

УДК 004.89

doi:10.38028/ESI.2022.25.1.008

Прогнозирование солнечной радиации и импутация данных для цифрового двойника солнечной электростанции

Лосев Алексей Сергеевич, Массель Алексей Геннадьевич

Институт систем энергетики им. Л.А. Мелентьева СО РАН,

Россия, Иркутск, *losev.as@isem.irk.ru*

Аннотация. В статье представлена реализация авторегрессионной модели для прогнозирования данных о солнечной радиации. Был проведен эксперимент с некоторыми подходами к импутации данных¹ и обоснован выбор подхода MICE (Multiple Imputation by Chained Equation – множественная импутация с помощью связанного уравнения). Также был проведен сравнительный эксперимент прогнозирования солнечной радиации на основе первоначальных данных и на основе данных, в которых была применена импутация.

Ключевые слова: цифровой двойник, импутация, солнечная радиация, авторегрессия, прогнозирование, ARIMA, SARIMA, MICE

Цитирование: Лосев А.С., Массель А.Г. Прогнозирование солнечной радиации и импутация данных для цифрового двойника солнечной электростанции // Информационные и математические технологии в науке и управлении. – 2022. – № 1 (25). – С. 91-101. – DOI:10.38028/ESI.2022.25.1.008.

Введение. Выделяют два этапа разработки цифровых двойников: 1) разработка прототипа с использованием научного инструментария исследователей, 2) отладка информационных взаимодействий с реальными информационными потоками для решения конкретных задач [1]. Работа с реальными данными допустима только на втором этапе, потому появляется необходимость в создании генератора правдоподобных синтетических данных, которые впоследствии могли бы использоваться во время тестирования цифрового двойника уже на первом этапе. В Институте систем энергетики им. Л.А. Мелентьева СО РАН было принято решение разработать модель генерации и импутации данных для цифрового двойника солнечной электростанции на основе имеющихся погодных данных, полученных от реального объекта.

В настоящее время продолжает возрастать роль солнечной энергии в производстве электроэнергии, поскольку возобновляемые источники энергии становятся многообещающей альтернативой для мира в целом [2]. Однако, если речь идет об установке солнечных электростанций в специфических регионах, например, на Крайнем севере, необходимо обосновать целесообразность такого решения. В качестве обоснования в статье предложен подход к прогнозированию солнечного излучения для лучшего представления потенциального генерируемого количества электроэнергии отдельно взятой системы солнечных панелей.

Существуют несколько моделей прогнозирования для предсказания будущего излучения, в том числе, включающие в себя технологии искусственного интеллекта, например, модель численного прогноза погоды (NWP), которая описывается нелинейными уравнениями в частных производных погодных переменных, с применением временных рядов в качестве фиксации динамики излучения [2, 3]. Эта технология, возможно, будет реализована в будущем, когда возникнет необходимость генерировать значения комплексно, с последующим вычислением мощности в зависимости не только от солнечного излучения, но и от влажности,

¹Процесс импутации - это замещение ошибочных, противоречивых и отсутствующих ответов в процессе редактирования данных другими ответами - значениями показателей.

скорости ветра и т.д. Приоритет первого прототипа будет больше сосредоточен на импутации пропущенных данных.

Оценка параметров в модели временных рядов определенно требует полного набора исторических данных. Данные зачастую содержат пропущенные значения. Более того, отсутствие данных может содержать пропуски за несколько дней подряд. Эту проблему возможно решить, заполнив значения записей с других датчиков [3, 2], однако нет никакой гарантии того, что запасные датчики также не выйдут из строя. Более продвинутые методы множественной импутации, например, анализ главных компонент, реализованы в MATLAB в качестве набора инструментов [4]. Однако, такие методы обычно предполагают, что данные являются статическими случайными величинами и не используют временную динамику данных применительно к временным рядам [3].

В этой работе сравниваются практические методы, подходящие для импутации большого объема пропущенных данных, реализованные на языке Python. Также предложен метод прогнозирования солнечной радиации как временного ряда с помощью модели авторегрессии ARIMA.

1. Прогнозирование солнечной радиации. Авторегрессионная интегрированная модель скользящего среднего (ARIMA) можно использовать для объяснения динамики нестационарных временных рядов. Эта модель популярна у исследователей за счет перспективы ее интеграции с нейронными сетями [5, 6].

Пусть $y(t)$ будет временным рядом солнечной радиации. Определим L как оператор запаздывания: $Ly(t) = y(t - 1)$, где шкала времени – каждый час. Модель ARIMA (p, d, q) описывается [7, 8]:

$$A(L)(1 - L)^d y(t) = C(L)v(t) \quad (1)$$

где полиномы авторегрессии и скользящего среднего

$$A(L) = 1 - (a_1L + \dots + a_pL^p), C(L) = 1 + c_1L + \dots + c_qL^q, \text{ соответственно.}$$

Член $(1 - L)^d$ является интегральной частью, где d – порядок разности, который может быть определен заранее из автокорреляционного анализа. Выполняем дифференцирование $\nabla y(t) = y(t) - y(t - 1)$ до тех пор, пока автокорреляционная функция (АКФ) результирующего временного ряда оказывается стационарной. Переменная $v(t) \in \mathbb{R}$ – белый шум с нулевым средним и дисперсией σ^2 . Интуитивно понятно, что солнечное излучение имеет суточный цикл. Предположим, что мы пренебрегаем данными ночных часов, поскольку значения освещенности равны нулю. Это приводит к временному ряду с периодом 12 часов. В качестве примера, который иллюстрирует метод определения сезонности данных с использованием декомпозиция команды в \mathbb{R} , необходимо построить тренд временного ряда

$$\hat{m}(t) = (1/16)[(1/2)y(t - 8) + y(t - 7) + \dots + y(t + 7) + (1/2)y(t + 8)]$$

и вычесть его значение из значения солнечной радиации: $\tilde{s}(t) = y(t) - \hat{m}(t)$ (измерения собираются ежечасно). Оценку суточного сезонного эффекта $\hat{s}(t)$ можно получить по формуле усреднения $\tilde{s}(t)$:

$$\hat{s}(t) = \frac{1}{N + 1} \sum_{k=0}^N \tilde{s}(t + kT)$$

где T – сезонный период. Рассмотрим модель, содержащую аддитивную сезонную индикаторную переменную согласно [7]:

$$A(L)(1 - L)^d y(t) = s(t) + \alpha + C(L)v(t),$$

где $s(t) = s(t - kT)$ – сезонный член, повторяющийся каждый цикл, k является целым числом, а α является константой. Если $y(t)$ имеет период T , то член $s(t)$ можно убрать операцией $y(t) - y(t - T) = (1 - L^T) y(t)$. Так как ARIMA используется для прогнозирования солнечной радиации с учетом сезонности, применяется сезонная модель ARIMA

$$\tilde{A}(L^T)A(L)(1 - L^T)D(1 - L)^d y(t) = \tilde{C}(L^T)C(L)v(t), \quad (2)$$

где $\tilde{A}(L)$ и $\tilde{C}(L)$ являются сезонными авторегрессионными полиномами скользящего среднего:

$$\begin{aligned} \tilde{A}(L) &= I - (\tilde{a}_1 L^T + \dots + \tilde{a}_p L^T), \\ \tilde{C}(L) &= I + \tilde{c}_1 L^T + \dots + \tilde{c}_q L^T \end{aligned}$$

D – интегрированный сезонный порядок. Описание сезонной модели ARIMA обычно обозначается SARIMA(p, d, q)(P, D, Q)_T.

Оценка моделей SARIMA требует выбора правильного значения модельных порядков p, d, q, P, D, Q. В ходе эксперимента было установлено, что T = 12, из предварительных данных о солнечной радиации. Другие параметры выбираются таким образом, чтобы преобразованный временной ряд имел стационарную характеристику, которая может наблюдаться на графиках АКФ, и оценка модели была сведена к минимуму. Когда параметры для нестационарных частей (d, D, P, Q) определены, остается оценить часть модели ARMA. В качестве наглядного примера можно привести график АКФ необработанных данных о солнечной радиации. На рис. 1 (сверху) видно, что данные не являются стационарным процессом, т.к. АКФ периодична каждые 12 лагов (сдвигов во времени), а также медленно затухает после 1 лага. Чтобы найти подходящий параметр d, разность на y применяется так, чтобы АКФ разностного ряда вела себя аналогично АКФ стационарного процесса по следующим характеристикам [8].

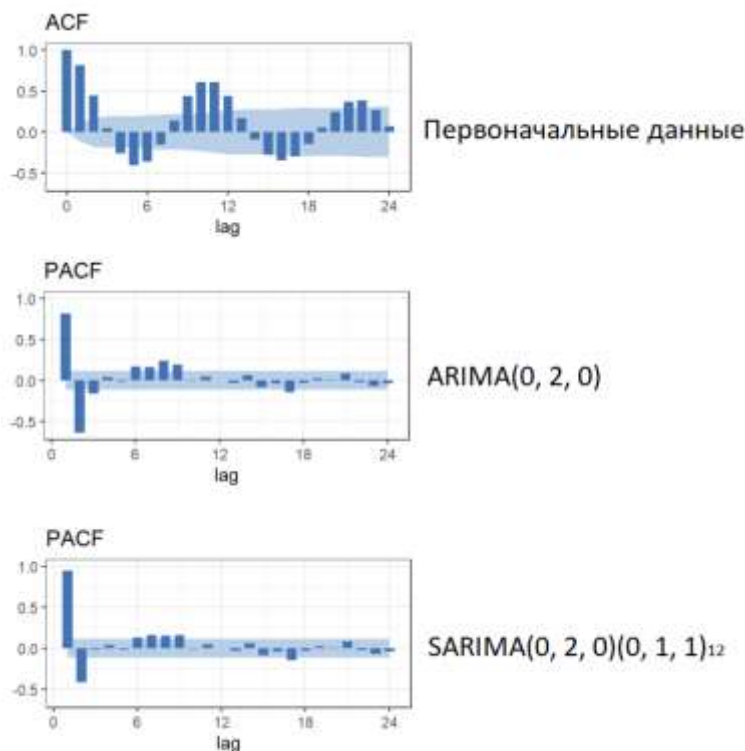


Рис. 1. Автокорреляционная функция (АКФ) первоначальных данных (вверху), АКФ остатка ARIMA (0,2,0) (в центре) и АКФ остатка SARIMA (0, 2, 0)(0, 1, 1)₁₂ (внизу)

Обнаружено, что $d = 2$ является подходящим выбором для этого. Набор данных и результаты АКФ, представленные на рис. 1 (в центре), показывают резкое снижение АКФ. Однако, некоторые лаги АКФ малы и лежат за пределами доверительного интервала. Далее задаем параметры $(P, D, Q)_T = (0, 1, 1)_{12}$ сезонных членов и это дает лишь несколько лагов АКФ, лежащих за пределами доверительного интервала, как показано на рис. 1 (внизу). Следовательно, соответствующий выбор p и q может быть найден путем минимизации информационного критерия Акаике (AIC) или байесовского информационного критерия (BIC) в качестве критериев выбора модели [8].

2. Импутация данных. После того, как определен метод прогнозирования временных рядов, встает вопрос о его работе с поврежденными данными. Модель SARIMA требует числовые входные значения, которые должны присутствовать для каждой строки и столбца в наборе данных. Таким образом, пропущенные значения могут вызвать проблемы во время получения прогноза.

Таким образом, обычно выявляют отсутствующие значения в наборе данных и заменяют их подходящим числовым значением. Это называется импутацией данных. В российских источниках термин «импутация» встречается в социологических исследованиях [9]. Вместо импутации можно также встретить термин «восстановление пропущенных значений», например [10].

Простой и популярный подход импутации данных включает использование статистических методов для оценки значения столбца на основе имеющихся значений, а затем замену всех отсутствующих значений в столбце рассчитанными статистическими. Он прост, потому что статистические данные подсчитываются быстро, и популярен, потому что часто оказывается очень эффективным [11].

Рассчитываемые общие статистические данные включают: среднее арифметическое значение столбца; медиану столбца; моду столбца; константу.

Необходимо было провести эксперимент для проверки каждой статистической стратегии и выяснить, что лучше всего работает для этого набора данных, сравнивая среднее значение, медиану, моду (наиболее частую) и постоянную (0) стратегии. Затем можно сравнить среднюю точность каждого подхода. Перед экспериментом было удалено 20% первоначальных погодных данных с различными интервалами, от одного часа (одна запись в базе данных) до суток (двадцать четыре записей подряд в базе данных). В качестве константы было выбрано значение «0» (отсутствие солнечной радиации, ночное время суток).

В результатах на рис. 2 видны смещения во всех четырех классических стратегиях. Самой лучшей стратегией модель выбрала вариант с константой, однако, в самих данных наблюдается неправдоподобность. Несмотря на хорошие результаты, провал эксперимента обусловлен потерей контекста данных – пропадает характер сезонности, а именно, нулевая радиация на протяжении нескольких суток.

Потому было принято решение использовать любой из двух специализированных методов: максимальное правдоподобие или множественную импутацию.

В статистике метод максимального правдоподобия – это статистическая оценка интересующего распределения, обладающего тем свойством, что он максимизирует функцию правдоподобия этих данных [12].

Статистическая оценка берет случайную выборку данных и пытается выяснить что-то об общем распределении, обобщая эту выборку. Например, $\sum(y)/len(y)$ является оценкой среднего значения набора данных y . Это метод максимального правдоподобия, потому что он не имеет смещения: он сходится к истинному среднему значению распределения (при достаточно большом количестве выборок). Для большинства задач метод максимального правдоподобия является простейшим оценщиком, но иногда оценка с его помощью невозможна, а в других случаях полезна некоторая степень смещения в оценке.

Импутация максимального правдоподобия – это оценка максимального правдоподобия, применяемая к отсутствующим данным.

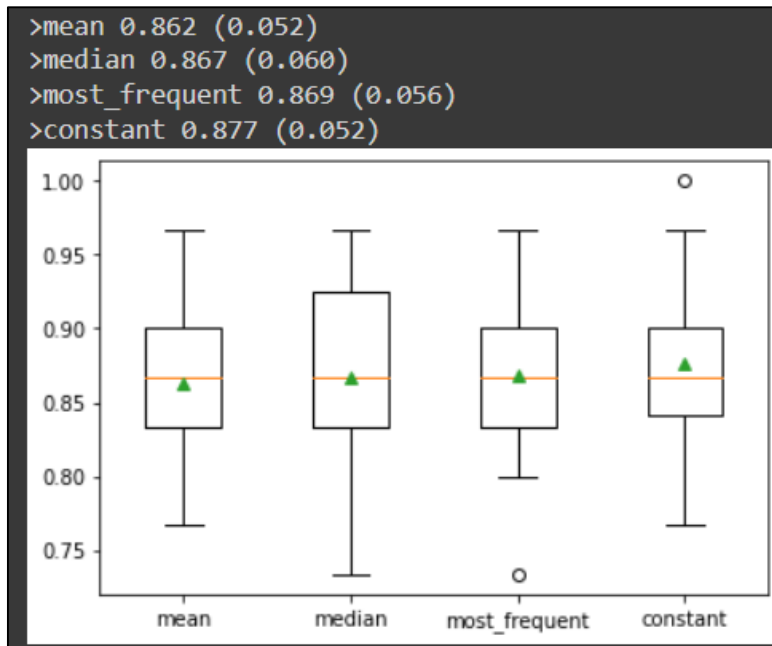


Рис. 2. Результаты импутации с использованием классических методов (среднее по столбцу, медиана, мода, константа)

Все методы, которые обсуждались до сих пор, можно назвать «одионой импутацией»: каждое значение в наборе данных заполняется ровно один раз. В общем, ограничение такой импутации заключается в том, что, поскольку эти методы находят максимально вероятные значения, они не генерируют записи, которые точно отражают распределение базовых данных.

Возьмем, к примеру, крайний случай замены отсутствующих значений в данных средним значением. Если бы мы могли наблюдать недостающие данные, мы, естественно, ожидали бы увидеть в них некоторую изменчивость: экстремальные значения, выбросы и записи, которые не полностью соответствуют «шаблону» данных. Этот шум присущ набору данных, но замена среднего значения не пытается представить его в своем результате. Это приводит к смещению в любых последующих моделях, которые подвержены тенденции (наличию среднего значения в наборе данных), которой нет в базовых данных. Это, в свою очередь, снижает точность.

В литературе [13, 14, 16], вероятно, наиболее продвинутой методологией выполнения отсутствующих данных является множественная импутация. При множественной импутации мы много раз генерируем пропущенные значения из набора данных. Затем отдельные наборы данных объединяются в окончательный импутированный набор данных, при этом значения, выбранные для замены отсутствующих данных, каким-то образом извлекаются из объединенных результатов. Другими словами, множественная импутация разбивается на три этапа: импутация (несколько циклов), анализ (постановка того, как результаты должны быть объединены) и объединение (интеграция результатов в окончательную матрицу импутации).

Любой метод, который следует этой общей схеме, является методом множественной импутации. Таким образом, существует множество доступных алгоритмов и реализаций импутации. Самый популярный алгоритм называется Multiple Imputation by Chained Equation – MICE (рус. – Множественная импутация с помощью связанного уравнения). Алгоритм импутации следующий [13, 15, 16]:

1. Простая импутация, например, среднего арифметического, выполняется для каждого отсутствующего значения в наборе данных. Эти средние значения можно рассматривать как «заполнители».

2. «Заполнитель» означает, что импутация для одной переменной («var») снова устанавливается как отсутствующее значение.
3. Наблюдаемые значения переменной «var» на шаге 2 регрессируют по другим переменным в модели импутации, которая может состоять или не включать все переменные в наборе данных. Другими словами, «var» является зависимой переменной в регрессионной модели, а все остальные переменные являются независимыми переменными в регрессионной модели.
4. Отсутствующие значения для «var» затем заменяются прогнозами из регрессионной модели. Когда «var» впоследствии используется в качестве независимой переменной в регрессионных моделях для других переменных, будут использоваться как наблюдаемые, так и эти заполненные значения.
5. Затем шаги 2–4 повторяются для каждой переменной, для которой отсутствуют данные. Перебор каждой из переменных составляет одну итерацию или «цикл». В конце одного цикла все пропущенные значения были заменены прогнозами регрессий, отражающими взаимосвязи, наблюдаемые в данных.
6. Шаги со 2 по 4 повторяются в течение нескольких циклов, при этом импутации обновляются в каждом цикле. В конце этих циклов окончательные импутации сохраняются, в результате чего получается один импутированный набор данных. Обычно выполняется десять циклов; однако необходимы исследования для определения оптимального количества циклов при импутации данных в различных условиях. Идея состоит в том, что к концу циклов распределение параметров, определяющих импутацию (например, коэффициентов в регрессионных моделях), должно сходиться в том смысле, что становится стабильным.

Применяя алгоритм MICE, на рис. 3 мы уже можем наблюдать график, отображающий качество модели. Условия для эксперимента остаются прежними. Точность алгоритма MICE составляет 0,881. Тем не менее, местами также наблюдается потеря сезонности. В случае применения константы отличие составляет 78,8%, MICE отличается от оригинала на 36%.

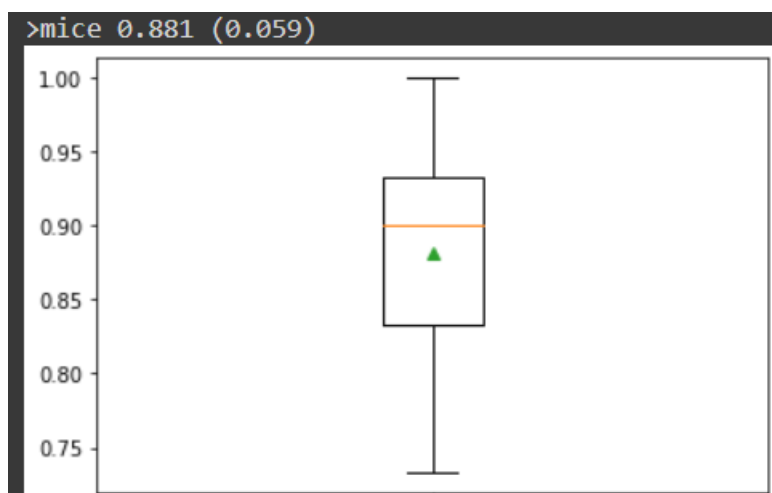


Рис. 3. Результаты импутации с использованием MICE

3. Результаты. Для тестирования результатов в прогнозировании, полученных в результате эксперимента по выбору стратегии импутации, был выбран четырехлетний отрезок, где будет применен MICE. Результаты применения MICE на четырехлетнем отрезке представлены на рис. 4, точность составляет 0,891. График выборки с примененной импутацией продемонстрирован на рис. 5.2.

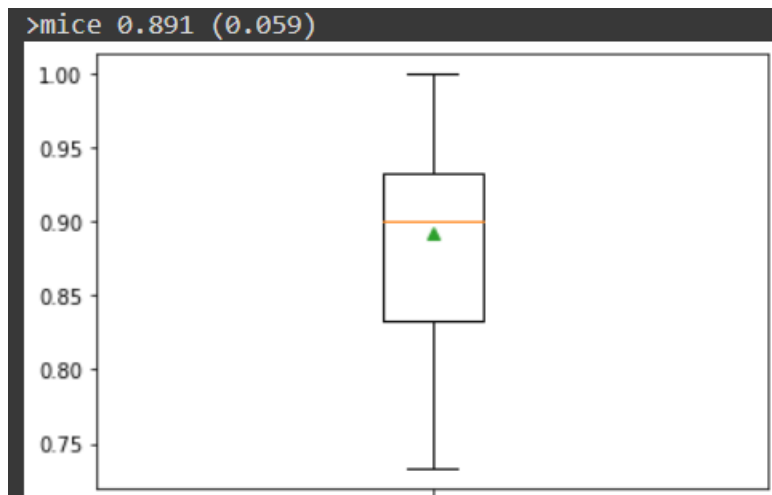


Рис. 4. Результаты импутации с использованием MICE на четырехлетнем отрезке

В результате исследования был испытан алгоритм SARIMA для прогнозирования погодных данных. На рис. 5.3 представлены результаты применения SARIMA: график слева – результат модели, обученной на четырехлетнем отрезке, где был применен MICE для импутации данных; график справа – результат модели, обученной на оригинальных данных, где не было потерь записи и не применялась импутация. Критерий RMSE (среднеквадратическое отклонение) для графика слева на рис. 5.3 составил 0,309, для графика справа RMSE составил 0,201. RMSE в обоих случаях не превышает 0,5, что говорит о способности модели точно предсказывать данные. На рис. 5.1 представлен график солнечной радиации, где синим цветом выделены оригинальные данные за семь лет.

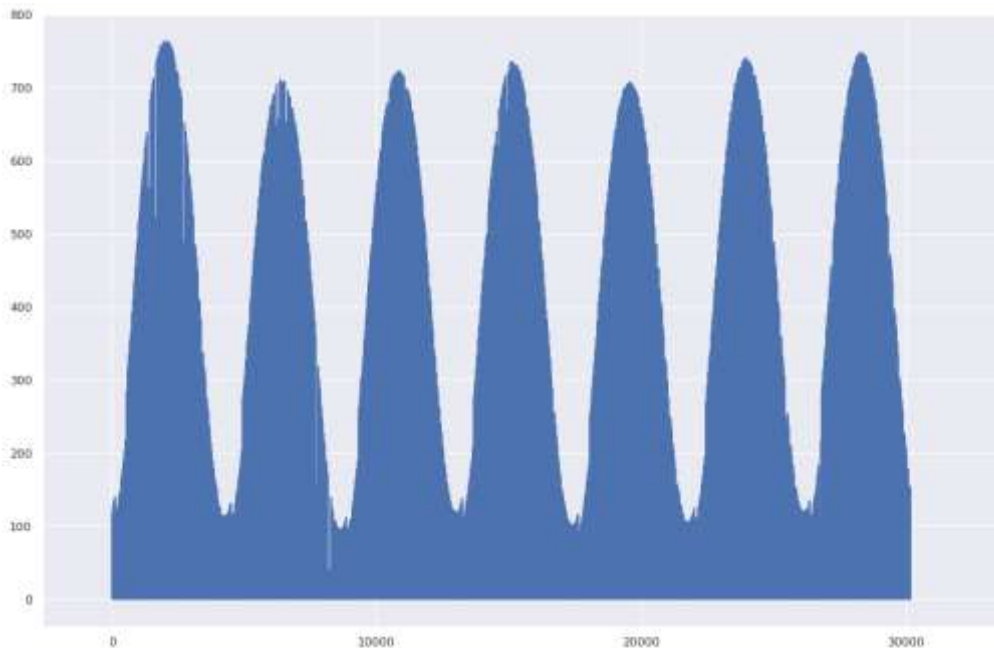


Рис. 5.1. График солнечной радиации, первоначальные данные

Вывод. Опытным путем были получены правдоподобные результаты и обосновано применение алгоритма SARIMA для прогнозирования солнечной радиации. Также было установлено, что в данном случае MICE является наиболее подходящим алгоритмом импутации. Исходя из обнаруженной проблемы потери сезонности данных после применения.

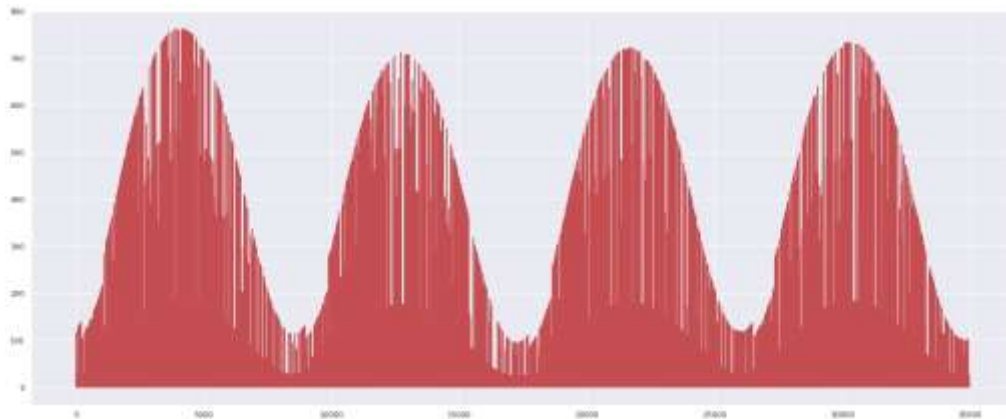


Рис. 5.2. График солнечной радиации, выборка с применением импутации

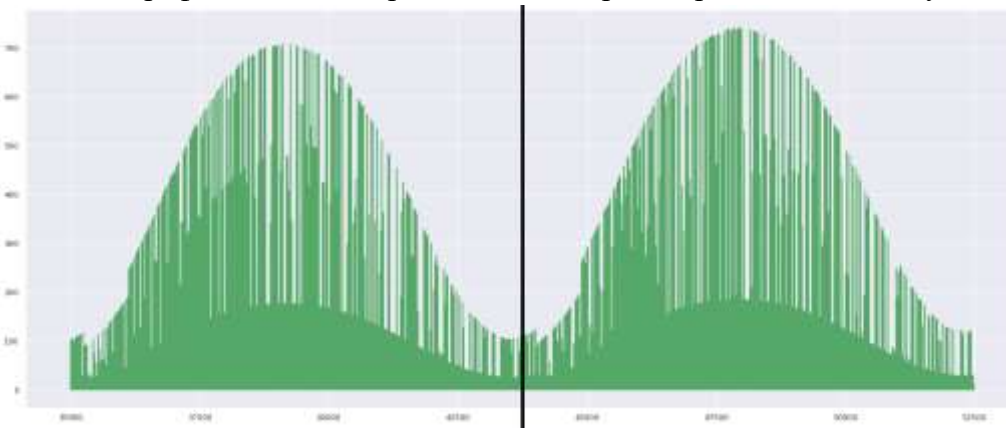


Рис. 5.3. Графики солнечной радиации, спрогнозированные моделью SARIMA на основе данных с применением импутации (слева) и на основе первоначальных данных (справа)

Комплексное рассмотрение параметров, таких, как влажность, скорость ветра и температура воздуха, позволит не только кардинально пересмотреть контекст нулевых значений, но и разработать более совершенную модель прогноза данных.

Благодарности. Результаты получены в рамках выполнения базового проекта ИСЭМ СО РАН АААА-А21-121012090007-7 по госзаданию FWEU-2021-0007.

СПИСОК ИСТОЧНИКОВ

1. Массель Л.В., Массель А.Г., Ворожцова Т.Н. Использование научной ИТ-инфраструктуры исследований энергетики для построения цифровых теней и цифровых двойников объектов и систем энергетики /Л.В. Массель, А.Г. Массель, Т.Н. Ворожцова // Применение технологий виртуальной реальности и смежных информационных систем в междисциплинарных задачах // Труды Международной научной конференции. – М.: Знание, 2020. – С. 163-169. – ISBN 978-5-907345-75-1. – DOI 10.38006/907345-75-1.2020.1.310
2. Huang R., Huang T., Gadh R., Li N. Solar generation prediction using the ARMA model in a laboratory-level micro-grid. 2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm). Tainan, Taiwan: IEEE, 11/5/2012 - 11/8/2012, pp. 528–533.
3. Badescu V. Modeling solar radiation at the earth's surface: Recent advances. Viorel Badescu, (ed.), Berlin, Springer, 2008, URL: <http://www.loc.gov/catdir/enhancements/fy0904/2007942168-d.html>.
4. Folch-Fortuny A., Arteaga F., Ferrer A. Missing Data Imputation Toolbox for MATLAB. Chemometrics and Intelligent Laboratory Systems. 2016, 154, pp. 93–100.

5. Fara L., Diaconu A., Craciunescu D., Fara S. Forecasting of Energy Production for Photovoltaic Systems Based on ARIMA and ANN Advanced Models. *International Journal of Photoenergy*. 2021, pp. 1–19, URL: <https://doi.org/10.1155/2021/6777488>
6. Huang, J., Liu, H. A hybrid decomposition-boosting model for short-term multi-step solar radiation forecasting with NARX neural network. *Journal of Central South University*, 28(2), pp. 507-526, URL: <http://doi.org/10.1007/s11771-021-4618-9>.
7. Metcalfe A. V., Cowpertwait P. S. *Introductory Time Series with R*. NY: Springer New York, 2009, 262p.
8. Shumway R. H., Stoffer D. S. *Time Series Analysis and Its Applications*. NY: Springer New York, 2011.
9. Фомина Е.Е. Обзор методов и программного обеспечения для восстановления пропущенных значений в массивах социологических данных // *Гуманитарный вестник* /Е.Е. Фомина. 2019. – №4 (78). – URL: <https://cyberleninka.ru/article/n/obzor-metodov-i-programmnogo-obespecheniya-dlya-vosstanovleniya-propuschnnyh-znacheniy-v-massivah-sotsiologicheskikh-dannyh>
10. Канищев И. С. Восстановление пропущенных значений в задачах классификации при несбалансированности данных / И. С. Канищев // *Естественные и Технические Науки*, 2021. – № 5. – С. 63–66.
11. Jamshidian M., Mata M. Advances in Analysis of Mean and Covariance Structure when Data are Incomplete. *Handbook of Latent Variable and Related Models*. Elsevier, 2007, pp. 21–44.
12. Hippel P. T. v., Bartlett J. Maximum likelihood multiple imputation: Faster imputations and consistent standard errors without posterior draws, 10/2/2012, 38p., URL: <http://arxiv.org/pdf/1210.0870v10>
13. Van Buuren S. *Flexible imputation of missing data*. Boca Raton: CRC Press, 2012, URL: <https://stefvanbuuren.name/fimd/>
14. Hayati Rezvan P., Lee K. J., Simpson J. A. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC medical research methodology*. 2015, no.15, 30p., DOI.org/10.1186/s12874-015-0022-1
15. Azur M. J., Stuart E. A., Frangakis C., Leaf P. J. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*. 2011, no. 20 (1), pp. 40–49.
16. Van Buuren S. Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011, no. 45 (3).

Лосев Алексей Сергеевич, инженер исследователь отдела систем искусственного интеллекта в энергетике, Институт систем энергетики им. Л.А. Мелентьева СО РАН, lossev.as@isem.irk.ru, Россия, Иркутск, Лермонтова 130

Массель Алексей Геннадьевич, кандидат технических наук, старший научный сотрудник отдела систем искусственного интеллекта в энергетике, Институт систем энергетики им. Л. А. Мелентьева СО РАН, amassel@gmail.com, AuthorID: 173012, SPIN-код: 2940-2510, Россия, Иркутск, Лермонтова 130

Time series forecasting and data imputation for the digital twin of a solar power plant

Aleksei S. Losev, Alexey G. Massel

Melentiev Energy Systems Institute Siberian Branch of the Russian Academy of Sciences, Russia, Irkutsk, *losev.as@isem.irk.ru*

Abstract. The article presents the implementation of an autoregressive model for predicting solar radiation data. An experiment was carried out with some approaches to data imputation and the choice of the MICE (Multiple Imputation by Chained Equation) approach was justified. A comparative experiment was also carried out for predicting solar radiation based on initial data and based on data in which imputation was applied.

Keywords: Digital twin, imputatuion, solar radiation, autoregression, forecast, ARIMA, SARIMA, MICE

Acknowledgements: The results were obtained as part of the implementation of the basic project ISEM SB RAS AAAA-A21-121012090007-7 under the state order FWEU-2021-0007.

REFERENCES

1. Massel L.V., Massel A.G., Vorozhtsova T.N. Ispol'zovaniye nauchnoy IT-infrastruktury is-sledovaniy energetiki dlya postroyeniya tsifrovyykh teney i tsifrovyykh dvoynikov ob'yektov i sistem energetik [Using the scientific IT infrastructure of energy research to build digital shadows and digital twins of energy objects and systems]. *Primeneniye tekhnologiy virtual'noy real'nosti i smezhnykh informatsionnykh sistem v mezhdistsiplinarnykh zadachakh* [Application of virtual reality technologies and related information systems in interdisciplinary tasks]. Proceedings of the International Scientific Conference. M., Znaniye., 2020, pp. 163-169, ISBN 978-5-907345-75-1, DOI 10.38006/907345-75-1.2020.1.310
2. Huang R., Huang T., Gadh R., Li N. Solar generation prediction using the ARMA model in a laboratory-level micro-grid. 2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm). Tainan, Taiwan: IEEE, 11/5/2012 - 11/8/2012, pp. 528–533.
3. Badescu V. Modeling solar radiation at the earth's surface: Recent advances. Viorel Badescu, (ed.), Berlin, Springer, 2008, URL: <http://www.loc.gov/catdir/enhancements/fy0904/2007942168-d.html>.
4. Folch-Fortuny A., Arteaga F., Ferrer A. Missing Data Imputation Toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems*. 2016, 154, pp. 93–100.
5. Fara L., Diaconu A., Craciunescu D., Fara S. Forecasting of Energy Production for Photovoltaic Systems Based on ARIMA and ANN Advanced Models. *International Journal of Photoenergy*. 2021, pp. 1–19, URL: <https://doi.org/10.1155/2021/6777488>
6. Huang, J., Liu, H. A hybrid decomposition-boosting model for short-term multi-step solar radiation forecasting with NARX neural network. *Journal of Central South University*, 28(2), pp. 507-526, URL: <http://.doi.org/10.1007/s11771-021-4618-9>.
7. Metcalfe A. V., Cowpertwait P. S. *Introductory Time Series with R*. NY: Springer New York, 2009, 262p.
8. Shumway R. H., Stoffer D. S. *Time Series Analysis and Its Applications*. NY: Springer New York, 2011.

9. Fomina E.E. Obzor metodov i programmno obespecheniya dlya vosstanovleniya propuschnykh znacheniy v massivakh sotsiologicheskikh dannykh [Review of methods and software for recovering missing values in arrays of sociological data]. Gumanitarnyy vestnik [Humanities Bulletin]. 2019, no. 4 (78), URL: <https://cyberleninka.ru/article/n/obzor-metodov-i-programmnogo-obespecheniya-dlya-vosstanovleniya-propuschnykh-znacheniy-v-massivakh-sotsiologicheskikh-dannykh>
10. Kanishchev I. S. Vosstanovleniye propuschnykh znacheniy v zadachakh klassifikatsii pri nesbalansirovannosti dannykh [Recovery of missing values in classification problems with data imbalance]. Yestestvennyye i Tekhnicheskiye Nauki [Natural and Technical Sciences]. 2021, no. 5, pp. 63–66.
11. Jamshidian M., Mata M. Advances in Analysis of Mean and Covariance Structure when Data are Incomplete. Handbook of Latent Variable and Related Models. Elsevier, 2007, pp. 21–44.
12. Hippel P. T. v., Bartlett J. Maximum likelihood multiple imputation: Faster imputations and consistent standard errors without posterior draws, 10/2/2012, 38p., URL: <http://arxiv.org/pdf/1210.0870v10>
13. Van Buuren S. Flexible imputation of missing data. Boca Raton: CRC Press, 2012, URL: <https://stefvanbuuren.name/fimd/>
14. Hayati Rezvan P., Lee K. J., Simpson J. A. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. BMC medical research methodology. 2015, no. 15, 30p., DOI.org/10.1186/s12874-015-0022-1
15. Azur M. J., Stuart E. A., Frangakis C., Leaf P. J. Multiple imputation by chained equations: what is it and how does it work? International journal of methods in psychiatric research. 2011, no. 20 (1), pp. 40–49.
16. Van Buuren S. Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011, no. 45(3).

Aleksei S. Losev, Research Engineer of the Department of Artificial Intelligence Systems in the Energy Industry of the Institute of Energy Systems named after. L. A. Melentiev SB RAS, losev.as@isem.irk.ru, Russia, Irkutsk, st. Lermontova d.130

Alexey G. Massel, Ph. D, Senior Research Fellow of the Department of Artificial Intelligence Systems in the Energy Industry of the Institute of Energy Systems named after. L. A. Melentiev SB RAS, amassel@gmail.com, AuthorID: 173012, SPIN: 2940-2510, Russia, Irkutsk, st. Lermontova d.130

Статья поступила в редакцию 05.03.2022; одобрена после рецензирования 25.03.2022; принята к публикации 29.03.2022.

The article was submitted 03.05.2022; approved after reviewing 03.25.2022; accepted for publication 03.29.2022.