

Herausgeber Prof. Dr. Barbara Dörsam

Schriftreihe Bachelor-Resümee

Forschungsbereich **Data Science**

Erforschung der Prognostizierbarkeit von Solarenergieerträgen anhand von Wettervorhersagen in Python

Immanuel Ricker

Studieren. Wissen. Machen.

Impressum

Hochschule der Medien

Nobelstrasse 10

70569 Stuttgart

www.hdm-stuttgart.de

0711 8923-0

Autor

Immanuel Ricker

Betreuer

Prof. Dr. Barbara Dörsam

Datum

August 2023

Wirtschaftsingenieurwesen Medien

www.hdm-stuttgart.de/wing

hitzges@hdm-stuttgart.de

0711/8923-2634

Layout

Jochen Riegg

Fotos und Illustrationen

Innenteil: Immanuel Ricker

Bachelor-Resümee

Erforschung der Prognostizierbarkeit von Solarenergieerträgen anhand von Wettervorhersagen

[Thema]

Immanuel Ricker

August 2023

Der Autor

Immanuel Ricker studierte Wirtschaftsingenieurwesen Medien mit dem Schwerpunkt Industrial Media Production an der Hochschule der Medien. Im Rahmen seiner Arbeit konnte signifikante Korrelationen zwischen Wettervorhersagen und Solarenergieerträgen nachgewiesen werden, die für verlässliche Vorhersagen von Photovoltaikleistungen genutzt werden konnten.

Inhaltsverzeichnis

| | |
|---------------------------------|----|
| 1. Einleitung..... | 5 |
| 2. Methoden und Material | 5 |
| ARMA-Modell | 5 |
| Erweiterung ARMA-Modell | 6 |
| Stationarität..... | 6 |
| Daten | 6 |
| Bewertung Prognosemodelle | 7 |
| 3. Ergebnisse..... | 7 |
| Stationarität..... | 7 |
| Korrelationen..... | 8 |
| Prognosewerte | 9 |
| 4. Diskussion..... | 9 |
| Ausblick | 11 |

1. Einleitung

Einer der wichtigsten Aspekte um die natürlichen Ressourcen der Erde zu schonen, ist die Umstellung von fossilen Energieträger hin zu regenerativer Energie [1]. Um diese Umstellung in privaten Haushalten zu vereinfachen, verfolgt das SmartEE Projekt der Hochschule der Medien das Ziel, die Nutzung von elektronischen Haushaltsgeräten mit der Erzeugung von eigenem Solarstrom zu synchronisieren. Im Rahmen dieser Bachelorarbeit wird der Zusammenhang zwischen Wettervorhersagen und Solarenergieerträgen erforscht. Anschließend werden die Erkenntnisse genutzt, um Solarenergieerträge vorherzusagen.

Dafür wurden dieser Arbeit Daten bereitgestellt. Diese umfassen die gemessene Leistung zweier Solaranlagen eines Haushalts in Stuttgart, als auch Wetterprognosen der letzten vier Jahre für die Messstation Stuttgart-Flughafen. Einzelne Werte der Solar- und Wetterdaten sind jeweils einem Zeitintervall zuzuordnen und somit als Zeitreihen anzusehen.

2. Methoden und Material

Mathematisch basieren Zeitreihen auf dem Modell der stochastischen Prozesse. Die Zeitreihenanalyse und insbesondere die Prognose von Zeitreihen setzt sich damit auseinander, mathematische Modelle an einen stochastischen Prozess anzupassen, der die gemessenen Werte einer Zeitreihe beschreibt [2]. Wird ein Prozess passend beschrieben, kann das Modell genutzt werden, um zukünftige, unbekannte Werte vorherzusagen. Um für den Rahmen dieser Abschlussarbeit ein passendes Modell auszuwählen, wurde sich nach der Richtlinie 2225 des Vereins der Deutschen Ingenieure (VDI) zwischen dem Autoregressive-Moving-Average (ARMA) Modell, dem Exponentiellen Glätten und Künstlichen Neuronale Netzen entschieden. Die Auswertung fiel zugunsten des ARMA-Modells und dessen Erweiterungen aus.

ARMA-Modell

ARMA-Modelle setzen sich aus einem Autoregressiven (AR) Modell und einem gleitenden Mittel (eng. Moving Average MA) Modell zusammen.

Ein AR-Modell nutzt das Prinzip der Autokorrelation. Im Gegensatz zur Korrelation zwischen zwei verschiedenen Variablen beschreibt die Autokorrelation den Zusammenhang zwischen verschiedenen Zuständen einer einzigen Variable zu unterschiedlichen Zeitpunkten. Für die Prognosen-Erstellungen zu einem Zeitpunkt t werden die unterschiedlich gewichteten Werte zu den Zeitpunkten $t - 1 \dots t - p$ genutzt, wobei p die Ordnung des AR-Modells angibt [3, p.25]. Die Gewichtung wird mit den Koeffizienten $\phi_{t-1} \dots \phi_{t-p}$ angegeben [2, p.79].

Das MA-Modell weist im Aufbau Ähnlichkeiten zum AR-Modell auf. Anstelle von vorangegangenen Werten, werden zeitversetzte Fehlerterme ε als Prädikatoren verwendet. Fehlerterme ε sind für MA-Modelle als gleitende Mittelwerte der zeitversetzten Residuen zu verstehen. Die Gewichtung der Fehlerterme wird durch θ ausgedrückt. Durch die Modellierung der angrenzenden Schätzfehler zu einem Zeitpunkt t wird so ein Wert zum Zeitpunkt t ausgedrückt. Die Anzahl der angrenzenden Fehlerterm wird durch die Ordnungszahl q angegeben [3, p.16].

Da ARMA-Modelle beide vorangegangenen Prinzipien vereinen, ergibt sich die Ordnung (p, q) . Die gewichteten Parameter, die ein Modell zur Beschreibung einer Zeitreihe nutzt, nennt man Prädikatoren [4, p.259].

Erweiterung ARMA-Modell

Für diese Abschlussarbeit waren insbesondere die ARMA-Erweiterung um exogene Inputs (ARMAX) von Bedeutung. Dabei werden die Kontexteffekte, die zum Zeitpunkt der Messung eines Wertes geherrscht haben zum Ausdruck gebracht. Diese Kontextvariablen werden wie andere Prädikatoren durch einen Koeffizienten unterschiedlich gewichtet, um den Zustand der Zielvariable Y_t zu beschreiben. Voraussetzung, um Prognosen für Y_t zu erstellen ist, dass die Kontextvariablen für den Zeitpunkt t bereits zum Zeitpunkt der Prognosen-Erhebung bekannt sind [5].

Stationarität

Eine Voraussetzung, um mit ARMA-Modellen Zeitreihenvorhersagen zu erstellen ist, dass die Zeitreihen schwache Stationarität aufweisen. Schwache Stationarität ist erreicht, wenn bestimmte mathematische Eigenschaften unabhängig von der Zeit sind. Dadurch kann charakteristisches Verhalten aus einem Beobachtungszeitraum auch für zukünftige Werte angenommen werden [6]. Allgemein wird im Folgenden von Stationarität gesprochen, gemeint ist die mathematische Eigenschaft der schwachen Stationarität. Die Stationarität einer Zeitreihe kann mithilfe eines Augmented-Dickey-Fuller (ADF) Test überprüft werden. Zeitreihen mit Trend- oder saisonalen Komponenten weisen zeitabhängiges Verhalten auf und sind daher grundsätzlich nicht stationär [3, p.5].

Für Zeitreihen mit Trend-, oder saisonaler Komponente kann schwache Stationarität durch Differenzieren erzeugt werden [7, p.18]. Bei einer einfachen Differenzierung wird von einem Prozess Y zum Zeitpunkt t der Wert zum Zeitpunkt $t - 1$ subtrahiert. Man erhält den differenzierten Prozess durch:

$$Y'_t = Y_t - Y_{t-1}$$

Um eine saisonale Komponente zu entfernen, wird nicht zum Zeitpunkt $t - 1$ differenziert, sondern zu dem Zeitpunkt, an dem sich eine saisonale Schwankung l wiederholt. Wird bspw. ein Klimadiagramm auf Monatebene betrachtet, würde man die Differenzierung wie folgt wählen:

$$Y'_t = Y_t - Y_{t-l} \text{ wobei } l = 12$$

Wenn durch Differenzieren Stationarität hergestellt wurde, können diese für ARMA-Modell-Prognosen verwendet werden. Um die Ergebnisse der Prognosen zu interpretieren, muss die Differenzierung für prognostizierte Werte wieder umgekehrt werden. Diesen Vorgang nennt man Integrieren [2, p.123].

Daten

Die zu untersuchenden Daten beinhalteten zu Beginn redundante Informationen und lagen in unterschiedlichen Dateiformat-, und Struktur vor. Nachdem die relevanten Informationen extrahiert wurden, konnten die Wetter- und Solardaten auf das kleinstmögliche gemeinsame Zeitintervall von drei Stunden skaliert werden, um jedem Messwert eine eindeutige Wetterprognose zuzuordnen zu können. Ein Messwert der Photovoltaikanlage entspricht der Leistung in kW/h, die innerhalb eines dreistündigen Zeitintervall in den Stromzähler des Haushaltes eingespeist wird. Im Weiteren mussten einige fehlende Messwerte der Solardaten imputiert werden. Für die erste Photovoltaikanlage (PV1) wurden 2% aller Werte als fehlend identifiziert. Diese konnten durch ein Mittelwerts-Verfahren

ergänzt werden [8, p.132]. Dabei wird zu jedem fehlenden Messzeitpunkt der Mittelwert zur gleichen Monat-Tag-Stunden Kombination über alle Jahre gebildet und anstelle des fehlenden Wertes eingesetzt [7, p.27]. Für die zweite Photovoltaikanlage (PV2) fehlen über 35% der Messwerte. Eine zuverlässige Imputation ist nur mit komplexen Verfahren zu bewerkstelligen, wobei die Methoden von Imputation und Vorhersagen verschwimmen [9]. Daher wurde sich für die folgenden Untersuchungen auf die Daten der PV1 beschränkt.

Bewertung Prognosemodelle

Um die Güte von Prognosen bewerten zu können, werden die zur Verfügung stehenden Daten in Trainings- und Testdaten geteilt. Das Prognosemodell passt die Koeffizienten für die Prädikatoren anhand der Trainingsdaten an und erstellt für den Zeitraum der Testdaten Prognosen [4, p.802]. Die Bewertung und Einordnung von Prognosewerten erfolgt mittels der Kenngrößen mittlerer absoluter Fehler (MAE) und mittlerer quadratischer Fehler (MSE). Diese errechnen sich wie folgt:

$$MAE = \frac{1}{h} \sum_{t=n+1}^{n+h} |\hat{y}_t - y_t| \quad ; \quad MSE = \frac{1}{h} \sum_{t=n+1}^{n+h} (\hat{y}_t - y_t)^2$$

Für MAE und MSE gilt:

n : Umfang der Testdaten; h : Umfang der Trainingsdaten; y : Messwert; \hat{y} : Prognosewert

[2, p.16]

3. Ergebnisse

Im ersten Schritt wurde die Zeitreihe der Solardaten für jede Uhrzeit getrennt. Das Ergebnis waren acht separate Zeitreihen, die für jeden Tag nur einen Messwert zu einem bestimmten Zeitintervall abbildeten. Dies war notwendig, da ca. 40% aller Werte 0,00 kW/h betragen. Besonders die Zeitintervalle zwischen 21 Uhr und 3 Uhr haben Nullwertanteile von 86% bis 100%. Die Vorhersagen der Beträge in diesen Zeiträumen sind trivial und für das SmartEE Projekt nicht von Interesse, da ohnehin kaum Energie produziert wird. Außerdem werden Korrelationen verzerrt, wenn 40% der gesamten Messwerte aufgrund der Uhrzeit und unabhängig von Wetterbedingungen 0,00 kW/h betragen. Auch die mittleren Fehlerwerte verlieren an Aussagekraft, da Vorhersagen der Nullwerte aller Wahrscheinlichkeit nach durchgängig sehr gut wären. Da die Abstände zwischen den Intervallen konstant bleiben müssen, wurden die Uhrzeiten getrennt. Somit konnten die relevanten Intervalle zwischen 9 Uhr und 15 Uhr gesondert untersucht, und die Prognosewerte durch MAE und MSE eingeordnet werden.

Stationarität

Stationarität wurde durch eine modifizierte Version der saisonalen Differenzierung hergestellt. Die Zeitreihen beinhalten pro Tag einen Messwert. Somit würde sich eine saisonale Länge von 365 und für Schaltjahre von 366 ergeben. Da die Werte im Gesamten zwar saisonale Schwankungen aufweisen, aber einzelne Monat-Tag Kombinationen verschiedener Jahre stark variieren können, wurde davon abgesehen jeden Wert mit einem einzelnen Wert des Vorjahres zu differenzieren. Stattdessen wurde für jeden Monat der Mittelwert gebildet. Einzelne Werte sind um den Mittelwert des zugehörigen Monats bereinigt, um so das saisonale Niveau der einzelnen Werte besser abzubilden. Um die Schwankungen weiter abzumildern, wurde die gesamte Zeitreihe vor der Monatsmittelwertbildung

logarithmiert. ADF-Tests ergeben, dass die Zeitreihe jeder Uhrzeit, nach dieser Transformation als schwach stationär angenommen werden kann.

Korrelationen

Um einen errechneten Korrelationskoeffizienten nach Pearson durch Hypothesentest für die Gesamtheit der Daten annehmen zu können, müssen die beiden untersuchten Variablen jeweils normalverteilt sein. Ansonsten kann der Korrelationskoeffizienten nach Spearman herangezogen werden [4, p.216]. Untersuchungen der Daten ergaben, dass eine Normalverteilung der zugrundeliegenden Daten nicht anzunehmen ist. Die Untersuchung erfolgte mit Shapiro Tests. Daher ist in Abbildung 1 eine Korrelationsmatrix nach Spearman zu sehen.

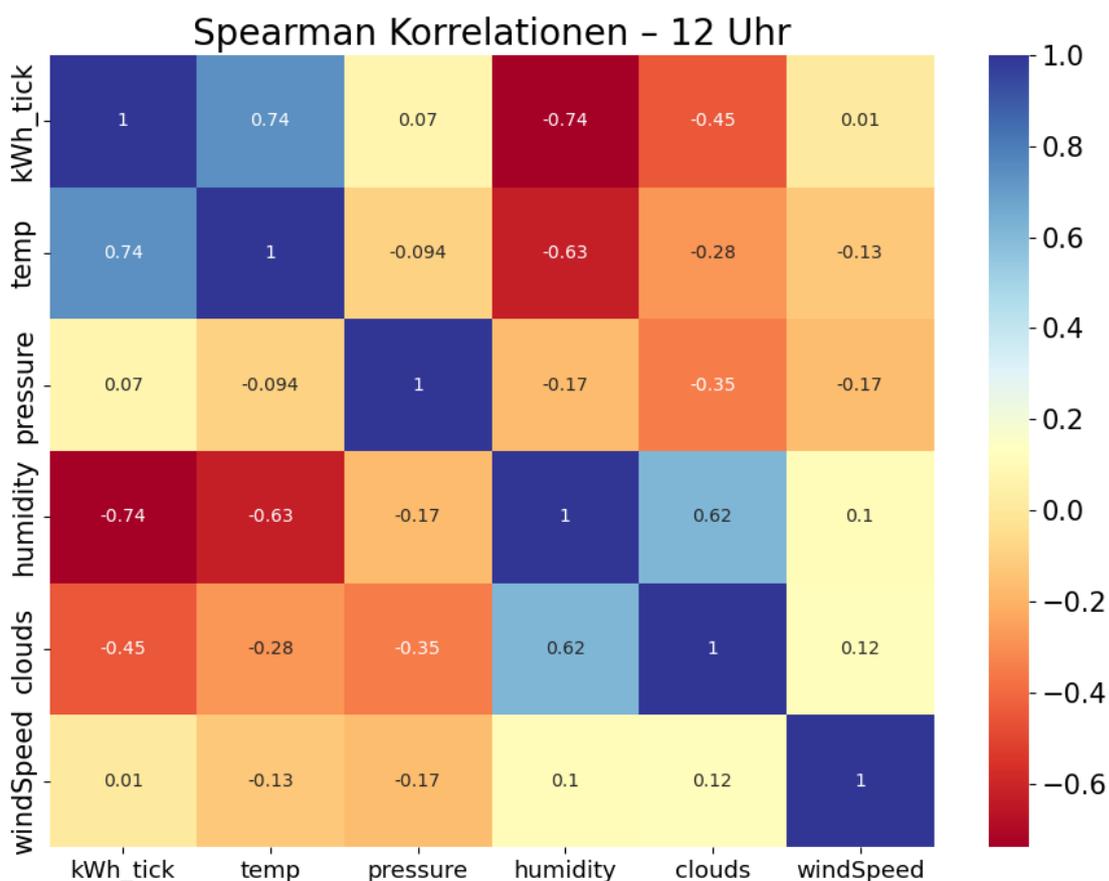


Abbildung 1 Korrelationsmatrix zwischen kWh-Erträge und Wetterprognosen

Aus der ersten Spalte in Abbildung 1 sind die Korrelationen zwischen den kW/h-Erträgen mit den einzelnen Wetterparameter abzulesen. Positive Korrelation sind in Blautönen, negative Korrelationen sind in Rottönen dargestellt. Hypothesentests ergaben, dass die abgebildete Korrelation für die Temperatur, die Luftfeuchtigkeit und die Bewölkung mit hoher Signifikanz für die Gesamtheit angenommen werden kann. Eine Korrelation zwischen der Windgeschwindigkeit und den kWh-Erträgen konnte für die Gesamtheit ausgeschlossen werden. Für den Luftdruck konnte eine Korrelation für die Gesamtheit nicht mit hoher Signifikanz angenommen werden, allerdings auch nicht eindeutig verworfen werden. Weitere Tests ergaben, dass die Beachtung des Luftdrucks nicht zur Verbesserung der Prognosemodelle führt.

Prognosewerte

Um Vergleichswerte zu schaffen, wird ein ARMA-Modell ohne Kontextvariablen und ein ARMAX-Modell mit Kontextvariablen angelernet. Das ARMA-Modell soll Prognosen nur anhand der Messwerte erstellen, während das ARMAX-Modelle die Wetterprognosen für einen vorherzusagenden Zeitpunkt als Prädiktoren mit anlernt. Dadurch wird der positive Einfluss der Wetterprognosen auf die Schätzwerte deutlich. Die beste Ordnung als auch die Fehlerwerte sind aus Tabelle 1 zu entnehmen.

| Modell | MAE | MSE |
|------------------|------|-------|
| ARMA((1,1),1) | 1,39 | 2,718 |
| ARMAX((1,0,1),1) | 0,84 | 1,38 |

Tabelle 1 Ergebnisse der Prognosemodelle

Die dargestellten Fehlerwerte beziehen sich für beide Modellen auf denselben Zeitraum.

4. Diskussion

Die Wetterprognosen konnten erfolgreich genutzt werden, um die Vorhersagegenauigkeit von Solarenergieerträgen zu erhöhen. Hoch signifikant dabei sind Temperatur, Bewölkung und Feuchtigkeit.

Um die Fehlerwerte genauer einschätzen zu können, sind in Abbildung 2 die Fehlerverteilungen von zwei Modellen aufgezeigt.

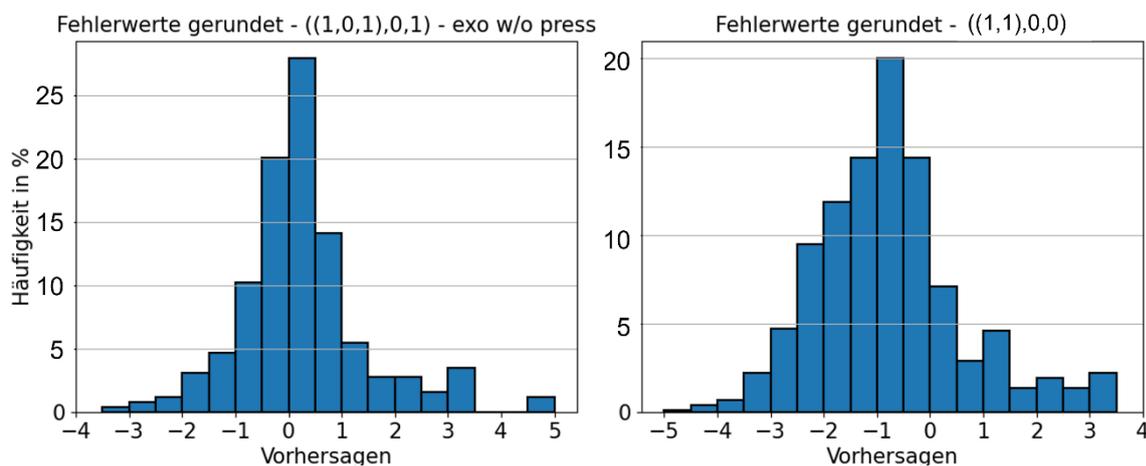


Abbildung 2 Fehlerwertverteilung für beste ARMAX und ARMA Modelle

Auf der linken Seite ist die Verteilung des besten ARMAX, auf der rechten Seite die Verteilung des besten ARMA-Modells, visualisiert. Man erkennt, dass auf der rechten Seite das größte Aufkommen von Fehlerwerten um den Wert -1 auftritt. Auf der linken Seite tritt das größte Aufkommen um den Wert 0 auf.

Lediglich größere Fehler am Rande des Spektrums scheinen kaum verändert zu sein. Auch kommt es bei der Verwendung des ARMAX Modells vereinzelt zu größeren Ausreißern, wie der Balken ganz rechts der linken Darstellung aufzeigt.

Weitere Untersuchungen ergeben, dass die zehn größten Fehler einen starken negativen Einfluss auf das Fehlermaß der ARMAX-Prognosen haben. Die ARMA-Prognosen dagegen reagieren weniger empfindlich darauf.

Eine Detailanalyse der größten Fehlerwerte für ARMAX Modelle ergab drei Hauptursachen.

Zum einen führen falsche Wettervorhersagen zu ungenauen Solarenergieprognosen. ARMA-Modelle, die blind für Wetterdaten sind, werden dadurch nicht beeinflusst. Dies erklärt die größeren Ausreißer des ARMAX-Modells gegenüber den ARMA-Modellen.

Eine weitere Ursache sind ungewöhnlich niedrige oder hohe kWh-Erträge, die sich nicht durch die Daten erklären lassen. Solche sind auf externe Einflüsse zurückzuführen und werden nicht in den Daten abgebildet.

Die letzte Ursache sind Prognosewerte, die unnatürlich hohe Erträge annehmen. Dies hat nicht mit der Ungenauigkeit der Schätzung zu tun, sondern mit der Rücktransformierung der Daten. Eine Analyse der Verläufe soll dies bestätigen. In Abbildung 3 ist als blaue gestrichelte Linie der tatsächliche stationäre Verlauf der kWh-Erträge abgebildet. Die grüne durchgezogene Linie entspricht den vorhergesagten Werten.

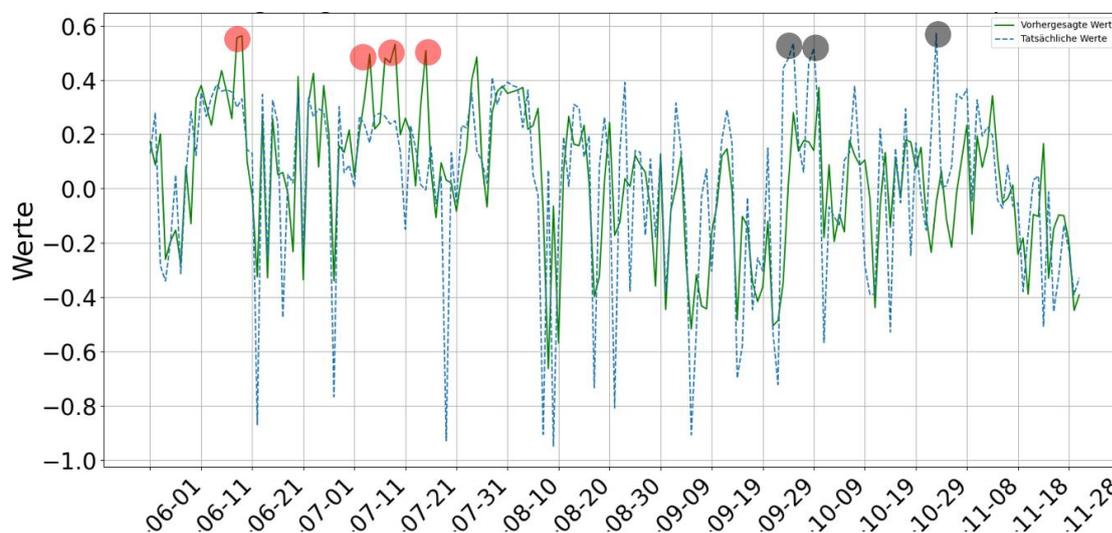


Abbildung 3 Verlauf der tatsächlichen und prognostizierten stationären Zeitreihe

Die roten Punkte in Abbildung 3 markieren sieben Werte, die nach der Umrechnung zu den Maxima der Vorhersagen werden. Im Vergleich zu anderen Werten der tatsächlichen Zeitreihe, sind es zwar hohe Werte aber keinesfalls ungewöhnlich hohe, wie die grauen Markierungen aufzeigen.

Die amerikanischen Datumsangaben auf der y-Achse der Abbildungen 3 und 4 beziehen sich jeweils auf das Jahr 2022.

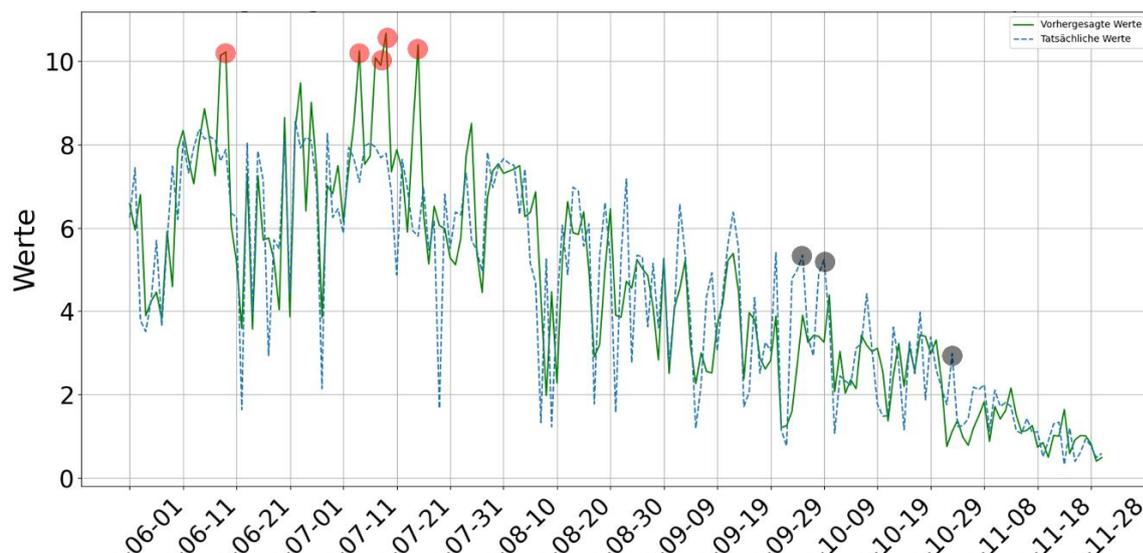


Abbildung 4 Verlauf der tatsächlichen und prognostizierten unbehandelten Zeitreihe

Abbildung 4 ist identisch zu Abbildung 3, nur dass die Werte rücktransformiert wurden. Durch das Addieren der Monatsmittelwerte und anschließende Exponenzieren übersteigen die rot markierten Werte die 10 kWh Marke. Da der Monatsmittelwert im Oktober und November deutlich geringer ist als im Juni bzw. Juli, ist der Effekt des Exponenzierens geringer. Die Ursache der Ausreißer ist demnach also nicht das Modell an sich, sondern die, für die Stationarität notwendigen Transformationschritte.

Ausblick

Neben den untersuchten Wetterparametern ergaben Recherchen, dass der Deutsche Wetterdienst Prognosewerte für Strahlungsdaten zur Verfügung stellt. Eine Untersuchung vergangener Ist-Werte der Strahlungsdaten ergaben, dass diese sehr stark mit den kW/h-Erträgen korrelieren. Diese für zukünftige Prognosemodelle mit einzubeziehen, wird die Güte der Prognosen steigern. Des Weiteren wird auch ein überarbeitetes Verfahren, um Stationarität zu erzeugen, zu besseren Prognosen führen.

Um die erarbeiteten Modelle für das SmartEE Projekt zu nutzen, müssen die Prognose-Werte jeder einzelnen Uhrzeit nacheinander gehängt werden, um den Verlauf eines Tages abzubilden.

5. Literaturverzeichnis

- [1] WWF: *Earth Overshoot Day*. URL <https://www.wwf.de/earth-overshoot-day> – Überprüfungsdatum 2023-07-24
- [2] VOGEL, Jürgen: *Prognose von Zeitreihen*. Wiesbaden : Springer Fachmedien Wiesbaden, 2015
- [3] PROF.DR.W.ZUCCHINI: *Zeitreihenanalyse : TeilIII:LineareZeitreihenmodelle* (2007). URL https://www.uni-goettingen.de/de/document/download/bced8d2cf092e3487839c365955c85f2.pdf/ZRSkript_Teil2_Okt09.pdf

- [4] KUCKARTZ, Udo ; RÄDIKER, Stefan ; EBERT, Thomas ; SCHEHL, Julia: *Statistik*. Wiesbaden : VS Verlag für Sozialwissenschaften, 2013
- [5] CHALEAMPONG KONGCHAROEN ; TAPANEE KRANGPRADIT: *Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export (2013)*. URL https://www.researchgate.net/profile/Chaleampong-Kongcharoen/publication/255731345_Autoregressive_Integrated_Moving_Average_with_Explanatory_Variable_ARIMAX_Model_for_Thailand_Export/links/0c9605209ac48013f6000000/Autoregressive-Integrated-Moving-Average-with-Explanatory-Variable-ARIMAX-Model-for-Thailand-Export.pdf – Überprüfungsdatum 2023-07-20
- [6] KREIB, Jens-Peter ; NEUHAUS, Georg: *Einführung in die Zeitreihenanalyse : Mit 8 Tabellen*. [Online-ausg.]. Berlin [u.a.] : Springer, 2006 (Statistik und ihre Anwendungen)
- [7] HIRSCHLE, Jochen: *Machine Learning für Zeitreihen : Einstieg in Regressions-, ARIMA- und Deep Learning-Verfahren mit Python*. München : Carl Hanser Verlag GmbH et Co. KG, 2021 (Hanser eLibrary)
- [8] GRAF, Sabrina: *Standardisierung im Konzerncontrolling*. Wiesbaden : Springer Fachmedien Wiesbaden, 2017
- [9] STEFFEN MORITZ ; THOMAS BARTZ-BEIELSTEIN: *Time Series Missing Value Imputation in R*. URL <https://pdfs.semanticscholar.org/cf38/7a44ef973ac37568a8ca482b4add12b646eb.pdf> – Überprüfungsdatum 2023-07-20
- [10] ALEXANDER ETZ: *Introduction to the Concept of Likelihood and Its Applications*. In: *SAGE journals* (2018). URL <https://journals.sagepub.com/doi/full/10.1177/2515245917744314> – Überprüfungsdatum 2023-07-16
- [11] YING ZHANG ; GONG MENG: *Simulation of an Adaptive Model Based on AIC and BIC ARIMA Predictions*. URL <https://iopscience.iop.org/article/10.1088/1742-6596/2449/1/012027/pdf> – Überprüfungsdatum 2023-07-16
- [12] WAGEMANN, Claudius ; GOERRES, Achim ; SIEWERT, Markus B.: *Handbuch Methoden der Politikwissenschaft*. Wiesbaden : Springer Fachmedien Wiesbaden, 2020