

МЕТОДИКА КОНТРОЛЯ ДОСТОВЕРНОСТИ ДАННЫХ ПРИБОРОВ УЧЁТА В АСКУЭ НА ОСНОВЕ ПРОГНОЗИРУЮЩЕЙ МОДЕЛИ ЭЛЕКТРОПОТРЕБЛЕНИЯ

Тымчук Алексей Игоревич

Соискатель, Кубанский Государственный
Технологический Университет
aleximchuck@gmail.com

RELIABILITY CONTROL METHOD FOR DATA FROM AMR METERING DEVICES, BASED ON A FORECASTING POWER CONSUMPTION MODEL

A. Tymchuck

Summary. Purpose. Methods that are currently used in AMR to control the reliability of data, for monitoring the reliability of data from metering devices have disadvantages in terms of redundancy of electricity metering devices, redundant computing power and growth in the volume of system databases, which increases AMR implementation and maintenance cost. The reliability control of these metering devices using a mathematical apparatus will eliminate these shortcomings. Target. The target is to develop a method for monitoring the reliability of these metering devices in AMR that does not have the disadvantages of currently used methods, as well as to evaluate its effectiveness by creating on its basis an information system for AMR of an apartment building. Methods: the following methods were used to develop the technique and the information system that implements it — the Box-Jenkins method for building ARIMA models, the Foster-Stewart method to check if power consumption time series have a trend, the analysis of autocorrelation and partial autocorrelation functions to check power consumption time series for the presence of periodic interdependencies, the Dolado-Jenkins-Sausville-Riviero method and the extended Dickey-Fuller test to test power consumption time series for stationarity, the gradient descent method for training a forecasting power consumption neural network model. Novelty. The novelty of the method lies in the use of a forecasting power consumption model for reliability control of data from AMR metering devices. Results. The use of this method will reduce the search time in the AMR for electricity metering devices that transmit false data, which in turn will reduce power losses and the cost of the system maintenance Practical relevance. The developed method can be put into practice in an information system that interacts with the AMR databases. The forecasting power consumption model is built separately for each type of AMR control object. This article provides an example of using the method for an apartment building equipped with an AMR system.

Keywords: information system, power consumption forecasting, neural networks, ARIMA, AMR, data reliability control.

Аннотация. Постановка задачи: используемые в настоящее время в автоматизированных системах контроля и учёта электроэнергии (АСКУЭ) для проверки достоверности данных, методы контроля достоверности данных приборов учёта имеют недостатки в плане избыточности приборов учёта электроэнергии, избыточности вычислительных мощностей и разрастании объёма баз данных системы, что увеличивает стоимость реализации и обслуживания АСКУЭ. Контроль достоверности данных приборов учёта с использованием математического аппарата позволит устранить данные недостатки. Целью работы является разработка методики контроля достоверности данных приборов учёта в АСКУЭ не имеющего недостатков, имеющихся у используемых в настоящее время методов, а также оценка её эффективности путём создания на её основе информационной системы для АСКУЭ многоквартирного дома. Используемые методы: для разработки методики и реализующей её информационной системы были использованы следующие методы — метод Бокса-Дженкинса для построения ARIMA моделей, метод Фостера-Стюарта для проверки временных рядов электропотребления на наличие тренда, анализ автокорреляционных и частных автокорреляционных функций для проверки временных рядов электропотребления на наличие периодических взаимозависимостей, метод Доладо-Дженкинса-Сосвилла-Ривьеро и расширенный тест Дики-Фуллера для проверки временных рядов электропотребления на стационарность, метод градиентного спуска для обучения прогнозирующей нейросетевой модели электропотребления. Новизна: новизна разработанной методики заключается в использовании для контроля достоверности данных приборов учёта в АСКУЭ прогнозирующей модели электропотребления. Результат: использование данной методики позволит сократить время поиска в АСКУЭ приборов учёта электроэнергии, передающих недостоверные данные, что в свою очередь уменьшит потери электроэнергии и снизит расходы на её обслуживание. Практическая значимость: разработанная методика может быть реализована на практике в информационной системе, которая взаимодействует с базами данных АСКУЭ. Прогнозирующая модель электропотребления строится отдельно для каждого типа объекта контроля АСКУЭ. В данной статье приведён пример использования разработанной методики для многоквартирного дома, оборудованного системой АСКУЭ.

Ключевые слова: информационная система, прогнозирование электропотребления, нейронные сети, ARIMA, АСКУЭ, контроль достоверности данных.

Актуальность

Актуальность данной работы обусловлена активным развитием технологий в сфере автоматизации процессов и производства. В области контроля и учёта потребления электроэнергии важную роль играет автоматизация процессов сбора хранения и передачи информации с приборов учёта. Автоматизированные системы контроля и учёта, созданные для решения этой задачи, эффективно справляются с поставленными целями, однако, используемые ими в настоящее время методики контроля достоверности информации, передаваемой приборами учёта, имеют недостаточную экономическую эффективность.

В настоящее время в существующих АСКУЭ применяются два основных метода контроля достоверности данных, передаваемых приборами учёта электроэнергии, которые могут применяться как совместно, так и по отдельности.

1 Кластеризация всех принадлежащих АСКУЭ приборов учёта и использование для обнаружения в полученных кластерах приборов учёта, которые передают недостоверные данные, контрольных точек между кластерами. Например, в случае если кластером является множество приборов учёта многоквартирного дома, то контрольной точкой будет являться общедомовой прибор учёта.

7. Увеличение частоты опроса приборов учёта. Основным недостатком данного метода является то, что уменьшение временного интервала опроса может помочь только в случае если производится преднамеренное вмешательство в работу прибора. Во всех остальных случаях, большинством из которых будут являться технические неисправности приборов учёта, данный метод не принесёт необходимого результата, однако серьёзно увеличит нагрузку на используемые АСКУЭ каналы связи и увеличит объём передаваемой информации, что в свою очередь увеличит скорость заполнения центральной базы данных [1].

Постановка задачи

Проведённый анализ предметной области показывает, что используемые в настоящее время в АСКУЭ для проверки достоверности данных, методы кластеризации множества приборов учёта и увеличения частоты опроса приборов учёта имеют недостатки. Данные недостатки заключаются в наличии избыточности приборов учёта электроэнергии, избыточности вычислительных мощностей и разрастании объёма баз данных системы, что увеличивает стоимость реализации

и обслуживания АСКУЭ, снижая тем самым её экономическую эффективность. В свою очередь аналитические методы контроля достоверности данных, передаваемых приборами учёта электроэнергии в АСКУЭ, имеют ряд существенных преимуществ по сравнению с методами, которые используются в настоящее время. В связи с этим была поставлена задача разработки методики контроля достоверности данных потребления электроэнергии в АСКУЭ и создания на её основе информационной системы для АСКУЭ многоквартирного дома.

Методика контроля достоверности данных потребления электроэнергии

Задачу контроля достоверности данных приборов учёта электроэнергии в АСКУЭ можно определить следующим образом:

Пусть O — объект контроля, оборудованный АСКУЭ.
 S — множество (кластер) приборов учёта электроэнергии объекта контроля O .

N — число приборов учёта электроэнергии во множестве S .

P_i — данные потребления электроэнергии i -го прибора учёта во множестве S на текущий момент времени.

ΔP_i — максимальный допустимый процент потерь электроэнергии для i -го прибора учёта.

D — подмножество приборов учёта во множестве S , передающих недостоверные данные.

R — подмножество приборов учёта во множестве S , отмеченных при использовании методики, как передающие недостоверные данные.

Необходимо для множества S объекта контроля O найти такое подмножество R , которое будет максимально возможным образом совпадать с подмножеством D .

Таким образом, на основании формальной постановки задачи, методику контроля достоверности данных приборов учёта электроэнергии в АСКУЭ можно представить следующим алгоритмом:

1. Если для каждого из приборов учёта электроэнергии выполняется, система неравенств (1), то перейти к выполнению второго пункта предложенной методики, иначе перейти к выполнению пятого пункта.

$$\begin{cases} P_i > P_i * (1 - \Delta P_i) \\ P_i < P_i * (1 + \Delta P_i) \end{cases} \quad (1)$$

2. Применить процедуру идентификации приборов учёта и выделить во множестве приборов учёта S подмножество приборов учёта R , максимально возможным образом совпадающее с подмножеством D приборов учёта, передающих недостоверные данные. Если подмножество $R = \emptyset$,

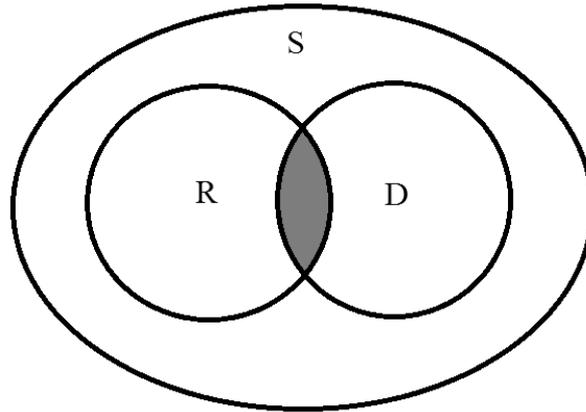


Рис. 1. Графическое представление задачи контроля достоверности данных приборов учёта электроэнергии в АСКУЭ

- то перейти к выполнению четвёртого пункта предлагаемой методики, иначе перейти к пункту три.
3. Провести техническим персоналом автоматизированной системы контроля и учёта проверку приборов учёта, попавших во множество *R*. Перейти к первому пункту данной методики.
 4. Провести техническим персоналом автоматизированной системы контроля и учёта проверку всех приборов учёта из множества *S*.
 5. Завершить выполнение проверки.

Идентификации приборов учёта с использованием прогнозирующей модели электропотребления

Поскольку хранящаяся в базе данных АСКУЭ информация о потреблении электроэнергии может быть представлена в виде временных рядов, то для решения задачи идентификации приборов учёта электроэнергии в АСКУЭ возможно применение методов статистического анализа временных рядов.

Для этого используем временной ряд значений каких-либо параметров исследуемого процесса за равные промежутки времени. Одной из основных целей анализа временных рядов является прогнозирование его поведения. Поэтому идентификация приборов учёта электроэнергии в АСКУЭ может быть произведена при помощи прогнозирующей модели потребления электроэнергии, использующей для решения поставленной задачи сравнение прогнозированного значения электропотребления для текущего момента времени с исходным значением электропотребления для того же момента времени.

Формально задачу идентификации приборов учёта электроэнергии в АСКУЭ с применением прогнозиру-

ющей модели потребления можно определить следующим образом:

Пусть X_{nm} — матрица из n различных временных рядов потребления электроэнергии, каждый из которых имеет одинаковый период контроля m .

$$\begin{pmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nm} \end{pmatrix}, \tag{2}$$

Необходимо проверить достоверность данных вектора $(X_{1m} \dots X_{nm})$ используя значения векторов $(X_{i1} \dots X_{im-1})$ для всех $i = 1 \dots n$.

Для решения этой задачи необходимо для каждого вектора $(X_{i1} \dots X_{im-1})$ $i = 1 \dots n$ построить прогнозирующую модель и с их помощью построить вектор прогнозных значений $(X'_{1m} \dots X'_{nm})$ и проведя сравнение его значений с соответствующими значениями вектора $(X_{1m} \dots X_{nm})$ сделать выводы о достоверности данных последнего.

Исходя из этого, условие достоверности данных приборов учёта можно записать в виде системы уравнений (3):

$$\begin{cases} X_{im} \leq X'_{im} * (1 + \Delta X_{im}) \\ X_{im} \geq X'_{im} * (1 - \Delta X_{im}), \\ i = 1 \dots n \end{cases} \tag{3}$$

Максимальное допустимое значение отклонения прогноза от исходных данных при оценке достоверности данных потребления для каждого конкретного прибора учёта ΔX_i вычисляется по формуле (4).

$$\Delta X_{im} = P_m + P_t, \tag{4}$$

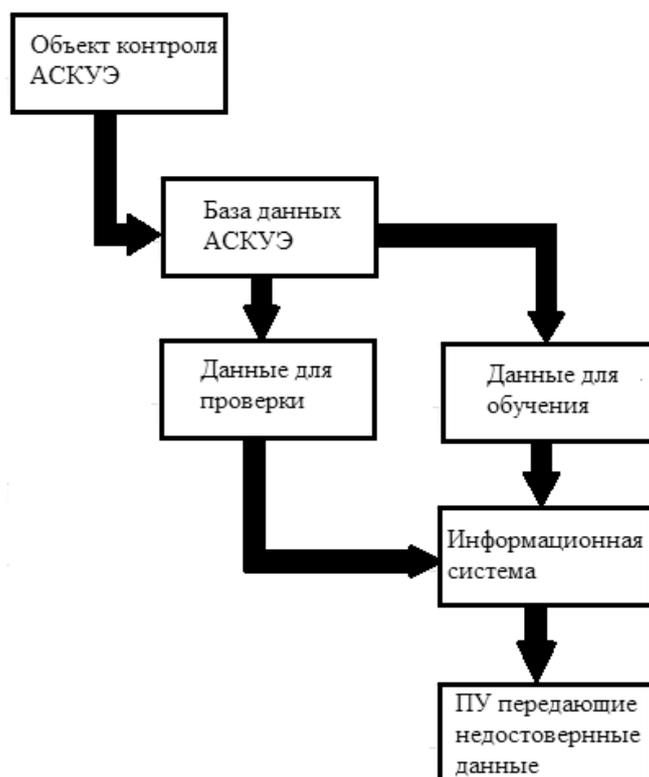


Рис. 2. Модель взаимодействия ИС с внешней средой

где:

P_m — погрешность прогноза модели потребления электроэнергии;

P_t — класс точности прибора учёта.

Погрешность прогноза модели потребления определяется при проведении оценки её адекватности путём вычисления среднего значения отклонения данных исходного ряда электропотребления от данных ряда, который был получен с использованием прогнозирующей модели.

Класс точности приборов учёта электроэнергии определяется с учётом задач, решаемых автоматизированной системой контроля электроэнергии и конкретным прибором учёта. Руководящим документом для определения класса точности прибора учёта является Предварительный национальный стандарт Российской Федерации ПНСТ 159–2016 «Автоматизированные информационно-измерительные системы коммерческого учёта электроэнергии» [2], а конкретно Таблица 1 этого документа.

Таким образом, для определения достоверности данных потребления электроэнергии в АСКУЭ с использованием прогнозирующей модели необходимо применение следующего алгоритма действий:

1. Учесть класс точности для каждого прибора учёта.
2. Определить погрешность прогнозирующей модели для каждого прибора учёта.
3. Вычислить максимальное допустимое значение отклонения прогноза от реальных данных для каждого прибора учёта.
4. Используя условие проверки достоверности оценить достоверность данных потребления электроэнергии для каждого прибора учёта.

Временной ряд потребления электроэнергии имеет равные промежутки между измерениями, следовательно, для построения прогнозирующей модели потребления электроэнергии на основе временных рядов могут быть применены «модели Бокса-Дженкинса или *ARIMA* модели» [3].

Структура информационной системы

Для решения задачи контроля достоверности данных потребления электроэнергии в АСКУЭ многоквартирного дома на основе рассмотренной выше методики была разработана информационная система. Структура информационной системы представлена моделями взаимодействия данной ИС с внешней средой на рис. 2 и внутреннего устройства ИС — на рис. 3.

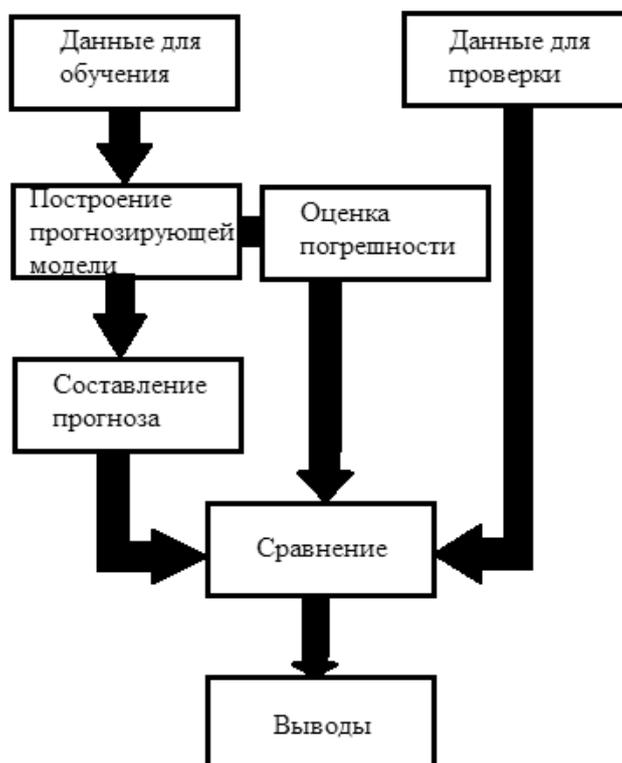


Рис. 3. Модель внутреннего устройства ИС

Источником исходных данных для информационной системы является база данных АСКУЭ объекта контроля. Информационная система получает из БД АСКУЭ данные потребления электроэнергии, зафиксированные приборами учёта объекта контроля, за некоторый период времени, достоверность которых была подтверждена путём проведения проверки приборов учёта, а также данные потребления электроэнергии, зафиксированные теми же приборами учёта, полученные после проведения проверки приборов учёта, достоверность которых необходимо проверить. Используя представленную ранее методику контроля достоверности данных, ИС проводит анализ полученной информации и предоставляет пользователю заключение о достоверности или недостоверности данных, передаваемых проверяемыми приборами учёта. На рис. 2 функционирование самой информационной системы представлено по принципу «чёрного ящика». Модель на рис. 3 раскрывает данный «чёрный ящик» и показывает объекты и взаимодействия между ними внутри информационной системы.

ARIMA модель электропотребления многоквартирного дома

Применение методики контроля достоверности данных потребления электроэнергии в АСКУЭ требу-

ет проведения идентификации приборов учёта электропотребления. Для идентификации приборов учёта согласно этой методике необходимо построить прогнозирующую *ARIMA* модель потребления электроэнергии. Для АСКУЭ многоквартирного дома данная представляет собой сумму отдельных прогнозирующих моделей каждого отдельного входящего в неё прибора учёта.

Для построения прогнозирующей *ARIMA* модели потребления отдельного прибора учёта необходимо определить тип данной модели. Для решения данной задачи были использованы данные потребления электроэнергии восьми разных приборов учёта за период в тринадцать месяцев, которые представлены в таблице 1, и применена методология построения *ARIMA* моделей Бокса-Дженкинса [4].

Согласно первому шагу методологии Бокса-Дженкинса, необходимо проверить данные на наличие в них линейного тренда, периодической зависимости и стационарности. Для проверки гипотезы о наличии в каждом из рядов из таблицы 1 тренда был проведён их анализ методом Фостера-Стюарта [5].

Согласно таблице критических значений *t*-критерия Стьюдента [6] для $n = 13$ при уровне значимости

Таблица 1. Исходные данные

№	Месяц	1	2	3	4	5	6	7	8
1	01.2020	332	1319	139	236	88	148	190	707
2	02.2020	307	1240	130	274	89	119	185	805
3	03.2020	306	1053	124	34	85	118	169	755
4	04.2020	266	956	135	118	87	138	190	638
5	05.2020	275	870	130	131	112	123	187	688
6	06.2020	316	854	119	125	140	126	257	638
7	07.2020	275	544	165	218	95	173	569	360
8	08.2020	329	433	218	195	97	159	662	283
9	09.2020	337	448	147	205	98	108	333	273
10	10.2020	297	482	197	114	76	238	256	340
11	11.2020	352	1201	166	134	81	70	201	533
12	12.2020	287	932	121	129	82	117	186	824
13	01.2021	282	1029	120	139	236	88	173	882

Таблица 2. Результаты анализа рядов из таблицы 1 методом Фостера-Стюарта

№ Ряда	s	d	t_s	t_d	Тренд в средних	Тренд в дисперсиях
1	-1	5	0,483	0,551	нет	нет
2	-7	7	3,383	2,084	да	нет
3	-1	5	0,483	0,551	нет	нет
4	0	2	0	1,748	нет	нет
5	2	6	2,416	1,317	да	нет
6	-2	6	2,416	1,317	да	нет
7	1	5	0,483	0,551	нет	нет
8	-1	7	0,483	2,084	нет	нет

Таблица 3. Результаты анализа АКФ и ЧАКФ рядов из таблицы 1

№ Ряда	№ Лага на котором достигнут максимум АКФ	№ Лага на котором достигнут максимум ЧАКФ
1	4	4
2	1	1
3	0	0
4	0	0
5	0	0
6	0	0
7	1	1
8	1	1

$p = 0,05$ значение равно 2,160. При сравнении данного значения с полученными результатами в таблице 2 видно, что для всех рядов за исключением второго они не превышают критического значения. Для второго ряда значение показателя t_s больше критического значения t -критерия Стьюдента при уровне значимости $p = 0,05$. Это позволяет сделать вывод о том, что ряды под номерами два, пять и шесть имеют тренд в средних значениях, также есть вероятность того, что ряды под номерами два и восемь имеют тренд в дисперсиях, по-

скольку значение показателя t_d для них очень близко к критическому значению.

Второй шаг методики Бокса-Дженкинса — проверка данных на наличие периодических зависимостей. Для проверки данных из таблицы 1 на наличие периодической зависимости, для всех рассматриваемых рядов, были построены автокорреляционная функция (АКФ) и частная автокорреляционная функция (ЧАКФ) с коэффициентами до двенадцатого порядка включительно,

Таблица 4. Результаты анализа рядов из таблицы 1 на стационарность

№	ADF с константой и трендом	ADF с константой, но без тренда	ADF без константы и тренда	Стационарность по ДФ критерию
1	Стационарный, $p = 0,08$	Стационарный, $p = 0,007$	-	Стационарный
2	Нестационарный, $p = 0,21$	Стационарный, $p = 0,08$	-	Стационарный
3	Стационарный, $p = 0,07$	Стационарный, $p = 0,05$	-	Стационарный
4	Стационарный, $p = 0,06$	Стационарный, $p = 0,02$	-	Стационарный
5	Нестационарный, $p = 0,25$	Нестационарный, $p = 0,18$	Нестационарный, $p = NA$	Нестационарный
6	Стационарный, $p = 0,002$	-	-	Стационарный
7	Нестационарный, $p = 0,21$	Стационарный, $p = 0,05$	-	Стационарный
8	Стационарный, $p = 0,04$	-	-	Стационарный

поскольку предположительно периодическая зависимость рядов связана с числом месяцев в году. Результаты анализа приведены в таблице 3.

Третий шаг методики Бокса-Дженкинса — проверка рядов на стационарность. Для проверки рядов потребления электроэнергии из таблицы 1 на стационарность был применён метод Доладо-Дженкинса-Сосвилла-Ривьеро [7], использующий комбинацию расширенных тестов Дики-Фуллера [8] результаты которых представлены в таблице 4.

Проведённый анализ временных рядов потребления электроэнергии из таблицы 1 показал, что временной ряд потребления электроэнергии на исследуемом промежутке:

1. может быть, как стационарным, так и не стационарным (не прогнозируемым);
2. может содержать периодическую зависимость;
3. может содержать линейный или нелинейный тренд, либо не иметь тренд вовсе.

С учётом присутствующего разнообразия в параметрах ряда построение некоей универсальной модели ряда потребления электроэнергии в многоквартирном доме не представляется возможным. Поэтому для каждого временного ряда потребления электроэнергии представляющая его математическая модель должна подбираться индивидуально с учётом его параметров.

Ряд под номером один на графиках АКФ и ЧАКФ имеет единственное значимое значение на четвёртом лаге [9], следовательно, он должен быть представлен моделью $ARMA(4,0)$ или $AR(4)$. Однако такой высокий, в отличие от остальных, параметр авторегрессионной модели неудобен при прогнозировании ряда нейронной сетью. Поэтому было произведено исследование конечных разностей данного ряда и было определено, что вторая конечная разность данного ряда может быть представлена моделью $AR(1)$, а следовательно, ряд под

номером один может также быть представлен моделью $ARMA(1,2,0)$.

Ряды под номерами два, семь и восемь на графиках АКФ и ЧАКФ имеют единственное значимое значение на первом лаге [9]. Следовательно, данные ряды могут быть представлены моделями $ARMA(1,0)$ или $AR(1)$.

Ряд номер три на графиках АКФ и ЧАКФ не имеет значимых значений, однако вторая конечная разность данного ряда на графике АКФ имеет единственное значимое значение на первом лаге [9] и на графике ЧАКФ значимые значения на первом и втором лагах и наибольшее из них на втором. Таким образом данный ряд является разностно-стационарным рядом второго порядка и может быть представлен моделью $ARIMA(0,2,2)$.

Ряд под номером четыре на графиках АКФ и ЧАКФ не имеет значимых значений, однако вторая конечная разность данного ряда на графиках АКФ и ЧАКФ имеет единственное значимое значение на первом лаге [9]. Таким образом данный ряд является разностно-стационарным рядом второго порядка и может быть представлен моделью $ARIMA(1,2,0)$.

Ряд под номером пять согласно ADF тестам нестационарный, первая и вторая разности данного ряда также нестационарные. Следовательно, ряд не является разностно-стационарным [9], а следовательно, является непрогнозируемым при помощи $ARIMA$ моделей. Такое поведение ряда обусловлено аномально высоким последним значением ряда, возникшим в результате форс-мажорного события, что обусловило резкий скачок потребления электроэнергии. Поскольку ряд является непрогнозируемым, то оценка достоверности данных с использованием предложенной ранее методики невозможна.

Ряд под номером шесть на графиках АКФ и ЧАКФ не имеет значимых значений, однако первая разность

Таблица 5. Количество нейронов входного и скрытого слоя для моделей $ARMA(p,q)$ до второго порядка включительно

Модель	Число параметров модели	Число нейронов во входном слое	Число нейронов в скрытом слое
ARMA(0,1)	3	1	3
ARMA(0,2)	4	2	4
ARMA(1,0)	3	1	3
ARMA(1,1)	4	1	4
ARMA(1,2)	5	2	5
ARMA(2,0)	4	2	4
ARMA(2,1)	5	2	5
ARMA(2,2)	6	2	6

данного ряда на графиках АКФ и ЧАКФ имеет единственное значимое значение на первом лаге [9]. Таким образом данный ряд является разностно-стационарным рядом первого порядка и может быть представлен моделью $ARIMA(1,1,0)$.

Таким образом ряды потребления электроэнергии в многоквартирном доме могут быть представлены множеством моделей $ARIMA(p,d,q)$, где $(p,d,q) \in [0,2]$, $p+q \geq 1$.

Прогнозирующая нейросетевая модель

Так как потребление электроэнергии отдельными приборами учёта в АСКУЭ многоквартирного дома может быть представлено при помощи множества моделей $ARIMA(p,d,q)$, где $(p,d,q) \in [0,2]$, $p+q \geq 1$, то построение прогнозирующей модели потребления электроэнергии может быть осуществлено при помощи набора искусственных нейронных сетей [10], реализующих данные $ARIMA(p,d,q)$ модели. Модели $ARIMA(p,d,q)$ приводятся к моделям $ARMA(p,q)$ посредством взятия конечных разностей ряда порядка d , которое происходит до обработки ряда нейронной сетью. Поэтому конфигурация нейронной сети зависит исключительно от параметров p и q . Для реализации этих моделей нейронная сеть должна иметь помимо входного и выходного слоя один скрытый слой, который представляет собой параметры уравнения $ARMA$ модели. Число нейронов в выходном слое равно одному, а число нейронов во входном и скрытом слое зависит от порядка модели и представлено в таблице 5. Общая структура данных нейронных сетей может быть представлена формулой:

$$(a * I, b * H, 1 * O), \quad (5)$$

где:

a — число нейронов входного слоя;

b — число нейронов скрытого слоя;

I, H, O — входной, скрытый и выходной слои нейронной сети.

Поскольку модели $ARMA(p,q)$ состоят из двух моделей $AR(p)$ и $MA(q)$, а максимальный порядок модели не превышает двух, то уравнения составных частей модели могут иметь вид:

1. $X_t = c + a_1 * X_{t-1} + \varepsilon_t$ для модели $AR(1)$;
2. $X_t = c + a_1 * X_{t-1} + a_2 * X_{t-2} + \varepsilon_t$ для модели $AR(2)$;
3. $X_t = c + b_1 * \varepsilon_{t-1} + \varepsilon_t$ для модели $MA(1)$;
4. $X_t = c + b_1 * \varepsilon_{t-1} + b_2 * \varepsilon_{t-2} + \varepsilon_t$ для модели $MA(2)$.

Для нейронной сети неважно в каких именно комбинациях данных уравнений будет представлено уравнение прогнозирующей модели ряда, важно лишь количество параметров в данных уравнениях как показано в таблице 5. Поскольку модели AR и MA имеют повторяющиеся параметры c и ε_t , то их возможно суммировать, тем самым уменьшив число нейронов в скрытом слое нейронной сети.

Из таблицы 5. видно, что существует восемь $ARMA$ моделей, имеющих порядок до второго включительно, однако, имеется три пары моделей, которые могут быть представлены одинаковыми нейронными сетями. Таким образом для представления $ARMA$ моделей, имеющих порядок до второго включительно, достаточно пяти различных конфигураций нейронной сети.

Каждая из данных нейронных сетей использует для обучения и работы наборы величин потребления электроэнергии. Количество величин в наборе зависит от конфигурации нейронной сети и равно числу нейронов в входном слое плюс один. Эти обучающие наборы получены путём нормирования и разбиения на части временных рядов потребления электроэнергии, полученных информационной системой, при этом

число выборок равно длине ряда минус число входных нейронов сети минус один. Нормирование ряда производится путём деления всех его членов на значение наибольшего члена ряда.

Скрытый слой нейронных сетей производит вычисление параметров уравнения *ARMA* модели. Функция активации для нейронов данного слоя должна удовлетворять следующим условиям:

1. Наличие аналоговых входных и выходных сигналов;
2. Область значений $(-1,1)$, параметры уравнения могут быть отрицательными и по причине нормированности входных параметров не могут превышать по модулю единицу.

Исходя из первого требования к функции активации она должна относиться к классу сигмоидальных функций активации поскольку только данный класс обладает аналоговыми выходными сигналами вместо бинарных. При проверке на практике из использованных сигмоидальных функций активации наибольшую эффективность показала функция активации в виде гиперболического тангенса, которая и была применена в качестве функции активации для скрытого слоя нейронных сетей.

Выходной слой производит оценку полученного значения ошибки. Для оценки результатов работы алгоритма необходимо построить функцию оценки. В качестве метода оценки для нейронной сети построения модели потребления электроэнергии был выбран метод наименьших квадратов — сумма квадратов расстояний от выходных сигналов нейронной сети до их требуемых значений. Функция оценки для данной нейронной сети имеет вид (6).

$$H = (Z - Z')^2, \tag{6}$$

где:

H — значение функции.

Z — требуемое значение.

Z' — выходное значение нейронной сети.

Выбор пороговой величины выхода нейронной сети оказывает существенное влияние на её работу. Слишком малое значение пороговой величины, увеличит время обучения нейронной сети или сделает её обучение невозможной, в тоже время слишком высокое значение пороговой величины снизит качество её работы. Поскольку идеальное выходное значение созданной нейронной сети должно быть равно последнему значению в обучающем наборе, то в качестве пороговой величины по результатам практических испытаний было выбрано значение 0,025.

Так как нейронная сеть является однослойным персептроном, то для её обучения был выбран наиболее часто применяемый в подобном случае метод обратного распространения ошибки (модификация метода градиентного спуска), так как он обладает простотой реализации и имеет высокую скорость сходимости.

Оценка эффективности информационной системы

Выбор критериев для оценки эффективности работы, информационной системы, использующей для контроля достоверности прогнозирующую нейросетевую модель электропотребления был проведён на основе следующих задач:

1. Минимизация ошибок идентификации приборов учёта.
2. Минимизация затрат времени на идентификацию приборов учёта.

Поскольку обе задачи имеют равнозначную важность, то критерием оценки эффективности работы (7) созданной информационной системы и сравниваемых методов контроля достоверности является среднее суммы критериев эффективности решения этих задач.

$$E_I = 0,5 * E_T + 0,5 * E_K, \tag{7}$$

где:

E_I — общая оценка эффективности рассматриваемого метода контроля достоверности.

E_T — оценка эффективности метода контроля достоверности при решении задачи минимизации затрат времени.

E_K — оценка эффективности метода контроля достоверности при решении задачи минимизации ошибок идентификации приборов учёта.

Задачу минимизации ошибок идентификации приборов учёта можно разбить на две подзадачи:

1. Минимизировать число ошибок первого рода.
2. Минимизировать число ошибок второго рода.

В контексте рассматриваемой задачи ошибка первого рода — прибор учёта, который передаёт недостоверные данные, но был отмечен методом контроля достоверности как передающий достоверные данные. Ошибка второго рода — прибор учёта электроэнергии, передающий достоверные данные, который был ошибочно отмечен методом контроля достоверности как передающий недостоверные данные. Таким образом, критерий эффективности решения задачи минимизации ошибок идентификации приборов учёта (8) состоит из двух компонент, которые имеют одинаковую значимость и может быть определён показанным ниже образом.

Таблица 6. Данные для обучения нейросетевой модели

№ ПУ	Янв.	Фев.	Мар.	Апр.	Май	Июн.	Июл.	Авг.	Сеп.	Окт.	Ноя.	Дек.	Янв.
1	332	307	306	266	275	316	275	329	337	297	352	287	282
2	1319	1240	1053	965	870	854	544	433	448	482	1201	932	1029
3	139	130	124	135	130	119	165	218	147	197	166	121	120
4	236	274	34	118	131	125	218	195	205	114	134	129	139
5	88	89	85	87	112	140	95	97	98	76	81	82	236
6	148	119	118	138	123	126	173	159	108	238	70	117	88
7	190	185	169	190	187	257	569	662	333	256	201	186	173
8	707	805	755	638	688	638	360	283	273	340	533	824	882

Таблица 7. Тестовые данные

№ ПУ	Реальные данные	Отклонение -5%
1	359	341
2	1112	1057
3	133	126
4	130	123
5	274	260
6	88	84
7	195	185
8	750	712

$$E_K = 0,5 * \frac{F_2}{F_1} + 0,5 * \frac{T_2}{T_1} \quad (8)$$

где:

E_K — критерий эффективности решения задачи минимизации ошибок идентификации приборов учёта.

F_1 — количество приборов учёта электроэнергии в выборке, которые передают недостоверные данные.

F_2 — количество приборов учёта электроэнергии в выборке, передающих недостоверные данные, которые были обнаружены.

T_1 — количество приборов учёта электроэнергии в выборке, передающих достоверные данные.

T_2 — количество приборов учёта электроэнергии в выборке, передающих достоверные данные, которые были отмечены правильно.

В качестве критерия оценки эффективности решения второй задачи (9) было выбрано отношение времени, затрачиваемого на проверку приборов учёта, отмеченных методом контроля достоверности, как передающие недостоверные данные, к времени, затрачиваемому на проверку всех приборов учёта объекта контроля АСКУЭ с учётом эффективности их поиска. Если взять некоторое среднее время проверки одного прибора учёта специалистом, то вместо переменных времени в критерии (9) возможно использовать переменные числа приборов учёта.

$$E_T = \frac{N_1 - N_2}{N_1} * E_f, \quad (9)$$

где:

N_1 — число всех приборов учёта электроэнергии.

N_2 — число приборов учёта, которые необходимо проверить;

E_f — процент найденных приборов учёта, передающих недостоверные данные.

Для сравнения оценки эффективности решения поставленных задач созданной информационной системой, в сравнении с другими методами, были взяты используемые в настоящее время в АСКУЭ, для решения аналогичных задач, методы «полная проверка всех приборов учёта специалистом» [1] — М1 и «увеличение частоты опроса приборов учёта» [1] — М2. Для проведения оценки были использованы данные электропотребления из таблицы 6 и таблицы 7.

Согласно критерия (8), была произведена сравнительная оценка информационной системы с методами М1 и М2. Результаты представлены в таблицах 8 и 9.

После проведения анализа данных из таблицы 8 и таблицы 9 были сделаны выводы о том, что созданная информационная система имеет, эффективность 75% на достоверных данных, эффективность 87,5% на недостоверных данных и среднюю эффективность 81,25%.

Ошибка в определении достоверности данных происходит, в случае если временной ряд электропотре-

Таблица 8. Идентификация ПУ, передающих достоверные данные, созданной информационной системой

№	Значение	Прогноз	Отклонение	Вывод ИС
1	1	0,98	1,91%	Достоверные
2	1	1	0,08%	Достоверные
3	1	0,99	0,68%	Достоверные
4	0,94	0,67	40,55%	Недостоверные
5	1	1	0%	Достоверные
6	0,75	0,85	13,03%	Недостоверные
7	1	0,99	0,75%	Достоверные
8	0,85	0,84	1,36%	Достоверные

Таблица 9. Оценка качества идентификации ПУ, передающих недостоверные данные с отклонением –5%, созданной информационной системой

№	Значение	Прогноз	Отклонение	Вывод ИС
1	1	0,97	3,28%	Недостоверные
2	1	0,87	15,52%	Недостоверные
3	1	0,82	21,31%	Недостоверные
4	0,88	0,67	31,64%	Недостоверные
5	1	1	0%	Достоверные
6	0,72	0,81	12,50%	Недостоверные
7	0,99	0,95	4,29%	Недостоверные
8	0,81	0,89	10,24%	Недостоверные

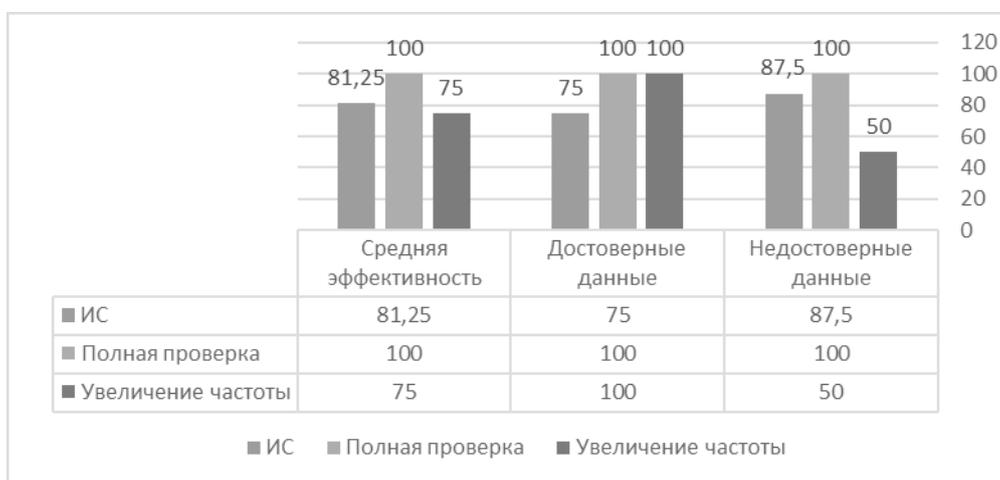


Рис. 4. Сравнительная диаграмма эффективности согласно критерию № 1

бления является нестационарным. Поскольку метод «полная проверка всех приборов учёта специалистом» заключается в проверке всех рассматриваемых приборов учёта экспертом человеком, то его эффективность согласно критерию (8) приближается к 100%.

Для информационной системы и метода М1, не имеет значения, по какой причине возникла недостовер-

ность данных, однако для метода М2 причина возникновения недостоверности данных существенным образом влияет на его эффективность. Проблема заключается в том, что данный метод рассчитан на поиск несанкционированных вмешательств в работу прибора учёта, которые не были замечены АСКУЭ. Поэтому согласно критерию (8) эффективность решения данным методом задачи минимизации ошибок идентификации

Таблица 10. Эффективность информационной системы по критерию (9)

№ Испытания	Число ПУ	Число ПУ с НД	Число ПУ, которые необходимо проверить	Затраченное время	Эффективность
1	10	1	3	3	61,25%
2	20	2	6	6	61,25%
3	30	4	11	11	55,39%
4	40	5	13	13	59,06%
5	50	5	15	15	61,25%
6	60	6	18	18	61,25%
7	70	7	21	21	61,25%
8	80	9	25	25	60,16%
9	90	10	28	28	60,27%
10	100	11	32	32	59,5%

Таблица 11. Эффективность метода М1 по критерию (9)

№ Испытания	Число ПУ	Число ПУ с НД	Число ПУ, которые необходимо проверить	Затраченное время	Эффективность
1	10	1	10	10	0%
2	20	2	20	20	0%
3	30	4	30	30	0%
4	40	5	40	40	0%
5	50	5	50	50	0%
6	60	6	60	60	0%
7	70	7	70	70	0%
8	80	9	80	80	0%
9	90	10	90	90	0%
10	100	11	100	100	0%

Таблица 12. Эффективность метода М2 по критерию (9)

№ Испытания	Число ПУ	Число ПУ с НД	Число ПУ, которые необходимо проверить	Затраченное время	Эффективность
1	10	1	1	1	45%
2	20	2	1	1	47,5%
3	30	4	2	2	46,65%
4	40	5	3	3	46,25%
5	50	5	3	3	47%
6	60	6	3	3	47,5%
7	70	7	4	4	47,15%
8	80	9	4	4	27,5%
9	90	10	5	5	47,2%
10	100	11	6	6	47%

Таблица 13. оценка эффективности по критерию (7)

Название	Критерий № 1	Критерий № 2	Общая эффективность
ИС	60,02%	81,25%	70,64%
«Полная проверка»	0%	100%	50%
«Увеличение частоты опроса»	47,5%	75%	61,25%



Рис. 5. Сравнительная диаграмма эффективности согласно критерию (9)

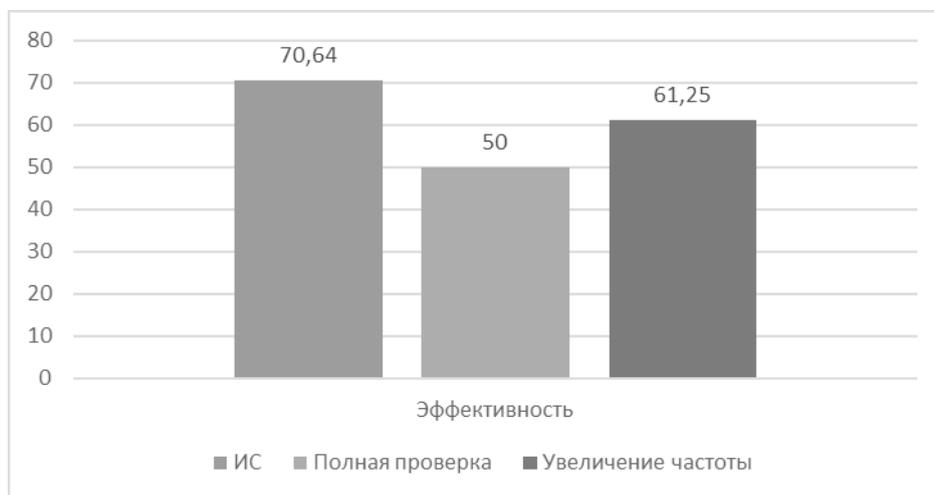


Рис. 6. Сравнительная диаграмма эффективности согласно критерию (7)

приборов учёта на данных из таблицы 6 и таблицы 7 равна 100% при идентификации достоверных данных, 50% при идентификации недостоверных данных. Средняя эффективность идентификации данных равна 75%.

На рис 4 приведена сравнительная диаграмма эффективности информационной системой и двух других методов согласно критерию (8).

Согласно критерию (9), оценки эффективности решения задачи минимизации затрат времени была произведена сравнительная оценка информационной системы с методами М1 и М2. Для проведения оценки были взяты выборки приборов учёта электроэнергии размером от десяти до ста единиц с известным числом приборов учёта, передающих недостоверные данные. За единицу времени была принята величина равная

средней продолжительности проверки одного прибора учёта. Результаты проведённых испытаний представлены в таблицах 10–12.

Информационная система выделяет среди множества всех исследуемых приборов учёта подмножество тех, которые в соответствии с используемой ею методикой отмечаются как передающие недостоверную информацию. Поэтому человеку эксперту необходимо проверить только отмеченные информационной системой приборы учёта. С учётом того что информационная система в среднем обнаруживает 87,5% приборов учёта, передающих недостоверные данные, это, для достаточно крупных выборок, даёт эффективность в границах от 55% до 62%, что отражено в таблице 10. Средняя же эффективность информационной системы равна 60,02%.

При использовании метода «полная проверка всех приборов учёта» [1] число приборов учёта, которые необходимо проверить человеку эксперту равно числу приборов учёта в выборке, что при любых выборках, несмотря на нахождение 100% приборов учёта, передающих недостоверные данные, даёт эффективность равную 0%, что отражено в таблице 11.

При использовании метода «увеличение частоты опроса приборов учёта» [1], эксперт человек, как

и в случае с использованием ИС должен проверить только отмеченные приборы учёта. Общая эффективность для большинства выборок находится в границах от 45% до 47,5%. Средняя же эффективность метода «увеличение частоты опроса приборов учёта» [1] согласно критерию (8) равна 46,87%.

После проведения сравнительной оценки эффективности созданной информационной системы и используемых в настоящее время в АСКУЭ методов контроля достоверности данных приборов учёта электроэнергии по критериям (8) и (9) была проведена их оценка эффективности согласно составного критерия (7). Результаты данной оценки приведены в таблице 13 и на рис. 6.

Как видно из таб. 13 и рис. 6 информационная система, использующая для контроля достоверности данных прогнозирующую нейросетевую модель электропотребления [10], имеет высокую эффективность при идентификации приборов учёта электроэнергии и эффективность выше среднего в плане минимизации затрат времени без потери качества поиска приборов учёта, передающих недостоверные данные. Её общая эффективность согласно критерию (7) равна 70,64% и превосходит по данному показателю два других метода, которые используются в настоящее время в АСКУЭ.

ЛИТЕРАТУРА

1. Тымчук А.И. К вопросу контроля достоверности показаний приборов учёта в автоматизированных системах коммерческого учёта // Наука и бизнес: пути развития. — 2018. — № 12(90). — С. 57–60.
2. Предварительный национальный стандарт Российской Федерации «Автоматизированные информационно-измерительные системы коммерческого учёта электроэнергии // Электронный фонд правовых и нормативно-технических документов — [Электронный ресурс]. URL: <https://docs.cntd.ru/document/1200141436>. (дата обращения 02.11.2022).
3. Алшимбеков А. Разработка проекта автоматизированной системы контроля и учета параметров электроснабжения поселка «Озерный» // Современные научные исследования и разработки. — 2018. — № 11(28) — С. 75–81.
4. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. Time Series Analysis, Forecasting and Control (4th ed.) // Hoboken, NJ: Wiley. (2008) ISBN9780470272848
5. Кобзарь А.И. Прикладная математическая статистика // М.: Физматлит, 2006.
6. Таблица критических значений t-критерия Стьюдента: [Электронный ресурс] // Statpsy.ru. URL: <https://statpsy.ru/t-student/t-test-tablica/> (дата обращения 10.11.2022).
7. Вакуленко Е.С. Анализ одномерных временных рядов: [Электронный ресурс] // URL: <https://pokrovka11.files.wordpress.com/2014/01/time-series-analysis-part1.pdf>. (дата обращения 02.11.2022).
8. Dickey D.A. and Fuller W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root // Journal of the American Statistical Association. — 74. — 1979.
9. Золоторёв А.А. Эконометрика // Электронное учебное пособие Института электронного обучения Санкт-Петербургского университета технологий управления и экономики. 2009 г. — [Электронный ресурс]. URL: <https://e.lanbook.com/book/63858>. (дата обращения 02.11.2022).
10. Атрощенко В.А., Тымчук А.И. Построение прогнозирующей модели потребления АСКУЭ с использованием нейросетевых технологий. // IX Международный конгресс «Энергосбережение и энергоэффективность. IT-технологии. Энергобезопасность. Экология».

© Тымчук Алексей Игоревич (aleximchuck@gmail.com).

Журнал «Современная наука: актуальные проблемы теории и практики»