УДК 519.246.8
doi:10.31799/1684-8853-2025-3-2-13
EDN: RAJEBP

Научные статьи
Articles

Метод минимизации продолжительности наблюдений и математические модели экстраполяции при малых объемах данных

А. Г. Лосев[a], *доктор физ.-мат. наук, профессор, orcid.org/0000-0002-1072-8375, alexander.losev@volsu.ru*
И. А. Медведев[a], *аспирант, orcid.org/0009-0001-2536-4359*
[a]*Волгоградский государственный университет, Университетский пр., 100, Волгоград, 400062, РФ*

Введение: *экстраполяция значений временных рядов с малым объемом данных — задача актуальная, особенно в областях, где требуется быстрое принятие решений на основе ограниченной информации. Одной из ключевых проблем является нахождение наименьшего отрезка исторических данных, достаточного для построения точного прогноза последующих значений временного ряда.* ***Цель:*** *разработать метод минимизации временных наблюдений, позволяющий осуществлять экстраполяцию будущих значений с заранее заданной точностью, и метод синтезации данных на основе параметров математической модели прогнозирования.* ***Результаты:*** *во-первых, разработан метод определения минимальных отрезков эффективного прогнозирования. Метод предполагает наличие обучающей выборки и основан на анализе динамики изменения значений временных рядов вокруг их максимума, что позволило автоматически определять отрезки данных минимальной длины, необходимые для построения прогнозов с заранее заданной точностью. Во-вторых, разработан метод формирования синтетических данных, использующий варьирование параметров математической модели. Синтез данных позволил повысить эффективность обучения искусственных нейронных сетей. Указанные методы были применены в анализе временных рядов напряжения микробных топливных элементов для оценки параметра загрязнения сточных вод — биохимического потребления кислорода. В результате удалось сократить время прогнозирования данного параметра с пяти суток до 12,68 часа с относительной погрешностью 8,1 %.* ***Практическая значимость:*** *предложенные методы, используемые в ходе экологического мониторинга, позволяют оперативно оценивать загрязнение воды и оптимизировать процессы очистки сточных вод.*

Ключевые слова *— математические модели экстраполяции, синтезация данных, временные ряды, искусственная нейронная сеть, логнормальное распределение, микробный топливный элемент, биохимическое потребление кислорода.*

Для цитирования: Лосев А. Г., Медведев И. А. Метод минимизации продолжительности наблюдений и математические модели экстраполяции при малых объемах данных. *Информационно-управляющие системы*, 2025, № 3, с. 2–13. doi:10.31799/1684-8853-2025-3-2-13, EDN: RAJEBP

For citation: Losev A. G., Medvedev I. A. Method of minimizing the duration of observations and mathematical extrapolation models at small data volumes. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 3, pp. 2–13 (In Russian). doi:10.31799/1684-8853-2025-3-2-13, EDN: RAJEBP

Введение

Разработка моделей экстраполяции имеет большую историю и не теряет своей актуальности до настоящего времени. В ряде случаев их практическое применение имеет ключевое значение для принятия обоснованных решений. В большинстве исследований, посвященных анализу временных рядов, рассматриваются вопросы выбора оптимальной модели прогнозирования, т. е. получения прогноза последующих значений с удовлетворительной погрешностью. Одной из важнейших задач при прогнозировании временных данных является определение начального участка последовательности, который будет использоваться для построения прогноза. Наиболее интересно определение такого отрезка данных, который не только позволяет адекватно отражать динамику исследуемого процесса, но и одновременно обеспечивает минимизацию продолжительности наблюдений, т. е. имеет минимальную длину относительно начальной точки. Далее будем называть его минимальным отрезком эффективного прогнозирования.

Различные методы определения минимальных отрезков эффективного прогнозирования рассматривались, например, в задачах сокращения продолжительности экспериментов по кинетике нагревания [1], оценки режимных значений временных рядов скорости ветра [2], анализа геофизических данных [3] и т. д. Эти методы, включая перебор отрезков фиксированной длины и усечение начальных данных, имеют существенные ограничения. Более эффективным решением является выявление характерных точек временных рядов, которые отражают ключевые изменения в данных и повышают точность экстраполяции. Характерные точки могут быть выявлены непосредственно из данных. Например, в работе [4] предложен статистический алгоритм на основе относительной дивергенции Пирсона для обнаружения точек изменений. Существуют

и другие способы определения характерных точек [5–7]. Автоматизация этого процесса возможна с использованием машинного обучения [8, 9]. Также предлагались методы сегментирования данных и выбора оптимальных моделей для локальных участков рядов [10]. Применение характерных точек и информативных сегментов для экстраполяции продемонстрировано в исследованиях [9, 11, 12]. Отметим, что анализ сегментов рядов важен не только для экстраполяции, но и в задачах классификации временных рядов и определения синхронизации их участков [13, 14].

Таким образом, проблема определения минимальных отрезков эффективного прогнозирования актуальна и применима в различных областях, например в экологии. Низкое качество воды — глобальная проблема современного человечества. Для оценки загрязненности воды широко используется показатель биохимического потребления кислорода (БПК; Biochemical Oxygen Demand, BOD), который количественно определяет уровень биоразлагаемых органических загрязнителей. Однако стандартный тест $БПК_5$ требует не менее пяти дней измерений при температуре 20 °C [15], что ограничивает его применение для мониторинга в реальном времени.

Для повышения эффективности определения $БПК_5$ разрабатывались различные биосенсоры, включая микробные топливные элементы (МТЭ), — устройства, преобразующие энергию органических веществ в электрический ток благодаря экзоэлектрогенным бактериям. Временные ряды напряжения МТЭ отражают динамику электрической активности системы как ответ на изменения в составе сточных вод. Общий заряд, рассчитанный по рядам напряжения, находится в линейной зависимости от $БПК_5$, так как количество электронов, переданных на анод, пропорционально концентрации органических веществ. Метод раннего определения $БПК_5$, основанный на линейной зависимости между общим зарядом и концентрацией $БПК_5$, может сократить время анализа до 0,5–4 сут с относительной погрешностью в пределах 10 % [16]. Искусственные нейронные сети (ИНС) могут применяться для прогноза напряжения МТЭ на основе значений напряжения, измеренных за 8–24 ч [17]. Это позволяет, если опираться на предсказанную динамику электрической активности системы, косвенно оценивать $БПК_5$ через линейную зависимость с рассчитанным общим зарядом. Однако применение ИНС предполагало использование фиксированного размера входных данных для всех рядов напряжения и разработку нескольких моделей ИНС для разного входного времени. При этом оптимальное время измерения может варьироваться для разных рядов напряжения. Это связано с тем, что динамика изменения напряжения может отличаться в зависимости от содержимого образцов воды.

Существующие математические модели, такие как модель Моно, не подходят для описания исследуемых функций напряжения [18, 19]. Численное моделирование требует создания новых моделей с учетом множества параметров. В работе предложен способ, близкий к суррогатному моделированию, использующий аппроксимацию и экстраполяцию функций напряжения семейством функций с малым числом параметров. На основе параметров этого семейства был предложен метод формирования синтетических данных, создающих обучающий набор, что позволило применять ИНС для улучшения прогнозирования и учета сложных зависимостей.

Условия на класс функций.

Для проблематики, связанной с определением минимальных отрезков эффективного прогнозирования, необходимо ввести некоторые обозначения и условия. Предположим, что мы имеем дело с набором временных рядов, которые можно интерпретировать как непрерывные функции, зависящие от параметров, описывающих динамику исследуемого процесса.

Обозначим через $V([a, b], A)$ некоторое подмножество неотрицательных непрерывных на $[a, b]$ функций, зависящих от параметра $\alpha \in A$, вообще говоря, бесконечномерного. Будем предполагать, что для любой функции $v(x, \alpha) \in V([a, b], A)$ отрезок $[a, b]$ можно разбить на два участка монотонности $[a, x_{\max}]$ и $[x_{\max}, b]$, т. е. существует единственный глобальный максимум $v_{\max}$, достигаемый в точке $x_{\max}$. Под $V([a_c, b_c], A)$ будем понимать множество сужений функций из $V([a, b], A)$ на отрезок $[a_c, b_c]$.

Пусть далее $U([a, b], S)$ — множество функций, заданных на $[a, b]$, зависящих от конечномерного вектора параметров $\mathbf{s}(\alpha, c)$ и обеспечивающих аппроксимацию функций из $V([a_c, b_c], A)$ с заданной точностью. Поясним последний тезис. Предположим, что нам задан параметр точности ε. Если для любой функции $v(x, \alpha) \in V([a_c, b_c], A)$ найдется функция $u(x, \mathbf{s}(\alpha, c)) \in U([a, b], S)$ такая, что $\rho_{[a_c,b_c]}(v, u) \le \varepsilon$, то будем говорить, что $U([a, b], S)$ обеспечивает аппроксимацию функций из $V([a_c, b_c], A)$ с точностью ε. Здесь $\rho_{[a_c,b_c]}(v, u)$ — некоторая мера близости функций v и u на отрезке $[a_c, b_c]$. В частности, это может быть метрика какого-нибудь функционального пространства, например:

$$\rho_{[a_c,b_c]}(v, u) = \left\| v(x, \alpha) - u(x, \mathbf{s}(\alpha, c)) \right\|_{L^2_{[a_c,b_c]}}. \quad (1)$$

Введенные обозначения и условия на класс функций позволяют формализовать задачу определения минимальных отрезков эффективного прогнозирования.

Метод определения минимальных отрезков эффективного прогнозирования

Первой задачей, решаемой в рамках данного исследования, является разработка метода нахождения отрезка $[a_c, b_c] \subset [a, b]$, минимизирующего величину $b_c - a$ и такого, что выполнено условие $p_{[a,b]}(v, u) \le \varepsilon_1$. Здесь $p_{[a,b]}(v, u)$ — некоторая мера близости функций v и u на отрезке $[a, b]$, а ε_1 — точность экстраполяции. Например, в качестве $p_{[a,b]}(v, u)$ можно использовать величину

$$p_{[a,b]}(v,\ u) = \frac{1}{b-a}\left\| 1 - \frac{u(x,\ \mathbf{s}(\alpha,\ c))}{v(x,\ \alpha)} \right\|_{L_{1[a,b]}}. \tag{2}$$

Введем параметр c, принимающий значения в интервале (0, 1). Каждому такому c поставим в соответствие точку $x_c \in (x_{\max}, b]$, в которой $v(x_c, \alpha) = cv_{\max}$. Если такой точки не существует, т. е. если для любого $x \in (x_{\max}, b]$ выполняется $v(x, \alpha) > cv_{\max}$, то положим x_c равной b. При таком определении x_c для любого $x \in (x_{\max}, x_c]$ будет выполнено $v(x, \alpha) \ge cv_{\max}$, что следует из монотонного убывания функций на отрезке $[x_{\max}, b]$.

Зафиксируем некоторую $v(x, \alpha) \in V([a, b], A)$ и рассмотрим множество отрезков $[a, x_c] \subset [a, b]$. Каждому отрезку $[a, x_c]$ можно поставить в соответствие функцию $u(x, \mathbf{s}(\alpha, c)) \in U([a, b], S)$, которая аппроксимирует $v(x, \alpha)$ с заданной точностью, например, в метрике (1). Векторы параметров $\mathbf{s}(\alpha, c)$, полученные при аппроксимации $v(x, \alpha)$ на $[a, x_c]$, позволяют восстановить значения $u(x, \mathbf{s}(\alpha, c))$ на отрезок $[a, b]$. Таким образом, на $[a, b]$ может быть получено множество $p_{[a,b]}(v, u, c)$ значений мер близости функций v и u [например, по формуле (2)], где c указывает, что параметры аппроксимации подбирались на $[a, x_c]$. На самом деле, в дальнейшем нами будут рассматриваться не произвольные пары (v, u), а произвольная v и аппроксимирующая u. Стало быть, на основе значений параметра c может быть получено множество отрезков для определения параметров аппроксимирующих функций и множество значений мер близости функций на $[a, b]$. Рассмотрим какую-нибудь агрегатную функцию, например: $\overline{p}_{[a,b]}(c) = \sup_{v,u}(p_{[a,b]}(v, u, c))$. Максимальное c, при котором выполнено $\overline{p}_{[a,b]}(c) \le \varepsilon_1$, будем обозначать c_r^* и называть оптимальным параметром формирования правых границ отрезков эффективного прогнозирования. А оптимальные правые границы отрезков каждой функции множества $V([a, b], A)$ обозначим как $x_{c_r^*}$. Аналогичным образом могут определяться $x_{c_l^*}$ — оптимальные левые границы отрезков эффективного прогнозирования и параметр формирования левых границ этих отрезков c_l^*.

В приложениях данные обычно рассматриваются в дискретной форме. Пусть на сетке: $\mathbf{x} = x_j$, где $x_j = a + (j - 1)h, j = 1, 2, \ldots, n$, с равномерным шагом задано множество из m функций $v_i(\mathbf{x})$, где $i = 1, 2, \ldots, m$. Значения функций можно представить в виде матрицы $\mathbf{V}$ с элементами $v_i(x_j)$. Назовем данный набор из m функций обучающим. Поскольку каждая функция $v_i(\mathbf{x})$ имеет единственный максимум, образуем $\mathbf{C}$, элементы которой имеют вид, представленный выражением

$$c_{ij} = \frac{v_i(x_j)}{\max\limits_{j=1,2,\ldots,n}(v_i(x_j))},\ i = 1,\ 2,\ \ldots,\ m;\ j = 1,\ 2,\ \ldots,\ n. \tag{3}$$

Для каждой i-й строки в матрице $\mathbf{C}$ определим номер столбца, в котором достигается максимальное значение. А именно, пусть $\mathbf{J}$ — вектор номеров столбцов, элементы которого зависят от i, и для любого i выполнено $c_{iJ(i)} = 1$.

Для любой i-й функции матрицы $\mathbf{V}$ можно рассмотреть множества значений $v_i(x_1)$, $v_i(x_2)$, …, $v_i(x_{J(i)+k})$, где $k = 1, 2, \ldots, n - J(i)$. Отметим, что для каждого i известна функция $u_i(x_1, x_2, \ldots, x_{J(i)+k}, \mathbf{s}_{ik})$, которая приближает $v_i(x_1, x_2, \ldots, x_{J(i)+k})$ с заданной точностью. Здесь $\mathbf{s}_{ik}$ — это конечномерный вектор параметров аппроксимации i-й функции, $k = 1, 2, \ldots, n - J(i)$. Вектор параметров $\mathbf{s}_{ik}$ позволяет восстановить значения $u_i(x_1, x_2, \ldots, x_{J(i)+k}, \mathbf{s}_{ik})$ на сетку $x_1, x_2, \ldots, x_n$. Таким образом, для каждого i можно сформировать матрицу значений функции приближения

$$\mathbf{U}_i = \begin{pmatrix} u_i(x_1,\ \mathbf{s}_{i1}) & u_i(x_2,\ \mathbf{s}_{i1}) & \ldots & u_i(x_n,\ \mathbf{s}_{i1}) \\ u_i(x_1,\ \mathbf{s}_{i2}) & u_i(x_2,\ \mathbf{s}_{i2}) & \ldots & u_i(x_n,\ \mathbf{s}_{i2}) \\ \ldots & \ldots & \ldots & \ldots. \\ u_i(x_1,\ \mathbf{s}_{i(n-J(i))}) & u_i(x_2,\ \mathbf{s}_{i(n-J(i))}) & \ldots & u_i(x_n,\ \mathbf{s}_{i(n-J(i))}) \end{pmatrix}. \tag{4}$$

Для любого i получаем вектор мер близости функций v_i и u_i на $[a, b]$

$$\mathbf{p}_i = \begin{pmatrix} p_{i[a,b]}(v_i(\mathbf{x}),\ u_i(\mathbf{x},\ \mathbf{s}_{i1})) \\ p_{i[a,b]}(v_i(\mathbf{x}),\ u_i(\mathbf{x},\ \mathbf{s}_{i2})) \\ \dots \\ p_{i[a,b]}(v_i(\mathbf{x}),\ u_i(\mathbf{x},\ \mathbf{s}_{i(n-J(i))})) \end{pmatrix}. \quad (5)$$

Формулу мер близости функций можно представить записью

$$p_{i[a,b]}(v_i(\mathbf{x}),\ u_i(\mathbf{x},\ \mathbf{s}_{ik})) = \frac{1}{n}\sum_{j=1}^{n}\left|1 - \frac{\left|u_i(x_j,\ \mathbf{s}_{ik})\right|}{\left|v_i(x_j)\right|}\right|. \quad (6)$$

Обозначим через $J' = \max\limits_{i=1,2,\dots,m}(n - J(i))$ максимальную размерность векторов $\mathbf{p}_i$. Далее рассмотрим агрегацию m векторов мер близости (5) в один вектор, например:

$$\overline{p}_j = \frac{1}{m}\sum_{i=1}^{m} p_{i,j}\ (p_{i,j} = 0,\ \text{если}\ j > J'),$$
$$\text{где}\ j = 1,\ 2,\ \dots,\ J', \quad (7)$$

или

$$\overline{p}_j = \max_{i=1,2,\dots,m} p_{i,j}\ (p_{i,j} = 0,\ \text{если}\ j > J'),$$
$$\text{где}\ j = 1,\ 2,\ \dots,\ J'. \quad (8)$$

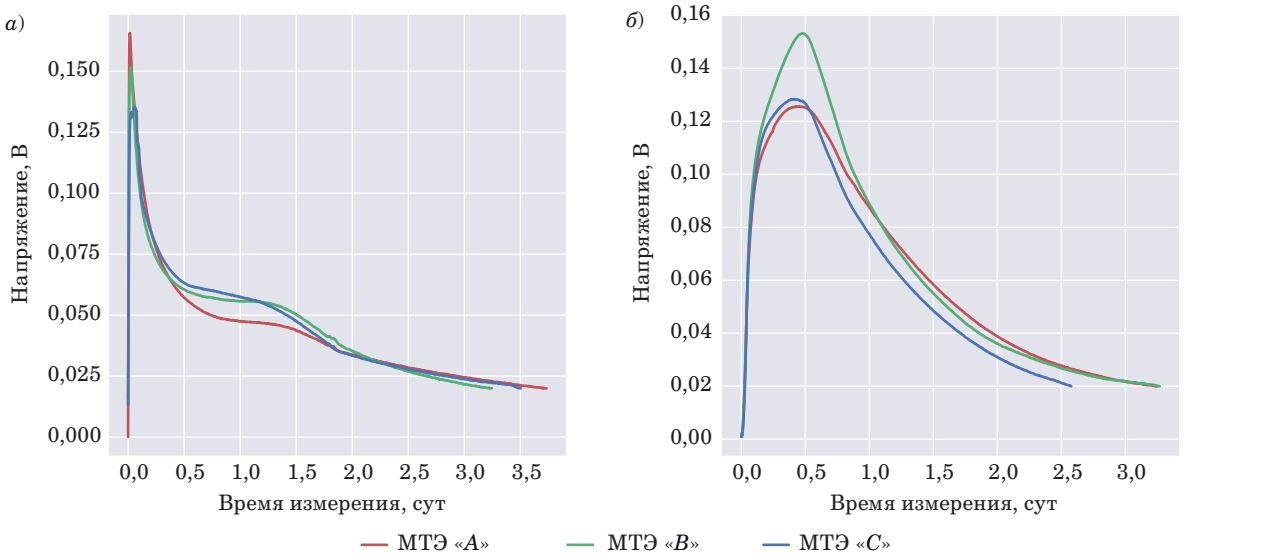
Обозначим через j' минимальный номер вектора $\overline{\mathbf{p}}$, при котором выполнено $\overline{p}_{j'} \le \varepsilon_1$. Здесь j' обозначает минимальное количество точек, на которое нужно отступить вправо от максимального значения функций, чтобы получить приемлемую погрешность экстраполяции. Зная j', на основе элементов матрицы $\mathbf{C}$ сформируем вектор

$$\overline{\mathbf{c}} = \begin{pmatrix} c_{1(J(1)+j')} \\ c_{2(J(2)+j')} \\ \dots \\ c_{m(J(m)+j')} \end{pmatrix}. \quad (9)$$

Усреднив элементы вектора $\overline{\mathbf{c}}$, получим оптимальный параметр формирования правых границ отрезков эффективного прогнозирования c_r^*. Аналогичным образом определяется c_l^* — оптимальный параметр формирования левых границ отрезков эффективного прогнозирования. Параметры c_l^*, c_r^* могут использоваться для формирования минимальных отрезков эффективного прогнозирования.

Отдельно стоит отметить, что реальные данные часто содержат шумы или ошибки измерений, что требует дополнительной обработки. Для решения этих вопросов применялись стандартные способы.

В исследовании был использован набор данных, описывающий результаты измерения напряжения МТЭ, из работы [16]. В качестве проб для МТЭ применялись два типа сточных вод с различным содержанием органических веществ: бытовые и от пивоваренных заводов. Три МТЭ, обозначенные «*A*», «*B*», «*C*», работали параллельно. Во время проведения измерений напряжение регистрировалось каждые 5 мин до тех пор, пока значения напряжения не достигали 0,02 В (рис. 1, *а* и *б*).



■ ***Рис. 1.*** Примеры изменения напряжения: *а* — бытовые сточные воды; *б* — сточные воды пивоваренных заводов

■ ***Fig. 1.*** Examples of voltage changes: *а* — domestic wastewater; *б* — wastewater from breweries

Набор данных был представлен 54 экспериментами измерения напряжения. Каждая функция напряжения имела разную длину. Самый долгий эксперимент длился примерно 4,5 сут, что соответствовало 1299 дискретным точкам напряжения. Для выравнивания размерности всех рядов напряжения недостающие значения были заполнены значением 0,02. Таким образом, набор данных с функциями напряжения был представлен матрицей **V** размерностью 54×1299 с элементами $v_i(x_j) \geq 0$, где $i = 1, 2, \ldots, 54; j = 1, 2, \ldots, 1299$. Бытовым сточным водам соответствовало 24 функции напряжения, а сточным водам пивоваренной фабрики — 30 функций.

Модель аппроксимации и прогнозирования

Биосенсоры на основе МТЭ перспективны для мониторинга качества воды, но компьютерное моделирование их данных затруднено из-за множества взаимосвязанных параметров. Можно применять ИНС для прогнозирования напряжения, плотности мощности и химических показателей [17, 20], однако такие модели не обладают достаточной интерпретируемостью. В данном исследовании предложен способ, близкий к суррогатному моделированию, где напряжение МТЭ аппроксимируется функцией плотности вероятности логнормального распределения с некоторыми дополнительными слагаемыми. Это упрощает моделирование и сохраняет интерпретируемость результатов.

Исследования показывают, что для некоторых проб воды функция напряжения МТЭ имеет симметричные логнормальные профили с выраженным пиком [21]. Логнормальное распределение широко используется в экологии, например для моделирования содержания загрязняющих веществ, металлов в биотканях, а также для описания геохимических процессов и динамики микроорганизмов в воде [22–25].

Функции напряжения МТЭ отвечают некоторым условиям для применения логнормального распределения. Также для лучшего соответствия экспериментальным данным были рассмотрены варианты добавления нелинейных функций $f(x, \mathbf{s})$ с конечным набором параметров **s**:

$$u(x, \sigma, \mu, f) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} + f(x, \mathbf{s}). \quad (10)$$

В исследовании проверялись различные варианты добавочных функций. Лучшие результаты экстраполяции напряжения были достигнуты при использовании

$$f(x, \lambda, \delta) = -\frac{\lambda}{\delta + x^2}. \quad (11)$$

Для подбора параметров функции (10) использовалась библиотека SciPy (Python), модуль optimize.curve_fit, предлагающий три метода нелинейной оптимизации: метод доверительной области отражения (Trust Region Reflective) [26], алгоритм изгиба с прямоугольными областями доверия и алгоритм Левенберга — Марквардта. Все три метода показали схожую точность аппроксимации, но метод Trust Region Reflective оказался предпочтительным для экстраполяции благодаря возможности задания ограничений на параметры функции. Этот метод является эффективным инструментом для решения задач нелинейной минимизации с ограничениями [27, 28].

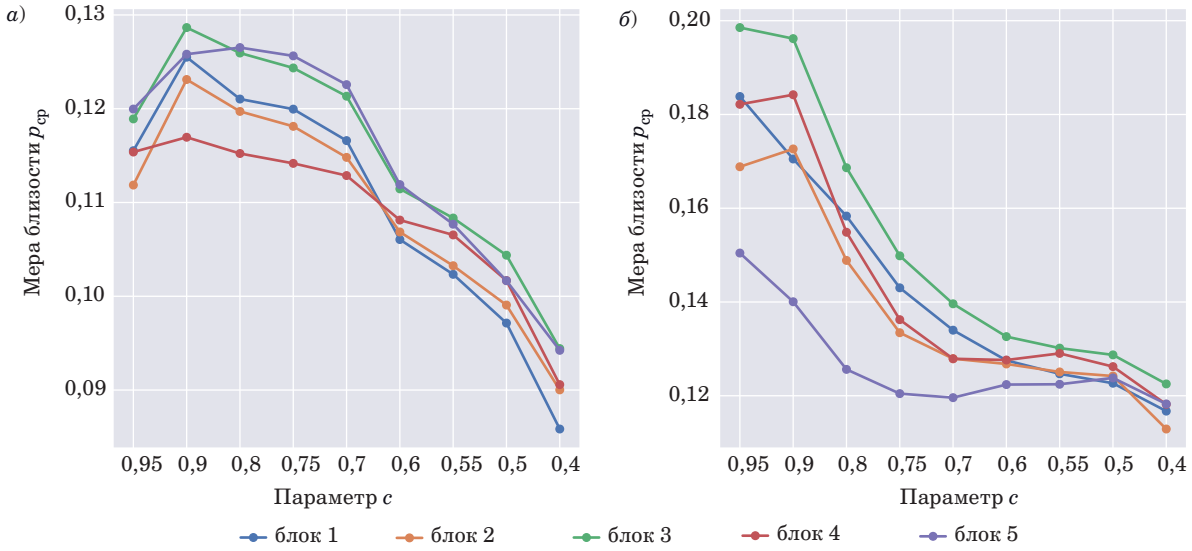
Применение метода определения минимальных отрезков эффективного прогнозирования

Предложенный метод был применен к 54 временным рядам напряжения МТЭ, включающим 24 ряда для бытовых сточных вод (первый подкласс) и 30 рядов для сточных вод пивоваренной фабрики (второй подкласс). Для каждого подкласса определялись свои параметры формирования границ отрезков эффективного прогнозирования. Использовалась кросс-валидация по пяти блокам отдельно для функций каждого подкласса. Вектор $\mathbf{x} = (a = 0{,}003472,\ 0{,}006944, \ldots, 4{,}510417 = b)$ задавал временную сетку измерений с шагом $h = 0{,}003472$ сут, что обеспечило удобное масштабирование данных. До подбора параметров функции (10) ряды напряжения были нормализованы по следующей формуле:

$$v_i(x_j)_{norm} = \frac{v_i(x_j) - 0{,}02}{\max\limits_{j=1,2,\ldots,1299}(v_i(x_j)) - 0{,}02},$$
$$j = 1,\ 2,\ \ldots,\ 1299;\ i = 1,\ \ldots,\ 54. \quad (12)$$

В данном случае 0,02 В — значение окончания большинства экспериментов. Применение нормализации по формуле (12) позволило устранить чувствительность метода к начальным величинам параметров и улучшить сходимость метода.

Начальные параметры для функции прогнозирования определялись в результате вычислительных экспериментов. Для первого подкласса были установлены параметры: $\sigma = 1{,}5$; $\mu = 0{,}5$; $\lambda = -0{,}1$; $\delta = 1$; для второго подкласса: $\sigma = 1{,}5$; $\mu = 0{,}5$; $\lambda = 0{,}1$; $\delta = 0{,}5$. На параметры были введены дополнительные ограничения, позволяющие более точно определить область поиска



■ ***Рис. 2.*** Средние значения меры близости функций в зависимости от изменения параметра c при формировании правых границ отрезков: *а* — первый подкласс; *б* — второй подкласс

■ ***Fig. 2.*** The average values of the function proximity measure depending on the change in parameter c when forming the boundaries of segments: *а* — first subclass; *б* — second subclass

оптимального решения: для первого подкласса: $0 \le \sigma \le 4$; $-0{,}3 \le \mu \le 3{,}5$; $-1 \le \lambda \le 0{,}3$; $0{,}8 \le \delta \le 1{,}2$; для второго подкласса: $0 \le \sigma \le 1{,}9$; $-1{,}5 \le \mu \le 0{,}55$; $-1{,}5 \le \lambda \le 0{,}15$; $0{,}05 \le \delta \le 1{,}2$.

Для формирования матрицы **C** был использован фиксированный набор значений параметра c, определяющего отношения к максимумам функций: (0,95, 0,9, 0,8, 0,75, 0,7, 0,6, 0,55, 0,5, 0,4). Затем были заданы пороговые значения ε_1 для мер близости, агрегированных по формулам (7) и (8). На рис. 2, *а* и *б* показаны изменения значений мер близости p, агрегированных по формуле (7), по каждому блоку при изменении значений параметра c.

В результате были определены следующие значения параметров формирования границ отрезков: $c_l^* = 0{,}95$, $c_r^* = 0{,}55$ для первого подкласса; $c_l^* = 0{,}9$, $c_r^* = 0{,}7$ для второго подкласса. На основе данных параметров были определены минимальные отрезки эффективного прогнозирования.

Анализ результатов экстраполяции и оценка БПК$_5$

Полученные правые границы отрезков эффективного прогнозирования в среднем соответствовали 13,05 ч для функций первого подкласса и 12,31 ч для функций второго подкласса. На полученных отрезках для каждой функции напряжения был определен набор параметров функции (10). Время завершения прогноза определено экспериментально, так как оно зависит от особенностей предметной области. Среднее значение используемой меры близости для первого подкласса составило 0,1038 и 0,1272 для второго подкласса. Примеры полученных результатов экстраполяции для некоторых функций приведены на рис. 3, *а* и *б*.

По прогнозным значениям напряжения были рассчитаны значения БПК$_5$ по формулам, полученным в работе [16], т. е. для каждого $i = 1, \ldots, 54$ было получено $\overline{b_i}$ — рассчитанное значение БПК$_5$ в i-м эксперименте. После расчета предсказанные значения БПК$_5$ сравнивались с реальными значениями, измеренными стандартным пятидневным способом. В качестве метрик были использованы средняя абсолютная

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|b_i - \overline{b_i}\right|, \qquad (13)$$

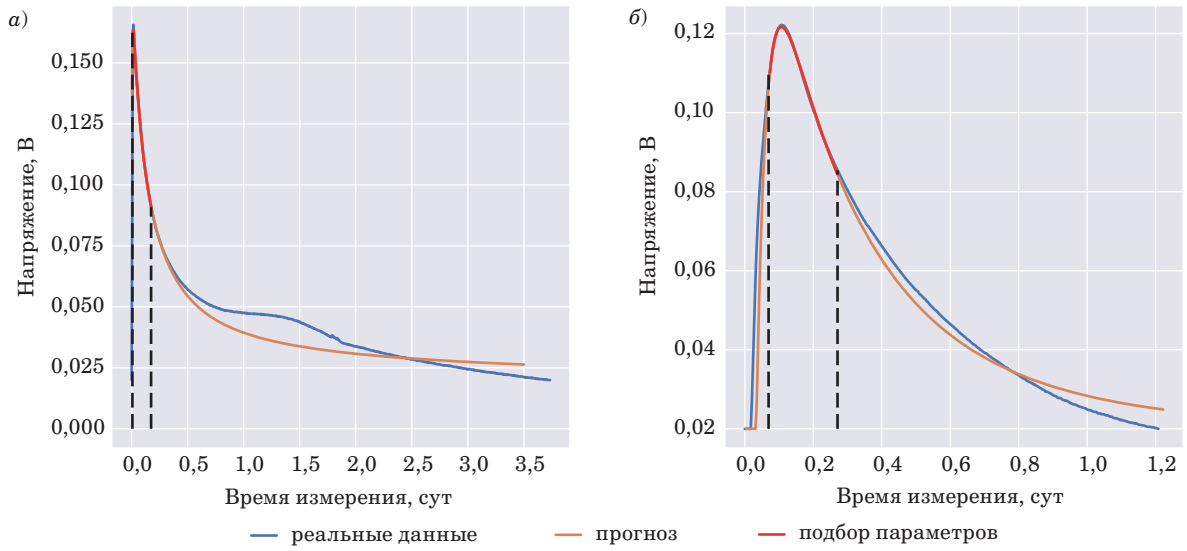
максимальная абсолютная

$$MAX = \max_{i=1,\ldots,m}\left|b_i - \overline{b_i}\right| \qquad (14)$$

и средняя относительная ошибки

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\frac{\left|b_i - \overline{b_i}\right|}{\left|b_i\right|}\cdot 100\%. \qquad (15)$$

Здесь $\mathbf{b} = (b_1, b_2, \ldots, b_{54})$ — это значения БПК$_5$, измеренные стандартным пятидневным способом и которые также были представлены в наборе данных. Полученные по формулам (13)–(15) оценки приведены в таблице.



■ ***Рис. 3.*** Примеры результатов аппроксимации и экстраполяции рядов напряжения: *а* — первый подкласс; *б* — второй подкласс

■ ***Fig. 3.*** Examples of the results of approximation and extrapolation of voltage series: *а* — first subclass; *б* — second subclass

■ Оценка прогнозирования $БПК_5$

■ Estimation of BOD_5 prediction

Подкласс	Среднее время, ч	MAE, мг	MAX, мг	MAPE, %
Подкласс 1	13,05	1,17	2,96	8,72
Подкласс 2	12,31	1,13	4,13	7,47
Среднее	12,68	1,15	3,54	8,1

Таким образом, комбинация данных МТЭ с представленной моделью прогнозирования и методом определения минимальных отрезков эффективного прогнозирования позволила сократить время измерения $БПК_5$ с пяти суток до 12,68 ч с относительной погрешностью 8,1 %, в то время как ИНС из работы [17] давала погрешность 8,42 % для фиксированного входного отрезка, соответствующего 16 ч измерений.

Специалисты и исследователи в области эксплуатации очистных сооружений указывают, что в зависимости от концентрации органических веществ в сточной воде погрешность применяемых методов контроля $БПК_5$ составляет 13–26 %. При этом отмечается, что применение МТЭ для раннего определения $БПК_5$ хотя и вносит дополнительную погрешность к погрешности стандартных методов контроля, считается допустимым, если погрешность МТЭ в среднем не превышает 10 % [16]. Представленные в данном исследовании погрешности моделей для прогнозирования $БПК_5$ приведены с учетом погрешности работы МТЭ. Таким образом, результаты определения $БПК_5$ с помощью представленных моделей считались приемлемыми, если полученная средняя относительная ошибка была в пределах 10 %. При этом отметим, что разработанный метод позволяет предсказывать концентрацию $БПК_5$ с заранее заданной погрешностью. Конечно, сокращение времени измерений может повысить суммарную относительную ошибку определения содержания $БПК_5$, но метод предоставляет возможность контролировать ее величину. Таким образом обеспечивается возможность нахождения «золотой середины» между точностью контроля сточных вод и временем проведения измерений.

Следует отметить, что математические зависимости между $БПК_5$ и функциями напряжения справедливы только для стационарных процессов, при которых состав сточных вод и технологические параметры остаются относительно постоянными. Значительные изменения в технологии производства, составе сточных вод или методах их очистки могут привести к изменению функциональных зависимостей между $БПК_5$ и напряжением. Это требует уточнения или пересмотра зависимостей на основе новых экспериментальных данных. Таким образом, практическое применение разработанных методов справедливо только в рамках данных ограничений.

Синтезация данных и применение искусственной нейронной сети

Предложенная выше модель прогнозирования может не учитывать некоторые особенно-

сти поведения рядов напряжения МТЭ. Опыт последних лет показывает, что для разрешения данной ситуации можно попытаться применить ИНС. Но в данном случае значимой трудностью является небольшой объем данных. В последние годы активно развиваются методы оптимизации ИНС для работы с малыми объемами данных, что позволяет повысить их эффективность даже в условиях ограниченных обучающих выборок [29]. И одним из используемых способов является генерация синтетических данных. Последние могут обеспечить количественную и пространственную сбалансированность данных, а также лишены ошибок, которые характерны для экспериментальных данных.

Некоторые комбинации предложенной модели прогнозирования могут использоваться для генерации данных, пригодных для обучения ИНС. Синтезация основана на параметрах функции (10), где добавочная функция имеет вид

$$f(x,\ \mu_1,\ \sigma_1,\ \lambda,\ \delta) = \frac{\lambda e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}}{\sigma_1\sqrt{2\pi}} + \delta. \qquad (16)$$

Функция (10) с добавлением функции (16) обеспечивает высокую точность аппроксимации экспериментальных данных, но неэффективна для экстраполяции на основе коротких начальных отрезков.

В процессе синтезации использовался метод варьирования параметров модели, полученных при аппроксимации экспериментальных функций. Часть результатов предложенного метода синтезации данных была анонсирована нами в тезисах [30]. Для 80 % функций из каждого подкласса были получены параметры аппроксимации (σ, μ, λ, μ_1, σ_1, δ). Для каждого параметра рассчитывались стандартные отклонения внутри подклассов. Создание отклонений параметров проводилось с использованием нормального распределения: $\varepsilon_2 \sim N(0,\ \beta \cdot \mathbf{std})$, где β — коэффициент, задающий масштаб отклонений; **std** — вектор стандартных отклонений каждого параметра. Сгенерированные наборы были получены следующим образом: $(\sigma, \mu, \lambda, \mu_1, \sigma_1, \delta)' = (\sigma, \mu, \lambda, \mu_1, \sigma_1, \delta) + \varepsilon_2$. Всего было синтезировано 2197 наборов параметров, по которым восстанавливались функции напряжения длиной 1299, представленные матрицей $\mathbf{V}_{gen}$ размерностью 2197×1299. Примем данный набор функций за эталон, который должна научиться предсказывать ИНС. Чтобы не обучать несколько ИНС для разных входных временных отрезков, предлагается использовать в качестве входных данных значения функций на отрезках эффективного прогнозирования. При этом, чтобы входные данные имели одинаковую размерность, предлагается применить функцию прогнозирования (10) с добавочной (11). Это позволит экстраполировать данные до нужной длины 1299 точек на основе отрезков эффективного прогнозирования.

Таким образом, с помощью полученных ранее значений c_l^*, c_r^* для функций матрицы $\mathbf{V}_{gen}$ были определены минимальные отрезки эффективного прогнозирования. На данных отрезках были получены параметры модели прогнозирования [комбинация функций (10) и (11)] и на основе данных параметров были восстановлены значения функций приближения длиной 1299. То есть значения приближения и экстраполяции синтезированных функций были представлены матрицей $\mathbf{U}_{gen}$ размерностью 2197×1299. Функции, хранящиеся в матрицах $\mathbf{V}_{gen}$ и $\mathbf{U}_{gen}$, являлись обучающей выборкой для ИНС. С помощью ИНС предлагается предсказывать функции матрицы $\mathbf{V}_{gen}$ (выходные значения сети) на основе функций матрицы $\mathbf{U}_{gen}$ (входные значения). Реальные значения функций напряжения не использовались в обучении, и все 54 функции были использованы в тестовой выборке.

Искусственная нейронная сеть была создана и обучена с использованием библиотеки Keras языка Python, в качестве метода обучения использован Adam со скоростью обучения 0,001. Для снижения вычислительной сложности данные усреднялись за каждые четыре значения, в результате размерность функций сократилась до 325. Матрицы $\mathbf{V}_{gen}$ и $\mathbf{U}_{gen}$ приняли размерность 2197×325, где 325 — количество нейронов во входном и выходном слоях сети. В отличие от анонсированных в тезисах [30] результатов применения синтезированных данных для обучения ИНС, в данном исследовании при формировании входных данных был применен разработанный метод поиска минимальных отрезков эффективного прогнозирования. Использовалась трехслойная архитектура ИНС, включающая еще и слой прореживания (20 %) между вторым и третьим слоями. Предварительно каждая функция была нормализована в диапазон от 0 до 1.

Тестирование ИНС проведено на выборке из 54 экспериментальных функций напряжения. По предсказанным функциям напряжения рассчитаны значения $БПК_5$. Средняя погрешность определения значений $БПК_5$ составила 8,44 % при среднем времени измерения 12,8 ч. При этом ИНС из работы [17] давала погрешность 10,66 % при использовании фиксированного входного времени 12 ч для всех функций. Для первого подкласса данных относительная погрешность выявления $БПК_5$ составила 7,39 %, а для второго подкласса 9,28 %. То есть удалось добиться улучшения для функций первого подкласса относительно использования предложенной модели прогнозирования. В частности, ИНС смог-

ла улавливать вторые скачки и широкие пики функций напряжения, чего не могла учитывать предложенная модель прогнозирования (10). За счет увеличения объема обучающих данных снизилась дисперсия ИНС и ее чувствительность к смене параметров. Кроме того, за счет определения минимальных отрезков эффективного прогнозирования отпала необходимость обучать несколько ИНС для разного входного времени.

Заключение

В исследовании разработан метод определения минимальных отрезков эффективного прогнозирования. Метод выявляет ключевые соотношения к максимальным значениям рядов на обучающей выборке, которые в дальнейшем применяются для определения границ отрезков эффективного прогнозирования на схожих рядах. Таким образом, метод определяет параметры формирования таких отрезков с учетом динамики изменений рядов. При этом отсутствует необходимость перебора фиксированных длин отрезков.

Применение метода продемонстрировало эффективность анализа временных рядов напряжения МТЭ, что важно для ранней оценки загрязнения вод. Сочетание метода с предложенной моделью прогнозирования позволило сократить время определения $БПК_5$ с пяти суток до 12,68 ч с приемлемой погрешностью 8,1 %, что открывает новые возможности для мониторинга качества сточных вод и управления процессами очистки. Метод ориентирован на однопиковые процессы (например, логнормальные), что ограничивает его применение для более сложных данных. Однако рассмотренные процессы часто встречаются в реальных задачах [21–25]. В дальнейшем может быть рассмотрен вопрос применения предложенного метода там, где временные ряды схожего вида играют ключевую роль.

Также был разработан метод формирования синтетических данных на основе параметров математической модели прогнозирования, который позволил расширить обучающую выборку и повысить эффективность обучения ИНС.

В дополнение отметим, что результаты исследования предоставляют еще один, достаточно качественный, инструмент, который, в том числе, можно использовать для автоматизации процессов обработки временных рядов. Последнее открывает перспективы для их практического применения в различных областях, включая экологию, где точный и быстрый анализ данных может способствовать эффективному мониторингу состояния окружающей среды.

Финансовая поддержка

Исследование выполнено за счет гранта Российского научного фонда (проект № 25-21-00330, https://rscf.ru/project/25-21-00330/).

Литература

1. **Павлов В. А., Шаповалов В. И., Шестаков Д. С., Кочин А. В., Рудаков А. В., Шабалин А. Е.** Сокращение продолжительности наблюдения за кинетикой нагревания подложки при магнетронном распылении медной мишени. *Известия СПбГЭТУ «ЛЭТИ»*, 2021, № 1, с. 12–17. EDN: EXFHQD
2. **Полников В. Г., Гоморев И. А.** Об оценке режимных значений временных рядов модуля скорости ветра и высоты ветрового волнения. *Метеорология и гидрология*, 2015, № 12, с. 64–73.
3. **Полников В. Г., Гоморев И. А.** Новый метод оценки режимных значений временных рядов геофизических величин. *Процессы в геосредах*, 2015, № 1, с. 54–60. EDN: VCHUIV
4. **Liu S., Yamada M., Collier N., Sugiyama M.** Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 2013, vol. 43, pp. 72–83. doi:10.1016/j.neunet.2013.01.012
5. **Ruusunen O., Jalli M., Jauhiainen L., Ruusunen M., Leiviskä K.** Identification of optimal starting time instance to forecast net blotch density in spring barley with meteorological data in Finland. *Agriculture*, 2022, vol. 12, no. 11, p. 1939. doi:10.3390/agriculture12111939
6. **Kawahara Y., Yairi T., Machida K.** Change-point detection in time-series data based on subspace identification. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 2007, pp. 559–564. doi:10.1109/ICDM.2007.78
7. **Kawahara Y., Sugiyama M.** Sequential change-point detection based on direct density-ratio estimation. *Statistical Analysis and Data Mining*, 2011, vol. 5, no. 2, pp. 114–127. doi:10.1002/sam.10124
8. **Zheng Y., Liu L., Wang L., Xie X.** Learning transportation mode from raw gps data for geographic applications on the web. *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 247–256. doi:10.1145/1367497.1367532
9. **Zhong Y., Ren Y., Cao G., Li F., Qi H.** Optimal starting point for time series forecasting. *Expert Systems with Applications*, 2025, vol. 273, p. 126798. doi:10.1016/j.eswa.2025.126798
10. **Лебедев И. С.** Адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации. *Информационно-управляющие системы*, 2022,

№ 3, с. 20–30. doi:10.31799/1684-8853-2022-3-20-30

11. **Парфёнов И. В., Поляков А. Н.** Развитие методики сокращения времени натурных тепловых испытаний станков. *Интеллект. Инновации. Инвестиции*, 2017, № 2, с. 64–69. EDN: YNLKDD
12. **Поляков А. Н., Кравцов А. Г.** Автоматизированная система прогнозирования тепловых характеристик станка. *Вестник Курганского государственного университета. Серия: Технические науки*, 2005, № 2 (2), с. 81–86. EDN: NUIBVB
13. **Li Z., Andreev A., Hramov A., Blyuss O., Zaikin A.** Novel efficient reservoir computing methodologies for regular and irregular time series classification. *Nonlinear Dynamics*, 2025, vol. 113, no. 5, pp. 4045–4062. doi:10.1007/s11071-024-10244-3
14. **Курбако А. В., Кульминский Д. Д., Боровкова Е. И., Киселев А. Р., Сказкина В. В., Пономаренко В. И., Прохоров М. Д., Безручко Б. П., Гриднев В. И., Караваев А. С.** Повышение чувствительности метода диагностики в реальном времени фазовой синхронизации автогенераторов по их нестационарным временным рядам. *Известия вузов. Прикладная нелинейная динамика*, 2021, т. 29, № 6, с. 892–904. doi:10.18500/0869-6632-2021-29-6-892-904
15. **Nagel B., Dellweg H., Gierasch L. M**. Glossary for chemists of terms used in biotechnology (IUPAC Recommendations 1992). *Pure and Applied Chemistry*, 1992, vol. 64, no. 1, pp. 143–168. doi:10.1351/pac199264010143
16. **Tardy G. M., Lóránt B., Gyalai-Korpos M., Bakos V., Simpson D., Goryanin I.** Microbial fuel cell biosensor for the determination of biochemical oxygen demand of wastewater samples containing readily and slowly biodegradable organics. *Biotechnology Letters*, 2021, vol. 43, no. 2, pp. 445–454. doi:10.1007/s10529-020-03050-5
17. **Medvedev I., Kornaukhova M., Galazis C., Lóránt B., Tardy G. M., Losev A., Goryanin I.** Using AI and BES/MFC to decrease the prediction time of BOD5 measurement. *Environmental Monitoring and Assessment*, 2023, vol. 195, no. 9, p. 1018. doi:10.1007/s10661-023-11576-0
18. **Lóránt B., Gyalai-Korpos M., Goryanin I., Tardy G. M.** Single chamber air–cathode microbial fuel cells as biosensors for determination of biodegradable organics. *Biotechnology Letters*, 2019, vol. 41, no. 4, pp. 555–563. doi:10.1007/s10529-019-02668-4
19. **Wang Y., Liu X., Wang M., Zhang P., Zong Y., Zhang Q.** A single-chamber microbial fuel cell for rapid determination of biochemical oxygen demand using low-cost activated carbon as cathode catalyst. *Environmental Technology*, 2018 vol. 39, no. 24, pp. 3228–3237. doi:10.1080/09593330.2017.1375998
20. **Oyedeji M. O., Alharbi A., Aldhaifallah M., Rezk H.** Optimal data-driven modelling of a microbial fuel cell. *Energies*, 2023, vol. 16, no. 12, p. 4740. doi:10.3390/en16124740
21. **Feng Y., Barr W., Harper Jr W. F.** Neural network processing of microbial fuel cell signals for the identification of chemicals present in water. *Journal of Environmental Management*, 2013, vol. 120, pp. 84–92. doi:10.1016/j.jenvman.2013.01.018
22. **Andersson A.** Mechanisms for log normal concentration distributions in the environment. *Scientific Reports*, 2021, vol. 11, no. 1, p. 16418. doi:10.1038/s41598-021-96010-6
23. **Ott W. R.** A physical explanation of the lognormality of pollutant concentrations. *Journal of the Air & Waste Management Association*, 1990, vol. 40, no. 10, pp. 1378–1383. doi:10.1080/10473289.1990.10466789
24. **Долгоносов Б. М., Корчагин К. А., Мессинева Е. М.** Модель флуктуаций бактериологических показателей качества речной воды. *Водные ресурсы*, 2006, т. 33, № 6, с. 686–700. EDN: HYVAWF
25. **Савичев О. Г., Луен Н. В.** О методике определения фоновых и аномальных значений гидрогеохимических показателей. *Известия Томского политехнического университета. Инжиниринг георесурсов*, 2015, т. 326, № 9, с. 133–142. EDN: UZCCDH
26. **Coleman T. F., Li Y.** An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 1996, vol. 6, no. 2, pp. 418–445. doi:10.1137/0806023
27. **Gao X., Hou Q., Yao S., Zhou K.** Opposite Normalized Trust-Region Reflective (ONTRR): A new algorithm for parameter extraction of single, double and triple diode solar cell models. *Applied Sciences*, 2023, vol. 13, no. 14, p. 8199. doi:10.3390/app13148199
28. **Le T. M., Fatahi B., Khabbaz H., Sun W.** Numerical optimization applying trust-region reflective least squares algorithm with constraints to optimize the non-linear creep parameters of soft soil. *Applied Mathematical Modelling*, 2017, vol. 41, pp. 236–256. doi:10.1016/j.apm.2016.08.034
29. **Гейдаров П. Ш.** Исследование устойчивости многослойного персептрона с вычисляемыми весами синапсов к меньшим объемам обучающей выборки. *Информационно-управляющие системы*, 2023, № 2, с. 2–14. doi:10.31799/1684-8853-2023-2-2-14, EDN: WOHMNE
30. **Медведев И. А., Корнаухова М. А.** Комбинация нейронной сети и математической модели прогнозирования напряжения микробных топливных элементов. *Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: сборник статей по материалам Девятой всероссийской научно-практической конференции с международным участием*, Пермь, 17–18 октября 2024 г. Пермь, Пермский государственный национальный исследовательский университет, 2024, с. 238–242. EDN: KLOGSC

UDC 519.246.8
doi:10.31799/1684-8853-2025-3-2-13
EDN: RAJEBP

Method of minimizing the duration of observations and mathematical extrapolation models at small data volumes

A. G. Losev[a], Dr. Sc., Phys.-Math., Professor, orcid.org/0000-0002-1072-8375, alexander.losev@volsu.ru
I. A. Medvedev[a], Post-Graduate Student, orcid.org/0009-0001-2536-4359
[a]Volgograd State University 100, Universitetsky Ave., 400062, Volgograd, Russian Federation

Introduction: Extrapolation of time series values with small amount of data is a topical problem, especially in areas where fast decision making based on limited information is required. One of the key problems is to find the smallest interval of historical data sufficient to make an accurate forecast of subsequent values of a time series. **Purpose:** To develop a method for minimizing time observations that allows extrapolation of future values with a predetermined accuracy, and to develop a data synthesis method based on the parameters of the mathematical forecasting model. **Results:** First, we develop a method for determining the minimum segments of effective forecasting. The method assumes the presence of a training sample and is based on the analysis of the dynamics of changes in the values of time series around its maximum, which makes it possible to automatically determine the minimum data segments required to make predictions with a predetermined accuracy. Second, we develop a method for generating synthetic data using the variation of the parameters of the mathematical model. Data synthesis based on varying the parameters of the mathematical model of forecasting makes it possible to increase the efficiency of training artificial neural networks. The developed methods have been applied in the analysis of microbial fuel cell voltage time series to estimate the parameter of wastewater pollution — biochemical oxygen demand. As a result, it has been possible to reduce the prediction time of this parameter from 5 days to 12.68 hours with a relative error of 8.1 %. **Practical relevance:** The developed methods used during environmental monitoring open the possibility to promptly assess water pollution and optimize wastewater treatment processes.

Keywords — mathematical extrapolation models, data synthesis, time series, artificial neural network, lognormal distribution, microbial fuel cell, biochemical oxygen demand.

For citation: Losev A. G., Medvedev I. A. Method of minimizing the duration of observations and mathematical extrapolation models at small data volumes. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 3, pp. 2–13 (In Russian). doi:10.31799/1684-8853-2025-3-2-13, EDN: RAJEBP

Financial support
The study was supported by the Russian Science Foundation (project No. 25-21-00330, https://rscf.ru/project/25-21-00330/).

References

1. Pavlov V. A., Shapovalov V. I., Shestakov D. S., Kochyn A. V., Rudakov A. V., Shabalin A. E. Shortening the duration of observation of the substrate heating kinetics uring magnetron sputtering of a copper target. *Izvestiya SPbGETU "LETI"*, 2021, no. 1, pp. 12–17 (In Russian). EDN: EXFHQD
2. Polnikov V. G., Gomorev I. A. On estimating the return values for time series of wind speed and wind wave height. *Russian Meteorology and Hydrology*, 2015, vol. 40, no. 12, pp. 820–827. doi:10.3103/S1068373915120079
3. Polnikov V. G., Gomorev I. A. A new method of estimating return values for time series of geophysical magnitudes. *Process in Geomedia*, 2015, no. 1, pp. 54–60 (In Russian). EDN: VCHUIV
4. Liu S., Yamada M., Collier N., Sugiyama M. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 2013, vol. 43, pp. 72–83. doi:10.1016/j.neunet.2013.01.012
5. Ruusunen O., Jalli M., Jauhiainen L., Ruusunen M., Leiviskä K. Identification of optimal starting time instance to forecast net blotch density in spring barley with meteorological data in Finland. *Agriculture*, 2022, vol. 12, no. 11, p. 1939. doi:10.3390/agriculture12111939
6. Kawahara Y., Yairi T., Machida K. Change-point detection in time-series data based on subspace identification. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 2007, pp. 559–564. doi:10.1109/ICDM.2007.78
7. Kawahara Y., Sugiyama M. Sequential change-point detection based on direct density-ratio estimation. *Statistical Analysis and Data Mining*, 2011, vol. 5, no. 2, pp. 114–127. doi:10.1002/sam.10124
8. Zheng Y., Liu L., Wang L., Xie X. Learning transportation mode from raw gps data for geographic applications on the web. *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 247–256. doi:10.1145/1367497.1367532
9. Zhong Y., Ren Y., Cao G., Li F., Qi H. Optimal starting point for time series forecasting. *Expert Systems with Applications*, 2025, vol. 273, p. 126798. doi:10.1016/j.eswa.2025.126798
10. Lebedev I. S. Adaptive application of machine learning models on separate segments of a data sample in regression and classification problems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 3, pp. 20–30 (In Russian). doi:10.31799/1684-8853-2022-3-20-30
11. Parfenov I. V., Polyakov A. N. Development of technique for reducing the time of machines full-scale thermal testings. *Intellect. Innovations. Investments*, 2017, no. 2, pp. 64–69 (In Russian). EDN: YNLKDD
12. Polyakov A. N., Kravtsov A. G. Automated system for predicting the thermal characteristics of the machine. *Vestnik Kurganskogo gosudarstvennogo universiteta*. Seriya: Tekhnicheskie nauki, 2005, no. 2 (2), pp. 81–86 (In Russian). EDN: NUIBVB
13. Li Z., Andreev A., Hramov A., Blyuss O., Zaikin A. Novel efficient reservoir computing methodologies for regular and irregular time series classification. *Nonlinear Dynamics*, 2025, vol. 113, no. 5, pp. 4045–4062. doi:10.1007/s11071-024-10244-3
14. Kurbako A. V., Kulminskiy D. D., Borovkova E. I., Kiselev A. R., Skazkina V. V., Ponomarenko V. I., Prokhorov M. D., Bezruchko B. P., Gridnev V. I., Karavaev A. S. Increasing the sensitivity of real-time method for diagnostic of autogenerators phase synchronization based on their non-stationary time series. *Izvestiya VUZ. Applied Nonlinear Dynamics*, 2021, vol. 29, no. 6, pp. 892–904 (In Russian). doi:10.18500/0869-6632-2021-29-6-892-904
15. Nagel B., Dellweg H., Gierasch L. M. Glossary for chemists of terms used in biotechnology (IUPAC Recommendations 1992). *Pure and Applied Chemistry*, 1992, vol. 64, no. 1, pp. 143–168. doi:10.1351/pac199264010143
16. Tardy G. M., Lóránt B., Gyalai-Korpos M., Bakos V., Simpson D., Goryanin I. Microbial fuel cell biosensor for the determination of biochemical oxygen demand of wastewater samples containing readily and slowly biodegradable organics. *Biotechnology Letters*, 2021, vol. 43, no. 2, pp. 445–454. doi:10.1007/s10529-020-03050-5
17. Medvedev I., Kornaukhova M., Galazis C., Lóránt B., Tardy G. M., Losev A., Goryanin I. Using AI and BES/MFC to decrease the prediction time of BOD5 measurement. *Environmental Monitoring and Assessment*, 2023, vol. 195, no. 9, p. 1018. doi:10.1007/s10661-023-11576-0
18. Lóránt B., Gyalai-Korpos M., Goryanin I., Tardy G. M. Single chamber air–cathode microbial fuel cells as biosensors for determination of biodegradable organics. *Biotechnology*

Letters, 2019, vol. 41, no. 4, pp. 555–563. doi:10.1007/s10529-019-02668-4
19. Wang Y., Liu X., Wang M., Zhang P., Zong Y., Zhang Q. A single-chamber microbial fuel cell for rapid determination of biochemical oxygen demand using low-cost activated carbon as cathode catalyst. *Environmental Technology*, 2018, vol. 39, no. 24, pp. 3228–3237. doi:10.1080/09593330.2017.1375998
20. Oyedeji M. O., Alharbi A., Aldhaifallah M., Rezk H. Optimal data-driven modelling of a microbial fuel cell. *Energies*, 2023, vol. 16, no. 12, p. 4740. doi:10.3390/en16124740
21. Feng Y., Barr W., Harper Jr W. F. Neural network processing of microbial fuel cell signals for the identification of chemicals present in water. *Journal of Environmental Management*, 2013, vol. 120, pp. 84–92. doi:10.1016/j.jenvman.2013.01.018
22. Andersson A. Mechanisms for log normal concentration distributions in the environment. *Scientific Reports*, 2021, vol. 11, no. 1, p. 16418. doi:10.1038/s41598-021-96010-6
23. Ott W. R. A physical explanation of the lognormality of pollutant concentrations. *Journal of the Air & Waste Management Association*, 1990, vol. 40, no. 10, pp. 1378–1383. doi:10.1080/10473289.1990.10466789
24. Dolgonosov B. M., Korchagin K. A., Messineva E. M. Model of fluctuations in bacteriological indices of water quality. *Water Resources*, 2006, vol. 33, no. 6, pp. 637–650. doi:10.1134/S0097807806060054, EDN: LJXQOV
25. Savichev O. G., Luyen N. V. The technique of determining background and extreme values of hydrogeochemical parameters. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2015, vol. 326, no. 9, pp. 133–142 (In Russian). EDN: UZCCDH
26. Coleman T. F., Li Y. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 1996, vol. 6, no. 2, pp. 418–445. doi:10.1137/0806023
27. Gao X., Hou Q., Yao S., Zhou K. Opposite Normalized Trust-Region Reflective (ONTRR): A new algorithm for parameter extraction of single, double and triple diode solar cell models. *Applied Sciences*, 2023, vol. 13, no. 14, p. 8199. doi:10.3390/app13148199
28. Le T. M., Fatahi B., Khabbaz H., Sun W. Numerical optimization applying trust-region reflective least squares algorithm with constraints to optimize the non-linear creep parameters of soft soil. *Applied Mathematical Modelling*, 2017, vol. 41, pp. 236–256. doi:10.1016/j.apm.2016.08.034
29. Geidarov P. Sh. Investigating stability of a multilayer perceptron with calculated synaptic weights to smaller training sample sizes. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 2, pp. 2–14 (In Russian). doi:10.31799/1684-8853-2023-2-2-14, EDN: WOHMNE
30. Medvedev I. A., Kornaukhova M. A. Combination of a neural network and a mathematical model for predicting the voltage of microbial fuel cells. *Iskusstvennyy intellekt v reshenii aktual'nykh sotsial'nykh i ekonomicheskikh problem XXI veka: Sbornik statey po materialam Devyatoy vserossiyskoy nauchno-prakticheskoy konferentsii s mezhdunarodnym uchastiem* [Artificial Intelligence in Solving Current Social and Economic Problems of the 21st Century: Proceedings of the 9th All-Russian Scientific-Practical Conference with International Participation]. Perm, 2024, pp. 238–242 (In Russian). EDN: KLOGSC

УВАЖАЕМЫЕ АВТОРЫ!

Научные базы данных, включая Scopus и Web of Science, обрабатывают данные автоматически. С одной стороны, это ускоряет процесс обработки данных, с другой — различия в транслитерации ФИО, неточные данные о месте работы, области научного знания и т. д. приводят к тому, что в базах оказывается несколько авторских страниц для одного и того же человека. В результате для всех по отдельности считаются индексы цитирования, что снижает рейтинг ученого.

Для идентификации авторов в сетях Thomson Reuters проводит регистрацию с присвоением уникального индекса (ID) для каждого из авторов научных публикаций.

Процедура получения ID бесплатна и очень проста, есть возможность провести регистрацию на 12 языках, включая русский (чтобы выбрать язык, кликните на зеленое поле вверху справа на стартовой странице): https://orcid.org