

УДК 621.31

Д. Е. Батуева, Я. Э. Шклярский, И. Е. Ревин
Санкт-Петербургский горный университет

Анализ данных генерации электроэнергии в ветродизельном комплексе с использованием алгоритма SSA

Проведено исследование разработанной авторами модели для анализа данных генерации электроэнергии в ветродизельном комплексе с использованием алгоритма SSA, позволяющего разложить временной ряд на простые компоненты: тренды, периодики, а также шумовые компоненты. Данные преобразования позволяют отобразить связь выработки электроэнергии комплексом электроснабжения, включающим в себя ветроустановку и дизель-генераторы, с климатическими факторами. В результате анализа графиков, полученных по шумовым компонентам, был сделан вывод, что в летний период наблюдается относительно стабильная выработка электроэнергии. Однако в весенне-осенний период наблюдаются помехи в данных, амплитуда шумовых компонент становится больше, что характеризует влияние резких климатических изменений на работу электрооборудования, а соответственно, и на неравномерность энергопотребления. Полученные результаты позволяют сделать поправки в системе управления на сезонность и, проанализировав шумовые компоненты, учесть влияние климатических условий на работу электрооборудования. Полученные восстановленные временные ряды корректно отображают динамику процесса. Для ряда генерации от ВЭУ коэффициент детерминации R^2 на 7 % больше по сравнению с рядом генерации от ДЭС. Это свидетельствует о том, что для анализа генерации электрической энергии с использованием ВЭУ, модель SSA корректна и более предпочтительна.

Метод главных компонент, анализ временных рядов, статистический анализ, машинное обучение, ветродизельный комплекс

В Энергетической стратегии России на период до 2035 г. [1] и в Стратегии экономической безопасности России на период до 2030 г. [2] отражена важность развития региональной энергетики на территории Арктики и Крайнего Севера, в том числе за счет возобновляемых источников энергии [3], [4]. Использование электротехнического комплекса, в составе которого присутствуют ветроэнергетическая установка, аккумуляторные батареи и дизель-генераторные установки [5], способно обеспечить надежное электроснабжение потребителей в соответствии с показателями качества электроэнергии [6]–[8]. В настоящий момент на объектах с изолированной системой электроснабжения не учитываются возможности повышения эффективности работы электротехнического комплекса посредством учета прогноза энергопотребления и анализа влияния изменения графиков

нагрузки в зависимости от продолжительности светового дня, сезонности, температуры окружающей среды [9], поэтому данное исследование посвящено анализу данных генерации электроэнергии с использованием алгоритма SSA (Singular spectrum analysis – анализ сингулярного спектра).

Из основ статистики известно, что временной (динамический) ряд – это статистические данные, собранные в различные периоды времени, и значения каких-либо параметров (в простейшем случае одного) исследуемых процессов [10]. Каждая единица таких статистических данных называется эталоном или мерой. Временной ряд существенно отличается от обычной выборки данных тем, что такой набор данных учитывает связь измерений со временем, а не только статистические разнообразие и характеристики выборки.

Тогда анализ временных рядов можно определить как совокупность методов математико-статистического анализа, предназначенных для определения структуры временных рядов и их последующего прогнозирования [11]. Анализ таких временных рядов также можно рассматривать как задачу машинного обучения. Имеется набор пар признаков x и целевых переменных $y\{(x_i, y_i), i = 1, \dots, n\}$ для получения искомой функциональной зависимости $y = f(x)$. Чтобы получить искомую функциональную зависимость $y = f(x)$, нужно минимизировать функцию потерь $L(y = f(x))$ [12].

Задачи в подобных исследованиях направлены на создание прогнозных значений необходимых измеряемых характеристик исследуемого объекта на заданный период времени. В настоящее время существует несколько различных методов прогнозирования, которые делятся на два основных класса (локальные и глобальные) по области параметров аппроксимирующей функции, при этом следующие значения временных рядов рекуррентно устанавливаются в соответствии с несколькими предыдущими.

Изначально были разработаны такие глобальные методы, в которых на основе статистического анализа предлагалось использовать авторегрессию, скользящее среднее и т. д. [13]. В дальнейшем в рамках нелинейной динамики были также разработаны некоторые практические методы:

- сингулярный спектральный анализ (SSA) – глобальный метод;
- локальная аппроксимация (LA);
- комбинация SSA–LA.

В данной статье для анализа временных рядов был выбран метод сингулярного спектрального анализа. Метод SSA, он же метод главных компонент, нашел широкое применение во многих технических сферах, в том числе и в электроэнергетике, как метод обработки и анализа большого количества данных, фиксируемых в электротехническом комплексе, на основе которого возможно прогнозировать выработку электроэнергии, учитывая влияние различных групп факторов.

Факторы, которые необходимо учитывать при построении прогнозных моделей, условно можно разделить на три группы: циклические, естественные и случайные. Такая классификация

отражает характер их возникновения и возможность с точки зрения наличия данных для их включения в прогнозную модель. По типу возникновения все факторы технологической среды делятся на метеорологические и социально-экономические. Каждый из них по-разному влияет на точность прогноза: температура внешней окружающей среды и порядковый номер недели в году и номер дня недели также повышают точность прогноза.

Для анализа временных рядов в качестве входного параметра алгоритма SSA в зависимости от задачи исследования выбирается длина окна L . Так, при существенно большей длине временного ряда и большом L получаемые после проведенного анализа результаты не будут зависеть от длины окна. Далее на основе полученного временного ряда строится траекторная матрица X , столбцами которой являются «скользящие» отрезки длины L : от X_1 до L , от X_2 до $(L + 1)$ и т. д. Исходный временной ряд преобразуется в последовательность многомерных векторов. Дальнейший анализ проводится на основе мультипликативного временного ряда, так как амплитуда колебаний изменяется в исследуемом диапазоне времени:

$$X = \alpha\beta\varepsilon,$$

где α – значение функции тренда; β – циклическая аперiodическая составляющая; ε – шум.

Затем проводится сингулярное разложение траекторной матрицы на сумму элементарных векторов (SVD). Каждая элементарная матрица раскладывается на «элементарные тройки», которые определяются набором собственных значений и двумя сингулярными векторами [14], [15]:

$$X_i = \sqrt{\lambda_i} U_i V_i^T,$$

где λ_i – собственные значения матрицы (сингулярные числа); U_i, V_i – сингулярные векторы матрицы.

Если представить исходный ряд как сумму нескольких рядов, то по форме собственных векторов и собственным значениям матрицы можно определить, что представляют собой эти члены и каким элементарным матрицам они соответствуют. Затем, суммируя элементарные матрицы внутри каждого множества и переходя от результирующих матриц к временному ряду, получим разбиение исходного ряда на аддитивные компо-

ненты, например сумму тренда, цикла и шума или сумму низкочастотных и высокочастотных составляющих.

Таким образом, целью данного метода служит разложение исходного ряда на совокупность составляющих его компонент. К тому же для использования этого метода не обязательна стационарность ряда, понимание пользователем тренда, сезонной и циклической составляющих ряда. Как раз в результате исследования могут быть обнаружены тренд, циклы и возможности сглаживания ряда.

Таким образом, перед анализом данных методом SSA для исходных данных необходимо решить следующие задачи:

1. Выбор длины окна L .
2. Наличие дополнительной информации о структуре данных исходного ряда и периодических компонент в нем.

Далее приведены некоторые рекомендации по решению этих задач.

При одновременном выборе различных компонент серии следует учитывать все аспекты выбора длины окна. Выбор длины окна необходимо начинать со значения, близкого к половине длины временного ряда и кратного периодической составляющей ряда, если таковая присутствует во временном ряду и его период известен. Для коротких рядов или рядов со сложной структурой такие рекомендации могут не подойти. Тогда выбор длины окна становится более сложной задачей.

Если отсутствует какая-либо дополнительная информация о данных ряда, из которых он состоит, то сам ряд может быть единственным источником информации для выбора длины окна. Наличие тренда или явная периодичность могут отражать свойства временного ряда. Однако наиболее правильным в данном случае будет выбор длины окна на основе анализа периодограмм только в случае, если длина ряда достаточно велика. К тому же наличие различных частотных диапазонов с высокими показателями амплитуды может отражать возможные естественные составляющие исходного ряда в соответствии с этими частотами.

Постановка задачи. В рамках исследования была разработана модель для анализа данных генерации электроэнергии в ветродизельном ком-

плексе с использованием алгоритма SSA, позволяющего разложить временной ряд на составляющие компоненты: тренды, периодики, а также шумовые компоненты. Данные преобразования позволят отобразить связь выработки электроэнергии с климатическими факторами. Анализ проводился в такой последовательности:

1. Преобразование одномерного ряда в многомерный на основе построения траекторной матрицы и многомерных векторов.
2. Применение сингулярного разложения для исследования полученной многомерной траекторной матрицы.
3. Получение исходного восстановленного ряда по выбранным главным компонентам.

На основе такого разложения можно разработать модель для прогнозирования как всего ряда, так и его компонент отдельно [16]. Выбор длины исходного ряда и числа его компонент (ширина окна) существенно влияет на получаемый результат.

Исходные данные загружались в программу в форме таблицы из 9 колонок и 366 строк, содержащей следующую информацию, отражающую объемы генерации электроэнергии и влияние окружающей среды на исследуемый объект:

- день, месяц, день недели;
- выработка электроэнергии ДЭС (дизель-электростанция), кВт · ч;
- выработка электроэнергии ВЭУ (ветроэнергетическая установка), кВт · ч;
- время работы ДЭС, ч;
- время работы ВЭУ, ч;
- скорость ветра максимальная, м/с;
- скорость ветра средняя, м/с;
- температура окружающей среды средняя, °С.

Для каждого объекта, имеющего в своем составе ветроэнергетическую установку, необходимо вводить поправочные коэффициенты на условия работы, связанные с климатическими условиями. Определяются максимальная и минимальная температуры окружающей среды, комфортные для работы установок. Задается диапазон скоростей ветра, при которых ветроустановка будет работать с максимальной выработкой электроэнергии. На данном объекте скорость страгивания ВЭУ составляет 2.5 м/с, номинальная скорость – 11 м/с, а максимальная – 25 м/с.

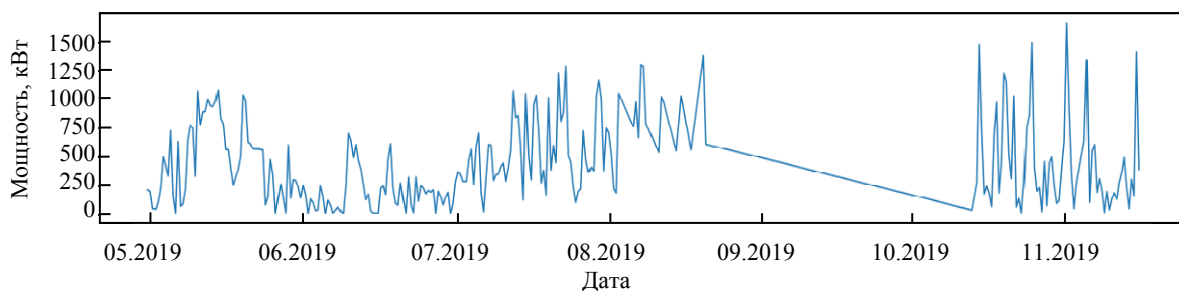


Рис. 1

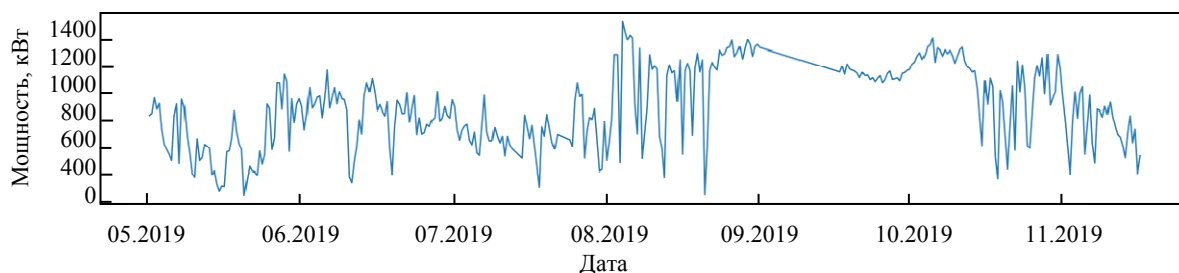


Рис. 2

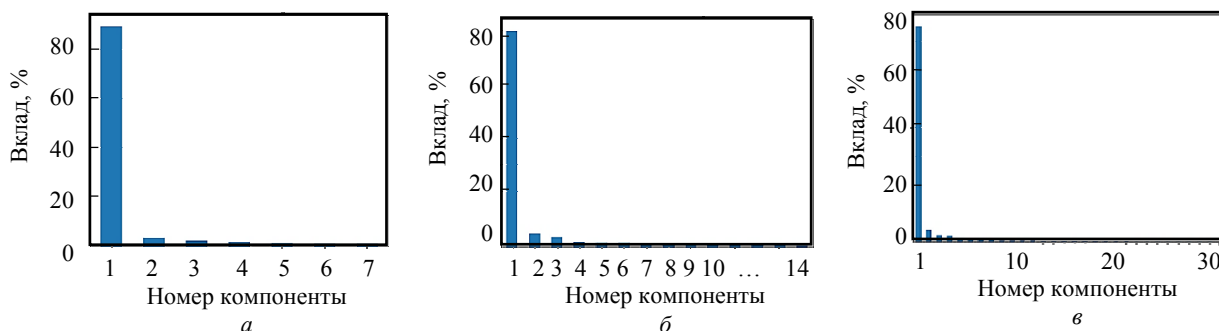


Рис. 3

Далее были построены графики выработки электроэнергии ВЭУ (рис. 1) и ДЭС (рис. 2), что показывает общий вид генерации электроэнергии в течение всего периода измерений, а при большом объеме исходной базы данных – присутствуют ли пропуски в данных, как на изучаемом объекте.

В данном исследовании для проведения анализа данных будет рассматриваться период с 01.05. до 31.12.2019 г., так как отсутствуют данные с 01.01.2020 г. по 31.01.2020 г. для ВЭУ и ДЭС. С 11.12.2019 г. по 23.03.2020 г. ВЭУ находилась в ремонте, однако задача восстановления потерянных значений в статье не стоит.

При выборе длины окна L руководствовались необходимой точностью схожести с исходным рядом при наименьшем количестве компонент. Чтобы понять, при какой ширине окна восстановленный ряд будет максимально соответствовать исходному, было проведено сравнительное моде-

лирование при различной длине окна – 7, 14, 30 компонент.

Анализ выработки ВЭУ. Далее приводятся результаты проведенного анализа для ветроэнергетической установки. После анализа компонент производился SSA разложения. С помощью команды `model.ssa_contrib()` можно посмотреть, насколько (в процентах) полученные компоненты восстановленного ряда схожи с исходным рядом (рис. 3).

Как видно из графиков, на первую компоненту приходится 82 % описания исходного ряда при $L = 7$ (рис. 3, а), 80 % при $L = 14$ (рис. 3, б), 75 % при $L = 30$ (рис. 3, в). Соответственно, можно сделать вывод, что чем больше собственное значение главной компоненты, тем больше вклад соответствующей восстановленной компоненты [17]. Данное значение влияет и на ошибку, и на соответствие полученной модели реальным данным.

Далее с помощью команды `model.plot_ssa_values()` можно отразить в виде графиков каждую компоненту восстановленного ряда. При длине окна $L = 7$ (рис. 4–10, а) первая компонента восстановленного ряда определяет тренд, по которому происходит выработка электроэнергии ветряком, вторая компонента отражает циклическую апериодическую составляющую, что позволяет обнаружить периодичности, провести дальнейшие корректировки на сезонность, а компоненты

с третьей по седьмую отражают шум, который, анализируя данные графики, можно сгладить или подавить. При $L = 14$ тренд также определяется 1-й и 2-й компонентами, которые в сумме дают 84 %, апериодическая составляющая – 3-й и 4-й компонентами, шум – 5–14-й компонентами. При $L = 30$ тренд определяется 1-й и 2-й компонентами, апериодическая составляющая – 3-й и 4-й компонентами, а шум – 5–30-й компонентами.

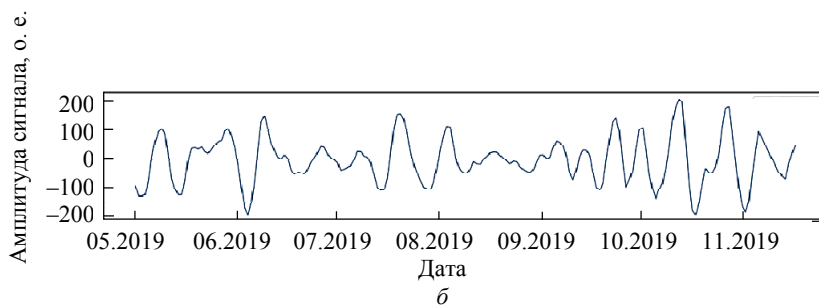
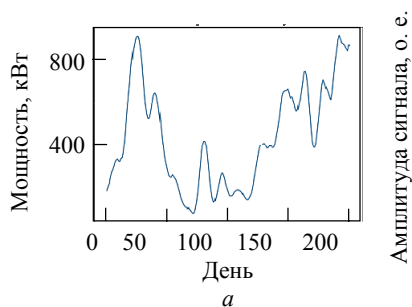


Рис. 4

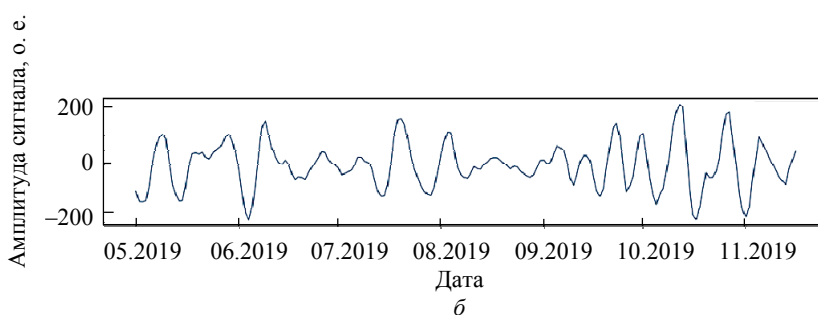
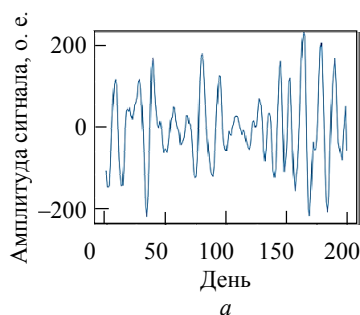


Рис. 5

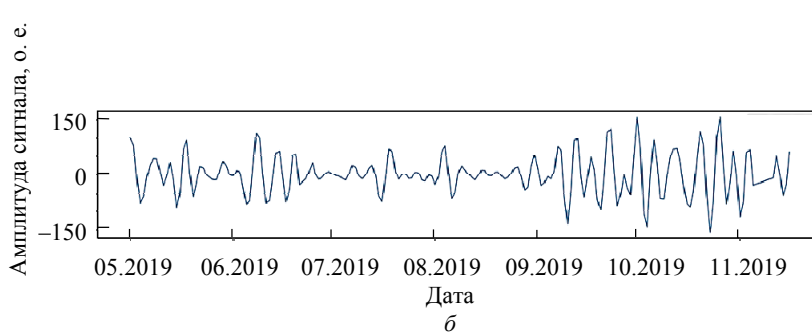
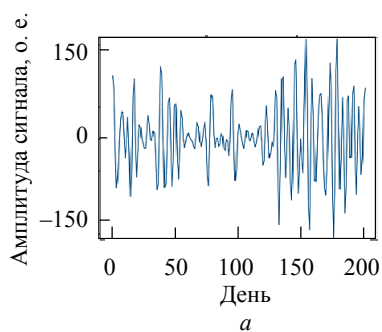


Рис. 6

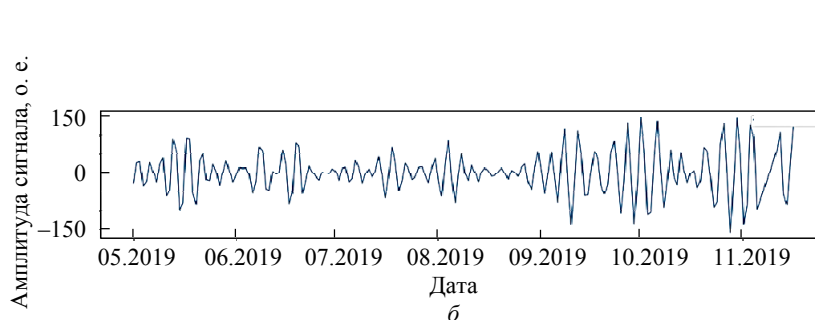
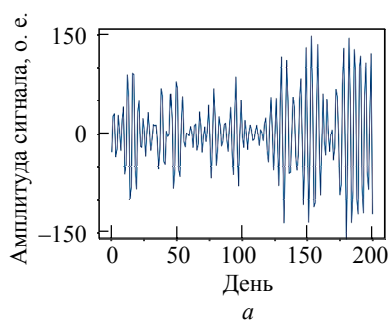


Рис. 7

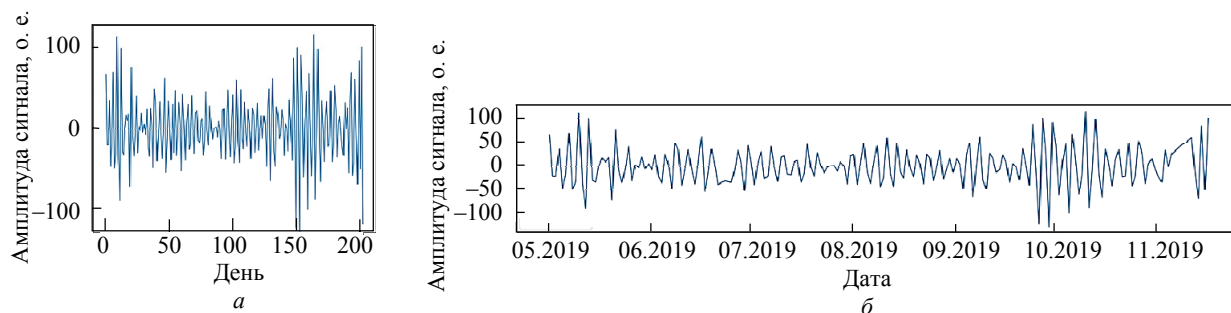


Рис. 8

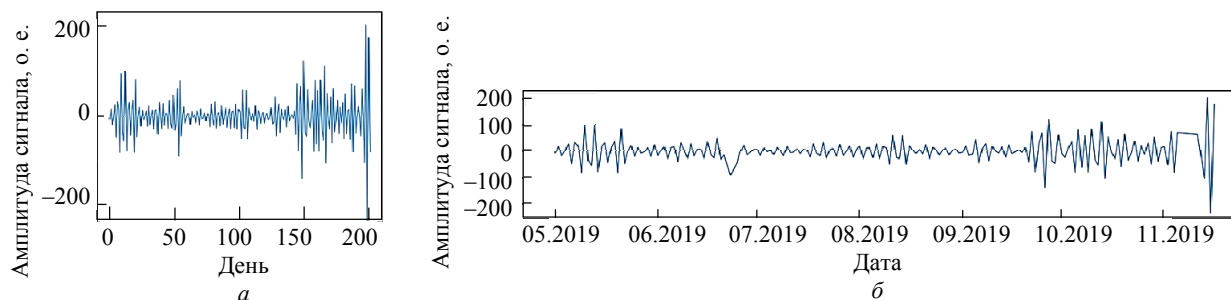


Рис. 9

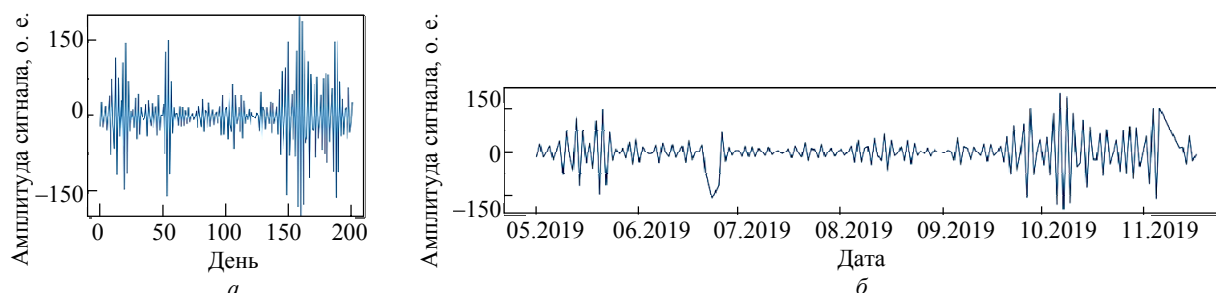


Рис. 10

Затем с помощью команды *model.plot_time_series()* сформированы графики восстановленного ряда выработки электроэнергии в рассматриваемый период времени (рис. 4–10, б).

Тот же алгоритм выполнялся и для анализа временного ряда выработки электроэнергии дизель-генератором.

Анализируя графики, полученные по шумовым компонентам, можно сделать вывод, что летом, с июня по август, наблюдается относительно стабильная выработка электроэнергии. В течение данного периода на выработку электроэнергии климатические параметры практически не влияют. Однако в весенне-осенний период наблюдаются помехи в данных, амплитуда шумовых компонент становится больше, что характеризует влияние резких климатических изменений на работу электрооборудования, и, соответственно, неравномерность энергопотребления. Отклонение реального энергопотребления может составлять от 11 до 56 % от планируемых графиков нагрузки.

Средняя температура с каждым месяцем существенно изменяется, что влечет за собой сдвиг графика нагрузки.

Как было установлено до проведения данного исследования, среднее отклонение месячного электропотребления на объекте составляет 14.5 %, что может отрицательно влиять на режимы работы комплекса энергоснабжения и приводить к недоотпуску электроэнергии потребителям без прогнозирования энергопотребления. Характерно также снижение электрической нагрузки в летний период времени. Разница в объемах потребления электроэнергии достигает 2.5 раз, если сравнивать самый теплый месяц (июль) с самым холодным (декабрь).

Данная методика анализа энергопотребления объекта позволит более эффективно планировать работу комплекса электроснабжения. Зная тренд и месяцы, когда отклонений в энергопотреблении не ожидается, и построив модель регрессии в периоды, когда ожидаются отклонения энергопо-

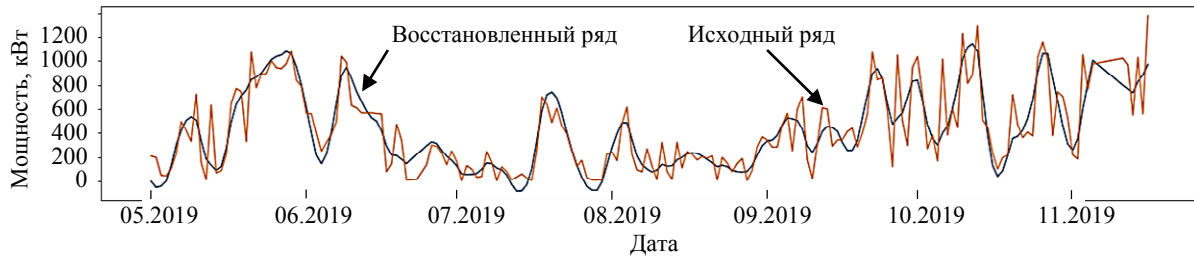


Рис. 11

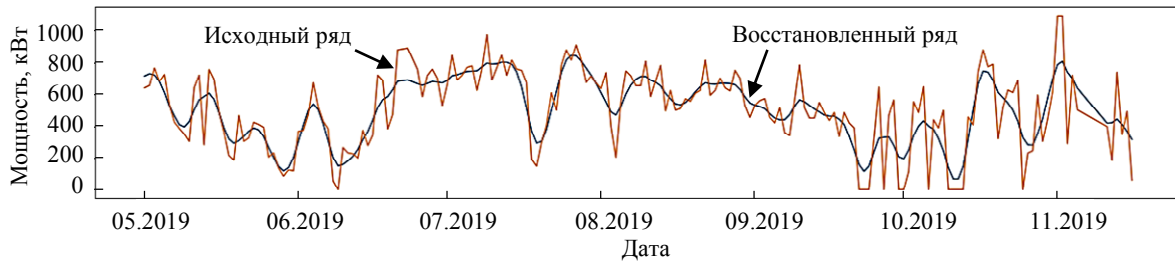


Рис. 12

требления от планируемых значений, загрузив на вход климатические параметры и данные прогноза, можно настраивать в системе автоматического управления режимы работы комплекса электрооборудования, распределять соотношения генерирующих мощностей в энергокомплексе. Машинное обучение позволит на основе базы данных, собранных в процессе эксплуатации комплекса, анализировать влияние климатических факторов как на генерацию электроэнергии, так и на энергопотребление.

Результаты. Сравнение полученных графиков выработки электроэнергии ДЭС и ВЭУ позволяет подтвердить правильность восстановленных данных, так как при выработке электроэнергии от ВЭУ, достаточной для покрытия спроса, ДЭС находится в резерве, а при нехватке выработки от ВЭУ подключается ДЭС. Полученные графики позволяют также сделать поправки в системе управления на сезонность, а проанализировав шумовые компоненты, учесть влияние климатических условий на работу электрооборудования.

На рис. 11, 12 показано сравнение исходных временных рядов ВЭУ и ДЭС и восстановленных по суммам их первых двух компонент.

Были посчитаны средняя абсолютная ошибка, средняя квадратичная ошибка и коэффициент детерминации R^2 . Коэффициент детерминации определяет соответствие полученной модели реальным данным и может принимать значения в диапазоне от 0 до 1. Чем ближе коэффициент к 1, тем больше зависимость. Полученные значения указаны в таблице.

Метрика качества модели	Восстановленный ряд ВЭУ	Восстановленный ряд ДЭС
Средняя абсолютная ошибка	125.577	103.324
Средняя квадратичная ошибка	166.985	138.524
Коэффициент детерминации (R^2)	0.753	0.685

В результате анализа полученных собственных векторов, собственных значений и метрик качества работы модели SSA можно сделать следующие выводы.

Полученные компоненты разложения исходного временного ряда интерпретируемы с точки зрения физики наблюдаемого процесса и позволяют проводить более детализированные исследования режимов работы комплекса электрооборудования.

Используя комбинации выбранных компонент разложения, получили восстановленные временные ряды, которые корректно отображают динамику процесса. Следует отметить, что для ряда ВЭУ коэффициент детерминации R^2 на 7 % больше по сравнению с рядом ДЭС. Это свидетельствует о том, что для анализа генерации электрической энергии с использованием ВЭУ модель SSA корректна и более предпочтительна.

Использованная техника анализа данных временных рядов предпочтительна для данной области по сравнению с другими классическими тех-

никами (авторегрессионные модели, скользящие окна и т. д.). Это связано с возможностью выбора числа компонент разложения, их детального

анализа в дальнейшем и получения скрытой информации об особенностях динамики наблюдаемого процесса во времени.

СПИСОК ЛИТЕРАТУРЫ

1. Энергетическая стратегия Российской Федерации на период до 2035 г. URL: <http://static.government.ru/media/files/w4sigFOiDjGVDYT4lgsApssm6mZRb7wx.pdf> (дата обращения 02.02.2021).
2. Стратегия экономической безопасности Российской Федерации на период до 2030 года. URL: <http://static.kremlin.ru/media/acts/files/0001201705150001.pdf> (дата обращения 02.02.2021).
3. Cherepovitsyn A., Tsvetkov P. Overview of the prospects for developing a renewable energy in Russia // Proc. of 2017 Intern. Conf. on Green Energy and Applications. URL: <https://www.semanticscholar.org/paper/Overview-of-the-prospects-for-developing-a-energy-Cherepovitsyn-Tsvetkov/65269c0eec7a37ab7560b9ce963d138a193d1b0c> (дата обращения 02.02.2021).
4. Kirsanova N. Y., Lenkovets O. M., Nikulina A. Y. The role and future outlook for renewable energy in the Arctic zone of Russian Federation // European Research Studies J. 2018. Vol. 21. P. 356–368.
5. Vasilkov O. S., Dobysh V. S. Features of application hybrid energy storage in power supply systems // Proc. of the 2019 IEEE Conf. of Russian Young Researchers in Electrical and Electronic Engineering. URL: https://www.researchgate.net/publication/331853935_Features_of_Application_Hybrid_Energy_Storage_in_Power_Supply_Systems (дата обращения 04.02.2021).
6. Lavrik A., Zhukovskiy Y., Buldysko A. Features of the optimal composition determination of energy sources during multi-criterial search in the russian arctic conditions // Proc. of the 2nd 2020 Intern. Youth Conf. on Radio Electronics, Electrical and Power Engineering, REEPE 2020. 9059215. URL: <https://iopscience.iop.org/article/10.1088/1742-6596/1753/1/012059/pdf> (дата обращения 04.02.2021).
7. Абрамович Б. Н., Бельский А. А. Выбор параметров ветродизельной установки для энергообеспечения минерально-сырьевого комплекса // Зап. Горного ин-та. 2012. Т. 195. С. 227–230.
8. Паньков И. А., Фролов В. Я. Повышение качества электроэнергии в автономных электроэнергетических системах // Зап. Горного ин-та. 2017. Т. 227. С. 563–568.
9. Energy resources and electricity generation in Arctic areas / M. de Witt, H. Stefánsson, A. Valfells, J. N. Larsen // Renewable Energy. 2021. Vol. 169. P. 144–156.
10. Мишулина О. А. Статистический анализ и обработка временных рядов / МИФИ. М., 2004.
11. Подкорытова О. А., Соколов М. В. Анализ временных рядов: учеб. пособие для бакалавриата и магистратуры. 2-е изд., перераб. и доп. М.: Юрайт, 2018.
12. Господариков А. П., Морозов К. В., Ревин И. Е. О методе обработки данных сейсмического и деформационного мониторинга при ведении подземных горных работ на примере Киксумчоррского месторождения АО «Апатит» // Горный информ.-аналит. бюл. 2019. № 8. С. 157–168.
13. Лоскутов А. Ю., Козлов А. А., Хаханов Ю. М. Энтропия и прогноз временных рядов в теории динамических систем // Изв. вузов. Прикладная нелинейная динамика. 2009. Т. 17, № 4. С. 98–113.
14. Golyandina N., Korobeynikov A. Basic singular spectrum analysis and forecasting with R // Computational Statistics and Data Analysis. 2014. Vol. 71. P. 934–954.
15. Golyandina N. On the choice of parameters in singular spectrum analysis and related subspace-based methods // Statistics and Its Interface. 2010. Vol. 3. P. 259–279.
16. Оценка показателя Ляпунова с использованием SSA-метода / Н. Д. Поляхов, А. В. Беспалов, О. Э. Якупов, Д. М. Филатов // Естественные и технические науки. 2011. № 4 (54). С. 23–28.
17. Леонтьева Л. Н. Многомерная гусеница, выбор длины и числа компонент // Машинное обучение и анализ данных. 2011. Т. 1, № 1. С. 5–15.

D. E. Batueva, Ya. E. Shklyarskiy, I. E. Revin
Saint Petersburg Mining University

ANALYSIS OF ELECTRIC POWER GENERATION DATA IN A WIND-DIESEL COMPLEX USING THE SSA ALGORITHM BASED

Studies a model developed by the authors for analyzing data on the generation of electricity in a wind-diesel complex using the SSA algorithm, which allows decomposing a time series into simple components: trends, periodicals, and noise components. These transformations will make it possible to display the relationship between the generation of electricity by the power supply complex, which includes a wind turbine and diesel generators, with climatic factors. After analyzing the graphs obtained for the noise components, it was concluded that relatively stable power generation is observed in the summer period. However, there is interference in the data in the spring-autumn period, the amplitude of the noise components becomes larger, which characterizes the influence of sharp climatic changes on the operation of electrical equipment, and, accordingly, the unevenness of energy consumption. The results obtained make it possible to make adjustments in the control system for seasonality, and by analyzing the noise components, it is possible to take into account the influence of climatic conditions on the operation of electrical equipment. The obtained reconstructed time series correctly reflect the dynamics of the process. The coefficient of determination R^2 for a series of generation from a wind turbine is 7 % higher compared to a series of generation from a diesel power plant. This indicates that the SSA model is correct and more preferable for analyzing the generation of electrical energy using wind turbines.

Principal component analysis, time series analysis, statistical analysis, machine learning, wind-diesel complex
