

<https://doi.org/10.21122/1029-7448-2023-66-4-305-321>

УДК 621.311

Повышение точности прогнозирования генерации фотоэлектрических станций на основе алгоритмов k -средних и k -ближайших соседей

П. В. Матренин¹⁾, А. И. Хальясмаа^{1, 2)}, В. В. Гамалей¹⁾, С. А. Ерошенко^{1, 2)},
Н. А. Попкова³⁾, Д. А. Секацкий³⁾, Я. В. Потачиц³⁾

¹⁾Новосибирский государственный технический университет
(Новосибирск, Российская Федерация),

²⁾Уральский федеральный университет имени первого Президента России Б. Н. Ельцина
(Екатеринбург, Российская Федерация),

³⁾Белорусский национальный технический университет (Минск, Республика Беларусь)

© Белорусский национальный технический университет, 2023
Belarusian National Technical University, 2023

Реферат. Возобновляемые источники энергии рассматриваются как средство снижения углеродного следа топливно-энергетического комплекса, при этом стохастический характер генерации осложняет их интеграцию с электроэнергетическими системами. Эта существенная трудность обуславливает необходимость создавать и совершенствовать методы прогнозирования генерации электрических станций, использующих энергию солнца, ветра и водных потоков. Наиболее важным направлением, обеспечивающим повышение точности прогнозных моделей, является глубокий анализ метеорологических условий как главного фактора, влияющего на выработку электроэнергии. В данной работе предложен и исследован метод адаптации прогнозных моделей под метеорологические условия работы фотоэлектрических станций на базе алгоритмов машинного обучения. При этом вначале выполняется обучение без учителя методом k -средних для формирования кластеров. Для этой задачи также предложено и исследовано использование алгоритма понижения размерности пространства признаков для визуализации оценки точности кластеризации. Затем для каждого кластера построена своя модель машинного обучения для формирования прогнозов и алгоритм k -ближайших соседей для отнесения текущих условий на этапе эксплуатации модели к одному из сформированных кластеров. Исследование было проведено на почасовых метеорологических данных за период с 1985 по 2021 г. Одной из особенностей этого подхода является кластеризация метеоусловий на часовых, а не суточных интервалах. В результате средний модуль относительной ошибки прогнозирования существенно снижается в зависимости от используемой модели прогнозирования. Для наилучшего варианта ошибка прогнозирования генерации фотоэлектрической станции на час вперед составила 9 %.

Ключевые слова: краткосрочное прогнозирование, генерация электроэнергии, фотоэлектрическая станция, возобновляемые источники энергии, метеорологические факторы, инсоляция, солнечная радиация, нейронные сети, кластеризация данных, прогнозная модель, предобработка данных, машинное обучение, метод главных компонент, адаптивный бустинг, линейная регрессия

Для цитирования: Повышение точности прогнозирования генерации фотоэлектрических станций на основе алгоритмов k -средних и k -ближайших соседей / П. В. Матренин [и др.] // *Энергетика. Изв. высш. учеб. заведений и энерг. объединений СНГ*. 2023. Т. 66, № 4. С. 305–321. <https://doi.org/10.21122/1029-7448-2023-66-4-305-321>

Адрес для переписки

Секацкий Дмитрий Александрович
Белорусский национальный технический университет
просп. Независимости, 65/2,
220013, г. Минск, Республика Беларусь
Тел.: +375 17 292-65-82
dsekatski@gmail.com

Address for correspondence

Sekatski Dzmitry A.
Belarusian National Technical University
65/2, Nezavisimosty Ave.,
220013, Minsk, Republic of Belarus
Tel.: +375 17 292-65-82
dsekatski@gmail.com

Improving of the Generation Accuracy Forecasting of Photovoltaic Plants Based on k -Means and k -Nearest Neighbors Algorithms

P. V. Matrenin^{1, 2)}, A. I. Khalyasmaa^{1, 2)}, V. V. Gamaley¹⁾, S. A. Eroshenko^{1, 2)},
N. A. Papkova³⁾, D. A. Sekatski³⁾, Y. V. Potachits³⁾

¹⁾Novosibirsk State Technical University (Novosibirsk, Russian Federation),

²⁾Ural Federal University named after the first President of Russia B. N. Yeltsin
(Ekaterinburg, Russian Federation),

³⁾Belarusian National Technical University (Minsk, Republic of Belarus)

Abstract. Renewable energy sources (RES) are seen as a means of the fuel and energy complex carbon footprint reduction but the stochastic nature of generation complicates RES integration with electric power systems. Therefore, it is necessary to develop and improve methods for forecasting of the power plants generation using the energy of the sun, wind and water flows. One of the ways to improve the accuracy of forecast models is a deep analysis of meteorological conditions as the main factor affecting the power generation. In this paper, a method for adapting of forecast models to the meteorological conditions of photovoltaic stations operation based on machine learning algorithms was proposed and studied. In this case, unsupervised learning is first performed using the k -means method to form clusters. For this, it is also proposed to use studied the feature space dimensionality reduction algorithm to visualize and estimate the clustering accuracy. Then, for each cluster, its own machine learning model was trained for generation forecasting and the k -nearest neighbours algorithm was built to attribute the current conditions at the model operation stage to one of the formed clusters. The study was conducted on hourly meteorological data for the period from 1985 to 2021. A feature of the approach is the clustering of weather conditions on hourly rather than daily intervals. As a result, the mean absolute percentage error of forecasting is reduced significantly, depending on the prediction model used. For the best case, the error in forecasting of a photovoltaic plant generation an hour ahead was 9 %.

Keywords: short-term forecasting, electricity generation, photovoltaic plant, renewable energy sources, meteorological factors, insolation, solar radiation, neural networks, data clustering, predictive model, data preprocessing, machine learning, principal component analysis, adaptive boosting, linear regression

For citation: Matrenin P. V., Khalyasmaa A. I., Gamaley V. V., Eroshenko S. A., Papkova N. A., Sekatski D. A., Potachits Y. V. (2023) Improving of the Generation Accuracy Forecasting of Photovoltaic Plants Based on k -Means and k -Nearest Neighbors Algorithms. *Energetika. Proc. CIS Higher Educ. Inst. and Power Eng. Assoc.* 66 (4), 305–321. <https://doi.org/10.21122/1029-7448-2023-66-4-305-321> (in Russian)

Введение

Процессы генерации на базе возобновляемых источников энергии (ВИЭ) напрямую определяются природными явлениями. Они имеют стохастическую природу и могут быстро изменяться, поэтому генерация ВИЭ нестабильна и труднопредсказуема. Для солнечных панелей и солнечных электростанций (СЭС) ключевым фактором, влияющим на выработку электрической энергии, является инсоляция, при этом рассматриваются различные горизонты планирования: от краткосрочных (10–30 минут) до долгосрочных (на месяц или год вперед). В зависимости от горизонта планирования и используемых данных применяются физические, статистические, вероятностные модели, а также методы машинного обучения [1–4]. В настоящее время при решении задач прогнозирования инсоляции и генерации СЭС для разных горизонтов планирования ошибка колеб-

лется от 3 % до 20 % [3]. Высокое качество прогнозирования генерации СЭС также необходимо и в контексте обеспечения электроэнергией промышленных объектов или производственных комплексов [5], при этом очень важна предобработка данных [6]. Одним из методов предобработки данных является кластеризация.

Алгоритмы кластеризации для предварительного анализа входных данных и последующего прогнозирования инсоляции или генерации СЭС на практике можно применять различными способами. Можно выделить несколько основных направлений.

Метеорологические параметры используются, чтобы разделить дни на кластеры и затем построить для каждого из них прогнозную модель. Подход, как правило, используется для прогноза на сутки вперед с применением алгоритмов на базе k -средних [7–9]. В работе [10] используются самоорганизующиеся карты Кохонена и k -средних для разделения дней на типовые дни (летние, весенние / осенние, зимние дни без / с постоянным током панели, дни с низкой инсоляцией).

В работе [11] кластеризация выполняется, чтобы разделить дни по параметрам облачности. Также применяются спутниковые данные и алгоритмы компьютерного зрения. Во время обработки изображений они разделяются на кластеры для определения различных типов облачности в течение суток [12]. В работе [13] используются данные для прогнозирования генерации СЭС через модельные значения инсоляции со спутников NASA (с учетом ясного неба). Авторы используют кластерный анализ коэффициента прозрачности атмосферы с помощью разных методов (смеси Гауссовых распределений, k -средних, нечеткой кластеризации C -средних).

Для краткосрочного прогнозирования инсоляции в работе [14] не используются метеорологические факторы. Авторы используют лишь краткосрочные измерения с дискретностью в полчаса за трехчасовой период с помощью модели, выполняя кластеризацию по методу PSA (Part and Select Algorithm). Одна из наиболее сложных моделей предложена в работе [15], где авторы используют модели на базе вейвлет-нейронных сетей, а входные погодно-климатические данные разбивают по временам года. Затем для каждой из четырех групп применяют алгоритм self-organizing map для выделения характерных шаблонов корреляции между инсоляцией и погодой. В работе [16], наоборот, методика очень проста. Используются лишь два кластера с простым правилом разделения: если инсоляция выше 150 Вт/м^2 , то день считается солнечным, если ниже – облачным. Прогнозные модели авторы создают на основе многослойного перцептрона.

Основная часть

Исследуемый объект и исходные данные

В качестве объекта рассмотрена Кемеровская область, источником данных является база метеорологических данных [17]. Средняя суммарная суточная инсоляция показана на рис. 1. На рис. 2 приводятся значения

инсоляции, измеренные для первых недель января, апреля, июля и октября 2021 г. Выработка солнечных панелей определяется метеорологическими факторами. Долгосрочное изменение климата влияет на них, поэтому важно исследовать работу моделей на длительных многолетних выборках данных.

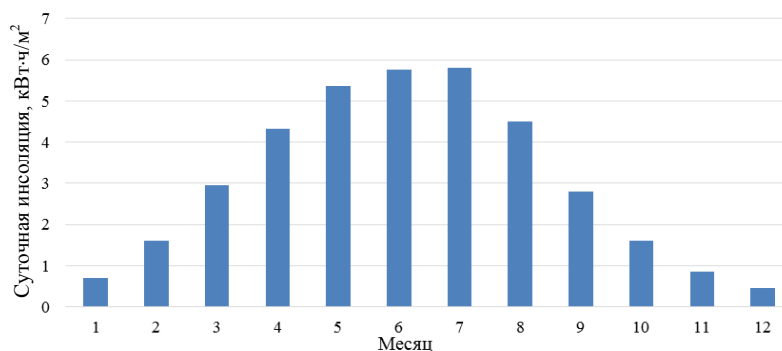


Рис. 1. Суммарная суточная инсоляция

Fig. 1. Total daily insolation

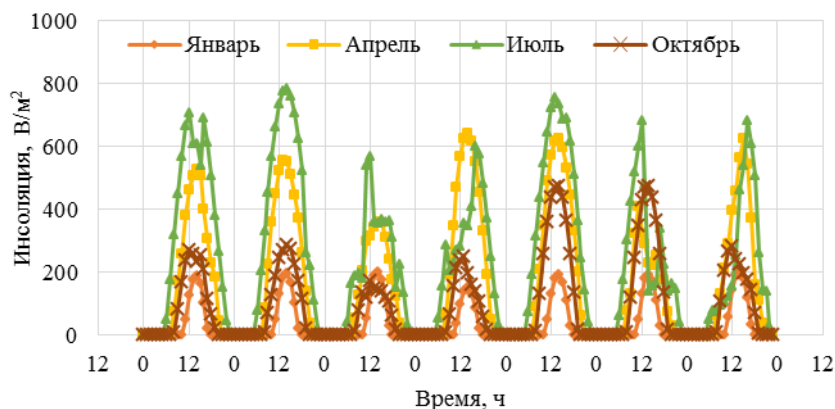


Рис. 2. Примеры почасовых значений инсоляции

Fig. 2. Examples of hourly insolation values

В работе использованы данные с 1 января 1985 г. по 31 декабря 2021 г., содержащие следующие параметры: продолжительность солнечного сияния (мин.); полная инсоляция на горизонтальную поверхность (В/м^2); прямая инсоляция (В/м^2); рассеянная инсоляция (В/м^2); количество осадков (мм); относительная влажность (%); облачность (%); температура ($^{\circ}\text{C}$); скорость ветра (км/ч). Дополнительно была рассчитана выработка солнечной панели для каждого часа с учетом значений инсоляции, угла наклона панели и угла склонения Солнца, а также учтено число солнечных панелей в электростанции (2500). В результате массив данных содержит 324 336 строк (записей) и 14 столбцов (признаков). В массив данных, кроме девяти метеорологических признаков, указанных выше, также входят величина вырабатываемой мощности, год, месяц, день месяца и час суток.

Постановка задачи и метрики точности прогноза

Необходимо построить модель краткосрочного (на один час вперед) прогнозирования выработки солнечной панели по имеющимся метеорологическим данным

$$P_i^* = f(W_{i-1}, W_{i-2}, \dots, W_{i-k}, P_{i-1}, P_{i-2}, \dots, P_{i-k}), \quad (1)$$

где P_i^* – прогнозная мощность солнечной панели в i -й час; f – модель; W_j – вектор значений метеорологических факторов в j -й час ($j = i-1, i-2, \dots, i-k$); P – истинное значение мощности в i -й час.

В работе не используются прогнозы метеорологических факторов, поскольку в этом случае система окажется зависимой от точности прогнозов метеопровайдеров. Для краткосрочного прогнозирования используются данные за k предыдущих часов.

Точность прогноза оценивается на тестовой выборке с использованием метрик mean absolute error (MAE), normalized mean absolute error (nMAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - P_i^*|;$$

$$nMAE = \sum_{i=1}^n |P_i - P_i^*| / \sum_{i=1}^n P_i \cdot 100\%,$$

где n – размер тестовой выборки в часах.

Кластеризация данных

Метеорологические условия отличаются высоким разнообразием, из-за чего построение одной модели, работающей для всех условий, может быть не столь эффективно, как разработка нескольких отдельных моделей, каждая из которых обучается давать прогноз для определенных условий. Под эффективностью здесь понимается не только точность прогноза, но и компактность модели, ведь единая общая модель будет более сложной, следовательно, менее интерпретируемой и иметь более высокий риск переобучения.

Чтобы разделить метеорологические условия, можно прописать ряд правил на основе экспертных знаний, что может быть затруднительно из-за большого числа факторов. Другим способом является применение методов машинного обучения. Алгоритмы обучения без учителя способны разбивать данные на кластеры, учитывая все имеющиеся в выборке данных признаки (параметры). При этом возникает проблема оценки точности кластеризации. Как и в других задачах обучения без учителя, отсутствие правильных ответов не позволяет получить объективную количественную оценку. В данной работе использован следующий подход:

- 1) выполняется кластеризация данных с помощью метода k -средних [18];
- 2) применяется метод главных компонент [19] для понижения размерности данных без учета результатов кластеризации, при этом число признаков уменьшается до двух главных компонент;

3) полученный результат понижения размерности визуализируется с помощью двумерной диаграммы рассеяния, а цвета точек на диаграмме выбираются в соответствии с результатами кластеризации.

Если кластеризация выполнена корректно, то на диаграмме рассеяния кластеры будут локализованы, в ином случае точки разных цветов окажутся перемешанными.

Построение модели прогнозирования

Для каждого кластера создается отдельная регрессионная модель, прогнозирующая выработку солнечных панелей. Для эксплуатации регрессионной модели необходимо определять принадлежность текущих метеорологических условий к какому-либо кластеру. Для этого в данной работе используется еще одна модель, которая выполняет классификацию. Она определяет, к какому кластеру нужно отнести текущие метеорологические условия. Так как для кластеризации используется алгоритм k -средних, то для классификации подходит близкий к нему по своей природе алгоритм k -ближайших соседей (kNN). На рис. 3 приведены укрупненные шаги построения и использования моделей.

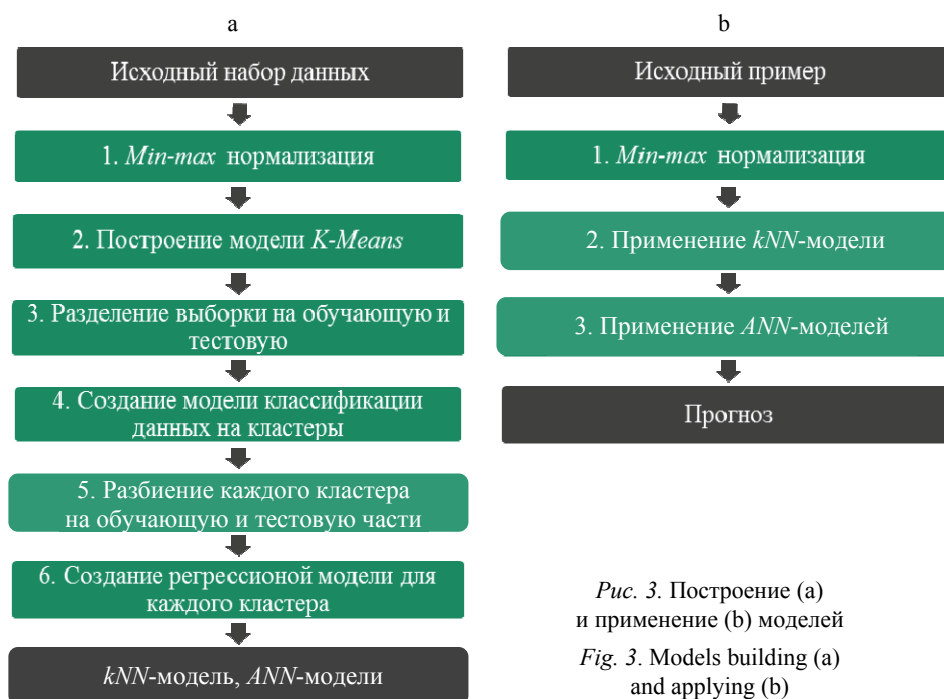


Рис. 3. Построение (а) и применение (б) моделей
Fig. 3. Models building (a) and applying (b)

Этап построения моделей состоит из шести шагов.

1. Поскольку признаки имеют различную природу, первым шагом является нормализация. Все признаки преобразуются с помощью линейной min-max нормализации.

2. Создается модель k -средних, разбивающая данные на кластеры. С помощью визуализации и метода главных компонент проверяется корректность модели.

3. Выборка разделяется на обучающую и тестовую в соотношении 80 на 20 произвольным образом для обучения модели kNN , выполняющей классификацию.

4. Создается модель для классификации данных по кластерам, которая нужна только на этапе использования, на этапе построения она не используется. На этом шаге предполагается обучение, настройка гиперпараметров и тестирование модели.

5. Для каждого кластера нужно обучить свою регрессионную модель, перед этим выполняется разбиение каждого кластера на обучающую и тестовую части.

6. Для каждого кластера создается своя регрессионная модель. Используемые компактные многослойные нейронные сети показывают высокую эффективность в задачах прогнозирования временных рядов в электроэнергетике [20, 21].

Этап применения моделей состоит из трех шагов.

1. Признаки входной порции данных преобразуются с помощью линейной min-max нормализации к шкале от 0 до 1. Используются те значения минимума и максимума, которые определены на этапе построения модели.

2. Применяется классификатор на базе kNN , чтобы определить, к какому кластеру относятся метеорологические условия. При этом используются только данные текущего (ближайшего к прогнозному) часа.

3. Применяется нейросетевая модель, соответствующая определенному на втором шаге кластеру для получения прогноза генерации на час вперед.

Особенности представленного подхода

Можно выделить следующие особенности предложенного подхода:

- большая интерпретируемость по сравнению с созданием единой модели за счет разделения процесса прогнозирования на части;
- использование алгоритмов k -средних и kNN способствует повышению интерпретируемости благодаря их простоте;
- построение разных моделей на базе машинного обучения для разных условий может приводить к более высокой точности прогноза, поскольку каждая модель оказывается сфокусированной на определенных условиях работы;
- применение кластеризации по исходным данным с проверкой по данным после снижения размерности позволяет визуально оценить корректность кластеризации;
- за счет использования k -средних и kNN для кластеризации и последующей классификации модель можно автоматически дообучать на новых данных без риска переобучения;
- использование компактных многослойных нейронных сетей возможно в режиме дообучения, так можно реализовать адаптацию моделей во времени, а с учетом метеорологической адаптации получить климатически-временную адаптацию моделей, что особенно важно в условиях глобального изменения климата.

Предварительный анализ данных и понижение размерности

Выработка электроэнергии СЭС зависит в первую очередь от инсоляции. На этапе предварительного анализа построены зависимости инсоляции от всех остальных признаков. Некоторые из них показаны на рис. 4. Видно, что ни один из признаков по отдельности не может быть использован для точной оценки инсоляции, тем более для прогнозирования. В табл. 1 приведены коэффициенты корреляции Спирмена между генерацией рассматриваемых солнечных панелей и другими параметрами (учитываются только светлые часы суток, когда генерация не равна нулю).

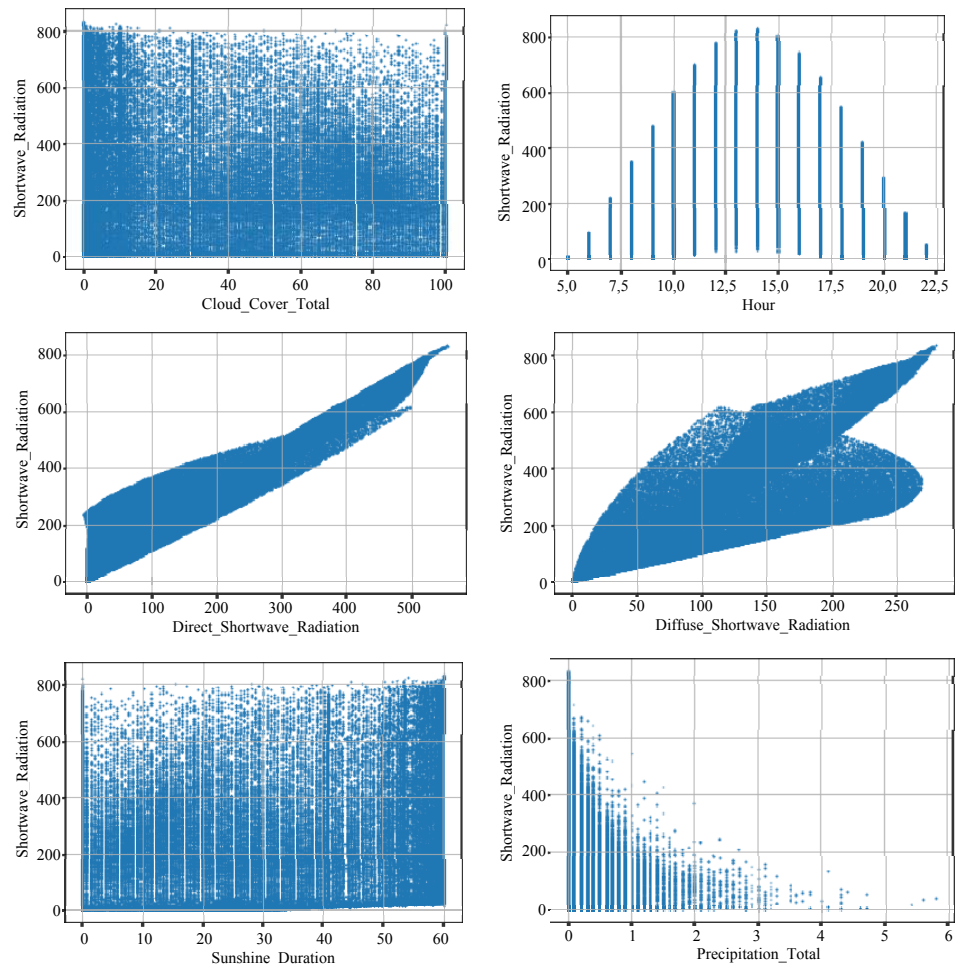


Рис. 4. Зависимости инсоляции от различных факторов

Fig. 4. Dependences of insolation on various factors

Очевидна зависимость факторов между собой. Так, инсоляция представляет собой сумму прямой и рассеянной солнечной радиации, которые, в свою очередь, зависят от облачности. Облачность коррелирует с осадками, осадки с влажностью и т. д. Более наглядно можно увидеть значимость различных факторов, выполнив понижение размерности и отобразив данные на плоскости, при этом используя цвет как третье измерение.

Таблица 1

**Коэффициенты корреляции Спирмена между генерацией солнечных панелей
и другими параметрами**

Spearman correlation coefficients between solar panel generation and other parameters

Признак	Коэффициент корреляции	Признак	Коэффициент корреляции
Продолжительность солнечного сия- ния	0,29	Температура	0,36
Полная инсоляция	0,86	Скорость ветра	0,06
Прямая инсоляция	0,77	Год	0,00
Рассеянная инсоляция	0,77	Месяц	-0,10
Количество осадков	-0,20	День	-0,01
Относительная влажность	-0,47	Час	-0,01
Облачность	-0,29		

В данной работе мы использовали метод главных компонент для понижения размерности. Данный алгоритм можно отнести к классу алгоритмов машинного обучения без учителя. Результаты приведены на рис. 5.

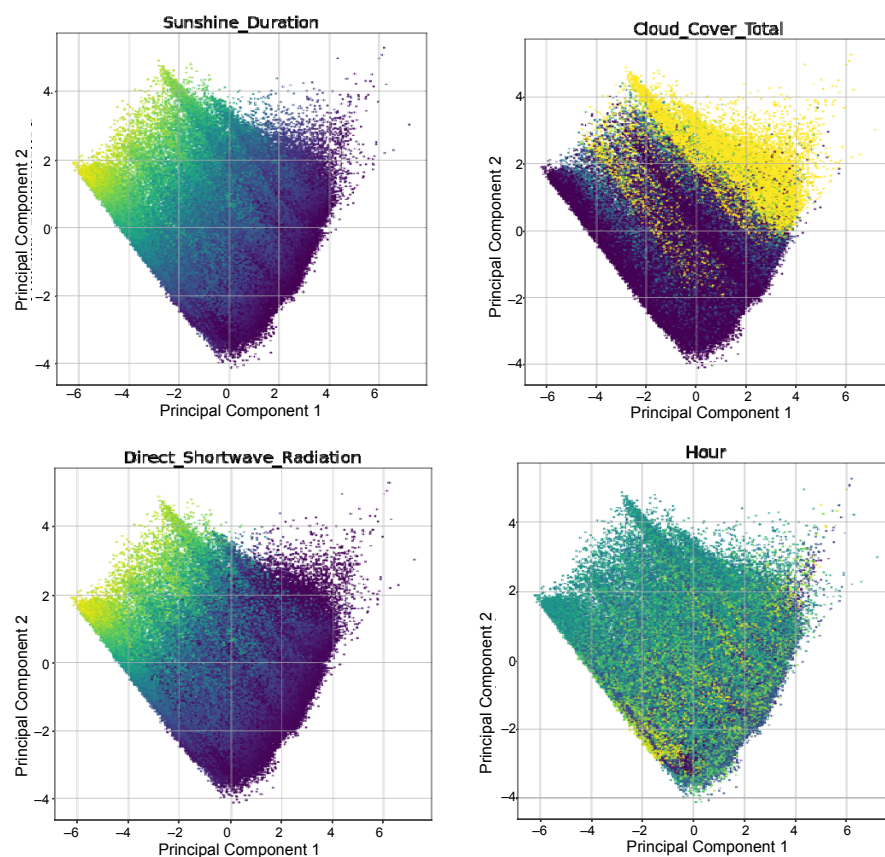


Рис. 5. Результаты визуализации методов главных компонент

Fig. 5. Visualization results of principal component analysis

Кластеризация

Алгоритм кластеризации k -средних является наиболее используемым за счет простоты, интерпретируемости и широкой области применения. Последний пункт отличает его от таких алгоритмов, как спектральная кластеризация, *DBSCAN*, *BIRCH* или иерархическая кластеризация, которые имеют свои определенные области применения [22]. Необходимо выбрать число кластеров, исходя из того, что для каждого будет создана своя прогнозная модель. На рис. 6 приведены результаты кластеризации при разном числе кластеров. При кластеризации используются только часы светлого времени суток.

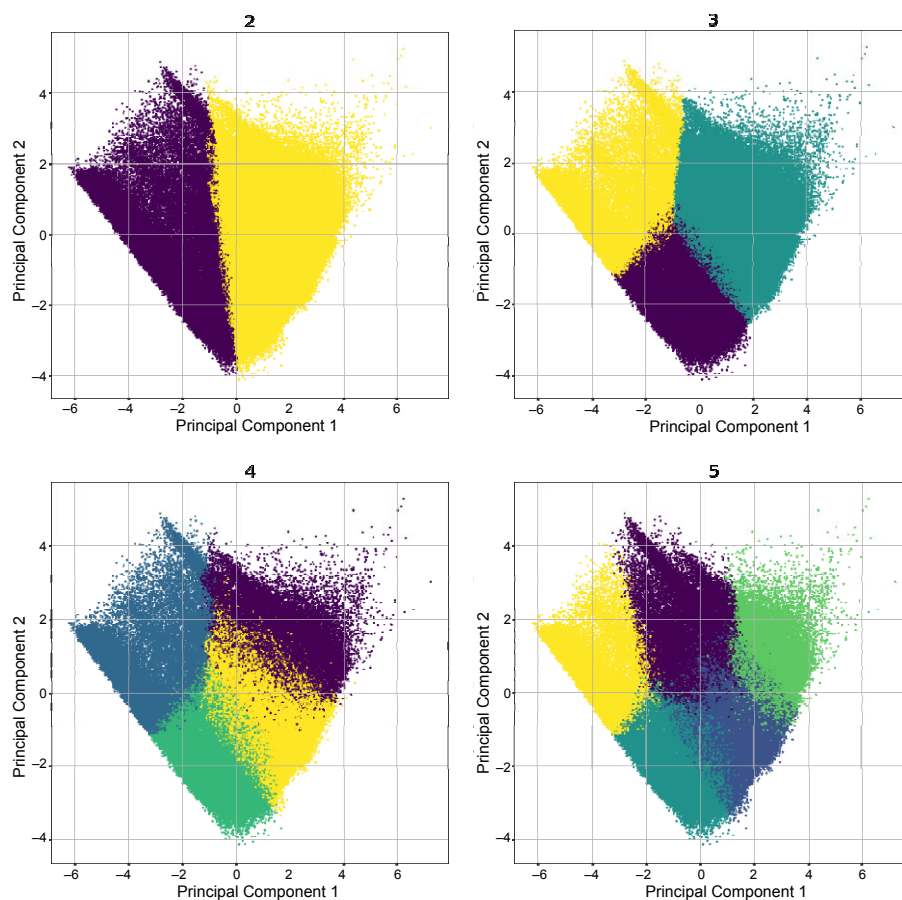


Рис. 6. Результаты кластеризации при разном числе кластеров

Fig. 6. Clustering results for different number of clusters

Было выбрано четыре кластера, поскольку в этом случае границы между кластерами получаются более четкими, чем при пяти. В то же время два и три кластера недостаточно с точки зрения разделения прогнозирования на несколько моделей для различных условий. Использование кластеризации на исходных данных с проверкой ее работы на данных после пони-

жения размерности позволяет оценить корректность кластеризации. Если выполнить кластеризацию после понижения размерности, то будут получены кластеры с четкими границами, независимо от того, отражают ли они внутренние зависимости между различными признаками в данных, как показано на рис. 7.

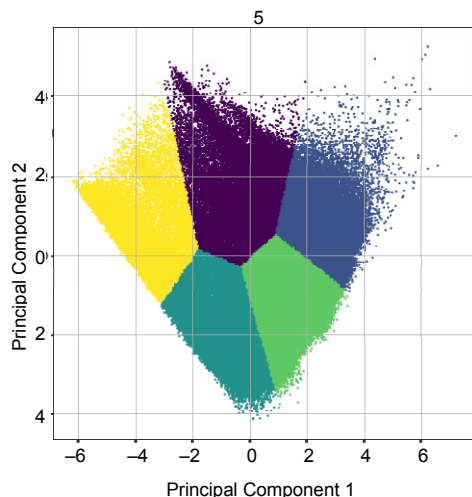


Рис. 7. Результаты кластеризации, выполненной после понижения размерности

Fig. 7. Results of clustering performed after dimensionality reduction

При тестировании разделения условий по ранее сформированным кластерам использованы следующие показатели точности:

$$Recall = \frac{1}{C} \sum_{i=1}^C \frac{TP_c}{TP_c + FN_c};$$

$$Precision = \frac{1}{C} \sum_{i=1}^C \frac{TP_c}{TP_c + FP_c},$$

где C – число кластеров (4); TP_c – число объектов кластера c , которые модель отнесла к кластеру c ; FN_c – число объектов кластера c , которые модель отнесла к другому кластеру; FP_c – число объектов любого кластера кроме c , которые модель отнесла к кластеру c .

Результаты приведены в табл. 2. При этом обучающая выборка содержит 81 % данных (1985–2014 гг.), тестовая – 19 % (2015–2021 гг.). При использовании трех кластеров точность совсем немного выше, чем при четырех, поэтому далее использована кластеризация с разделением на четыре кластера. Таким образом, точность распределения метеорологических условий по кластерам высокая (95 %). При этом ошибки классификации не являются критичными, поскольку они происходят на границах кластеров, то есть в пограничных метеорологических условиях. Поэтому выбор не той модели, которая построена для этого кластера, а иной, не является серьезной ошибкой.

Таблица 2

Результаты кластеризации метеорологических факторов по кластерам

Clustering results of meteorological factors by clusters

C	Кластер	Обучающая выборка, Recall, %	Обучающая выборка, Precision, %	Тестовая выборка, Recall, %	Тестовая выборка, Precision, %
4	1	96	97	92	93
4	2	97	97	95	97
4	3	97	98	96	97
4	4	97	98	95	92
4	Среднее	97	97	95	95
3	1	98	97	96	96
3	2	97	97	96	95
3	3	99	99	98	98
3	Среднее	98	98	97	97

Построение регрессионных моделей

В работе рассмотрены три вида моделей машинного обучения: линейная регрессия, адаптивный бустинг деревьев решений [23] и многослойная компактная нейронная сеть. Модели построены в соответствии с выражением (1), при этом использованы данные за шесть предыдущих часов ($k = 6$). Настройка параметров выполнялась экспериментально. Для адаптивного бустинга использовано 10 деревьев максимальной глубиной 10, остальные параметры использованы по умолчанию [24]. Для нейронной сети использована следующая архитектура из пяти слоев: входной, три скрытых слоя (32, 32, 8 нейронов, функция активации *ReLU*), выходной слой с сигмоидальной функцией активации.

Использован метод обучения *Adam*. Реализация нейронной сети выполнена с помощью библиотеки *Keras*. В табл. 3 приведены результаты обучения указанных моделей с использованием кластеризации и без нее. На рис. 8 и 9 показано наложение прогноза лучшей модели на истинные значения для зимнего и летнего месяцев тестовой выборки.

Таблица 3

Результаты краткосрочного прогнозирования генерации

Results of short-term generation forecasting

Модель	Используется ли кластеризация	MAE, кВт	nMAE, %
LR	нет	28,55	18,12
LR	да	21,50	13,46
Эффект от кластеризации, %	24,7		
AB	нет	19,68	12,45
AB	да	16,25	10,31
Эффект от кластеризации, %	17,1		
ANN	нет	17,65	11,20
ANN	да	13,62	8,58
Эффект от кластеризации, %	22,8		

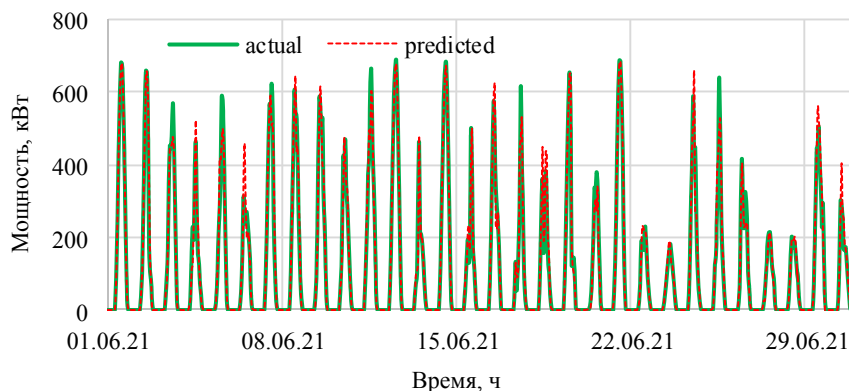


Рис. 8. Сравнение истинных и прогнозных значений выработки СЭС, июнь 2021 г.

Fig. 8. Comparison of actual and predicted solar power generation, June 2021

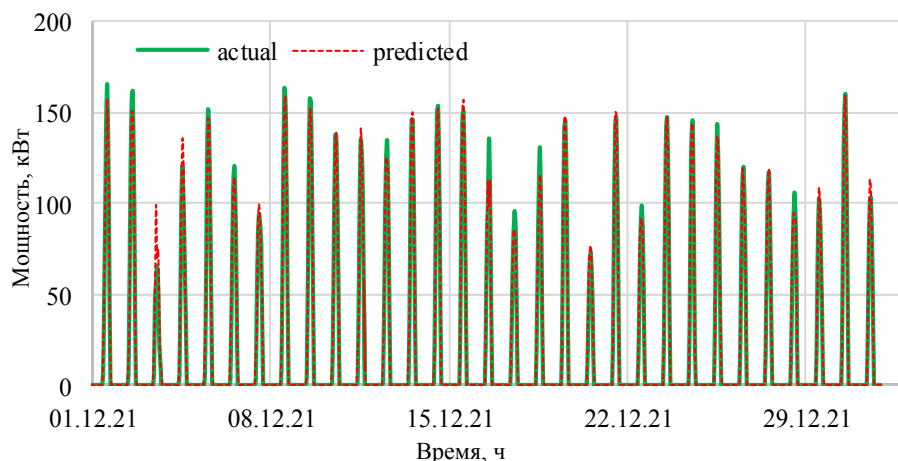


Рис. 9. Сравнение истинных и прогнозных значений выработки СЭС, декабрь 2021 г.

Fig. 9. Comparison of actual and predicted solar power generation, December 2021

ВЫВОД

Предложен и апробирован на данных многолетних метеорологических наблюдений метод краткосрочного прогнозирования генерации фотоэлектрических станций с адаптацией климатической модели краткосрочного прогнозирования. Особенностью метода является кластеризация часовых значений метеоусловий и создание отдельных моделей машинного обучения для каждого кластера. В результате отдельные модели оказываются более простыми и точными, поскольку каждая сфокусирована на определенных условиях работы. В результате экспериментов определено, что предложенная кластеризация снижает ошибку прогнозирования регрессионных

моделей на 17–25 %. Итоговая ошибка прогнозирования при использовании четырех кластеров и компактных многослойных нейронных сетей составила 8,6 % при коэффициенте детерминации $R^2 = 0,96$.

За счет использования простых, надежных и не склонных к переобучению алгоритмов k -средних, kNN и компактных многослойных нейронных сетей систему прогнозирования можно автоматически дообучать на новых данных без риска переобучения. Таким образом, в дальнейшем предполагается реализовать адаптацию моделей во времени. С учетом метеорологической адаптации будет реализована климатически-временная адаптация моделей, что особенно важно для создания интеллектуальных информационно-аналитических систем, которые будут надежно работать при изменении условий функционирования.

Исследование выполнено за счет гранта Российского научного фонда (проект № 22-79-00181).

ЛИТЕРАТУРА

1. El hendouzi, A. Solar Photovoltaic Power Forecasting / A. El hendouzi, A. Bourouhou // Journal of Electrical and Computer Engineering. 2020. Vol. 2020. P. 1–21. <https://doi.org/10.1155/2020/8819925>.
2. Review of photovoltaic power forecasting / J. Antonanzas [et al.] // Solar Energy. 2016. Vol. 136. P. 78–111. <https://doi.org/10.1016/j.solener.2016.06.069>.
3. Machine Learning Methods for Solar Radiation Forecasting: a Review / C. Voyant [et al.] // Renewable Energy. 2017. Vol. 105. P. 569–582. <https://doi.org/10.1016/j.renene.2016.12.095>.
4. Оптимизация топологии сети с ВИЭ-генерацией на основе модифицированного адаптированного генетического алгоритма / А. М. Брамм [и др.] // Энергетика. Известия высших учебных заведений и энергетических объединений СНГ. 2022. Т.65, № 4. С. 341–354. <https://doi.org/10.21122/1029-7448-2022-65-4-341-354>.
5. Industry Experience of Developing Day-Ahead Photovoltaic Plant Forecasting System Based on Machine Learning / A. I. Khalyasmaa [et al.] // Remote Sensing. 2020. Vol. 12, Iss. 20. P. 3420. <https://doi.org/10.3390/rs12203420>.
6. The Impact of Data Filtration on the Accuracy of Multiple Time-Domain Forecasting for Photovoltaic Power Plants Generation / S. A. Eroshenko [et al.] // Applied Sciences. 2020. Vol. 10, Iss. 22. P. 8265. <https://doi.org/10.3390/app10228265>.
7. Rana, M. Solar Power Forecasting Using Weather Type Clustering and Ensembles of Neural Networks / M. Rana, I. Koprinska, V. G. Agelidis // International Joint Conference on Neural Networks (IJCNN). Vancouver, BC, Canada: IEEE, 2016. P. 4962–4969. <https://doi.org/10.1109/IJCNN.2016.7727853>.
8. Solar Radiation Intensity Probabilistic Forecasting Based on K-Means Time Series Clustering and Gaussian Process Regression / Z. Zhang [et al.] // IEEE Access. 2021. Vol. 9. P. 89079–89092. <https://doi.org/10.1109/ACCESS.2021.3077475>.
9. Unsupervised Clustering-Based Short-Term Solar Forecasting / C. Feng [et al.] // IEEE Transactions on Sustainable Energy. 2021. Vol. 10, Iss. 4. P. 2174–2185. <https://doi.org/10.1109/TSTE.2018.2881531>.
10. Unsupervised Clustering of Battery Waveforms in Off-Grid PV Installations / I. Sanz-Gorrachategui [et al.] // 2020 Fifteenth International Conference on Ecological Vehicles and

- Renewable Energies (EVER). Monte-Carlo, Monaco, 2020. P. 1–6. <https://doi.org/10.1109/EVER48776.2020.9242942>.
11. Development of Algorithm for Day Ahead PV Generation Forecasting Using Data Mining Method / M. C. Kang [et al.] // Proc. Int. IEEE 54th Midwest Symposium on Circuits and Systems (MWSCAS). Seoul, Korea (South): IEEE, 2011. P. 1–4 <https://doi.org/10.1109/MWSCAS.2011.6026333>.
 12. Pattern Classification and PSO Optimal Weights Based Sky Images Cloud Motion Speed Calculation Method for Solar PV Power Forecasting / F. Wang [et al.] S // IEEE Industry Applications Society Annual Meeting (IAS). Portland, OR, USA: IEEE, 2018. P. 1–9. <https://doi.org/10.1109/IAS.2018.8544468>.
 13. Daily Clearness Index Profiles Cluster Analysis for Photovoltaic System / C. S. Lai [et al.] // IEEE Transactions on Industrial Informatics. 2017. Vol. 13, Iss. 5. P. 2322–2332. <https://doi.org/10.1109/TII.2017.2683519>.
 14. Severiano, C. Very Short-Term Solar Forecasting Using Multi-Agent System Based on Extreme Learning Machines and Data Clustering / C. Severiano, F. G. Guimarães, M. W. Cohen // 2016 IEEE Symposium Series on Computational Intelligence (SSCI). Athens, Greece: IEEE, 2016. P. 1–8. <https://doi.org/10.1109/SSCI.2016.7850162>.
 15. Yang, Y. Short-Term PV Generation System Direct Power Prediction Model on Wavelet Neural Network and Weather Type Clustering / Y. Yang, L. Dong // 2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics. Hangzhou, China: IEEE, 2013. P. 207–211. <https://doi.org/10.1109/IHMSC.2013.56>.
 16. Day-Ahead Photovoltaic Forecasting: A Comparison of the Most Effective Techniques / N. Alfredo [et al.] // Energies. 2019. Vol. 12, Iss. 9. P. 1621. <https://doi.org/10.3390/en12091621>.
 17. Meteoblue [Electronic Resource]. Mode of access: <https://www.meteoblue.com/en/weather>. Date of access: 13.05.2022.
 18. Principal Manifolds for Data Visualization and Dimension Reduction / A. N. Gorban [et al.] // Berlin: Springer, 2008. 364 p. (Lecture Notes in Computational Science and Engineering, Vol. 58). https://doi.org/10.1007/978-3-540-73750-6_5.
 19. Hartigan, J. A. Algorithm AS 136: A K-Means Clustering Algorithm / J. A. Hartigan, M. A. Wong // Journal of the Royal Statistical Society. Series C (Applied Statistics). 1979. Vol. 28, No 1. P. 100–108. <https://doi.org/10.2307/2346830>.
 20. Improving Accuracy and Generalization Performance of Small-Size Recurrent Neural Networks Applied to Short-Term Load Forecasting / P. V. Matrenin [et al.] // Mathematics. 2020. Vol. 8. Iss. 12. P. 2169. <https://doi.org/10.3390/math8122169>.
 21. Матренин, П. В. Устойчивое краткосрочное прогнозирование скорости ветра с помощью адаптивных компактных нейронных сетей / П. В. Матренин, В. З. Манусов, Е. А. Игумнова // Проблемы региональной энергетики. 2020. Т. 7, №47. С. 69–80. <https://doi.org/10.5281/zenodo.4018960>.
 22. Clustering [Electronic Resource] // Scikit-Learn. Machine Learning in Python. Mode of access: <https://scikit-learn.org/stable/modules/clustering.html>. Date of access: 13.04.2022.
 23. Drucker, H. Improving Regressors using Boosting Techniques / H. Drucker // Proceedings of the 14th International Conference on Machine Learning (ICML). Citeseer, 1997. P. 107–115.
 24. Adaptive Boosting [Electronic Resource] // Scikit-Learn. Machine Learning in Python. Mode of access: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html?highlight=adaptive+busting. Date of access: 13.04.2022.

Поступила 18.10.2022 Подписана в печать 28.12.2022 Опубликовано онлайн 31.07.2023

REFERENCES

1. El hendouzi A., Bourouhou A. (2020) Solar Photovoltaic Power Forecasting. *Journal of Electrical and Computer Engineering*, 2020, 1–21. <https://doi.org/10.1155/2020/8819925>.

2. Antonanzas J., Osorio N., Escobar R., Urraca R., Martinez-de-Pison F. J., Antonanzas-Torres F. (2016) Review of Photovoltaic Power Forecasting *Solar Energy*, 136, 78–111. <https://doi.org/10.1016/j.solener.2016.06.069>.
3. Voyant C., Notton G., Kalogirou S., Nivet M. L., Paoli C., Motte F., Fouilloy A. (2017) Machine Learning Methods for Solar Radiation Forecasting: a Review. *Renewable Energy*, 105, 569–582. <https://doi.org/10.1016/j.renene.2016.12.095>.
4. Bamm A. M., Khalyasmaa A. I., Eroshenko S. A., Matrenin P. V., Papkova N. A., Sekatski D. A. (2022) Topology Optimization of the Network with Renewable Energy Sources Generation Based on a Modified Adapted Genetic Algorithm. *Energetika. Izvestiya Vysshikh Uchebnykh Zavedenii i Energeticheskikh Ob'edinenii SNG = Energetika. Proceedings of CIS Higher Education Institutions and Power Engineering Associations*, 65 (4), 341–354. <https://doi.org/10.21122/1029-7448-2022-65-4-341-354> (in Russian).
5. Khalyasmaa A. I., Eroshenko S. A., Tashchilin V. A., Ramachandran H., Piepur Chakravarthi T., Butusov D. N. (2020) Industry Experience of Developing Day-Ahead Photovoltaic Plant Forecasting System Based on Machine Learning. *Remote Sensing*, 12 (20), 3420. <https://doi.org/10.3390/rs12203420>.
6. Eroshenko S. A., Khalyasmaa A. I., Snegirev D. A., Dubailova V. V., Romanov A. M., Butusov D. N. (2020) The Impact of Data Filtration on the Accuracy of Multiple Time-Domain Forecasting for Photovoltaic Power Plants Generation. *Applied Sciences*, 10 (22), 8265. <https://doi.org/10.3390/app10228265>.
7. Rana M., Koprinska I., Agelidis V. G. (2016) Solar Power Forecasting Using Weather Type Clustering and ensembles of Neural Networks *International Joint Conference on Neural Networks (IJCNN)*. Vancouver, BC, Canada: IEEE, 4962–4969. <https://doi.org/10.1109/IJCNN.2016.7727853>.
8. Zhang Z., Wang C., Peng X., Qin H., Lv H., Fu J., Wang H. (2021) Solar Radiation Intensity Probabilistic Forecasting Based on K-Means Time Series Clustering and Gaussian Process Regression *IEEE Access*, 9, 89079–89092. <https://doi.org/10.1109/ACCESS.2021.3077475>.
9. Feng C., Cui M., Hodge B. M., Lu S., Hamann H. F., Zhang J. (2021) Unsupervised Clustering-Based Short-Term Solar Forecasting. *IEEE Transactions on Sustainable Energy*, 10 (4), 2174–2185. <https://doi.org/10.1109/TSTE.2018.2881531>.
10. Sanz-Gorrachategui I., Pastor-Flores P., Guillén-Asensio A., Artal-Sevil J. S., Bono-Nuez A., Martín-del-Brio B., Bernal-Ruiz C. (2020) Unsupervised Clustering of Battery Waveforms in Off-Grid PV Installations. *2020 Fifteenth International Conference on Ecological Vehicles and Renewable Energies (EVER)*. Monte-Carlo, Monaco, 1–6. <https://doi.org/10.1109/EVER48776.2020.9242942>.
11. Kang M. C., Sohn J. M., Park J. Y., Lee S. K., Yoon Y. T. (2011) Development of Algorithm for Day Ahead PV Generation Forecasting Using Data Mining Method. *IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Seoul, Korea (South), IEEE, 1–4. <https://doi.org/10.1109/MWSCAS.2011.6026333>.
12. Wang F., Pang S., Zhen Z., Li K., Ren H, Shafie-Khah M., Catalão J. P. S. (2018) Pattern Classification and PSO Optimal Weights Based Sky Images Cloud Motion Speed Calculation Method for Solar PV Power Forecasting *IEEE Industry Applications Society Annual Meeting (IAS)*. Portland, OR, USA, IEEE, 1–9. <https://doi.org/10.1109/IAS.2018.8544468>.
13. Lai C. S., Jia Y., McCulloch M. D., Xu Z. (2017) Daily Clearness Index Profiles Cluster Analysis for Photovoltaic System *IEEE Transactions on Industrial Informatics*, 13 (5), 2322–2332 <https://doi.org/10.1109/TII.2017.2683519>.
14. Severiano C., Guimarães F. G., Cohen M. W. (2016) Very Short-Term Solar Forecasting Using Multi-Agent System Based on Extreme Learning Machines and Data Clustering. *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. Athens, Greece, IEEE, 1–8. <https://doi.org/10.1109/SSCI.2016.7850162>.
15. Yang Y., Dong L. (2013) Short-Term PV Generation System Direct Power Prediction Model on Wavelet Neural Network and Weather Type Clustering *2013 5th International Conference*

- on *Intelligent Human-Machine Systems and Cybernetics*. Hangzhou, China, IEEE, 207–211. <https://doi.org/10.1109/IHMSC.2013.56>.
16. Alfredo N., Ogliari E., Leva S., Pavan A. M., Mellit A., Lughi V., Dolara A. (2019) Day-Ahead Photovoltaic Forecasting: A Comparison of the Most Effective Techniques *Energies*, 12 (9), 1621. <https://doi.org/10.3390/en12091621>.
 17. *Meteoblue*. Available at: <https://www.meteoblue.com/en/weather>.
 18. Gorban, A. N., Kegl B., Wunsch D. C., Zinovyev A. (2008) *Principal Manifolds for Data Visualization and Dimension Reduction. Lecture Notes in Computational Science and Engineering, Vol 58*. Berlin, Springer. 364. https://doi.org/10.1007/978-3-540-73750-6_5.
 19. Hartigan J. A., Wong M. A. (1979) Algorithm AS 136: A K-Means Clustering Algorithm *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28 (1), 100–108. <https://doi.org/10.2307/2346830>.
 20. Matrenin P. V., Manusov V. Z., Khalyasmaa A. I., Antonenkov D. V., Eroshenko S. A., Butusov D. A. (2020) Improving Accuracy and Generalization Performance of Small-Size Recurrent Neural Networks Applied to Short-Term Load Forecasting *Mathematics*, 8 (12), 2169. <https://doi.org/10.3390/math8122169>.
 21. Matrenin P. V., Manusov V. Z., Igumnova E. A. (2020) Robust Short-Term Wind Speed Forecasting Using Adaptive Shallow Neural Networks. *Problems of The Regional Energetics*, 3 (47), 69–80. <https://doi.org/10.5281/zenodo.4018960> (in Russian).
 22. Clustering. *Scikit-Learn. Machine Learning in Python*. Available at: <https://scikit-learn.org/stable/modules/clustering.html>.
 23. Drucker H. (1997) Improving Regressors using Boosting Techniques. *Proceedings of the 14th International Conference on Machine Learning (ICML)*. Citeseer, 107–115.
 24. Adaptive Boosting. *Scikit-Learn. Machine learning in Python*. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDRegressor.html?highlight=adaptive+busting.