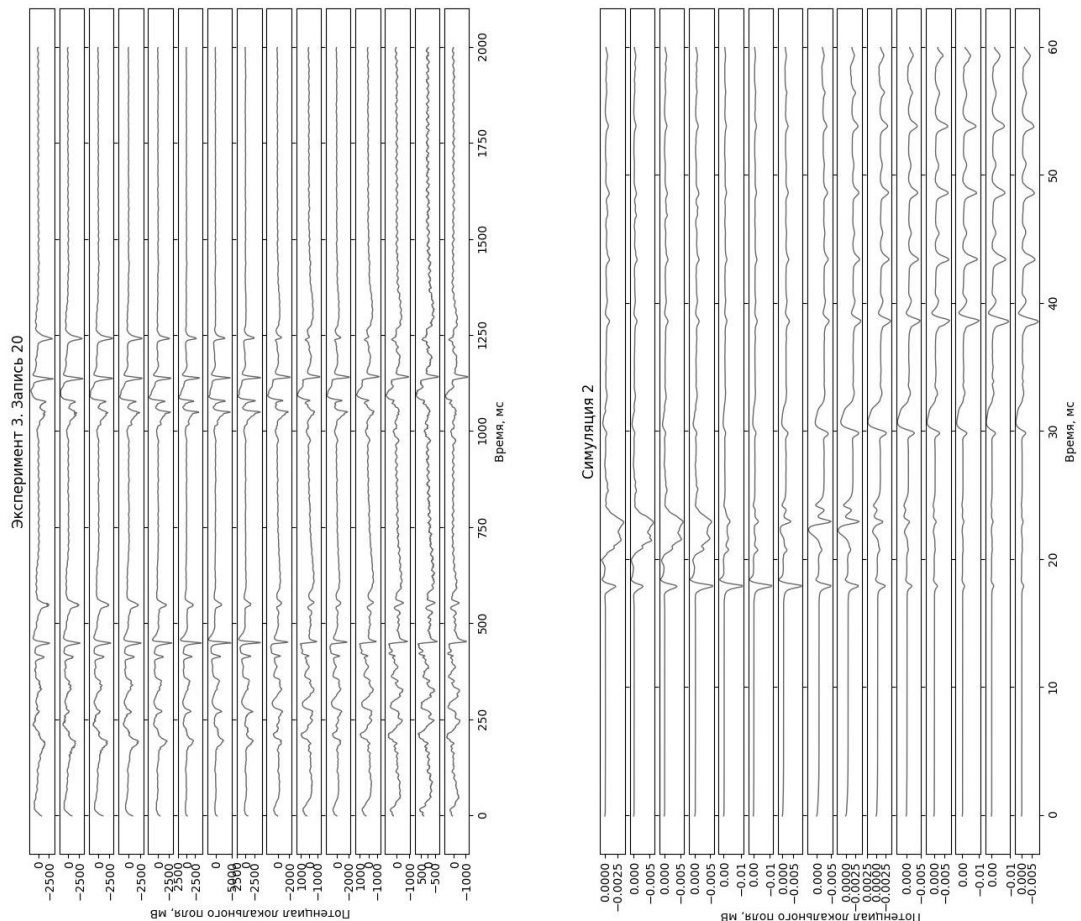
Однако исследования показывают, что зачастую эпилепсия характеризуется распространением судорожной активности от патологической (очаговой) кортикальной колонки к другим колонкам, находящимся в окружающей коре [5, с. 46]. На основании данного предположения была разработана симуляционная модель колонки кортекса мыши, валидация которой описывается ниже.

1. Постановка задачи

Биологические данные были представлены в виде 15 сенсоров тетрода, отражающих суммарный внеклеточный потенциал (lfp) около нейрона в определённый момент времени (рис. 1а) в течение 2000 мс. К аналогичному виду были приведены симуляционные данные (рис. 1б), изначально представленные

в виде двумерной матрицы размерностью (n, b) , где n - количество смоделированных нейронов кортикальной колонки, b - количество параметров нейрона. Продолжительность симуляции составляла 60 мс, исходные данные были записаны с интервалом 0,1 мс.

После анализа исходных данных была поставлена следующая задача: проверить соответствие симуляционных и биологических данных с помощью статистического критерия Колмогорова-Смирнова и метода ядерной оценки плотности вероятности, а также протестировать эквивалентность внутри различных биологических экспериментов, после чего выявить наиболее подходящий метод валидации симуляционной модели кортикальной колонки и получить итоговые результаты с использованием данного метода.



(а) Исходные биологические данные (б) Модифицированные симуляционные данные

Рис. 1(а)-(б). Примеры исходных биологических (а) и симуляционных (б) данных. Лфр на 15 сенсорах

2. Методы исследования

В математической статистике различают несколько групп статистических критериев - критерии значимости, критерии согласия и критерии однородности. Последние исследуют, подчиняются ли 2 заданные случайные величины одному и тому же распределению [6; 7]. Очевидно, что для сравнения законов распределения биологических и симуляционных данных нейрональной активности кортикальной колонки необходимо использование статистических критериев однородности.

2.1. Критерий однородности Смирнова

Различают критерий согласия Колмогорова и критерий однородности Смирнова [8]. Первый применяется для того, чтобы проверить гипотезу о совпадении функции распределения заданной выборки с распределением, когда оно также полностью задано. Второй же необходим, чтобы проверить гипотезу о совпадении функций распределения 2 независимых выборок, когда они заранее не заданы явно. Иногда эти два критерия объединяют в один и называют критерием Колмогорова-Смирнова [8]. Таким образом, в случае валидации симуляционных и биологических данных нейрональной активности кортикальной колонки необходимо использовать критерий однородности Смирнова.

Нулевая гипотеза H_0 данного критерия заключается в совпадении функций распределения двух независимых выборок [8, с. 83; 9, с. 45]. Существует 2 алгоритма проверки правильности H_0 - алгоритм

по критической области и алгоритм с использованием p -value. В дальнейшей работе использовался алгоритм с использованием p -value, поскольку он обладает большей информативностью [10, с.59]. Уровень значимости α принимался равным 0,05 - это значение является одним из наиболее популярных на практике [8, с.19; 9, с.8]. Статистика критерия Смирнова находится по следующей формуле:

$$(1) T = \sup |F_n(x) - G_m(x)|,$$

где x - значение выборки, $F_n(x)$ и $G_m(x)$ - эмпирические функции распределения выборок, n и m - объёмы выборок [8, с.84].

Статистика критерия однородности Смирнова имеет распределение Колмогорова, а критическая область имеет вид $T > C$ [8, с. 84; 9, с. 46]. Очевидно, что p -value будет находиться по формуле:

$$(2) p = 1 - G_T(t),$$

где G_T - функция распределения Колмогорова F_{kolm} , t - найденное по формуле (1) значение статистики. Однако в большинстве пакетов вычислительной статистики отсутствуют методы для вычисления функции распределения Колмогорова [9, с.46], поэтому при математических расчётах можно пользоваться аппроксимацией:

$$(3) F_{kolm}(t) \approx 1 - 2e^{-2t^2},$$

где: t - найденное по формуле (1) значение статистики.

На основании найденного критического уровня значимости делается вывод о правильности или ошибочности нулевой гипотезы.

2.2. Критерий ядерной оценки плотности вероятности

Ядро представляет собой весовую функцию, которая используется при оценке распределений, параметров, а также при анализе временных рядов. Нулевая гипотеза H_0 критерия kde test заключается в совпадении плотностей вероятностей двух независимых выборок [11]. Плотность вероятности в случае использования данного критерия находится методом ядерной оценки по следующей формуле:

$$(4) f(x) = \frac{1}{n} * \sum_{i=1}^n K_h(x - x_i),$$

где: x - точка, в которой происходит оценка плотности, n - объём выборки, (x_1, x_2, \dots, x_n) - значения исходной выборки, K - используемое ядро, h - ширина окна (сглаживающий параметр) [11]. Ширина окна задаётся свободно исследователем. В данном исследовании использовались 4 вида ядер (прямоугольное, косинусное, Гауссово, Епанечникова), формулы расчёта которых основывались на научных исследованиях, посвящённых применению ядерной оценки плотности вероятности [12, с.37].

Статистика критерия kde test находится по следующей формуле:

$$(5) T = \psi_1 + \psi_2 - (\psi_{1,2} + \psi_{2,1}),$$

где: ψ_1 и ψ_2 - внутрипараметрические различия выборок, а $\psi_{1,2}$ и $\psi_{2,1}$ - межпараметрические различия выборок [13, с.25; 14, с.2]. Параметры для расчёта формулы (5) находятся следующим образом:

$$(6) \psi_1 = \frac{1}{n_1^2} * \sum_{i=1}^{n_1} \sum_{i_2=1}^{n_1} K_{h_1}(x_{i_1} - x_{i_2}),$$

$$(7) \psi_2 = \frac{1}{n_2^2} * \sum_{i=1}^{n_2} \sum_{i_2=1}^{n_2} K_{h_2}(y_{i_1} - y_{i_2}),$$

$$(8) \psi_{1,2} = \frac{1}{n_1 n_2} * \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} K_{h_1}(x_i - y_j),$$

$$(9) \psi_{2,1} = \frac{1}{n_1 n_2} * \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} K_{h_2}(x_i - y_j),$$

где: n_1 и n_2 - объёмы выборок, K - значение выбранного ядра, h_1 и h_2 - сглаживающие параметры выборок, x_i и y_j - данные соответственно 1-й и 2-й выборки [13, с.25; 14, с.3].

Если внутрипараметрические различия превышают межпараметрические (т.е. $T > 0$), H_0 принимается, в противном случае - отклоняется. Критическая область kde test имеет вид $T < 0$, а статистика имеет асимптотически нормальное распределение [13, с.26; 14, с.4]. Таким образом, p -value критерия kde test будет находиться по формуле:

$$(10) p = G_T(t) = \frac{1}{\sqrt{2\pi}} * \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

где: G_T - функция нормального распределения (функция Гаусса), t - найденное по формуле (5) значение статистики.

Сравнивая полученное значение p -value с заданным уровнем значимости, принимаем решение о правильности нулевой гипотезы.

3. Результаты исследования

3.1. Результаты применения критерия Колмогорова-Смирнова

Из 15 имеющихся биологических экспериментов (крыс) были выбраны 3 с наиболее интересной и разноплановой нейрональной активностью - эксперименты №3, 10, 23. Поскольку биоданные были записаны с частотой 1 мс, а симуляционные - 0,1 мс, для проведения валидации было взято каждое 10-е значение исходных симуляционных данных с целью приведения данных к одинаковому формату для достоверности результатов валидации. Результаты валидации симуляционной и биологической моделей с помощью критерия однородности Смирнова приведены в Таблице 1.

Таблица 1. P-value валидации биологических и симуляционных данных по методу Колмогорова-Смирнова

Сенсор	Симуляция 1 / Эксперимент Запись 0	Симуляция 1 / Эксперимент 5. Запись 5	Симуляция 2 / Эксперимент 7. Запись 30	Симуляция 2 / Эксперимент 23. Запись 15
Сенсор#1	3.5504e ⁻¹³	2.9782e ⁻¹³	2.5697e ⁻¹⁵	1.3026e ⁻¹³
Сенсор#2	3.3486e ⁻¹³	2.8085e ⁻¹³	4.5661e ⁻¹⁶	2.8918e ⁻¹⁴
Сенсор#3	2.2195e ⁻¹³	4.4237e ⁻¹⁴	7.2766e ⁻¹⁷	2.9782e ⁻¹³
Сенсор#4	1.1564e ⁻¹³	2.3542e ⁻¹³	3.0539e ⁻¹⁷	3.7641e ⁻¹³
Сенсор#5	2.497e ⁻¹³	6.3539e ⁻¹⁴	3.4926e ⁻¹⁷	9.1073e ⁻¹⁴
Сенсор#6	1.4741e ⁻¹³	4.1637e ⁻¹⁴	1.0827e ⁻¹⁶	4.2302e ⁻¹³
Сенсор#7	1.2692e ⁻¹⁷	4.5281e ⁻¹⁵	4.8801e ⁻¹⁷	5.3031e ⁻¹⁴
Сенсор#8	8.8182e ⁻²⁰	2.6204e ⁻¹⁸	2.5429e ⁻¹⁶	8.9911e ⁻¹⁵
Сенсор#9	1.9169e ⁻²⁰	8.0793e ⁻¹⁴	3.4553e ⁻¹⁸	1.9724e ⁻¹³
Сенсор#10	1.7525e ⁻¹³	2.9782e ⁻¹³	1.0184e ⁻¹⁹	2.0042e ⁻¹⁴
Сенсор#11	3.9664e ⁻¹⁸	4.4237e ⁻¹⁴	4.5648e ⁻¹⁷	3.3486e ⁻¹³
Сенсор#12	1.1796e ⁻²⁴	7.9411e ⁻¹⁵	7.77602e ⁻¹⁷	7.4623e ⁻¹⁵
Сенсор#13	4.4841e ⁻¹³	2.2195e ⁻¹³	2.1253e ⁻¹⁵	2.8918e ⁻¹⁴
Сенсор#14	8.0793e ⁻¹⁴	3.5504e ⁻¹³	7.1755e ⁻¹⁶	4.8208e ⁻¹⁵
Сенсор#15	1.0263e ⁻¹³	3.7641e ⁻¹³	1.3861e ⁻¹⁴	1.1986e ⁻¹⁵
Обобщённые данные	3.5444e⁻²²⁴	4.3359e⁻¹⁹⁸	4.1573e⁻²⁴⁴	1.5316e⁻¹⁹⁵

Исходя из данных Таблицы 1 можно увидеть, что метод Колмогорова-Смирнова не подходит для валидации биологических и симуляционных данных нейрональной активности кортикальной колонки из-за получения низких значений *p-value*.

При валидации данным методом записей биологических данных между собой в некоторых сравнениях были получены высокие *p-value* (в данном контексте высокими считаются все значения критического уровня значимости выше 0,05), выделенные в Таблице 2 серым цветом. Однако они наблюдались только при валидации данных по сенсорам; при анализе же обобщённых данных критерий также не дал положительных результатов.

Таблица 2. P-value валидации биологических данных между собой по методу Колмогорова-Смирнова. Серым цветом выделены высокие значения *p-value* ($p > 0.05$)

Сенсор	Эксперимент 3. / Запись 4 / Эксперимент 10. Запись 10	Эксперимент 10. / Запись 20 / Эксперимент 10. Запись 37	Эксперимент 23. / Запись 40 / Эксперимент 23. Запись 45	Эксперимент 3. / Запись 1 / Эксперимент 23. Запись 10
Сенсор#1	0.12011	0.14974	0.44071	3.1806e ⁻¹¹
Сенсор#2	0.01489	0.12939	0.08827	4.562e ⁻¹⁹
Сенсор#3	6.2844e ⁻⁵	0.09548	0.39402	4.505e ⁻²⁴
Сенсор#4	7.4776e ⁻⁹	0.10317	0.19813	1.88002e ⁻³¹
Сенсор#5	3.4799e ⁻¹⁰	0.08827	0.18923	2.2188e ⁻⁴⁵
Сенсор#6	2.5758e ⁻¹²	0.00301	0.24187	5.4639e ⁻⁵³
Сенсор#7	0.00149	6.2844e ⁻⁵	0.08827	8.2318e ⁻⁶³
Сенсор#8	0.12011	1.3442e ⁻⁸	0.05873	3.7936e ⁻⁵⁵
Сенсор#9	0.00422	5.7494e ⁻¹⁷	0.06384	1.2275e ⁻²⁷
Сенсор#10	0.08152	4.6052e ⁻¹⁸	0.19813	1.2102e ⁻³⁷
Сенсор#11	0.00653	3.4799e ⁻¹⁰	0.13927	1.1315e ⁻²⁹
Сенсор#12	2.205e ⁻⁷	6.8815e ⁻¹⁴	0.06933	2.3166e ⁻²³
Сенсор#13	0.08152	6.1816e ⁻¹¹	0.08827	3.4799e ⁻¹⁰
Сенсор#14	0.00896	3.4799e ⁻¹⁰	0.27489	6.2137e ⁻¹³
Сенсор#15	0.03484	4.7024e ⁻⁵	0.03811	2.8883e ⁻⁶
Обобщённые данные	2.0627e⁻²⁰	5.8063e⁻⁴⁵	1.6127e⁻⁷	7.3643e⁻²⁶⁶

При анализе релевантности симуляционных данных между собой было использовано 2 метода валидации - данные с частотой записи 0,1 мс (исходные) и 1 мс (преобразованные для сравнения с биологическими). Как и в случае с биологическими, анализ обобщённых данных не привёл к высоким значениям *p-value* в большинстве случаев (Таблица 3).

Таблица 3. P-value валидации симуляционных данных между собой по методу Колмогорова-Смирнова. Серым цветом выделены высокие значения p-value ($p > 0.05$)

Сенсор	Сим. 1 / Сим. 2 По 0.1 мс	Сим. 1 / Сим. 2 По 1 мс	Сенсор	Сим. 1 / Сим. 2 По 0.1 мс	Сим. 1 / Сим. 2 По 1 мс
Сенсор#1	3.1469e ⁻⁴⁵	1.0577e ⁻⁵	Сенсор#9	7.8579e ⁻¹¹	0.37775
Сенсор#2	1.1261e ⁻⁵⁶	4.2291e ⁻⁶	Сенсор#10	1.0235e ⁻¹¹	0.04704
Сенсор#3	1.1372e ⁻⁵⁶	2.2135e ⁻⁷	Сенсор#11	7.9007e ⁻¹⁶	0.01619
Сенсор#4	3.2847e ⁻³³	0.00013	Сенсор#12	1.2639e ⁻¹⁹	0.00903
Сенсор#5	2.4464e ⁻²⁰	0.00255	Сенсор#13	4.595e ⁻²¹	0.00488
Сенсор#6	2.4279e ⁻¹⁷	0.00903	Сенсор#14	1.4822e ⁻²¹	0.00923
Сенсор#7	1.4822e ⁻²¹	0.01619	Сенсор#15	5.7644e ⁻²⁶	0.00225
Сенсор#8	2.4279e ⁻¹⁷	0.07627	Обобщённые данные	3.5203e ⁻⁶¹	3.296e ⁻⁸

Поскольку была выявлена неприменимость критерия однородности Смирнова к валидации имеющихся данных, в дальнейшем при работе был использован статистический критерий kde test.

3.2. Результаты применения критерия kde test

Для сравнительного анализа методов валидации с использованием kde (применение к различным модификациям данных) и определения наиболее подходящего из них были взяты 10 биологических записей различных экспериментов (№3, 4, 5, 8, 15, 21). Записи были подобраны таким образом, чтобы показать различную степень активности кортикальной колонки.

Биологические данные находятся в отрезке $\sim [-5000; 5000]$ в зависимости от записи, у симуляционных же данных минимальное и максимальное значения не превосходят 1 по модулю, поэтому исходные данные были нормализованы для приведения к эквивалентному размаху, после чего провалидированы с помощью Гауссова ядра (рис. 2). Нормализация данных производилась в 2 шага - деление на среднее выборки и вычитание стандартного отклонения. Далее в работе использовались исключительно нормализованные данные для точности исследования.

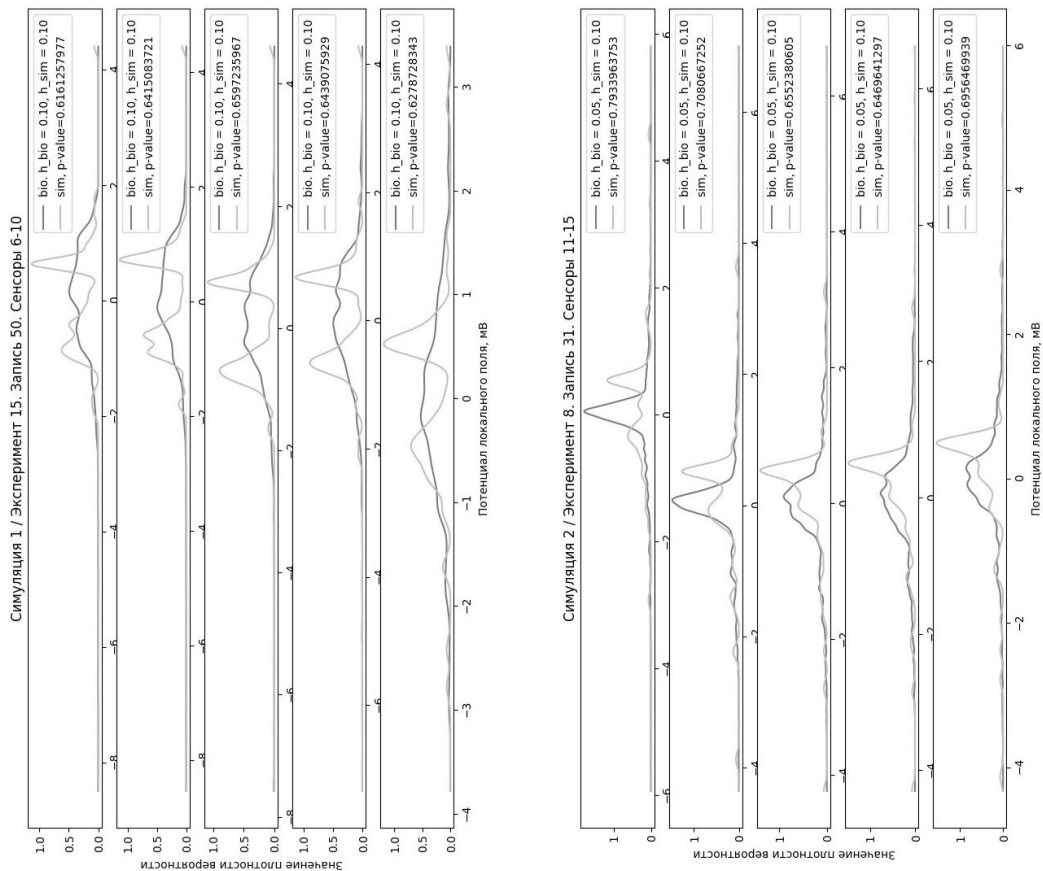


Рис. 2. Результаты применения kde test и нормального ядра к валидации нормализованных био- и симуляционных данных

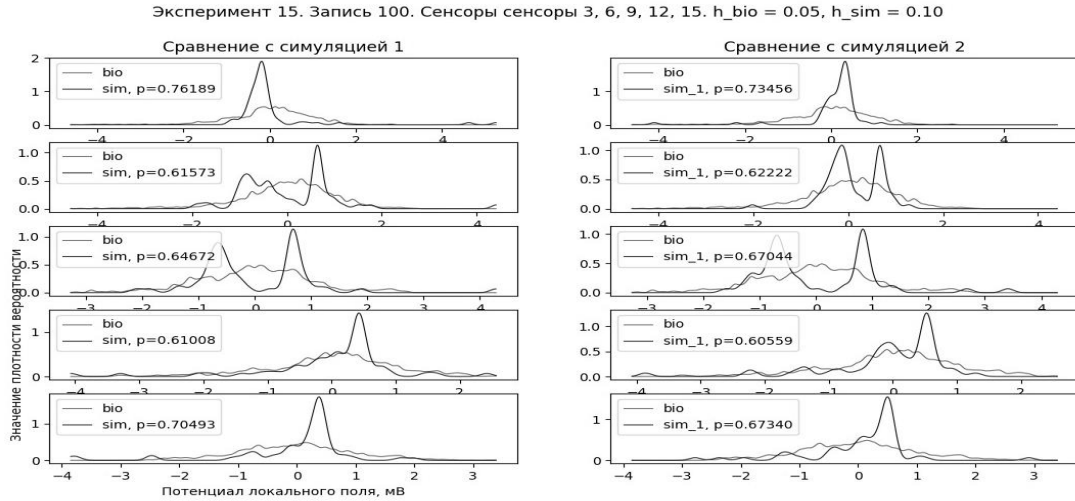


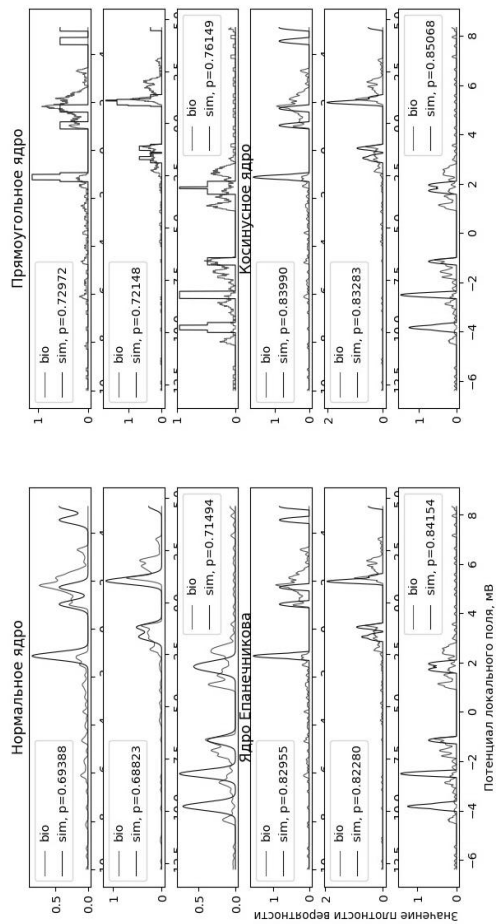
Рис. 3. Сравнение валидации записи 100 биологического эксперимента 15 с обеими симуляциями

Следующим шагом было сравнение валидации одной и той же биологической записи с разными симуляционными записями при одинаковых сглаживающих параметрах (нормальное ядро, нормализованные данные). Для валидации были взяты данные 5 сенсоров, помещённых на разной глубине кортикальной колонки. Результаты сравнений, отражающие графики плотности вероятности, заданные сглаживающие параметры h и полученные p -value, представлены на рис. 3. Исходя из рис. 3, можно увидеть, что валидация с разными симуляциями даёт близкие результаты (различие в p -value не превышает $0,05$), что говорит о релевантности симуляционных экспериментов между собой. С полным перечнем полученных результатов с использованием нормального ядра можно ознакомиться по [ССЫЛКЕ](#).

При валидации нормализованных данных плотность вероятности чрезмерно высокая в нуле, из-за чего мы не видим точные значения плотности в других значениях выборок. Данные, близкие к нулю, означают так называемый “белый шум” – в них не наблюдается явно видимой нейрональной активности, поэтому они были исключены из исходного датасета методом фильтрации по пороговому значению, так как по графику биоданных мы можем явно определить интервал “белого шума” и его границы (пороговые значения). Для нормализованных био- и симуляционных данных таким интервалом является отрезок $[-0.1; 0.1]$.

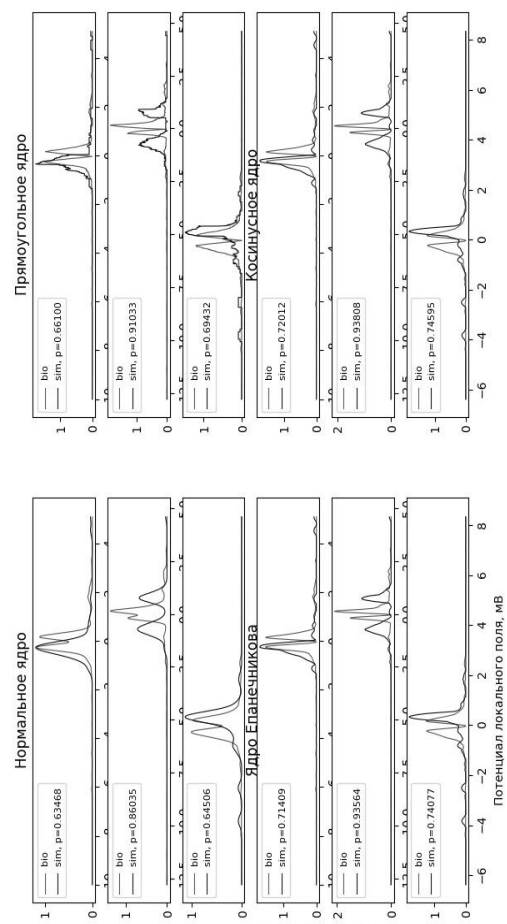
На рис. 4 показан пример сравнения 4-х видов ядер при валидации нормализованных фильтрованных данных с различными пороговыми значениями – $[-1; 1]$ (4a) и $[-0.1; 0.1]$ (4b). Было получено, что наиболее высокие значения p -value показывает косинусное ядро, которое и было применено для получения итоговых результатов. С полным перечнем всех полученных графиков при применении \cos -ядра к нормализованным фильтрованным данным можно ознакомиться по [ССЫЛКЕ](#).

Симуляция 1 / Эксперимент 3. Запись 0. Сенсоры 1, 8, 15. $h_{bio} = 0.07$, $h_{sim} = 0.15$



(a) Валидация нормализованных данных с фильтрацией $[-1; 1]$

Симуляция 1 / Эксперимент 3. Запись 0. Сенсоры 1, 8, 15. $h_{bio} = 0.10$, $h_{sim} = 0.20$



(b) Валидация нормализованных данных с фильтрацией $[-0.1; 0.1]$

Рис. 4(a)-(b). Сравнение применения разных видов ядер к валидации симуляционных и биологических данных кортикальной колонки

Итоговые результаты были также получены с помощью анализа 3-х биологических экспериментов - №3, 10, 23, в каждом из которых были взяты по 2 эксперимента с наиболее интересной и разноплановой нейрональной активностью.

Таблица 4. P-value валидации биологической крысы № 3 с биологической крысой №10 / крысой №23 по методу *kde test*

Сравнение с Био 10 / с Био 23	Эксперимент 3. Нормализованные данные	Эксперимент 3. Фильтрация $[-0.1; 0.1]$	Эксперимент 10. Фильтрация $[-1; 1]$
Сенсор#1	0.50009 / 0.50011	0.500102 / 0.50017	0.500203 / 0.50041
Сенсор#2	0.50012 / 0.50017	0.50011 / 0.50015	0.50021 / 0.50036
Сенсор#3	0.50013 / 0.50025	0.50014 / 0.50015	0.50016 / 0.50022
Сенсор#4	0.500107 / 0.500404	0.50013 / 0.50013	0.50014 / 0.500187
Сенсор#5	0.50008 / 0.50047	0.500104 / 0.50018	0.500175 / 0.500184
Сенсор#6	0.50004 / 0.500716	0.500101 / 0.50011	0.50017 / 0.50019
Сенсор#7	0.500078 / 0.500714	0.50009 / 0.50012	0.50019 / 0.50015
Сенсор#8	0.50006 / 0.50061	0.50011 / 0.50017	0.500157 / 0.50017
Сенсор#9	0.50013 / 0.50021	0.50024 / 0.50019	0.50023 / 0.50096

Сенсор#10	0.50015 / 0.50019	0.500109 / 0.50028	0.50038 / 0.50168
Сенсор#11	0.50014 / 0.50019	0.50027 / 0.50019	0.50025 / 0.50086
Сенсор#12	0.50017 / 0.50023	0.500097 / 0.50032	0.50036 / 0.50174
Сенсор#13	0.5001 / 0.50017	0.50013 / 0.50052	0.50058 / 0.51918
Сенсор#14	0.500121 / 0.50011	0.500124 / 0.50046	0.50045 / 0.51307
Сенсор#15	0.500085 / 0.50014	0.500122 / 0.50028	0.500403 / 0.50932
Усреднённое <i>p-value</i>	0.50011 / 0.50031	0.50013 / 0.50022	0.50027 / 0.50324

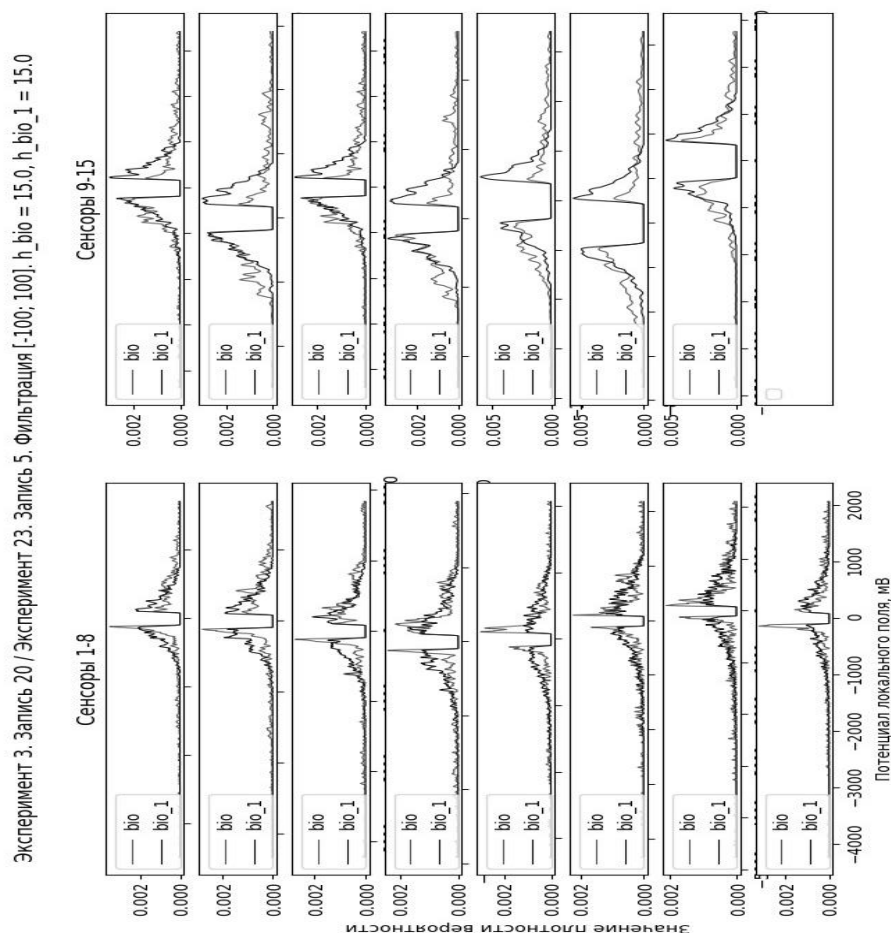


Рис. 5. Валидация биологических экспериментов №3 и 23 с фильтрацией [-100; 100]

В ходе валидации биоданных между собой были получены результаты сравнения взятых крыс по принципу “каждая с каждой”. Таблица 4 и рис. 5 показывают пример сравнения 3-й крысы с 10-й и 23-й крысой. Была сформирована сводная таблица с *p-значениями* по сенсорам и рассчитан усреднённый показатель *p-value* по каждому сравнению, показывающий соответствие биоданных между собой.

Таблица 5. *P-value* валидации биологической крысы №10 с симуляцией 1 / симуляцией 2 по методу *kde test*

Сравнение Симуляцией 1 / Симуляцией 2	с 1/	Эксперимент 10. Нормализованные данные	Эксперимент 10. Фильтрация [-0.1; 0.1]	Эксперимент 10. Фильтрация [-1; 1]
Сенсор#1		0.76068 / 0.83274	0.70436 / 0.75866	0.83122 / 0.77029
Сенсор#2		0.91021 / 0.89857	0.81272 / 0.86752	0.88366 / 0.84233
Сенсор#3		0.90718 / 0.89397	0.75597 / 0.82573	0.87954 / 0.82093
Сенсор#4		0.70732 / 0.86141	0.72718 / 0.69782	0.82946 / 0.81548
Сенсор#5		0.94999 / 0.74985	0.74654 / 0.71169	0.72676 / 0.86115
Сенсор#6		0.96109 / 0.88416	0.76348 / 0.70588	0.72256 / 0.9467
Сенсор#7		0.97383 / 0.98019	0.81203 / 0.82895	0.67778 / 0.70023
Сенсор#8		0.96715 / 0.97619	0.79679 / 0.84158	0.78911 / 0.75171
Сенсор#9		0.91617 / 0.92442	0.78512 / 0.81051	0.65376 / 0.71649

Сенсор#10	0.88119 / 0.88659	0.77323 / 0.75017	0.68171 / 0.66201
Сенсор#11	0.84913 / 0.86327	0.75648 / 0.6753	0.69105 / 0.65216
Сенсор#12	0.84944 / 0.81166	0.72317 / 0.71701	0.65479 / 0.71997
Сенсор#13	0.79425 / 0.77646	0.66593 / 0.72613	0.70696 / 0.64581
Сенсор#14	0.79644 / 0.78714	0.69875 / 0.74337	0.7348 / 0.70747
Сенсор#15	0.81531 / 0.79894	0.70332 / 0.73307	0.8409 / 0.72953
Обобщённые данные	0.7953 / 0.79137	0.69765 / 0.69227	0.58572 / 0.59039

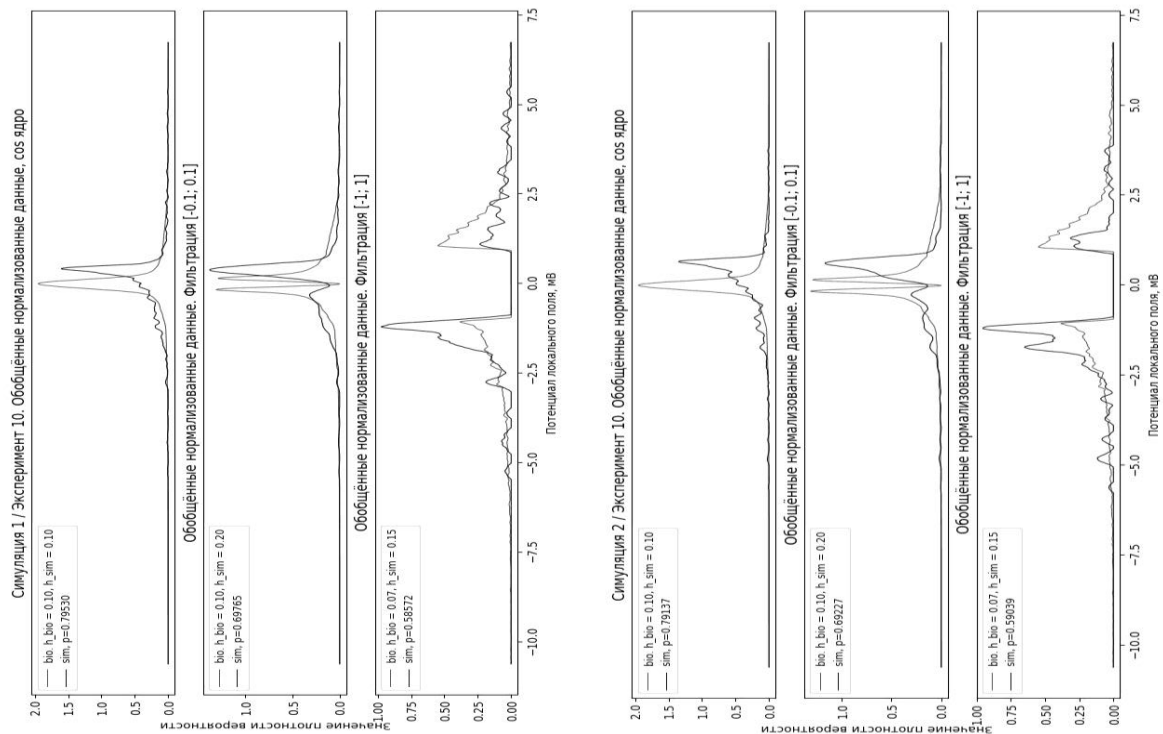


Рис. 6(a)-(b). Результаты валидации 10-й биологической крысы с симуляцией 1 (a) и симуляцией 2 (b)

Первая симуляционная модель, как и вторая, более всего соответствует 10-й биологической крысе, результаты сравнения с которой приведены в Таблице 5 и на рис. 6(a)-(b) выше. Из данных результатов следует, что значения p -value существенно превышают 0,05, необходимые для принятия H_0 о совпадении плотностей вероятности выборок, что говорит о биореалистичности симуляционной модели. С полным перечнем результатов валидации био- и симуляционных данных по методу kde test можно ознакомиться по [ССЫЛКЕ](#).

Заключение

Поставленная задача полностью выполнена в ходе исследования. Валидация данных проводилась с использованием критериев Колмогорова-Смирнова и ядерной оценки плотности вероятности с применением различных видов ядер. Была протестирована эквивалентность как внутри биологической группы данных, так и на предмет соответствия симуляционным.

В ходе работы была доказана неприменимость критерия однородности Смирнова к валидации симуляционных и биологических данных кортикальной колонки мыши в связи с получением низких значений p -value. На основании сравнительного анализа был выявлен наиболее подходящий метод валидации разработанной модели, которым является kde test при применении косинусного ядра.

На основании анализа с помощью kde test и косинусного ядра был сделан вывод о том, что все биологические крысы соответствуют друг другу примерно в равной степени (различия между усреднёнными p -value < 0,01), при этом никакие две биологические крысы не показывают в точности одинаковые результаты (поскольку значения p -value не достигают 1).

Оба симуляционных эксперимента показали наибольшую релевантность с 10-й биологической крысой; в то же время каждая биологическая крыса соответствует обоим симуляционным результатам примерно в равной степени (различия между p -value ~ 0,01 и менее). Таким образом, в ходе исследования была доказана биореалистичность разработанной симуляционной модели, о чём говорят высокие значения p -value (>0,5), полученные при валидации симуляционных данных с биологическими по методу kde test.

Список литературы / References

1. Epilepsy. Key facts // World health organization [Электронный ресурс], 2019. Режим доступа: <https://www.who.int/news-room/fact-sheets/detail/epilepsy/> (дата обращения: 25.03.2021).
2. Эпилепсия в России и мире // Бюджет.RU [Электронный ресурс], 2019. Режим доступа: <https://bujet.ru/article/378591.php/> (дата обращения: 25.03.2021).
3. *Represa Alfonso*. Why Malformations of Cortical Development Cause Epilepsy // *Frontiers in neuroscience*. 2019. № 3. С. 54-64.
4. *Leventer Richard J., Guerrini Renzo, Dobyns William B.* Malformations of cortical development and epilepsy // *Dialogues in Clinical Neuroscience*, 2008. № 10 (1). С. 47-62.
5. *Peterson Andre D.H., Burkitt Anthony N. and Mareels Iven M.Y.* The effect of network structure on epileptic dynamics: analysis of the synchronisation properties of an inter-network of cortical columns // *BMC Neurosci*, 2011. № 12 (1). С. 46.
6. *Gillard Jonathan*. A first course in statistical inference. / под ред. М. А. J. Chaplain, Angus Macintyre, Simon Scott. Berlin etc.: Springer, 2020. 174 с.
7. *Wasserman Larry*. All of statistics. A concise course in statistical inference. / под ред. George Casella, Stephen Fienberg, Ingram Olkin. Berlin etc.: Springer, 2004. 462 с.
8. *Симушкин С.В.* Теоретические основы выполнения курсовой работы по математической статистике / под ред. С.В. Симушкин, А.А. Заикин, И.А. Кареев, Р.Ф. Салимов. Казань: Казан. ун-т, 2020. 106 с.
9. *Симушкин С.В.* Теоретические аспекты заданий курсового проекта по математической статистике. Казань: Изд-во КГУ, 2004. 65 с.
10. *Капитанова Е.К.* Математическая статистика. Казань: Системы дистанционного обучения Казанского федерального университета, 2017. 97 с.
11. *Антонов А.В., Зюляева Н.Г., Чепурко В.А.* Исследование метода ядерной оценки плотности распределения [Электронный ресурс], 2007. Режим доступа: https://www.researchgate.net/publication/279852423_Issledovanie_metoda_adernoj_ocenki_plotnosti_raspredelenia/ (дата обращения: 18.04.2021).
12. *Węglarczyk Stanisław*. Kernel density estimation and its application // *ITM Web of Conferences*, 2018. № 23. С. 36-44.
13. *Tarn Duong, Schauer Kristine, Goud Bruno*. Closed-form density-based framework for automatic detection of cellular morphology changes // *Proceedings of the National Academy of Sciences*, 2012. № 109 (22). С. 24-30.
14. *Pavia Jose M.* Testing Goodness-of-Fit with the Kernel Density Estimator: GoFKernel // *Journal of Statistical Software*, 2015. № 66. С. 1-27.