

КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ ЭЛЕКТРОПОТРЕБЛЕНИЯ НА ОСНОВЕ УСТОЙЧИВЫХ МОДЕЛЕЙ

АВТОР:

А.В. КЫЧКИН, К.Т.Н.
НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ
ВЫСШАЯ ШКОЛА
ЭКОНОМИКИ
ФИЛИАЛ В Г. ПЕРМЬ
(НИУ ВШЭ - ПЕРМЬ)

Управление спросом на электроэнергию позволяет существенно снизить пиковые нагрузки в энергосистеме. Кроме того, адекватное управление спросом инициирует снижение энергопотребления ко-

нечными потребителями, которым становится экономически выгодно добровольно изменять график энергопотребления по результатам рынка «на сутки вперед» без дополнительных указаний от системного оператора.

Ключевые слова: машинное обучение; управление спросом на электроэнергию; электропотребление; краткосрочное прогнозирование; устойчивые модели; авторегрессионные модели; модель Хольта — Винтерса; сезонные модели.

ВВЕДЕНИЕ

Сегодня большинство участников рынка электроэнергии используют все возможности для повышения эффективности своей деятельности. Вовлечение компаний в процессы управления спросом на электроэнергию могло бы существенно снизить затраты на энергию и себестоимость производства продукции или совершения полезной работы. Мировая практика показывает высокую активность участия потребителей при управлении спросом на электроэнергию на сутки вперед, при котором компании получают существенное снижение тарифа [5, 24]. В этом случае эффективность оптимизации в значительной степени зависит от точности прогнозирования про-

филя нагрузки. Как правило, такой прогноз основывается на предыдущих измерениях энергопотребления, регистрируемых через регулярные промежутки времени, как показано на рис. 1. Вид и горизонт прогнозов могут варьировать [16] в диапазоне от 1 до 24 ч вперед. Методы построения таких прогнозов называются методами краткосрочного прогнозирования нагрузки (Short Term Load Forecasting, STLF) [2].

Известно, что прогнозы для больших агрегированных групп потребителей могут быть построены с более высокой точностью, поскольку содержат в себе значительную долю повторяющихся и накладывающихся друг на друга событий. Однако в случае краткосрочного прогнозирования нагрузки для отдельных потребителей

точность прогноза значительно снижается [12]. Это может быть связано с быстрыми изменениями в работе оборудования или поведении пользователей. Кроме того, прогнозы часто строятся с высокой степенью детализации (например, 15-минутные интервалы), что создает дополнительные трудности в обеспечении высокой точности предсказания.

С учетом трудностей, связанных с построением прогнозов электрической нагрузки в краткосрочном периоде (на сутки вперед), предлагаются так называемые устойчивые модели. Будем считать, что устойчивая модель прогнозирования формирует адекватный и интерпретируемый профиль нагрузки для любых выбранных суток из исследуемого горизонта времени, при этом обеспечивает стабильную

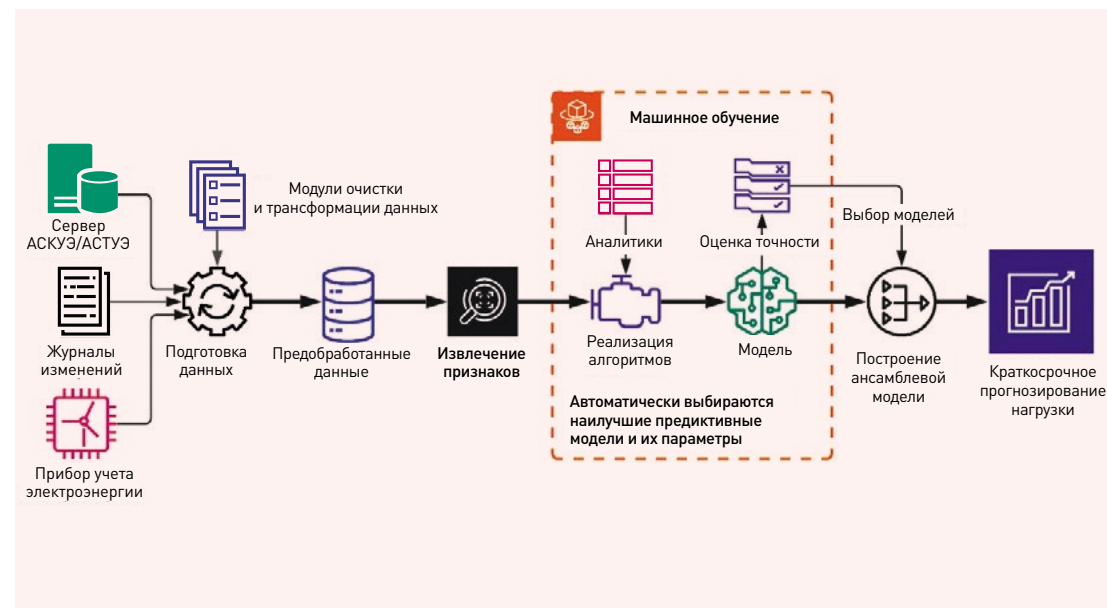


Схема краткосрочного прогнозирования энергопотребления на основе ансамблевых моделей машинного обучения

ПОТРЕБЛЕНИЕ ЭЛЕКТРОЭНЕРГИИ В ЗДАНИИ А (а) И ЗДАНИИ Б (б) В ТЕЧЕНИЕ ТРЕХ МЕСЯЦЕВ (С НОЯБРЯ 2016 ПО ФЕВРАЛЬ 2017 Г.)

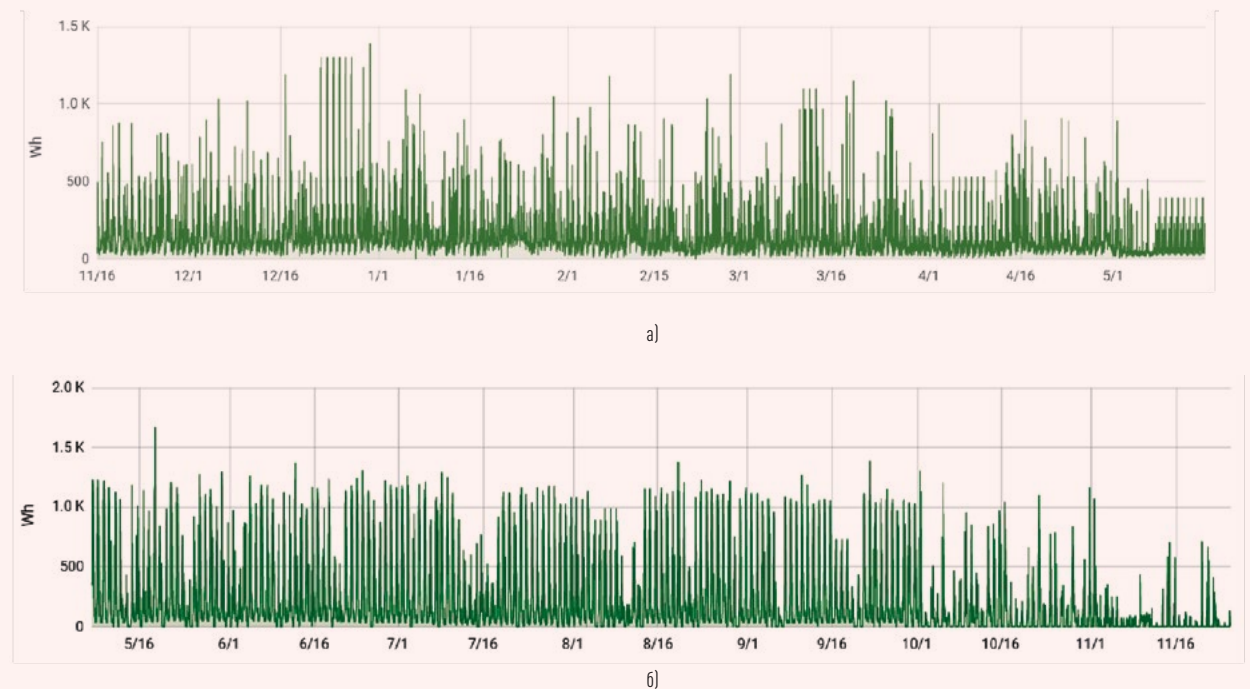


Рис. 1

ошибку, не превышающую ошибку стандартных методов прогнозирования и методов на основе базовых линий.

В статье предлагается семейство регрессионных моделей, реализуемых на основе методов машинного обучения, разработанных специально для прогнозирования электрической нагрузки в зданиях, проводится сравнительный анализ этих моделей со стандартными методами прогнозирования. Подобно экспертному подходу к обучению [4], целью исследования являются извлечение признаков из профилей нагрузок, интеграция этих характеристик в состав регрессионных моделей, а также определение их относительного влияния друг на друга при построении прогнозов. На основе реальных данных демонстрируется, что предлагаемые модели значительно превосходят стандартные модели и модели на основе базовых линий. В табл. 1 приведены используемые в статье обозначения.

АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ, ИСПОЛЬЗУЕМЫХ ПРИ КРАТКОСРОЧНОМ ПРОГНОЗИРОВАНИИ ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ

В литературе встречается множество известных методов прогнозирования, их можно разделить на три группы:

- 1) стандартные модели;
- 2) модели, специально разработанные для краткосрочного прогнозирования нагрузки;

ИСПОЛЬЗОВАННЫЕ ОБОЗНАЧЕНИЯ И АББРЕВИАТУРЫ

Обозначение/аббревиатура	Значение
d	Индекс дня
t	Индекс времени
$y_d(t)$	Нагрузка в период t в день d
$\hat{y}_d(t)$	Прогноз нагрузки в период t в день d
AR	Авторегрессионная модель
CLD	Устойчивая модель, учитывающая N дней одного и того же дня недели
HW	Модель Хольта — Винтерса
N-day	Устойчивая модель для N предыдущих дней (график базовой линии)
PAR	Устойчивая авторегрессионная модель
PM	Устойчивая модель
SPR	Сезонная устойчивая регрессионная модель
STLF	Краткосрочное прогнозирование нагрузки

Таблица 1

- 3) оптимальное сочетание моделей для краткосрочного прогнозирования нагрузки.

Первая группа может быть представлена стандартными методами усреднения и авторегрессионными моделями. Например, в работах С. Хабена и соавт. [13], А.В. Кычкина [17] обсуждаются простые модели, которые рассчитывают средние значения энергопотребления на сутки вперед на основе значений за предыдущие дни. Модели, основанные на авторегрессии (AR/ARMA/ARIMA/SARIMA) [12, 10], а также модели, основанные на экспоненциальном сглаживании, например метод Хольта — Винтерса [2], широко использовались для прогнозирования электрической нагрузки [1]. Кроме того, SARIMA и метод Хольта — Винтерса более эффективны, если данные имеют сезонность.

Вторая группа включает модели краткосрочного прогнозирования, специально разработанные для прогно-

зирования электрической нагрузки. В работе Дж.Р. Кансело и соавт. [3] показана реализация краткосрочного прогнозирования с учетом экзогенных данных, в частности, приведено описание влияния данных о погоде на прогноз. Регрессионный анализ как инструмент для определения связи между переменными активно используется для краткосрочного прогнозирования нагрузки [15]. Набор подходящих признаков может быть извлечен из временного ряда энергоданных [9, 8]. В упомянутых работах также описываются задача выбора атрибутов данных временного ряда и алгоритм извлечения признаков. В своей работе А.В. Кычкин и Г.П. Микрюков [18] предлагают метод, основанный на линейных регрессионных моделях, для анализа данных по потреблению энергии в зданиях с многосекционными системами освещения и климат-контролем/кондиционированием. Искусственные нейронные сети также часто используются для

краткосрочного прогнозирования нагрузки [7, 12, 14]. Более того, можно использовать нечеткую логику и модели, основанные на знаниях [2].

Третья группа методов прогнозирования оптимально сочетает в себе несколько других базовых моделей. Например, в работе Х. Чена и соавт. [6] представлена методология Wavelet-ARMAX-Winters, которая включает в себя три стратегии моделирования: ARMAX-модели, тригонометрические регрессии, чувствительные к сезонности, и модель Хольта — Винтерса. Р. Йе и соавт. [25] предлагают гибридный алгоритм прогнозирования временных рядов, основанный на трансферном обучении, а именно online sequential extreme learning machine, сочетающий в себе методы ансамблевого обучения. При использовании таких комбинаций важно адаптивно обновлять значимость (или вес) базовых моделей. В работе Л.Дж. Соареса и М.С. Медейроса [22] представлена двухуровневая сезонная авторегрессионная модель (two-level seasonal autoregressive model, TLSAR), которая объединяет расчеты потенциальной и нерегулярной нагрузки, и модель dummy-adjusted SARIMA, которая модифицирует стандартную модель SARIMA с помощью фиктивных переменных, отражающих тип дня. Предсказания, полученные этими методами, лучше по сравнению с моделями, основанными на нейронных сетях, но они были протестированы только для агрегированных энергоданных. Наконец, в работе Х. Хипперта и соавт. [14] описаны комбинации из наивных, сглаживающих и регрессионных моделей, приводится их сравнение с нейронными сетями, имеющими большое количество параметров, демонстрируется их преимущество на тестовом наборе данных.

Таким образом, методы первой группы направлены на сглаживание данных, из-за чего они не могут

предсказать нагрузку с нерегулярными пиками. Методы второй группы помогают выявить скрытые закономерности в суточном энергопотреблении и могут определить пики или просадки нагрузки. Некоторые из этих моделей также могут идентифицировать нелинейные зависимости в энергоданных (как нейронные сети), но обучение таких моделей обычно занимает больше времени и требует больших наборов данных по сравнению с линейной регрессией. Наконец, модели третьей группы имеют наибольший потенциал, комбинируя базовые модели и подбирая их относительную значимость (вес). Однако построение ансамблевых моделей обычно увеличивает вычислительную сложность.

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОСТРОЕНИЯ УСТОЙЧИВЫХ ПРОГНОЗНЫХ МОДЕЛЕЙ

Одна из самых больших трудностей при краткосрочном прогнозировании электрической нагрузки заключается в непредсказуемом поведении наблюдаемых объектов, на которые оказывают влияние различные внешние факторы, в том числе действия пользователей. Подобная неопределенность возникает даже на относительно коротком горизонте прогнозирования, составляющем один час. Вариабельность энергопотребления обоснована многочисленными периодическими (повторяемыми) и случайными по времени событиями использования электрооборудования. Тем не менее, каким бы случайным ни было поведение пользователей или влияние факторов на режимы работы оборудования, существуют

некие зависимости, которые, как правило, позволяют оценить нагрузку с большой точностью. В данной работе мы формулируем признаки, характерные для электрической нагрузки в зданиях, чтобы свести к минимуму связанную с этим неопределенность. Использование методов машинного обучения в этом случае является критерием оптимизации при комбинировании прогнозов различных моделей, подобно экспертному обучению [4]. Такой подход дополнительно снижает неопределенность, вызванную непрерывностью некоторых моделей непрерывно обеспечивать стабильность прогнозов (например, в связи с изменениями в условиях эксплуатации). Кроме того, он позволяет определить относительную значимость моделей и используемых ими признаков, что важно для интерпретации полученных результатов.

ФОРМАЛИЗАЦИЯ ЗАДАЧИ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ

Предположим, были выполнены замеры энергопотребления здания за время $(d - 1) > 0$ дней, при этом замеры выполнялись с периодом в 15 мин. На рис. 1 приведены два примера таких профилей нагрузки за период времени, равный трем месяцам. В результате за одни сутки накапливается 96 последовательных замеров, и каждый замер в момент времени t представляет собой общее значение потребления электрической энергии за последние 15 мин. Используя доступные измерения за все предыдущие $(d - 1)$ дни, предлагается предсказать нагрузку на следующий день d (прогноз на сутки вперед) со степенью детализации, составляющей 15 мин.

МОДЕЛИ НА ОСНОВЕ БАЗОВЫХ ЛИНИЙ (BASELINE)

Устойчивые методы обычно основаны на моделях базовых линий (нагрузок)

и могут быть использованы для сравнения с другими методами. Во многих случаях полезно знать, может ли разработанная модель прогнозирования дать лучшие прогнозы, чем Baseline-метод. Такие модели считаются самыми очевидными, поскольку они основаны на предположении, что «все остается по-прежнему», т. е. прогноз всегда определяется несколькими последними точками данных.

По мнению Г. Ноттона и К. Войанта [20], устойчивая модель на основе базовых линий подразумевает, что электрическая нагрузка в момент времени $(t + 1)$ равна нагрузке в момент времени t . Однако как именно должны быть определены моменты времени $(t + 1)$ и t ? С учетом того, что прогноз выполняется на сутки вперед с детализацией в 15 мин, устойчивая модель, которая предполагает, что нагрузка остается постоянной в течение следующего дня и равна текущей нагрузке, вероятнее всего, будет давать ошибку. Устойчивая модель была бы более точной, если бы предполагала, что электрическая нагрузка в момент времени t дня d (кратко (t, d)) совпала бы с соответствующей нагрузкой в то же время t в предыдущий день $(d - 1)$ или в предыдущий день неделю назад $(d - 7)$. Дополнительная вариация такой модели могла бы учитывать и более чем один предыдущий день (например, среднее потребление в одно и то же время в N предыдущих дней).

Поэтому в работе рассматривается устойчивая модель, использующая N последних дней одного и того же дня недели. Например, если необходимо спрогнозировать нагрузку на следующий день, а этот день понедельник, тогда прогноз будет равен средней нагрузке в одно и то же время за предыдущие N понедельников. Для обозначения этих моделей используется аббревиатура CLD (англ. copy-last-day).

Формальное определение будет следующим. Пусть $y_d(t)$ — электрическая нагрузка в момент времени t в день d . Устойчивая модель (PM), учитывающая один предыдущий день, предполагает, что:

$$\hat{y}_d^{PM}(t) = y_{d-1}(t).$$

Аналогично можно определить устойчивую модель, учитывающую N дней, следующим образом:

$$\hat{y}_d^{PM}(t) = \frac{1}{N} \sum_{i=d-N}^{d-1} y_i(t).$$

Другими словами, устойчивая модель, учитывающая N дней, использует в качестве прогноза среднюю нагрузку за N предыдущих дней в то же самое время.

Поскольку электрическая нагрузка сильно зависит от действий пользователей (т. е. от их расписания использования энергопотребителей), можно дополнительно улучшить устойчивую модель, учитывающую N дней, рассмотрев только N предыдущих дней одного и того же дня недели. Так, если день d соответствует понедельнику, то для прогноза на время t нужно определить среднюю нагрузку в то же время за предыдущие N понедельников. Эта модель обозначена как устойчивая модель, учитывающая N дней одного и того же дня недели. Согласно этой модели, прогноз определяется следующим образом:

$$\hat{y}_d^{PM}(t) = \frac{1}{N} \sum_{i=d-7N}^{d-7} y_i(t).$$

АВТОРЕГРЕССИОННЫЕ МОДЕЛИ

Как было показано в предыдущем подразделе, устойчивые модели пытаются передать временные зависимости, частота которых составляет несколько дней или недель (например, в связи со сходством в расписании использования энергопотребителей пользователями

в одни и те же дни). Такие зависимости обозначены как низкочастотные временные зависимости.

Тем не менее временные зависимости могут также существовать и в пределах одного и того же дня. Вполне вероятно, что нагрузка в момент времени t зависит от нагрузки в предыдущий момент времени $(t - 1)$ того же дня. Такие временные зависимости негибкой нагрузки обозначены как высокочастотные временные зависимости.

Авторегрессионные модели могут быть использованы для учета высокочастотных временных зависимостей нагрузки в пределах одного и того же календарного дня. Возможно, самой простой такой моделью является авторегрессионная модель (AR-модель), согласно которой предсказание нагрузки в момент времени t определяется линейной комбинацией нагрузки за предыдущие моменты времени. Так, имеется:

$$\hat{y}_d^{AR}(t) = a_1 \times y_d(t - 1) + \dots$$

$$\dots + a_n \times y_d(t - n)$$

(что является результатом оценки апостериорного максимума белого случайного процесса (белого шума) [19]).

В случае когда необходимо создать прогноз на сутки вперед, можно реализовать вариацию этой модели, обычно называемую псевдорегрессионной, которая может быть представлена следующим образом:

$$\hat{y}_d^{AR}(t) = a_1 \times \hat{y}_d^{AR}(t - 1) + \dots$$

$$\dots + a_n \times \hat{y}_d^{AR}(t - n).$$

Другими словами, если нагрузка в момент времени $(t - j)$, $j = 1 \dots n$ не известна, она заменяется на доступное предсказание нагрузки в этот момент времени.

ПРИЗНАКИ, ПОЛУЧЕННЫЕ ИЗ ЭНЕРГОДАНЫХ, ДЛЯ РАЗНЫХ ДНЕЙ

Обозначение	День, для которого строится прогноз	Предыдущий день	День одну неделю назад
L	–	w_0	w_1
rs	–	w_2	w_3
d	w_4	–	–
Lh	–	w_5	w_6
Ld	–	w_7	w_8
DLh	–	w_9	w_{10}
LC	–	w_{11}	w_{12}
PC	–	w_{13}	w_{14}

Таблица 2

Следует отметить, что можно определить альтернативные модели, основанные на авторегрессии. Например, часто используются модели скользящего среднего, которые учитывают белый шум и пытаются улавливать влияние низкочастотных искажений на профиль нагрузки.

Одним из главных ограничений таких методов является то, что они больше подходят для краткосрочных прогнозов в диапазоне нескольких часов. В действительности нетрудно заметить, что даже небольшие ошибки прогнозирования на один момент времени вперед могут разрастись до непредсказуемых размеров при формулировании долгосрочных прогнозов более чем на сутки вперед.

ТРЕХПАРАМЕТРИЧЕСКОЕ ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ (МЕТОД ХОЛЬТА — ВИНТЕРСА)

В отличие от предыдущих моделей модель Хольта — Винтерса предназначена для выявления не только

тренда, но и сезонности. Сезонная составляющая модели объясняет повторяющиеся колебания и описывается длиной сезона — периодом, после которого начинается повторение колебаний [23]. Для каждого наблюдения в течение сезона формируется своя собственная сезонная составляющая. Для описанных ранее данных продолжительность сезона установлена равной 96×7 (96 формирует ежедневную сезонность, 7 соответствует еженедельной сезонности), в результате получают 672 сезонных компонента, по одному на каждый 15-минутный интервал дня недели.

В частности, оценка будущего значения y в момент времени t в соответствии с моделью Хольта — Винтерса формируется следующим образом:

$$\hat{y}_d^{HW}(t) = L(t - k) + kP(t - k) + S(t - T),$$

где $L(t)$ — компонент уровня, который задается по формуле:

$$L(t) = \alpha(y(t) - S(t - T)) +$$

$$+ (1 - \alpha)(L(t - 1) + P(t - 1));$$

$P(t)$ — компонент тренда, который задается по формуле:

$$P(t) = \beta(L(t) - L(t - 1)) +$$

$$(1 - \beta)P(t - 1);$$

$S(t)$ — сезонный компонент, который задается по формуле:

$$S(t) = \gamma(y(t) - L(t)) + (1 - \gamma)S(t - T),$$

где k — диапазон прогнозирования, $k = 96$;

$y(t)$ — фактическое значение электрической нагрузки в момент времени t ;

T — период временного ряда;

α — фактор сглаживания данных;

β — фактор сглаживания тренда;

γ — фактор сглаживания сезонных изменений. Также $\alpha, \beta, \gamma \in [0, 1]$. В случае последовательного прогнозирования $y(t)$ заменяется соответствующим прогнозом в то же самое время.

В указанной выше модели компонент уровня отражает уровень базовой нагрузки (или опорный уровень), а компонент тренда аппроксимирует с помощью фильтрации нижних частот, как компонент уровня изменяется со временем (в течение нескольких дней). Наконец, сезонный компонент фиксирует более низкочастотные зависимости (в течение более длительных периодов времени). Таким образом, в целом модель Хольта — Винтерса можно рассматривать как комбинацию высоко- и низкочастотных временных зависимостей.

Для обучения этой модели в качестве функции потерь может быть использована среднеквадратическая ошибка (RMSE). Эта функция показывает качество обобщения моделью зависимостей, выявленных на обучающем наборе данных. Затем с помощью кросс-валидации должна быть выполнена оценка значения функции потерь на основе параметров α, β, γ , после чего параметры будут изменены в соответствии

с градиентом. Поскольку существует ограничение на значения параметров сглаживания, которые должны оставаться в пределах $[0, 1]$, для обновления параметров может использоваться метод сопряженных градиентов Ньютона.

СЕЗОННАЯ МОДЕЛЬ АВТОРЕГРЕССИИ — СКОЛЬЗЯЩЕГО СРЕДНЕГО

Модель ARIMA (интегрированная модель авторегрессии — скользящего среднего) является одним из наиболее распространенных методов анализа и прогнозирования временных рядов. Это расширение моделей ARMA для нестационарных временных рядов, которые можно сделать стационарными, перейдя от исходного временного ряда к разности определенного порядка.

В модели ARIMA используются параметры p, d, q , выраженные целыми числами. Эти три параметра вместе учитывают сезонность, тренд и шум в наборах данных. В частности:

- p — это порядок авторегрессии, который позволяет учитывать предыдущие значения временного ряда;
- d — это порядок интегрированности, позволяющий включить предыдущие разности временного ряда;
- q — это порядок скользящего среднего, который позволяет задать ошибку модели как линейную комбинацию ранее наблюдаемых значений ошибки.

Основным ограничением данной модели является то, что она

не учитывает сезонность временных рядов, что делает невозможным ее использование для прогнозирования временных рядов энергопотребления, для которых характерна сильная сезонность, например сезонность продолжительностью в один день или одну неделю.

Вариация модели ARIMA — модель SARIMA может быть использована для отслеживания сезонного компонента временного ряда.

В этой модели параметры p, d, q рассматриваются как несезонные параметры и остаются такими же, как и в предыдущей модели. Дополнительно к этим параметрам вводятся параметры P, D, Q , определение которых аналогично определению параметров p, d, q , но параметры P, D, Q применяются к сезонной составляющей временного ряда. Кроме того, добавляется параметр S , описывающий длину сезона временного ряда (96 , если сезон соответствует одному дню, 7×96 , если сезон соответствует одной неделе, и т. д., при этом 96 — это количество точек, т. е. замеров датчика, снятых за один день).

Подобно модели Хольта — Винтерса, модель SARIMA — это универсальная модель, учитывающая сезонность ряда, и в исследовании она используется для оценки работы производных авторегрессионных моделей.

Выбор параметров p, d, q и $(P, D, Q)S$ основан на рекомендациях, представленных в работе М. Акпинара и Н. Юмусака [1], которые включают следующие ограничения: $d + D \leq 2, P + Q \leq 2$. Для расчета наиболее подходящих параметров должен выполняться перебор всех вариантов параметров модели, а выбор наилучшего варианта при этом может быть выполнен в результате сравнения

параметров по информационному критерию Акаике (AIC). В данной работе для модели SARIMA были экспериментально установлены следующие параметры: $1, 1, 1$ и $(1, 1, 1)96$.

УСТОЙЧИВЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ

В этом разделе представлены два класса устойчивых моделей. Модели первого класса — устойчивые авторегрессионные модели (PAR) предназначены для комбинирования нескольких моделей. Модели второго класса — сезонные устойчивые регрессионные модели (SPR) нацелены на комбинирование большого количества устойчивых параметров, что позволяет лучше учитывать сезонную составляющую ряда. В следующих подразделах представлено подробное описание таких моделей.

Все модели в этом разделе были обучены путем постановки задачи оптимизации линейной регрессии, которая была итеративно решена с использованием рекурсивного метода наименьших квадратов RLS [21].

УСТОЙЧИВАЯ АВТОРЕГРЕССИОННАЯ МОДЕЛЬ

Как было сказано ранее, устойчивые модели могут выявлять низкочастотные временные зависимости в профиле нагрузки (зависимости, проявляющиеся в течение нескольких дней или недель), в то время как авторегрессионные модели могут улавливать высокочастотные временные зависимости (проявляющиеся в течение одного календарного дня). Кроме того, авторегрессионные модели

работают хорошо только при краткосрочном прогнозировании. Поэтому в данном подразделе рассматривается возможность оптимального сочетания этих двух типов моделей. Такой подход хорошо согласуется с экспертными методами прогнозирования [4] и методами трансферного обучения [11].

При подобном сочетании двух методов прогнозирования итоговый прогноз будет представлен в виде:

$$\hat{y}_d^{PAR}(t|a_1, \dots, a_n, b_0) = a_1 \hat{y}_d^{AR}(t-1) + \dots + a_n \hat{y}_d^{AR}(t-n) + b_0 \hat{y}_d^{PM}(t).$$

В этом случае необходимо выполнить расчет нового набора весов $a_1, a_2, \dots, a_n, b_0$, соответствующего оптимальной комбинации высокочастотных временных зависимостей (выявленных авторегрессионной моделью) и низкочастотных временных или сезонных зависимостей (выявленных устойчивой моделью).

СЕЗОННАЯ УСТОЙЧИВАЯ РЕГРЕССИОННАЯ МОДЕЛЬ

Стандартные устойчивые модели формируют прогнозы путем усреднения нагрузки в течение тех же

СРАВНЕНИЕ ПРОГНОЗА НАИВНЫХ УСТОЙЧИВЫХ МОДЕЛЕЙ С ФАКТИЧЕСКИМИ ДАННЫМИ ПО ЭНЕРГОПОТРЕБЛЕНИЮ В ЗДАНИИ А

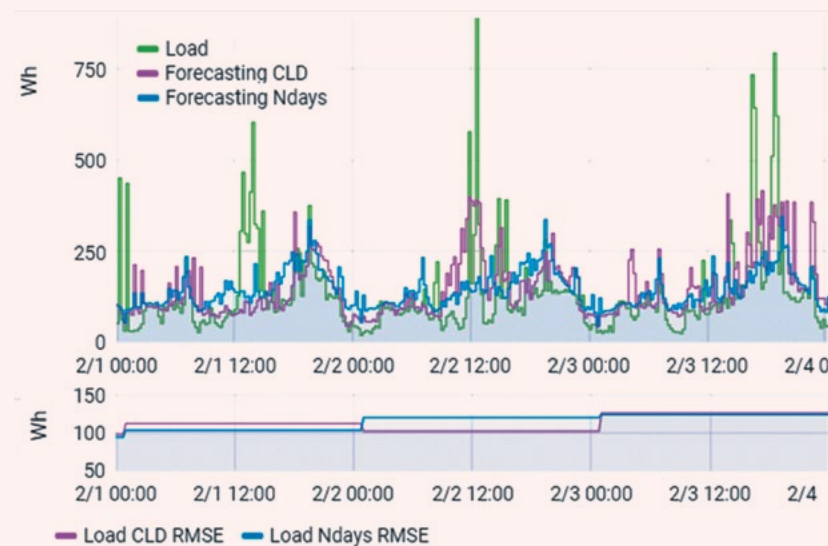


Рис. 2

СРАВНЕНИЕ ПРОГНОЗА, ПОЛУЧЕННОГО ПО МОДЕЛИ ТРОЙНОГО ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ (МЕТОД ХОЛЬТА — ВИНТЕРСА), С ФАКТИЧЕСКИМИ ДАННЫМИ ПО ЭНЕРГОПОТРЕБЛЕНИЮ В ЗДАНИИ А

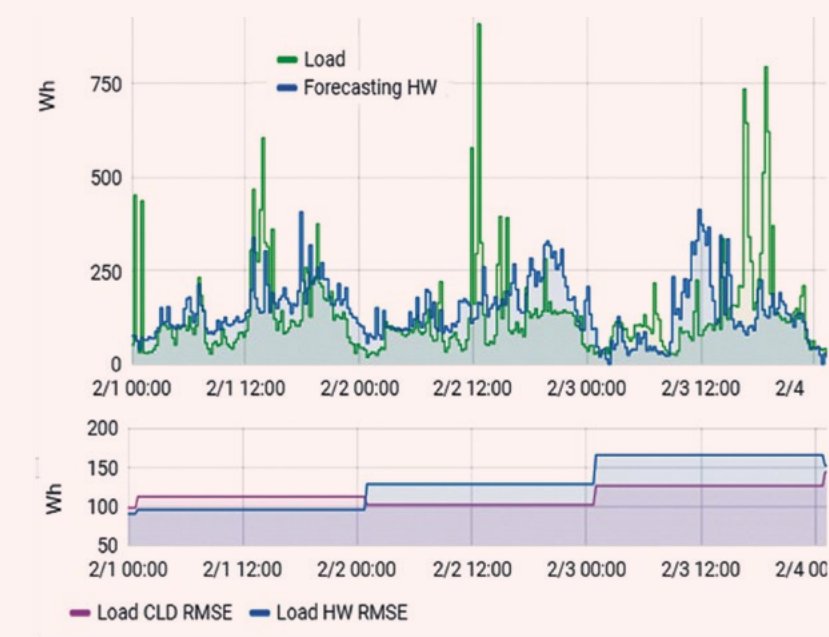


Рис. 3

ОЦЕНКА МОДЕЛЕЙ, ОБУЧЕННЫХ НА ДАННЫХ ПО ЭНЕРГОПОТРЕБЛЕНИЮ ЗДАНИЯ А, С ПОМОЩЬЮ МЕТРИКИ RMSE

Модель	Один день 03.02.2017	Одна неделя 01.02.2017– 07.02.2017	Один месяц 17.11.2016– 16.12.2016	Полгода 16.11.2016– 15.05.2017
CLD	124	122	116	126
N-day	123	118	107	115
HW	164	131	130	135
SARIMA	127	118	112	119
PAR	126	118	107	114
SPR	115	107	102	100

Таблица 3

ОЦЕНКА МОДЕЛЕЙ, ОБУЧЕННЫХ НА ДАННЫХ ПО ЭНЕРГОПОТРЕБЛЕНИЮ ЗДАНИЯ В, С ПОМОЩЬЮ МЕТРИКИ RMSE

Модель	Один день 22.10.2016	Одна неделя 17.10.2016– 23.10.2016	Один месяц 01.10.2016– 31.10.2016	Полгода 16.10.2016– 15.04.2017
CLD	153	163	189	128
N-day	153	188	205	120
HW	380	289	284	138
SARIMA	211	141	163	117
PAR	174	121	159	110
SPR	103	65	82	69

Таблица 4

временных интервалов за предыдущие сутки. Основное предположение заключается в том, что пользователи в плане эксплуатации энерг потребителей в зданиях ведут себя почти одинаково в одно и то же время каждого дня. Однако когда прогнозы запрашиваются с детализацией до 15-минутных интервалов, даже небольшие изменения в расписании пользователей могут оказать существенное влияние на точность прог-

нозов. Так, даже получасовая разница в выполнении утреннего сценария приведет к большим прогнозным ошибкам на следующий день.

По этой причине необходимо создать модель, более устойчивую к малейшим изменениям в расписании пользователей. Например, вместо того чтобы основывать прогноз на нагрузке в тот же момент времени в предыдущие сутки, можно

использовать усредненное значение нагрузки в увеличенном окне времени в предыдущий день (продолжительность окна может составить, например, один час). Таким образом, прогноз будет более устойчив, даже если расписание пользователей будет выполняться с отклонениями. Кроме того, общее энергопотребление может быть еще одним устойчивым фактором, который может сократить неопределенность прогнозов. Например, несмотря на то что расписание пользователей может немного различаться в разные дни, обычно люди используют примерно одинаковое количество энергии каждый день.

С этой целью были введены признаки профилей нагрузок, предназначенные для снижения уровня неопределенности, связанной с небольшими отклонениями в расписании пользователей. Так, рассматриваемый набор состоит из следующих факторов:

- L — электрическая нагрузка;
- R_s — скользящая сумма электрической нагрузки в течение временного окна, равного одному часу (четыре 15-минутных интервала);
- d — индикатор типа дня — рабочий или выходной (целочисленная переменная);
- L_h — общее потребление энергии в течение временного окна, равного одному часу (четыре 15-минутных интервала);
- L_d — доля, которую составляет текущее потребление энергии за 15-минутный интервал, по отношению к среднему общему потреблению энергии за один день;
- DL_h — разница в часовом потреблении энергии за последние два часа;

- LC — индикатор низкого энергопотребления (бинарная переменная) показывает, что общая нагрузка составляет менее 20% среднего значения потребления нагрузки в течение одного дня;
- PC — индикатор высокого энергопотребления (бинарная переменная) показывает, что общая нагрузка составляет более 150% среднего значения потребления нагрузки в течение одного дня.

На основе этих устойчивых факторов была определена линейная регрессионная модель, признаки которой показаны в табл. 2.

ОЦЕНКА МОДЕЛЕЙ

В качестве метрики для оценки точности предсказания была выбрана среднеквадратическая ошибка (RMSE). В табл. 3 и 4 показаны оценки RMSE рассмотренных ранее моделей, полученные в результате обучения моделей на реальных данных нагрузок при энергопотреблении двух зданий (здание А и здание В).

На рис. 2 показано фактическое энергопотребление здания А по сравнению с прогнозными значениями, полученными с помощью устойчивой модели CLD, использующей N дней одного и того же дня недели, и модели N-days, использующей среднее значение нагрузки за 10 предыдущих последовательных дней.

Видно, что данные устойчивые модели часто переоценивали уровень электрической нагрузки. Кроме того, эти модели не смогли определить основные пики нагрузки в течение одного дня.

Модель Хольта — Винтерса оказалась очень чувствительной к сезонности, при этом точность была лучше, когда длина сезона составляла одну неде-

лю по сравнению с длиной сезона, равной одним суткам (рис. 3). Однако почти все прогнозы оказались завышенными, что, вероятно, было вызвано попыткой модели отследить тренд. Несмотря на изменение среднесуточного потребления энергии, в данном временном ряду тренда нет.

Прогноз с помощью модели SARIMA (рис. 4) привел к сглаживанию профиля нагрузки, в результате модель практически не предсказала пики нагрузки и случаи быстрого снижения энергопотребления в течение дня. Следует отметить, что SARIMA очень хорошо описала ночной режим энергопотребления, а также падение нагрузки в период с вечернего пика в 18:00 и приблизительно до 00:00.

На рис. 5 показаны прогнозы, полученные с помощью модели

PAR, сформированной в результате объединения устойчивой модели, использующей среднее значение нагрузки за трое суток в течение трех недель, и авторегрессионной модели. На рисунке прогнозы сравниваются с фактическими наблюдениями по энергопотреблению здания А. Видно, что комбинированная PAR-модель значительно улучшает прогностические качества отдельных устойчивых моделей. Намного сократилась вероятность прогнозирования ложных пиков, также модель хорошо повторяла некоторые закономерности в профиле нагрузки, например небольшое увеличение нагрузки утром с 6:00 до 8:00.

На рис. 6 показаны прогноз, полученный с помощью модели SPR, и его сравнение с фактическими наблюдениями здания А в этот день. Также

СРАВНЕНИЕ ПРОГНОЗА, ПОЛУЧЕННОГО С ПОМОЩЬЮ МОДЕЛИ SARIMA, С ФАКТИЧЕСКИМИ ДАННЫМИ ПО ЭНЕРГОПОТРЕБЛЕНИЮ В ЗДАНИИ А

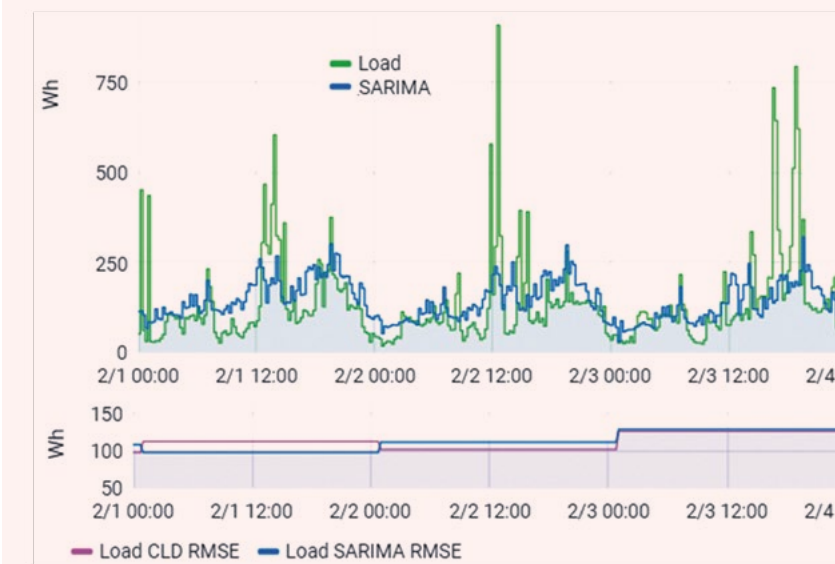


Рис. 4

СРАВНЕНИЕ УСТОЙЧИВОЙ АВТОРЕГРЕССИОННОЙ МОДЕЛИ И УСТОЙЧИВОЙ МОДЕЛИ С ФАКТИЧЕСКИМИ ДАННЫМИ ПО ЭНЕРГОПОТРЕБЛЕНИЮ В ЗДАНИИ А

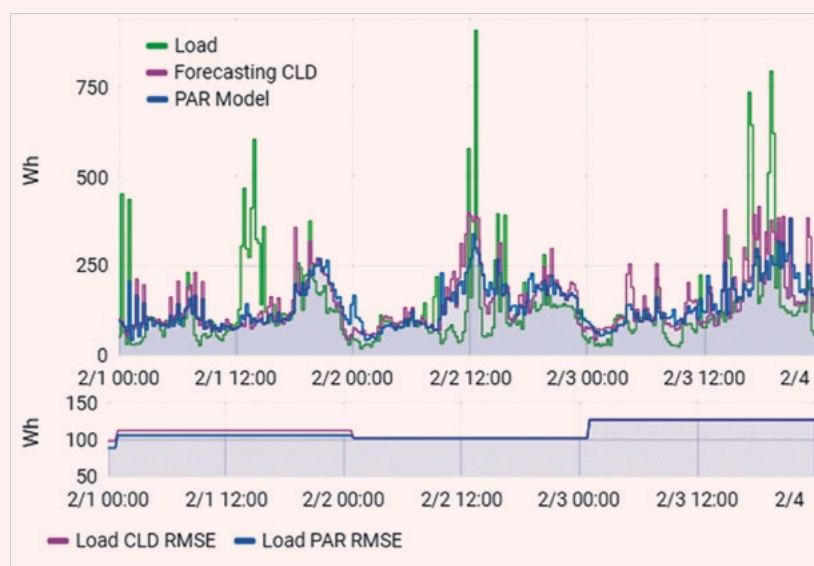


Рис. 5

рассчитывается RMSE, показывающий разность между прогнозами PAR и SPR в этом периоде. Также в табл. 3 и 4 показано, что модель SPR работает значительно лучше, чем модель PAR, стандартные методы HW и SARIMA. Это объясняется тем, что модель SPR использует больший набор устойчивых факторов, которые значительно снижают влияние неопределенности, вызванной изменениями в расписании пользователей.

ЗАКЛЮЧЕНИЕ

В статье представлены стандартные методы для прогнозирования электрической нагрузки, такие как модели на основе базовых линий, модели Хольта — Винтерса и SARIMA,

а также новый класс регрессионных моделей, разработанных специально для решения задачи прогнозирования нагрузки: устойчивая авторегрессионная модель (PAR) и сезонная устойчивая регрессионная модель (SPR). Обе модели вычисляют относительную значимость между низкочастотными и высокочастотными устойчивыми и временными характеристиками профиля нагрузки. На основе реальных данных двух зданий оценивается точность моделей PAR и SPR, выполняется сопоставление этих моделей со стандартными методами прогнозирования.

Предложенная модель SPR, объединяющая различные признаки профиля нагрузки, показывает высокую точность прогнозирования по сравнению со стандартными моделями

и может быть использована в системах энергомониторинга и управления энергоресурсами [26, 27]. Модель SPR также является вычислительно эффективной, так как основана на рекурсивном расчете коэффициентов линейной регрессии. Однако потенциальным ограничением является то, что алгоритм не предполагает наличие нелинейных зависимостей между предикторами и признаками.

В перспективе планируется исследовать возможные направления оптимизации модели, которые можно реализовать за счет добавления нелинейных зависимостей в модель SPR. Кроме того, предложенные методы могут быть скомбинированы с использованием ансамблевого подхода.

ЛИТЕРАТУРА

1. Akpinar M., Yumusak N. Year Ahead Demand Forecast of City Natural Gas Using Seasonal Time Series Methods//Energies. 2016. № 9 (9). P. 727.
2. Alfares H., Mohammad N. Electric load forecasting: Literature survey and classification of methods//International Journal of Systems Science. 2002. № 33. P. 23–34.
3. Cancelo J.R., Espasa A., Grafe R. Forecasting the electricity load from one day to one week ahead for the Spanish system operator//International Journal of Forecasting. 2008. № 24 (4). P. 588–602.
4. Cesa-Bianchi N., Lugosi G. Prediction, Learning, and Games. Cambridge University Press, 2006.
5. Chasparis G.C., Pichler M., Spreitzhofer J., Esterl T.A cooperative demand-response framework for day-ahead optimization in battery pools//Energy Informatics. 2019. № 2. P. 1–17.
6. Chen H., Vidakovic B., Mavris D. Multiscale forecasting method using armax models//Current Development in Theory and Applications of Wavelets. 2004. № 4.
7. Chitsaz H., Shaker H., Zareipour H., Wood D., Amjady N. Short-term electricity load forecasting of buildings in microgrids//Energy and Buildings. 2015. № 99. P. 50–60.
8. Christ M., Braun N., Neuffer J., Kempa-Liehr A.W. Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh —

СРАВНЕНИЕ СЕЗОННОЙ УСТОЙЧИВОЙ РЕГРЕССИОННОЙ МОДЕЛИ С ДРУГИМИ МОДЕЛЯМИ И ФАКТИЧЕСКИМИ ДАННЫМИ ПО ЭНЕРГОПОТРЕБЛЕНИЮ В ЗДАНИИ А

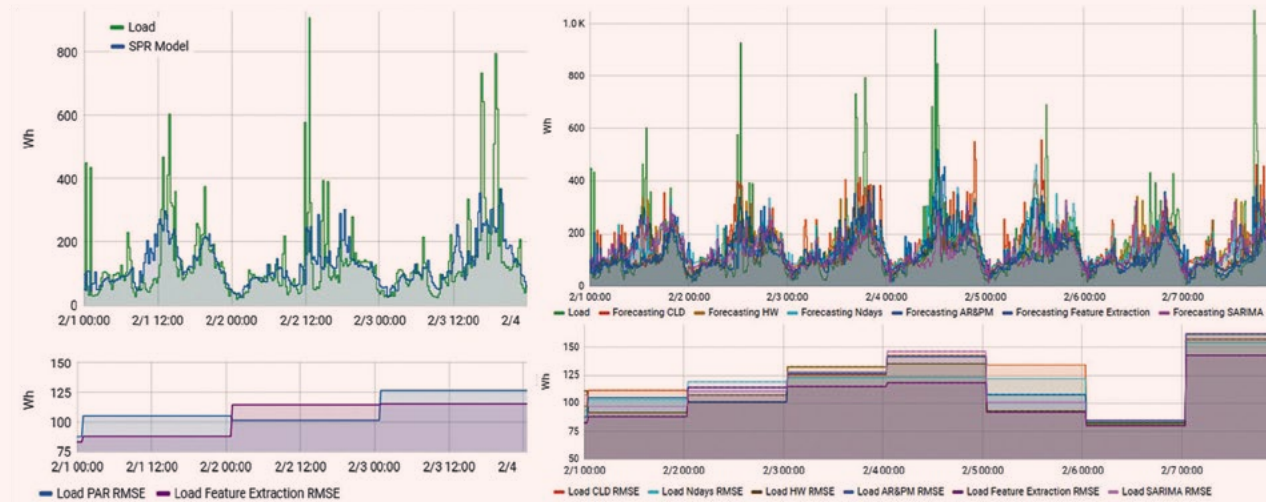


Рис. 6

9. Christ M., Kempa-Liehr A.W., Feindt M. Distributed and parallel time series feature extraction for industrial big data applications. 2017. ArXiv:1610.07717 [cs].
10. Clements A., Hurn A., Li Z. Forecasting day-ahead electricity load using a multiple equation time series approach//European Journal of Operational Research. 2016. № 251 (2). P. 522–530.
11. Grubinger T., Chasparis G.C., Natschläger T. Generalized online transfer learning for climate control in residential buildings//Energy and Buildings. 2017. № 139. P. 63–71.
12. Haben S., Giasemidis G., Ziel F., Arora S. Short term load forecasting and the effect of temperature at the low voltage level//International Journal of Forecasting. 2019. № 35 (4). P. 1469–1484.
13. Haben S., Ward J., Greetham D.V., Singleton C., Grindrod P. A new error measure for forecasts of household-level, high resolution electrical energy consumption//International Journal of Forecasting. 2014. № 30 (2). P. 246–256.
14. Hippert H., Bunn D., Souza R. Large neural networks for electricity load forecasting: Are they

- over fitted?//International Journal of Forecasting. 2005. № 21 (3). P. 425–434.
15. Hong T., Gui M., Baran M.E., Willis H.L. Modeling and forecasting hourly electric load by multiple linear regression with interactions. In: IEEE PES General Meeting. 2010. P. 1–8.
16. Hong T., Fan S. Probabilistic electric load forecasting: A tutorial review//International Journal of Forecasting. 2016. № 32 (3). P. 914–938.
17. Kychkin A. Synthesizing a system for remote energy monitoring in manufacturing//Metallurgist. 2016. № 59. P. 752–760.
18. Kychkin A.V., Mikriukov G.P. Applied data analysis in energy monitoring system//Journal of Problems Regional Energy. 2016. № 2 (31). P. 84–92.
19. Ljung L. System Identification: Theory for the User. 2nd ed. NJ: Prentice Hall, 1999.
20. Notton G., Voyant C. Forecasting of Intermittent Solar Energy Resource. In: Advances in Renewable Energies and Power Technologies, 2018. P. 77–114. Elsevier.
21. Sayed A. Fundamentals of Adaptive Filtering. New Jersey: John Wiley & Sons, Inc., 2003.

22. Soares L.J., Medeiros M.C. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data//International Journal of Forecasting. 2008. № 24 (4). P. 630–644.
23. Szmit M., Adamus S., Buga la S., Szmit A. Implementation of Brutlag's algorithm in Anomaly Detection 3.0.7. 2012.
24. Xu Y., Li N., Low S.H. Demand Response With Capacity Constrained Supply Function Bidding//IEEE Transactions on Power Systems. 2016. № 31 (2). P. 1377–1394.
25. Ye R., Dai Q. A novel transfer learning framework for time series forecasting//Knowledge-Based Systems. 2018. № 156. P. 74–99.
26. Faizrahmanov R.A., Frank T., Kychkin A.V., Fedorov A. B. Sustainable energy consumption control using the MY-JEVIS energy management data system//Russian Electrical Engineering. 2011. № 82 (11). P. 607–611.
27. Kostygov A.M., Kychkin A.V., Artemov S.A. An automated system for remote energy monitoring of mobile objects with electric drives//Russian Electrical Engineering. 2015. Vol. 86. № 11. P. 667–669.