

2.4.3 Электроэнергетика

Научная статья

УДК 621.31

<https://doi.org/10.37493/2307-907X.2025.6.5>

ВЛИЯНИЕ СПОСОБОВ ВОССТАНОВЛЕНИЯ ПРОПУЩЕННЫХ ДАННЫХ НА ТОЧНОСТЬ ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ АЛГОРИТМОМ ГРАДИЕНТНОГО БУСТИНГА

Арсен Гумарович Шидов^{1*}, Юрий Григорьевич Кононов², Дмитрий Александрович Костюков³,
Мурат Рустемович Куршев⁴, Беслан Гумарович Шидов⁵

¹⁻⁵ Северо-Кавказский федеральный университет (д. 1, ул. Пушкина, Ставрополь, 355017, Российская Федерация)

¹ shidovarsen@gmail.com; <https://orcid.org/0009-0007-8777-4680>

² iukononov@ncfu.ru; <https://orcid.org/0009-0000-4960-4647>

³ d-kostjukov@mail.ru; <https://orcid.org/0009-0007-2183-5217>

⁴ murat.kurshev64@gmail.com; <https://orcid.org/0009-0007-4632-2522>

⁵ bgshidov@ncfu.ru; <https://orcid.org/0009-0005-4675-2536>

* Автор, ответственный за переписку

Аннотация. Введение. В низковольтных сетях активно внедряют интеллектуальные системы учета электроэнергии (ИСУЭ), данные которых могут использоваться для прогнозирования потребления. Однако наличие пропусков в данных увеличивает погрешность результатов прогноза. **Цель.** Сравнительная оценка влияния различных методов восстановления пропущенных значений в данных ИСУЭ на точность прогнозирования электропотребления. **Материалы и методы.** Исследование проведено на основе реального набора данных, содержащего почасовые значения активной энергии 132 однофазных бытовых потребителей одного из регионов Северного Кавказа за 25-месячный период. Применены четыре способа восстановления пропусков в данных: средним значением, медианным значением, интерполяцией и медианой по каждому часу суток. Для прогнозирования часовых графиков электропотребления на месяц, неделю и сутки вперед использована модель машинного обучения XGBoost, оценка качества прогноза выполнялась по метрике RMSE. **Результаты и обсуждение.** На основе анализа средних значений показателя RMSE сделан вывод о том, что выбор способа усреднения оказывает некоторое влияние на месячный интервал прогноза, для которого медианные способы заполнения недостающих данных дают меньшее на 0,066 кВт·ч, или 12,6 %, значение RMSE по сравнению со способом заполнения отсутствующих данных средними значениями и на 0,053 кВт·ч, или 10,2 %, по сравнению со способом заполнения отсутствующих данных интерполяцией. При недельном горизонте прогнозирования лучший результат дает медианный по часам с преимуществом RMSE над остальными способами 0,013–0,021 кВт·ч, а для суточного прогноза наибольшую эффективность показывает медианный способ с RMSE 0,012–0,023 кВт·ч. **Заключение.** На месячном горизонте прогноза целесообразно применять медианные способы заполнения недостающих данных. При прогнозе на неделю и сутки все рассмотренные способы практически равноценны.

Ключевые слова: интеллектуальный учет, восстановление недостающих данных, прогнозирование, электропотребление, XGBoost

Для цитирования: Влияние способов восстановления пропущенных данных на точность прогнозирования электропотребления алгоритмом градиентного бустинга / А. Г. Шидов [и др.] // Вестник Северо-Кавказского федерального университета. 2025. № 6(111). С. 46–55. <https://doi.org/10.37493/2307-907X.2025.6.5>

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 13.10.2025;

одобрена после рецензирования 17.11.2025;

принята к публикации 24.11.2025.

Research article

IMPACT OF MISSING DATA RECOVERY METHODS ON THE ACCURACY OF ELECTRICITY CONSUMPTION FORECASTING BY GRADIENT BOOSTING ALGORITHM

Arsen G. Shidov^{1*}, Yuri G. Kononov², Dmitry A. Kostyukov³, Murat R. Kurshev⁴, Beslan G. Shidov⁵

¹⁻⁵ North-Caucasus Federal University (1, Pushkin str., Stavropol, 355017, Russian Federation)

¹ shidovarsen@gmail.com; <https://orcid.org/0009-0007-8777-4680>

² iukononov@ncfu.ru; <https://orcid.org/0009-0000-4960-4647>

³ d-kostjukov@mail.ru; <https://orcid.org/0009-0007-2183-5217>

⁴ murat.kurshev64@gmail.com; <https://orcid.org/0009-0007-4632-2522>

⁵ bgshidov@ncfu.ru; <https://orcid.org/0009-0005-4675-2536>

* Corresponding author

Abstract. Introduction. In low-voltage networks, smart electricity metering systems (ISMS) are being actively implemented, and their data can be used to predict consumption. However, the presence of missing values in the data increases the error in the prediction results. **Goal.** To compare the impact of different methods for recovering missing values in SES data on the accuracy of electricity consumption forecasting. **Materials and methods.** The

study was based on a real dataset containing hourly values of active energy from 132 single-phase household consumers in one of the regions of the North Caucasus over a 25-month period. Four methods were used to fill in missing data: the mean, the median, interpolation, and the median for each hour of the day. The XGBoost machine learning model was used to predict hourly electricity consumption patterns for the next month, week, and day, and the quality of the prediction was evaluated using the RMSE metric. **Results and discussion.** Based on the analysis of the average values of the RMSE indicator, it was concluded that the choice of the averaging method has some effect on the monthly forecast interval, for which the median methods of filling in missing data result in a lower RMSE value by 0,066 kWh, or 12,6 %, compared to the method of filling in missing data with average values, and by 0,053 kWh, or 10,2 % compared to the method of filling in missing data with interpolation. With a weekly forecast horizon, the best result is obtained by the hourly median with an advantage of RMSE over the rest of the methods of 0,013-0,021 kWh, and for the daily forecast, the median method with RMSE of 0,012-0,023 kWh shows the greatest efficiency. **Conclusion.** On a monthly forecast horizon, it is advisable to use median methods to fill in the missing data. When forecasting for a week or a day, all of the methods discussed are almost equivalent.

Keywords: intelligent accounting, recovery of missing data, forecasting, power consumption, XGBoost

For citation: Shidov AG, Kononov YuG, Kostyukov DA, Kurshev MR, Shidov BG. Impact of missing data recovery methods on the accuracy of electricity consumption forecasting by gradient boosting algorithm. Newsletter of North-Caucasus Federal University. 2025;6(111):46-55. (In Russ.). <https://doi.org/10.37493/2307-907X.2025.6.5>

Conflict of interest: The authors declare that there is no conflict of interest.

The article was submitted 13.10.2025;

approved after reviewing 17.11.2025;

accepted for publication 24.11.2025.

Введение / Introduction. Одной из актуальных задач электросетевых компаний на современном этапе является обеспечение наблюдаемости при распределении электроэнергии [1]. Интеллектуальные системы учёта электроэнергии (ИСУЭ) в условиях высокой варьированности потерь электроэнергии в субъектах России – от 2,87 до 52 % – и растущей нагрузки на низковольтные сети являются ключевым элементом цифровизации отрасли [2, 3].

Основным компонентом ИСУЭ являются интеллектуальные приборы учёта электроэнергии (ИПУЭ), которые обеспечивают автоматический сбор и передачу данных о потреблении энергии с высоким разрешением, что открывает новые возможности для мониторинга и прогнозирования электропотребления. Эффективность прогнозирования зависит от качества данных, получаемых с ИПУЭ [4].

Наиболее распространенными проблемами ИСУЭ являются недостаточная надежность каналов передачи данных и уязвимость приборов учета от стороннего вмешательства в их работу, из-за чего возникают пропуски (недостающие значения) в наборах данных об электроэнергии на интервалах профиля нагрузки [5]. Кроме того, причиной пропусков в данных являются плановые и аварийные отключения потребителей. Наличие пропусков снижает точность прогноза электропотребления, который может использоваться в том числе и для выявления аномалий в электропотреблении, основной причиной которых являются умышленное воздействие на систему учета с целью хищения электроэнергии.

За последнее время для прогноза электропотребления разработаны достаточно эффективные алгоритмы, основанные на модели машинного обучения XGBoost [6]. В работе [7] показано, что погрешность почасового прогноза электропотребления по таким алгоритмам не превышает 4 %. Однако исследования по влиянию пропусков на погрешность прогноза часовых профилей нагрузки отсутствуют.

Целью настоящего исследования является сравнительный анализ существующих способов заполнения недостающих значений в данных ИСУЭ и оценка влияния этих способов на точность прогнозирования электропотребления.

Материалы и методы исследований / Materials and methods of research. Существуют различные способы восстановления недостающих данных. Например, в статье [5] для обработки данных используют интерполяцию или вовсе удаляют данные в случае дубликатов или пропущенных значений. В работе [8] авторы применили простой метод обработки данных на основе среднего значения по строкам (строка соответствует одному потребителю в исходной таблице данных) и выполнили нормализацию данных для приведения набора данных к определенному диапазону. Метод был протестирован на реальном наборе данных, собранном Государственной электросетевой корпорацией Китая (SGCC) и содержащем информацию о ежедневном потреблении электроэнергии 42 372 пользователями за 2 года, и показал свою эффективность. Наибольшей популярностью для заполнения пропущенных данных пользуется интерполяционный подход [9, 10].

В настоящей статье применяется предварительная обработка данных представленными выше способами, а также восстановление пропущенных данных медианными значениями. Затем производится обучение модели машинного обучения XGBoost и анализируется влияние различных способов восстановления данных на точность прогнозирования.

Для проведения исследований были созданы четыре выборки с часовым потреблением электроэнергии для 132 потребителей, отличающиеся способами заполнения недостающих значений:

1. Заполнение средним значением часового потребления:

$$\hat{x}_i = \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, \quad (1)$$

где \hat{x}_i – заполненное значение часового потребления для i -го потребителя; \bar{x} – среднее арифметическое всех непустых значений часового потребления для i -го потребителя; n – количество непустых значений в данных профиля i -го потребителя с 01.01.2022 по 31.12.2023; x_j – непустое значение часового потребления для i -го потребителя.

2. Заполнение медианным значением часового потребления:

$$x_i^{\wedge} = \text{median}(x_1, x_2, \dots, x_j), \quad (2)$$

где $\text{median}()$ – медианное значение часового потребления для i -го потребителя с 01.01.2022 по 31.12.2023, т.е. значение часового потребления, соответствующее потреблению центрального элемента упорядоченного по возрастанию массива непустых значений часового потребления для i -го потребителя с 01.01.2022 по 31.12.2023; x_1, x_2, \dots, x_j – все непустые значения часового потребления для i -го потребителя с 01.01.2022 по 31.12.2023.

3. Линейная интерполяция:

$$\hat{x}_i = x_{i-1} + (t_i - t_{i-1}) / (t_{i+1} - t_{i-1}) \times (x_{i+1} - x_{i-1}), \quad (3)$$

где t_i – временная метка i -го часа; x_{i-1}, x_{i+1} – ближайшие непустые значения часового потребления для i -го потребителя; t_{i-1}, t_{i+1} – временные метки ближайших непустых значений часового потребления для i -го потребителя.

4. Заполнение медианой по значениям часового потребления для i -го потребителя в соответствующий час суток:

$$\hat{x}_i = \text{median}(\{x_j : \text{hour}(x_j) = \text{hour}(x_i) \wedge x_j \neq \emptyset\}), \quad (4)$$

где $\text{hour}(x_i)$ – час суток для i -го потребителя; $\{x_j : \text{hour}(x_j) = \text{hour}(x_i) \wedge x_j \neq \emptyset\}$ – множество всех непустых значений часового потребления для i -го потребителя в час суток $\text{hour}(x_i)$.

На основе выборок о часовом потреблении электроэнергии была обучена модель машинного обучения XGBoost [11], которая известна своей эффективной обработкой структурированных табличных данных и способности к обучению на неполных данных после восстановления недостающих данных. Известные авторам статьи результаты исследований по применению данного алгоритма для прогнозирования часовых профилей электропотребления показывают эффективность XGBoost-алгоритма по сравнению с другими методами прогнозирования [11].

Признаками модели были временные характеристики и значение потребления электроэнергии. Для оценки качества прогнозирования потребления электроэнергии использовалось временное разделение данных, которое обеспечивает реалистичность эксперимента. Выборка была разделена на обучающую и тестовую. Обучающая выборка включала все исторические данные за исключением последнего месяца, тестовая – последний месяц (январь 2024 года). Для каждого потребителя модель обучалась отдельно, а затем использовалась для прогнозирования потребления электроэнергии за последний месяц. Полученные прогнозы сравниваются с реальными значениями из тестовой выборки. Далее рассчитываются метрики качества для каждого потребителя отдельно, затем метрики усредняются, находятся потребители с лучшей и худшей метриками.

Для оценки качества прогнозирования применяется корень среднеквадратической ошибки (RMSE) [11] как одна из стандартных метрик регрессии.

Характеристика набора реальных данных потребления электроэнергии одного из регионов Северного Кавказа приведена в таблице 1.

Таблица 1 / Table 1

Исследуемый набор данных / The data set under study

Количество приборов учета	132
Формат данных	Почасовые значения активной энергии
Характер потребителей	Однофазное подключение
Временной период	01.01.2022–31.01.2024
Количество значений	2 410 848
Количество пропусков, %	1,34

*Источник: составлено авторами / Source: compiled by the authors

Приведенный в таблице набор данных был создан на основе выборки из профилей 1 770 приборов учета энергии. В наборе пропуски данных составляют 23 % от общего количества значений, что свидетельствует о длительных промежутках пропуска значений от одного до нескольких месяцев. В связи с этим был введен критерий отбора, который подразумевает сохранение недельной цикличности путем отбора приборов учета, у которых отсутствовали пропуски не более 168 часов, что соответствует ровно 7 суткам. При таких условиях имеется возможность обеспечить репрезентативность выборки. Это позволяет корректно применять методы восстановления данных, а также избегать увеличения количества пропусков и их длительности в случае установления критериев отбора на 2-4 недельных интервалах времени. Нулевые значения не относились к пропускам данных, т.к. по сведениям электросетевой компании, предоставившей информацию, нулевые значения являются достоверными данными и в базе данных представляются численным значением в отличие от пропущенных данных, представляемых пустым полем.

На рисунке 1 приведена гистограмма количества суммарных пропусков в данных для каждого из 132 ИПУЭ за весь период наблюдения.

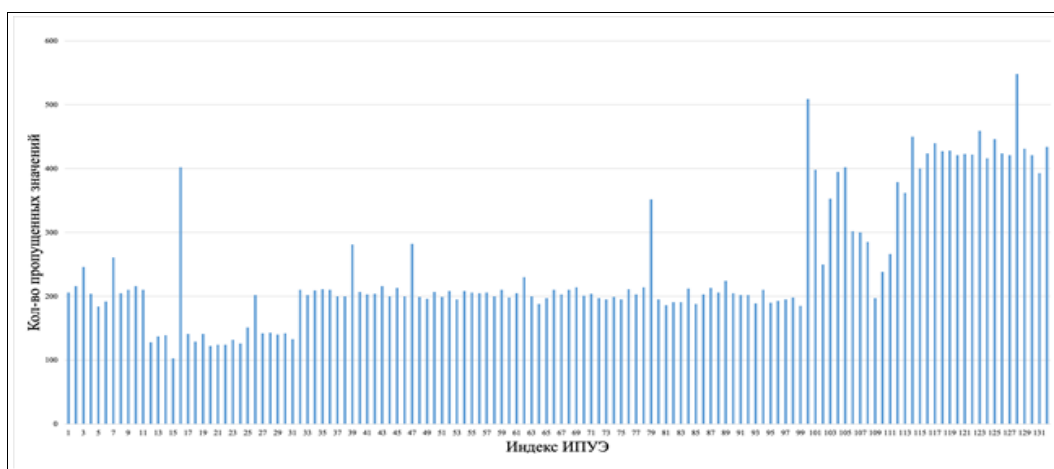


Рис. 1. Величины количества суммарных пропусков в данных по каждому ИПУЭ / Fig. 1. The number of total misses in the data for each IEMD

*Источник: составлено авторами / Source: compiled by authors

Наибольшее число пропусков в данных (548) имеет ИПУЭ с номером 128, а минимальное (102) имеет ИПУЭ с номером 14. Среднее число пропусков для всех ИПУЭ равно 245, что соответствует 10 дням и 5 часам пропусков. Всего количество недостающих значений для всех потребителей составляет 32 306.

Результаты исследований и их обсуждение / Research results and their discussion. Пример графика часового потребления одного из 132 абонентов на интервале 3 месяца с пропущенными значениями приведен на рисунке 2.

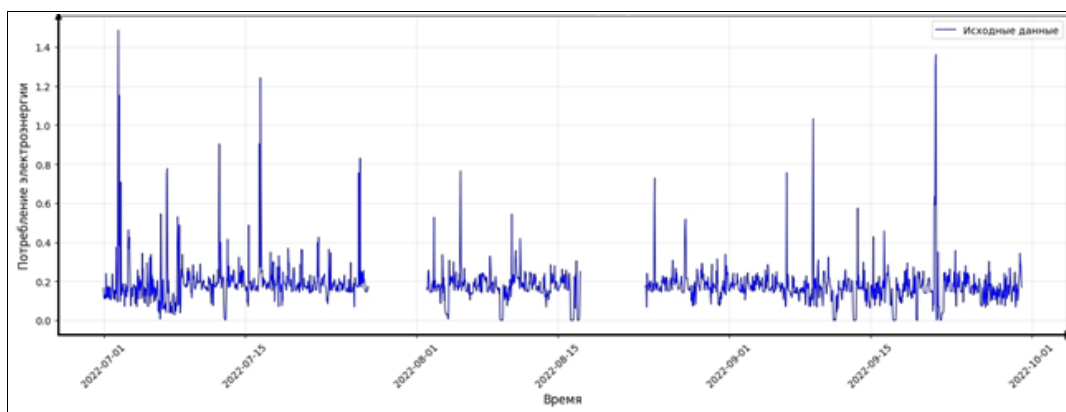


Рис. 2. График потребления абонента с пропущенными значениями данных / Fig. 2. Subscriber consumption graph with missing data values

*Источник: составлено авторами / Source: compiled by authors

Как следует из представленного на рисунке графика, на рассматриваемом трехмесячном периоде имеются 2 интервала недостающих значений продолжительностью около 5 дней каждый.

Результаты восстановления пропущенных значений 4 рассмотренными выше способами представлены на рисунке 3.

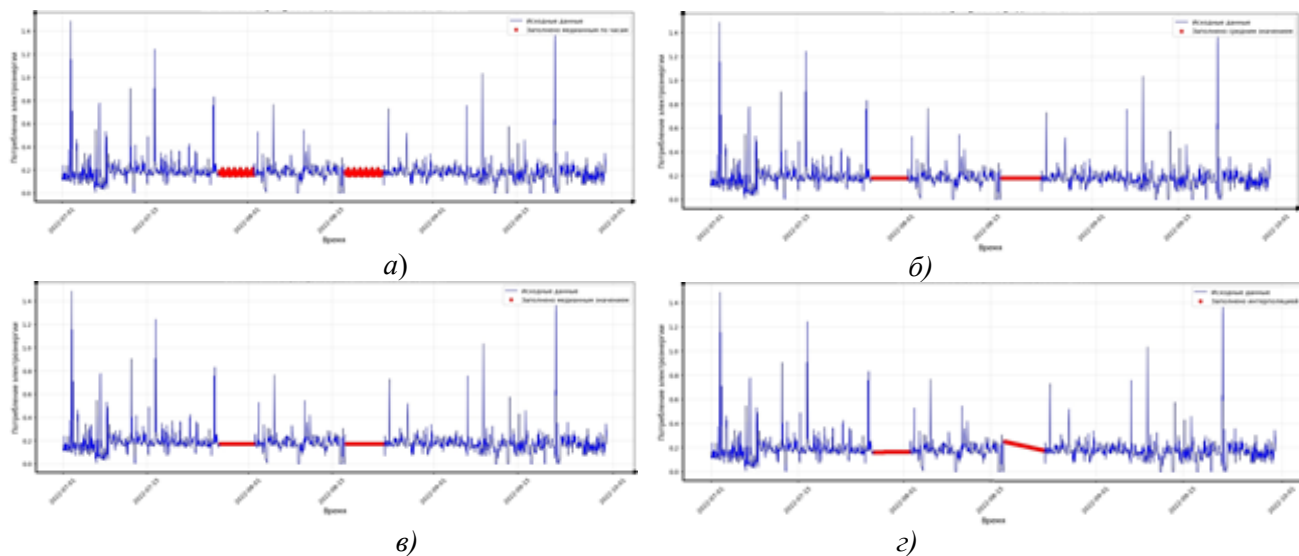


Рис. 3. Графики часового потребления абонента при заполнении недостающих значений различными способами (а – медианным значением по часам; б – средним значением; в – медианным значением; г – интерполяцией) / Fig. 2. Graphs of subscriber consumption when filling gaps in different ways (a – filling gaps with the median value by hour; b – filling gaps with the average value; c – filling gaps with the median value; d – filling gaps with interpolation)

*Источник: составлено авторами / Source: compiled by authors

Восстановленные участки графиков электропотребления средним (рисунок 3б) и медианным (рисунок 3в) значениями визуально не отличаются, т. к. численно восстановленные значения часовых потреблений очень близки: 0,1795 кВт·ч и 0,1710 кВт·ч соответственно.

Более детальное сравнение результатов заполнения недостающих значений исследуемыми способами продемонстрировано на рисунке 4.

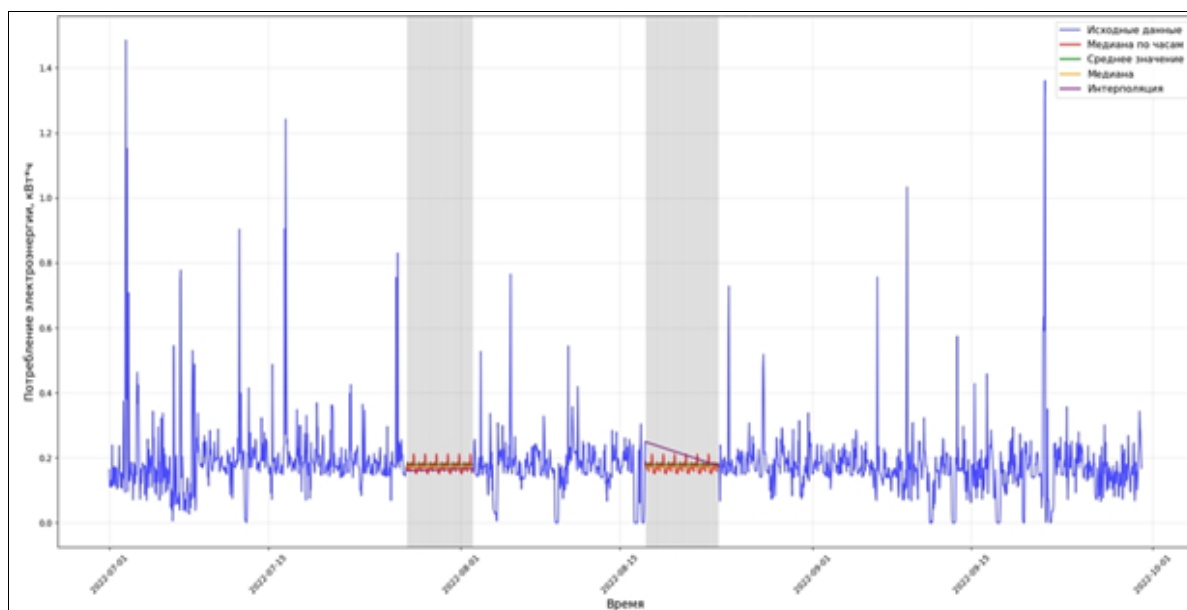


Рис. 4. Детальное сравнение результатов заполнения недостающих значений исследуемыми способами / Fig. 4. Detailed comparison of methods for filling in missing values

*Источник: составлено авторами / Source: compiled by authors

Обобщенные результаты расчета среднеквадратичных ошибок прогноза (RMSE) на горизонтах месяц, неделя и сутки представлены на рисунке 5.

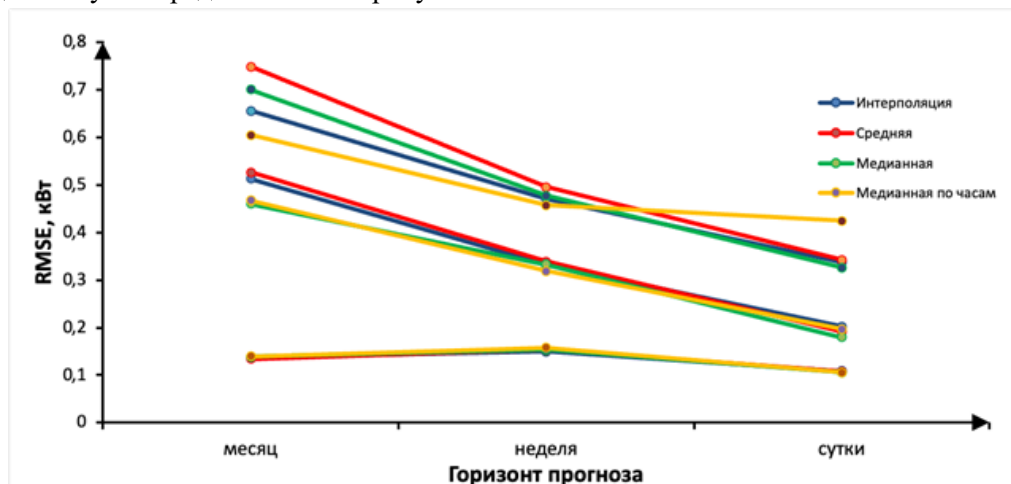


Рис. 5. Влияние длительности прогноза на его погрешность (RMSE) / Fig. 5. Effect of forecast duration on its error (RMSE)

*Источник: составлено авторами / Source: compiled by authors

На этом рисунке центральная группа графиков представляет значения RMSE, усредненные для всех 132 ИПУЭ, крайняя нижняя группа – значения для потребителя с наилучшим прогнозом, а крайняя верхняя – значения для потребителя с наихудшим прогнозом. На основе анализа средних значений RMSE можно сделать вывод, что выбор способа усреднения оказывает некоторое влияние на месячный интервал прогноза, для которого медианный способ заполнения недостающих данных дает меньшее на 0,066 кВт·ч, или 12,6 %, значение RMSE по сравнению со способом заполнения отсутствующих данных средними значениями и на 0,053 кВт·ч, или 10,3 %, по сравнению со способом заполнения отсутствующих данных интерполяцией. При недельном горизонте прогнозирования лучший результат дает медианный по часам способ с преимуществом по RMSE над остальными способами 0,013–0,021 кВт·ч, а для суточного прогноза наибольшую эффективность показывает медианный способ (улучшение RMSE на 0,012–0,023 кВт·ч). При месячном и недельном прогнозировании преимущество предлагаемого способа заполнения медианным заполнением по часам по сравнению с остальными способами более выражено для потребителя с наихудшим прогнозом. Однако явное отставание этого способа при суточном прогнозировании требует проведения дополнительных исследований. Следует отметить, что только для суточного интервала прогнозирования среднее значение RMSE ниже средней потребляемой мощности: 0,363 кВт·ч. Это подтверждает известный факт о повышении точности прогноза с уменьшением горизонта прогнозирования.

Для наглядной оценки качества прогноза на рисунке 6 проведены графики лучших и худших прогнозных результатов и измеренных значений часового электропотребления для каждого горизонта прогноза: месяца, недели, суток.

Дополнительно приведенные эксперименты по включению температуры воздуха от ближайшей метеостанции в качестве дополнительного входного параметра в модели прогнозирования показали, что во всех проанализированных случаях, за исключением двух (способов заполнения пропусков интерполяцией и средним значением), учет температурных данных приводил не к снижению, а к увеличению погрешности прогноза. Такой результат свидетельствует о вероятно слабой взаимосвязи электропотребления с погодными условиями на выбранном временном интервале (январь 2024 года).

Для оценки качества полученных результатов проведём сравнение полученных показателей RMSE с его значениями, приводимыми в статье [12] вьетнамскими авторами, которые применили алгоритм XGBoost для прогнозирования почасового потребления электроэнергии на данных одного домохозяйства в Ирландии и одного во Вьетнаме в период с 23 сентября 2023 года по 6 июля 2024 года. Для суточного интервала прогнозирования авторы достигли значений RMSE 0,0712 кВт по данным ирландского домохозяйства и 0,0265 кВт по данным вьетнамского домохозяйства, что в 2,5 и 6,8 раза меньше значения RMSE 0,179 кВт, полученного в настоящем исследовании в качестве среднего значения для всех 132 потребителей. Для лучшего прогноза различия в показателях менее значи-

тельны и составляют всего 33 % для показателя RMSE, рассчитанного по данным ирландского домохозяйства и 75 % для показателя, вычисленного по данным вьетнамского домохозяйства. Лучшие значения показателя RMSE в работе [12] могут быть объяснены более сложным алгоритмом предварительной подготовки данных, включающим фильтрацию аномальных выбросов, разделение данных по сезонам и видам погоды (сухая и влажная), и ряд других улучшений. Целесообразность учета этих факторов при реализации алгоритмов прогнозирования электропотребления бытовыми абонентами на территории России требует проведения дополнительных исследований в будущем.

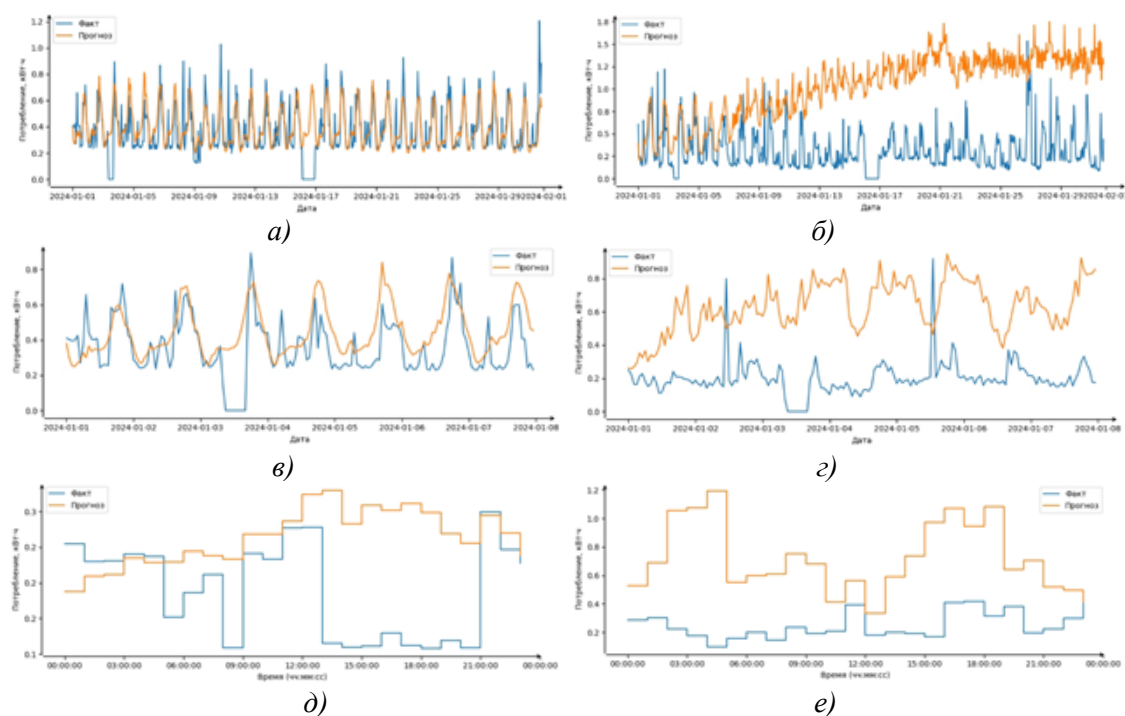


Рис. 6. Графики лучшего и худшего прогноза потребителей на различных горизонтах прогноза (а – лучший месячный, б – худший месячный, в – лучший недельный, г – худший недельный, д – лучший суточный, е – худший суточный) / Fig. 5. Graphs of the best and worst consumer forecasts at different forecast horizons (a – best monthly, b – worst monthly, c – best weekly, d – worst weekly, e – best daily, f – worst daily)

*Источник: составлено авторами / Source: compiled by authors

Заключение / Conclusion. В рамках настоящего исследования проведена сравнительная оценка эффективности различных способов заполнения недостающих значений в наборах данных ИСУЭ при решении задачи прогнозирования часового электропотребления на базе современной модели машинного обучения XGBoost.

Анализ влияния четырёх различных способов (замена средним значением, медианой, интерполяция и медианой по часам) к заполнению недостающих данных часового электропотребления показал, что на месячном горизонте прогноза медианные способы заполнения недостающих данных дают в среднем меньшие на 0,053–0,066 кВт·ч значения показателя RMSE по сравнению со способами заполнения отсутствующих данных средними значениями или интерполяцией. При недельном и суточном горизонте прогноза преимущества медианных способов сохраняются, но сравнительная эффективность по показателю RMSE снижается до 0,012–0,023 кВт·ч.

Перспектива развития полученных в настоящей работе результатов связана с повышением эффективности алгоритма градиентного бустинга при решении задачи выявления аномалий в электропотреблении бытовых абонентов за счет более качественной обработки входных данных и их кластеризации, оптимизации макропараметров XGBoost-модели.

СПИСОК ИСТОЧНИКОВ

1. Концепция «Цифровая трансформация 2030» / ПАО «Россети». М., 2018. 31 с. URL: https://rosseti.ru/investment/Kontsepsiya_Tsifrovaya_transformatsiya_2030.pdf.

2. О порядке предоставления доступа к минимальному набору функций интеллектуальных систем учета электрической энергии (мощности): постановление Правительства РФ от 19 июня 2020 г. № 890 (в ред. от 29 марта 2024 г.) // Официальный интернет-портал правовой информации. URL: <http://publication.pravo.gov.ru/Document/View/0001202006230034>.
3. О внесении изменений в отдельные законодательные акты Российской Федерации в связи с развитием систем учета электрической энергии (мощности) в Российской Федерации: федеральный закон РФ от 27 дек. 2018 г. № 522-ФЗ : принят Гос. Думой Федер. Собр. РФ 19 дек. 2018 г.: одобр. Советом Федерации Федер. Собр. РФ 21 дек. 2018 г. // Рос. газ. 2018. 29 дек. № 295; Собр. законодательства Рос. Федерации. 2018. 31 дек. № 53 (ч. I). Ст. 8448. URL: <http://publication.pravo.gov.ru/Document/View/0001201812290002>
4. Shimmari M. Al. and Wallom D. Short-term load forecasting using UK non-domestic businesses to enable demand response aggregators' participation in electricity markets // IEEE PES Grid Edge Technologies Conference & Exposition (Grid Edge), San Diego, CA, USA. 2023. pp. 1-5, <https://doi.org/10.1109/GridEdge54130.2023.10102712>.
5. Gao H.-X., S. Kuenzel and X.-Y. Zhang. A Hybrid ConvLSTM-Based Anomaly Detection Approach for Combating Energy Theft // IEEE Transactions on Instrumentation and Measurement. 2022. Vol. 71. P. 2517110. <https://doi.org/10.1109/TIM.2022.3201569>.
6. Mark Ryan, Luca Massaron. Machine Learning for Tabular Data: XGBoost, Deep Learning, and AI. Manning, 2025. 504 p.
7. Sobrino E. M., Santiago A. V., González A. M. Forecasting the Electricity Hourly Consumption of Residential Consumers with Smart Meters using Machine Learning Algorithms // IEEE Milan PowerTech: proceedings. Milan, Italy, 2019. P. 1–6. <https://doi.org/10.1109/PTC.2019.8810902>.
8. Munawar S. Novel FDI-based data manipulation and its detection in smart meters' electricity theft scenarios / S. Munawar, Z. A. Khan, N. I. Chaudhary, N. Javaid, M. A. Z. Raja, A. H. Milyani, A. A. Azhari // Frontiers in Energy Research. 2022. Vol. 10, article 1043593. <https://doi.org/10.3389/fenrg.2022.1043593>.
9. Chen Z. Electricity Theft Detection Using Deep Bidirectional Recurrent Neural Network / Z. Chen, D. Meng, Y. Zhang, T. Xin, D. Xiao // 22nd International Conference on Advanced Communication Technology (ICACT). Phoenix Park, Korea (South), 2020. P. 401-406. <https://doi.org/10.23919/ICACT48636.2020.9061565>.
10. Маниковский А. С., Мухопад А. Ю. Методы восстановления пропущенных измерений во временных рядах в системе прогнозирования электропотребления // Инженерный вестник Дона. 2022. № 7. URL: <http://cyberleninka.ru/article/n/metody-vosstanovleniya-propuschnnyh-znacheniy-vo-vremennyh-ryadah-v-sisteme-prognozirovaniya-elektropotrebleniya>.
11. Raviprabhakaran V. P. Pusuluri, B. Nendrilla [et al.] Household Power Consumption Analysis using Machine Learning // IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET): proceedings. 2024. <https://doi.org/10.1109/SEFET61574.2024.10718254>.
12. Asghar, Ehtisham & Hill, Martin & Şengör, Ibrahim & Lynch, Conor & Quang An, Phan. Validation of a 24-hourahead Prediction model for a Residential Electrical Load under diverse climate. 2025. <https://doi.org/10.48550/arXiv.2505.00348>.

REFERENCES

1. Concept «Digital Transformation 2030». PAO Rosseti. – Moscow, 2018. 31 p. – Available from: https://rosseti.ru/investment/Kontseptsiya_Tsifrovaya_transformatsiya_2030.pdf. (In Russ.).
2. On the procedure for providing access to the minimum set of functions of intelligent systems for metering electrical energy (capacity): Government of the Russian Federation Resolution №. 890 of June 19, 2020: as amended on March 29, 2024. Official Internet Portal of Legal Information. Available from: <http://publication.pravo.gov.ru/Document/View/0001202006230034>. (In Russ.).
3. On Amending Certain Legislative Acts of the Russian Federation in Connection with the Development of Electric Energy (Power) Metering Systems in the Russian Federation: Federal Law of the Russian Federation of December 27, 2018 №. 522-FZ: adopted by the State Duma of the Federal Assembly of the Russian Federation on December 19, 2018: approved by the Council of the Federation of the Federal Assembly of the Russian Federation on December 21, 2018. Rossijskaya Gazeta. 2018. December 29. №. 295; Collection of Legislation of the Russian Federation. 2018. December 31. No. 53 (Part I), Art. 8448. (In Russ.).
4. Shimmari MAL, Wallom D. Short-term load forecasting using UK non-domestic businesses to enable demand response aggregators' participation in electricity markets. IEEE PES Grid Edge Technologies Conference & Exposition (Grid Edge), San Diego, CA, USA; 2023. P. 1-5, <https://doi.org/10.1109/GridEdge54130.2023.10102712>.
5. Gao H-X, S. Kuenzel and Zhang X-Y. A Hybrid ConvLSTM-Based Anomaly Detection Approach for Combating Energy Theft. IEEE Transactions on Instrumentation and Measurement. 2022;(71):2517110. <https://doi.org/10.1109/TIM.2022.3201569>.
6. Mark Ryan, Luca Massaron. Machine Learning for Tabular Data: XGBoost, Deep Learning and AI. Manning; 2025. 504 p.
7. Sobrino EM, Santiago AV, González AM. Forecasting the Electricity Hourly Consumption of Residential Consumers with Smart Meters using Machine Learning Algorithms. IEEE Milan PowerTech: proceedings. Milan, Italy; 2019. P. 1-6. <https://doi.org/10.1109/PTC.2019.8810902>.

8. Munawar S, Khan ZA, Chaudhary NI, Javaid N, Raja MAZ, Milyani AH, Azhari AA. Novel FDI-based data manipulation and its detection in smart meters' electricity theft scenarios. *Frontiers in Energy Research*. 2022;(10): article 1043593. <https://doi.org/10.3389/fenrg.2022.1043593>.
9. Chen Z. Electricity Theft Detection Using Deep Bidirectional Recurrent Neural Network / Z. Chen, D. Meng, Y. Zhang, T. Xin, D. Xiao // 22nd International Conference on Advanced Communication Technology (ICACT). Phoenix Park, Korea (South), 2020. P. 401-406. <https://doi.org/10.23919/ICACT48636.2020.9061565>.
10. Manikovsky AS, Mukhopad AYu. Methods of recovery of missed measurements in time series in the power consumption forecasting system. *Engineering Bulletin of the Don*. 2022;(7). Available from: <http://cyberleninka.ru/article/n/metody-vosstanovleniya-propuschnykh-znacheniy-vo-vremennyh-ryadah-v-sisteme-prognozirovaniya-elektropotrebleniya>. (In Russ.).
11. Raviprabhakaran V, Pusuluri P, Nendrilla B. [et al.]. Household Power Consumption Analysis using Machine Learning. *IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET): proceedings*. 2024. <https://doi.org/10.1109/SEFET61574.2024.10718254>.
12. Asghar, Ehtisham & Hill, Martin & Şengör, Ibrahim & Lynch, Conor & Quang An, Phan. Validation of a 24-hour-ahead Prediction model for a Residential Electrical Load under diverse climate. 2025. <https://doi.org/10.48550/arXiv.2505.00348>.

ИНФОРМАЦИЯ ОБ АВТОРАХ

- Арсен Гумарович Шидов** – аспирант кафедры АЭСиЭ, факультет нефтегазовой инженерии, Северо-Кавказский федеральный университет, Researcher ID: GQP-6570-2022.
- Юрий Григорьевич Кононов** – доктор технических наук, профессор, заведующий кафедрой автоматизированных электроэнергетических систем и электроснабжения, факультет нефтегазовой инженерии, Северо-Кавказский федеральный университет, Scopus ID: 56985782200, Researcher ID: J-3033-2015.
- Дмитрий Александрович Костюков** – кандидат технических наук, доцент кафедры автоматизированных электроэнергетических систем и электроснабжения, факультет нефтегазовой инженерии, Северо-Кавказский федеральный университет, Scopus ID: 58029421300, Researcher ID: OFN-4011-2025.
- Мурат Рустемович Куршев** – аспирант кафедры автоматизированных электроэнергетических систем и электроснабжения, факультет нефтегазовой инженерии, Северо-Кавказский федеральный университет, Researcher ID: OFN-4850-2025.
- Беслан Гумарович Шидов** – магистр кафедры АЭСиЭ, факультет нефтегазовой инженерии, Северо-Кавказский федеральный университет, Researcher ID: LUA-0241-2024.

ВКЛАД АВТОРОВ

- Арсен Гумарович Шидов**
Проведение исследования – сбор данных электропотребления и их обработка. Утверждение окончательного варианта – принятие ответственности за все аспекты работы, целостность всех частей статьи и ее окончательный вариант.
- Юрий Григорьевич Кононов**
Проведение исследования – интерпретация и анализ полученных результатов модели. Утверждение окончательного варианта – принятие ответственности за все аспекты работы, целостность всех частей статьи и ее окончательный вариант.
- Дмитрий Александрович Костюков**
Проведение исследования – анализ современной российской и зарубежной литературы по теме исследования.
- Мурат Рустемович Куршев**
Проведение исследования – разработка моделей прогнозирования в наборе данных, оценка их эффективности.
- Беслан Гумарович Шидов**
Подготовка и редактирование текста – составление черновика рукописи и формирование его окончательного варианта, участие в научном дизайне.

INFORMATION ABOUT THE AUTHORS

- Arsen G. Shidov** – Postgraduate Student of the Department of AESiE, Faculty of Oil and Gas Engineering, North-Caucasus Federal University, Researcher ID: GQP-6570-2022.
- Yuri G. Kononov** – Dr. Sci. (Techn.) Professor, Head of the Department of Automated Electric Power Systems and Power Supply, North-Caucasus Federal University, Scopus ID: 56985782200, Researcher ID: J-3033-2015.
- Dmitry A. Kostyukov** – Cand. Sci. (Techn.), Associate Professor of the Department of Automated Electric Power Systems and Power Supply, North-Caucasus Federal University, Scopus ID: 58029421300, Researcher ID: OFN-4011-2025.
- Murat R. Kurshev** – Postgraduate Student, Department of AECiE, Faculty of Oil and Gas Engineering, North-Caucasus Federal University, Researcher ID: OFN-4850-2025.

Beslan G. Shidov – Master Student of the AESiE, Faculty of Oil and Gas Engineering, North-Caucasus Federal University, Researcher ID: LUA-0241-2024.

CONTRIBUTION OF THE AUTHORS

Arsen G. Shidov

Conducting a study is collecting power consumption data and processing it. The approval of the final version is the acceptance of responsibility for all aspects of the work, the integrity of all parts of the article and its final version.

Yuri G. Kononov

Conducting a study is the interpretation and analysis of the obtained model results. The approval of the final version is the acceptance of responsibility for all aspects of the work, the integrity of all parts of the article and its final version.

Dmitry A. Kostyukov

The research is an analysis of modern Russian and foreign literature on the research topic.

Murat R. Kurshev

The purpose of the research is to develop forecasting models in the data set, and evaluate their effectiveness.

Beslan G. Shidov

Text preparation and editing – drafting of the manuscript and its final version, contribution to the scientific layout.