

# PROGNOSEMODELL EINER PV-ANLAGE BASIEREND AUF EINEM KURZZEITMESSSYSTEM, WETTERDATEN UND MACHINE-LEARNING VERFAHREN

Marta Sophia Potyka, Hannes Seefluth, Prof. Peter Schegner

TU Dresden, Mommsenstraße 10 01069 Dresden, +49 351 463-43203, marta\_sophia.potyka@tu-dresden.de, <https://tu-dresden.de/ing/elektrotechnik/ieeh/ev/die-professur>

**Kurzfassung:** Aufgrund des steigenden Umweltbewusstseins der Bevölkerung werden vermehrt dezentrale und zumeist regenerative Energieerzeugungsanlagen errichtet. Um eine hohe Versorgungszuverlässigkeit weiterhin gewährleisten zu können, ist es unerlässlich, volatile Erzeuger so genau wie möglich zu prognostizieren. Ebenso kann der Betreiber dieser Anlagen die Prognosen für ein Energiemanagement und die bestmögliche ökonomische Verwertung nutzen. Hierfür wird ein Prognosemodell für PV-Anlagen erstellt, dessen Eingangsparameter sich aus Messdaten der erzeugten elektrischen Wirkleistung und Wetterdaten über wenige Wochen zusammensetzen. Gängige physikalische Modelle bilden meist nur lineare Zusammenhänge ab, wohingegen Machine-Learning- oder Regressionsverfahren auch nichtlineare Abhängigkeiten erlernen können. Daher werden verschiedene Verfahrensarten auf ihre Eignung für die genannte Problemstellung untersucht.

**Keywords:** Prognose, PV-Anlage, Machine-Learning, Regression, regenerative und dezentrale Erzeugungsanlagen

## 1 Einleitung

Heutzutage werden zunehmend dezentrale und regenerative Energieerzeugungsanlagen (DEA) installiert, um die Umwelt und ihre Ressourcen zu schonen. Diese sind jedoch meist volatile Erzeuger. Das bedeutet, dass deren Leistungserbringung im Gegensatz zu bisherigen Anlagen, beispielsweise einem Kohlekraftwerk, nicht frei bestimmbar ist. Ziel ist es, trotz Nutzung der DEA weiterhin eine hohe Versorgungszuverlässigkeit des elektrischen Netzes zu gewährleisten. Dafür bedarf es genauer Prognosen der erzeugten elektrischen Wirkleistung. Ebenfalls kann der Betreiber der Anlagen mit den Vorhersagen ein Energiemanagement implementieren und die ökonomische Verwertung der bereitgestellten elektrischen Energie optimieren.

## 2 Literaturrecherche

Bisher wurden schon einige Ansätze zur Prognose von PV-Anlagen untersucht. Neben physikalischen Modellen werden zunehmend auch Modelle mit Machine Learning (ML) Verfahren integriert. Denn diese können im Gegensatz zu den physikalischen Modellen auch nichtlineare Zusammenhänge in einer überschaubaren Komplexität abbilden [1]. Beispielsweise wird das Modell Seasonal Autoregressive Integrated Moving Average (SARIMA) in [2] verwendet. In dieser Veröffentlichung wird mit über einen Monat lang aufgenommenen Messdaten von der erzeugten elektrischen Leistung gearbeitet, welche in einem Intervall von 20 Minuten vorlie-

gen. Die Prognose wird für den nächsten 20-Minutenwert durchgeführt. Der geringste Prognosefehler liegt bei 3,4 %, wobei der Mean Absolute Percentage Error (MAPE) angewendet wurde. Dieser Wert bezieht sich auf ein Panel, dessen Bemessungsleistung 60 W beträgt. Allerdings werden nur sonnige Tage betrachtet.

Ein zweiter Ansatz ist die Regression. In [3] wird die multiple lineare Regression (MLR) und die multiple polynomielle Regression (MPR) der 2. und 3. Ordnung untersucht. Als Eingangsdaten liegen die erzeugte elektrische Leistung, einige Kennwerte der PV-Anlage und die Globalstrahlung durch ein selbst installiertes Pyranometer vor. Es wird immer der nächste 10-Minutenwert prognostiziert. Der geringste Prognosefehler liegt für die polynomielle Regression der 3. Ordnung vor. Dieser beträgt 3,19 kW bezogen auf 25 kW Spitzenleistung. Dabei wird der Root Mean Square Error (RMSE) verwendet.

In [4] wird ein künstliches neuronales Netz (KNN) mit einem Clear-Sky-Modell verknüpft. Dabei soll immer das Modell angewendet werden, was im jeweiligen Moment den geringeren Prognosefehler liefert. Typischerweise wird das Clear-Sky-Modell vorwiegend an wolkenlosen Tagen benutzt. Die Prognose bezieht sich auf den nächsten Tag, wobei eine Datengrundlage von einem Monat gegeben ist. Die Eingangsparameter für das KNN ist die erzeugte elektrische Leistung, der Tag, die Viertelstunde des Tages, die Umgebungstemperatur, die Luftfeuchtigkeit, die Windgeschwindigkeit sowie der Wolkenbedeckungsgrad. Beim Clear-Sky-Modell wird eine physikalische Gleichung für das PV-Modul verwendet, welche abhängig von der Globalstrahlung ist. Der geringste Prognosefehler beträgt 1,9 kW bezogen auf 20 kW Spitzenleistung. Es wurde der RMSE angewendet.

Die Veröffentlichung [5] beschäftigt sich mit einer Verknüpfung von einem Support-Vector-Regression (SVR)-Modell und Random Forest. Zuerst wird die SVR angewendet, deren Prognosegüte in einem zweiten Schritt durch Random Forest verbessert wird. Die Datengrundlage besteht aus der erzeugten elektrischen Leistung und 14 Wetterparametern über 26 Monate in stündlicher Auflösung. Es wird der nächste Tag prognostiziert. Der geringste Prognosefehler (RMSE) beträgt 0,07 kW bezogen auf 1,56 kW Bemessungsleistung.

In keiner der genannten Veröffentlichungen wird ein Prognosehorizont von über einem Tag untersucht. Dies wäre jedoch für eine zuverlässige Einsatzplanung in elektrischen Netzen sowie für ein optimales Energiemanagement von großem Nutzen. Daher soll in dem hier vorgestellten Beitrag ein Modell erstellt werden, welches die erzeugte elektrische Leistung zehn Tage im Voraus in hinreichender Qualität prognostizieren kann. Eingangsparameter sind Verläufe der erzeugten elektrischen Wirkleistung sowie der Wetterdaten. Welche Wetterdaten herangezogen werden, wird in einem ersten Schritt untersucht.

Da eine Datengrundlage von einem Monat vorliegt, ist es nicht sinnvoll KNN-Methoden zu untersuchen. Diese eignen sich erst bei einer größeren Datenmenge, sodass andere Verfahren getestet werden. Es werden vier Machine Learning bzw. Regressionsverfahren untersucht: die MLR, MPR, SVR und der k-nächste-Nachbarn-Algorithmus.

### 3 Methode zur Erstellung des Prognosemodells

#### 3.1 Eingangsparameter für das Modell

Zuerst erfolgt die Auswahl der Eingangsparameter für die Modellbildung. Es sollen nur diejenigen Parameter herangezogen werden, welche einen physikalischen Zusammenhang zur erzeugten elektrischen Wirkleistung aufweisen. Abbildung 1 zeigt das Ersatzschaltbild einer realen Solarzelle nach [6-9].

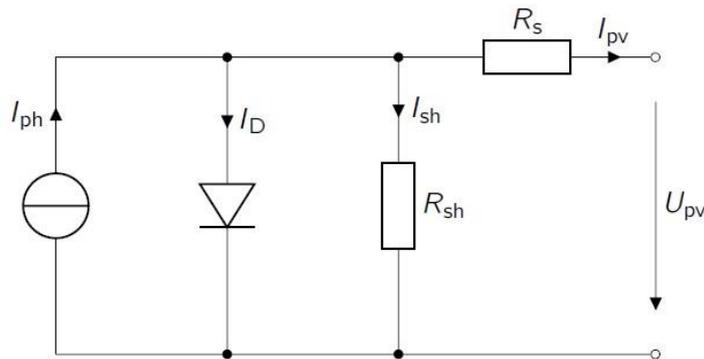


Abbildung 1: Ersatzschaltbild einer Solarzelle nach [6-9]

Gleichung (1) folgt aus dem Kirchhoff'schen Gesetz,

$$I_{PV} = I_{ph} - I_D - I_{sh} \quad (1)$$

mit folgenden Zusammenhängen nach [6]:

$$I_{ph} = \frac{G}{G_{STC}} \cdot (I_{KS\ STC} + \alpha_{KS\ STC} \cdot (T_Z - T_{Z\ STC})) \quad (2)$$

$$I_D = I_{T0} \cdot \left( \exp\left(\frac{e \cdot U_{Diode}}{n \cdot k_B \cdot T_Z}\right) - 1 \right) \quad (3)$$

Die erzeugte elektrische Wirkleistung einer Solarzelle auf der Gleichspannungsseite  $P_{PV\ DC}$  setzt sich aus dem Produkt des Stroms  $I_{PV}$  und der Spannung  $U_{PV}$  zusammen. Dieser Strom kann gemäß Gleichung (1) aus dem photogenerierten Strom  $I_{ph}$  (siehe Gleichung (2)), dem Diodenstrom  $I_D$  (siehe Gleichung (3)) und dem Nebenschlussstrom  $I_{sh}$  gebildet werden.  $G$  steht dabei für die Globalstrahlung,  $\alpha$  für den Temperaturkoeffizienten,  $T_Z$  für die Zelltemperatur,  $I_{T0}$  für den temperaturabhängigen Sperrstrom,  $e$  für die Elementarladung,  $n$  für den Idealitätsfaktor und  $k_B$  für die Boltzmann-Konstante. Der Index STC kennzeichnet die Standardtestbedingungen und KS den Kurzschluss. Die Standardtestbedingungen sind auf  $G = 1000\ W/m^2$ ,  $T_Z = 25\ ^\circ C$  und eine Astronomische Luftmasse von  $AM = 1,5$  festgelegt.

Aus diesen Gleichungen ist die Abhängigkeit der Leistung der Solarzelle von der Globalstrahlung und der Zelltemperatur ersichtlich. Die Globalstrahlung kann direkt vom Deutschen Wetterdienst (DWD) bezogen werden, die Zelltemperatur jedoch nicht. Es kann angenommen werden, dass diese ungefähr der Modultemperatur  $T_M$  entspricht, da ein Modul die Zusammenschaltung mehrerer Zellen ist [10]. Mit [11] und [12] kann sie wie folgt in Gleichung (4) beschrieben werden.

$$T_Z \approx T_M = T_U + \frac{G}{\mu_0 + \mu_1 \cdot v_W} \quad (4)$$

Damit ist die Zelltemperatur neben der Globalstrahlung abhängig von der Umgebungstemperatur  $T_U$  und der Windgeschwindigkeit  $v_W$ . Diese beiden Parameter können ebenfalls über den DWD bezogen werden.  $\mu_0$  und  $\mu_1$  sind Konstanten.

Ebenfalls übt die Ausrichtung der PV-Anlage einen erheblichen Einfluss auf die Höhe der erzeugten elektrischen Leistung aus. Ist sie nach Süden ausgerichtet, erzeugt sie die meiste Leistung, weil die Globalstrahlung mittags am höchsten ist. Daher wird als Eingangsparameter noch der Azimutwinkel  $\gamma$  der Sonne hinzugenommen. Dieser liegt in einem Wertebereich von  $0 \dots 360^\circ$  vor und ist mittags bei  $180^\circ$ . Für die Umrechnung der Leistung von der Gleichstromleistung zur Wechselstromleistung kann folgende Gleichung verwendet werden:

$$P_{PVAC} = P_{PVDC} - P_V = P_{PVDC} \cdot \eta_{WR} \quad (5)$$

$P_{PVAC}$  steht für die PV-Leistung auf der Wechselstromseite,  $P_V$  für die Verlustleistung und  $\eta_{WR}$  für den Wirkungsgrad des Wechselrichters.

## 3.2 Untersuchte Verfahren zur Prognose

Für alle genannten Verfahren wird auf dem Pythonmodul *sklearn* aufgebaut [13]. In diesem gibt es mehrere Submodule. Für die MLR wird *linear\_model* mit *LinearRegression* und *Ridge*, für die MPR *preprocessing* mit *PolynomialFeatures*, für k-nächste Nachbarn *neighbors* mit *KNeighborsRegressor* und für die SVR *svm* mit *SVR* verwendet.

### 3.2.1 Multiple linear Regression

Die MLR eignet sich für lineare oder annähernd lineare Zusammenhänge, wobei die Ausgangsgröße von mehreren Variablen abhängt. Das Ziel ist es, dass die gefundene Funktion möglichst exakt den zugrundeliegenden Zusammenhang widerspiegelt. Dieses Verfahren wird nach der Methode der kleinsten Quadrate optimiert [14]. Das dazugehörige Gleichungssystem kann ausgehend von [15] wie folgt formuliert werden. Die lineare Modellgleichung lautet:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (6)$$

$$\mathbf{y} = \begin{pmatrix} y_0 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_0 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (7)$$

Die Modellgleichung soll hinsichtlich des Parameters  $\boldsymbol{\beta}$  bestimmt werden. Der Ansatz für die Methode der kleinsten Quadrate dafür sieht folgendermaßen aus:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (8)$$

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}} \quad (9)$$

$\mathbf{y}$  steht dabei für die Ausgangsgröße,  $\mathbf{X}$  für die Eingangsgrößen,  $\boldsymbol{\beta}$  für die Parameter des Modells und  $\boldsymbol{\varepsilon}$  für den Fehler. Der Akzent  $\hat{\phantom{x}}$  symbolisiert geschätzte Größen. Bei der MLR wird der Hyperparameter  $\Psi$  angewendet. Ist  $\Psi > 0$ , werden die Parameter  $\boldsymbol{\beta}$  kleiner und das Modell robuster gegenüber Kollinearität. Die Daten werden für diese sogenannte Ridge-Regression normalisiert. Dafür wird der Mittelwert von dem aktuellen Wert abgezogen und durch die L2-Norm dividiert.

### 3.2.2 Multiple polynomielle Regression

Oft unterliegen Vorgänge in der Natur nichtlinearen Zusammenhängen. Daher kann die MPR diese besser als eine MLR abbilden. Jedoch muss darauf geachtet werden, dass das Modell noch allgemeingültig genug ist, damit es auf andere Datensätze ebenfalls angewendet werden kann. Die Regressionsgleichung für beispielsweise die zweite Ordnung mit zwei Eingangsgrößen lautet:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 + \hat{\beta}_4 X_1^2 + \hat{\beta}_5 X_2^2 \quad (10)$$

Auch hier gilt es wieder, die Schätzparameter  $\hat{\beta}$  zu ermitteln, um ausgehend davon die gesuchte Größe  $\hat{y}$  zu schätzen.

### 3.2.3 k-nächste Nachbarn

Diese Methode verfolgt den Ansatz, dass ähnliche Eingabewerte ähnliche Ausgabewerte hervorrufen. Aufgrund dessen wird die Zielgröße basierend auf dem lokalen Verhalten der Eingangsparameter ermittelt. Dabei spielt der Abstand zwischen zwei Punkten  $A$  und  $B$  eine erhebliche Rolle. Als Distanzmaß wird vorwiegend entweder der euklidische Abstand (siehe Gleichung (11)) oder die Manhattan-Distanz (siehe Gleichung (12)) gewählt [16].  $k$  symbolisiert die Anzahl der berücksichtigten Nachbarn und  $p$  die Anzahl der Prädiktoren. Der Algorithmus weist dem zu untersuchenden Punkt einen Wert zu, der sich aus den Werten der  $k$ -nächsten Nachbarn ergibt.

$$d_{\text{eukl}} = \sqrt{\sum_{v=1}^p (B_v - A_v)^2} \quad (11)$$

$$d_{\text{manh}} = \sum_{v=1}^p |B_v - A_v| \quad (12)$$

Der Algorithmus lässt sich wie folgt nach [16] darstellen:

1. Wähle für  $k$  einen Wert.
2. Berechne die Distanz zwischen allen Punkten  $A$  und  $B$ .
3. Ordne die Punkte  $A$  nach aufsteigender Distanz.
4. Wähle  $k$  Punkte mit der geringsten Distanz zum Punkt  $B$ .
5. Berechne die Zielgröße  $\hat{y}$  am Punkt  $B$  aus dem Durchschnitt nach Gleichung (13).

$$\hat{y} = \frac{1}{k} \sum_{u=1}^k y_u \quad (13)$$

Ebenfalls kann die Methode gewichtet ausgeführt werden. Dabei unterscheiden sich die Punkte 4 und 5. Diese lauten nach [17]:

4. Wähle  $k$  Punkte mit der geringsten Distanz  $d_u$  zum Punkt  $B$  und weise jedem Punkt ein Gewicht  $w_u$  nach Gleichung (14) zu.
5. Berechne die Zielgröße  $\hat{y}$  am Punkt  $B$  aus dem Durchschnitt nach Gleichung (15).

$$w_u = \begin{cases} \frac{1}{d_u}, & \text{wenn } A \neq B \\ 1, & \text{sonst} \end{cases} \quad (14)$$

$$\hat{y} = \frac{\sum_{u=1}^k w_u y_u}{\sum_{u=1}^k w_u} \quad (15)$$

Üblicherweise werden die Eingangsdaten normiert, da es zu unkorrekten Ergebnissen bei der gewichteten k-nächste-Nachbarn-Methode aufgrund der unterschiedlichen Wertebereiche der Eingangsdaten kommen kann. Die Normierung wird wie in Gleichung (16) durchgeführt.  $\mu$  stellt den Mittelwert,  $\sigma$  die Standardabweichung und  $x$  den Eingabewert dar. Nach der Normierung weist  $x_{\text{normiert}}$  einen Mittelwert von 0 und eine Standardabweichung von 1 auf.

$$x_{\text{normiert}} = \frac{x - \mu}{\sigma} \quad (16)$$

### 3.2.4 Support Vector Regression

Die SVR ermöglicht lineare, polynomielle oder radial-basierte Regressionen. Es wird der Parameter  $\varepsilon$  definiert, welcher ein Toleranzband für die Instanzen der gewählten Funktion darstellt. Ziel ist es, die Ordnung der Regressionsgleichung möglichst gering zu halten. Neben  $\varepsilon$  wird noch der Hyperparameter  $C$  berücksichtigt. Dieser wird bei der Optimierung der Ordnung verwendet und bildet einen Kompromiss zwischen einer möglichst flachen Funktion und der Berücksichtigung außerhalb des Toleranzbandes liegender Instanzen. Die Eingangsdaten werden wie bei der k-nächste-Nachbarn-Methode normiert. Für weitere Informationen sei auf [18] und [19] verwiesen.

## 4 Ergebnisse

Die untersuchte PV-Anlage weist eine Bemessungsleistung von 84 kW auf und die Leistung kann nicht abgeregelt werden. Die Fläche der Module beträgt 523,6 m<sup>2</sup> und der Wirkungsgrad von diesen 16,2 %. Der Wechselrichter hat einen Wirkungsgrad von 98,2 %, der als konstant angenommen wird. Ziel ist es, den Prognosefehler der Wirkungsleistung so weit wie möglich zu reduzieren. Dafür werden in einem ersten Schritt Trainingsdatensätze untersucht, die sich aus Daten bestehend aus Mittelwerten über 10-Minuten-Intervalle für einen Monat zusammensetzen. Für die Zeit zwischen Sonnenuntergang und Sonnenaufgang wird die elektrische Leistung auf 0 kW gesetzt. Für die Eingangsparameter wird unterschieden, ob der Azimutwinkel mit hinzugezogen wird. Da nur eine nach Süden ausgerichtete PV-Anlage untersucht wurde, kann nicht sicher gesagt werden, ob eine anders ausgerichtete Anlage die gleichen Ergebnisse mit dem Azimutwinkel als Eingangsparameter liefert. Daher wird der Azimutwinkel für künftige, andere Anlagen vernachlässigt. Ansonsten wird stets die Globalstrahlung als Eingangsparameter verwendet, da die PV-Leistung zweifach von dieser abhängig ist, wie in den Gleichungen (1) bis (5) gezeigt wurde. Für die Umgebungstemperatur und die mittlere Windgeschwindigkeit wird untersucht, ob sie den Prognosefehler verringern. Die Ergebnisse werden anhand der Standardabweichung und des Gütekriteriums  $MAE\%$  evaluiert. Dies ist der auf die Bemessungsleistung  $P_r$  der PV-Anlage bezogene Mean Absolute Error (MAE).

$$MAE\% = \frac{\frac{1}{N} \sum_{n=1}^N |\hat{P}_n - P_n|}{P_r} \quad \text{mit } P_r = 84 \text{ kW} \quad (17)$$

#### 4.1.1 Ergebnisse der multiplen linearen Regression

Die Ergebnisse der MLR für die jeweiligen Eingangsparameter sind in Tabelle 1 dargestellt:

Tabelle 1: Ergebnisse der MLR

Eingangsparameter	$MAE\%_{\text{Min}}$ in %	$\sigma$ in kW
$G$	1,17	0,065
$G, T_U$	1,25	0,124
$G, v_W$	1,22	0,126
$G, T_U, v_W$	1,26	0,125
$G, \gamma$	1,12	0,130
$G, T_U, \gamma$	1,25	0,124
$G, v_W, \gamma$	1,22	0,126
$G, T_U, v_W, \gamma$	1,26	0,125

Die ermittelten  $MAE\%_{\text{Min}}$  liegen nahe beieinander. Das beste Ergebnis liefert das Modell mit der Globalstrahlung und dem Azimutwinkel als Eingangsparameter. Wenn der Azimutwinkel nicht berücksichtigt wird, ist es am besten, nur die Globalstrahlung als Eingangsparameter zu verwenden. Wird die Messungenauigkeit mit einbezogen, gibt es keinen signifikanten Unterschied, ob der Azimutwinkel hinzugezogen werden sollte oder nicht.

#### 4.1.2 Ergebnisse der multiplen polynomiellen Regression

Bei der MPR wurden die Ordnungen 2 bis 14 untersucht. Die besten Ergebnisse der jeweiligen Kombination der Eingangsparameter finden sich in Tabelle 2. Es sind die dazugehörigen Ordnungen notiert.

Tabelle 2: Ergebnisse MPR

Eingangsparameter	Ordnung	$MAE\%_{\text{Min}}$ in %	$\sigma$ in kW
$G$	8	1,06	0,121
$G, T_U$	7	1,04	0,111
$G, v_W$	6	1,07	0,112
$G, T_U, v_W$	6	1,04	0,088
$G, \gamma$	4	1,08	0,104
$G, T_U, \gamma$	4	1,08	0,100
$G, v_W, \gamma$	5	1,13	0,101
$G, T_U, v_W, \gamma$	4	1,07	0,102

Auch bei dieser Regressionsart liegen die Werte nahe beieinander. Zu bemerken ist, dass die optimale Ordnungszahl zwischen 4 und 8 liegt. Bei der Variante ohne Azimutwinkel ist diejenige mit den Eingangsparametern Globalstrahlung und Umgebungstemperatur oder zusätzlich noch mit der Windgeschwindigkeit am besten. Da im letzten Fall die Standardabweichung etwas geringer ist, wird die Variante mit allen drei Wetterparametern als bestes Ergebnis gewählt. Wird der Azimutwinkel berücksichtigt, liegt der geringste Prognosefehler unter Verwendung aller untersuchten Eingangsparameter vor. Da auch hier kein signifikanter Unterschied vorliegt, kann keine eindeutige Aussage getroffen werden, welche Variante die beste ist.

#### 4.1.3 Ergebnisse der k-nächste-Nachbarn-Methode

Der Parameter  $k$  wird zwischen 1 und 120 variiert. Die Ergebnisse der k-nächste-Nachbarn-Methode sind in Tabelle 3 dargestellt. Die Anzahl der Nachbarn  $k$  ist unterschiedlich ausgefallen und reicht von 3 bis 37. Zumeist ist die gewichtete Variante mit dem Manhattan-Distanzmaß die genauere. Das beste Ergebnis wird unter Hinzunahme des Azimutwinkels inklusive der drei anderen Wetterparameter erzielt. Wenn der Azimutwinkel nicht beachtet werden soll, ist die Variante mit den drei verbliebenen Parametern die beste. Allerdings ist diese um rund 0,9 % ungenauer als die zuvor genannte.

Tabelle 3: Ergebnisse der k-nächste-Nachbarn-Methode

Eingangsparameter	$k$	Distanzmaß	gewichtet/ ungewichtet	$MAE\%_{\text{Min}}$ in %	$\sigma$ in kW
$G$	33	Manhattan	ungewichtet	1,09	0,118
$G, T_U$	20	Manhattan	gewichtet	0,98	0,087
$G, v_W$	14	Euklidisches	ungewichtet	1,05	0,136
$G, T_U, v_W$	6	Manhattan	gewichtet	0,90	0,099
$G, \gamma$	37	Manhattan	gewichtet	1,05	0,120
$G, T_U, \gamma$	7	Manhattan	gewichtet	0,85	0,077
$G, v_W, \gamma$	11	Manhattan	gewichtet	1,00	0,099
$G, T_U, v_W, \gamma$	3	Manhattan	gewichtet	0,81	0,075

#### 4.1.4 Ergebnisse der Support Vector Regression

Bei der SVR gibt es drei verschiedene Einstellparameter. Es wird in einen linearen, polynomiellen und radialen-Basisfunktion-Kernel unterschieden. Die geringsten Prognosefehler traten immer nur bei Verwendung des radialen-Basisfunktion-Kernel auf. Der Wertebereich für den Parameter  $C$  wird zwischen 5 und 24 und für  $\varepsilon$  wird zwischen 0 und 0,6 gewählt. Die geringsten Fehler traten bei einem  $\varepsilon$  von 0 auf. Die Ergebnisse dieser Methode sind in Tabelle 4 dargestellt. Das beste Ergebnis liefert die Variante mit allen Eingangsparametern und dem Azimutwinkel. Wird der Azimutwinkel vernachlässigt, ist der Prognosefehler nur um 0,01 % größer, sodass auch hier kein signifikanter Unterschied vorliegt.

Tabelle 4: Ergebnisse der SVR

Eingangsparameter	$C$	$MAE\%_{\text{Min}}$ in %	$\sigma$ in kW
$G$	20	1,01	0,128
$G, T_U$	15	0,97	0,120
$G, v_W$	20	0,99	0,099
$G, T_U, v_W$	15	0,90	0,100
$G, \gamma$	15	0,99	0,121
$G, T_U, \gamma$	20	0,95	0,111
$G, v_W, \gamma$	20	0,98	0,099
$G, T_U, v_W, \gamma$	20	0,89	0,094

#### 4.1.5 Auswahl des Modells

Generell liegen die Prognosefehler alle nah beieinander. Die geringsten Prognosefehler liefert die k-nächste-Nachbarn-Methode, sodass nur diese im Weiteren betrachtet wird. Fast genauso gut ist die SVR. Die MLR und MPR folgen in einigem Abstand, wobei die MLR am schlechtesten abschneidet. Das beste Ergebnis von der k-nächste-Nachbarn-Methode liegt bei 0,81 % und bei der MLR bei 1,12 %. Auffällig ist, dass die zwei zuletzt untersuchten Modelle am besten prognostizieren, wenn alle Eingangsparameter hinzugezogen werden und die MPL und MPR auch schon mit wenigen Parametern ihr optimales Ergebnis erzielen. Die Standardabweichungen liegen bei allen untersuchten Varianten im ähnlichen Intervall von 0,065 bis 0,136.

Wird die k-nächste-Nachbarn-Methode für die Variante ohne Azimutwinkel durchgeführt, erhält man einen  $MAE\%$  von 6,98 % auf die vorhergesagten und 1,79 % auf die tatsächlich eingetretenen Wetterdaten. Bei der Variante mit dem Azimutwinkel liegen die Werte bei 7,32 % bzw. 2,11 %. Diese sind um einiges größer als in Kap. 4.1.3, jedoch sei darauf verwiesen, dass in Kap. 4.1.1. bis 4.1.4. nur Trainingsdatensätze über einen Monat und hier Testdatensätze über zehn Tage untersucht werden. Ebenso ist ersichtlich, dass die Prognose erwartungsgemäß für die wirklich eingetretenen Wetterparameter besser abschneidet. Im Testdatensatz ist das Modell ohne den Azimutwinkel etwas besser als mit diesem. Die Ergebnisse des ausgewählten Modells sind in den Abbildung 2 bis 5 dargestellt. Es ist zu erkennen, dass die Verläufe für die vorhergesagten Wetterdaten mehr von der eingetretenen Leistung abweichen als wenn die Prognose mit den eingetretenen Wetterdaten vorgenommen wird. Am 5. und 11. Dezember 2021 sind die Abweichungen am größten. Eine Ursache hierfür könnte sein, dass die Wolkenzüge an diesen Tagen stärker von der Vorhersage bzw. Messung abweichen, da diese nur für einen größeren Radius angegeben werden und nicht speziell für die Fläche der PV-Anlage.

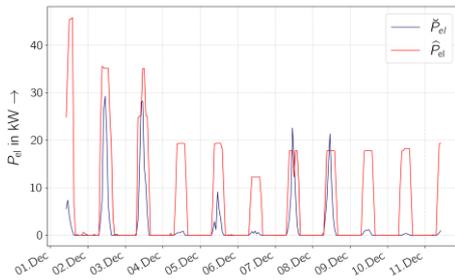


Abbildung 2: Modell ohne Azimutwinkel für vorhergesagte Wetterdaten

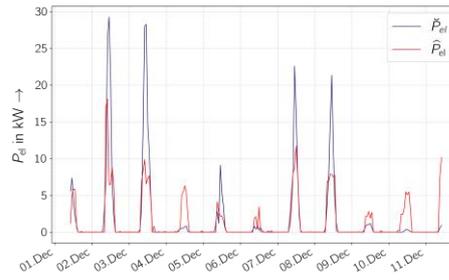


Abbildung 3: Modell ohne Azimutwinkel für eingetretene Wetterdaten

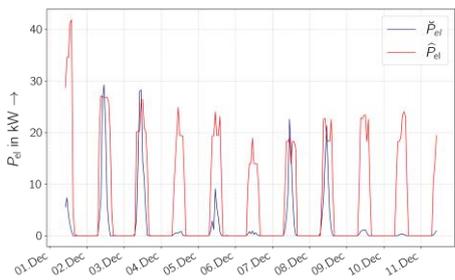


Abbildung 4: Modell mit Azimutwinkel für vorhergesagte Wetterdaten

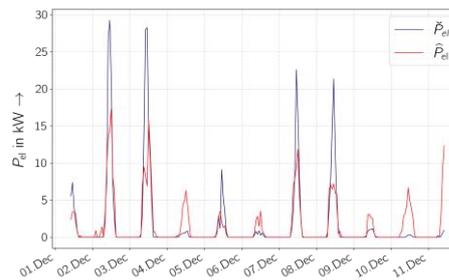


Abbildung 5: Modell mit Azimutwinkel für eingetretene Wetterdaten

## 5 Zusammenfassung

Es konnte ein Modell der erzeugten elektrischen Wirkleistung einer PV-Anlage für einen Prognosehorizont von zehn Tagen erstellt werden. Die Messdaten wurden über einen Monat hinweg aufgenommen. Der maximale Prognosefehler liegt unter 7,5 %, wobei der  $MAE\%$  als Gütekriterium verwendet wird. Für das Modell werden am Anfang mögliche Wetterdaten als Eingangsparameter ausgewählt. Als geeignete Parameter haben sich die Globalstrahlung, die mittlere Windgeschwindigkeit, die Außentemperatur auf 2 m Höhe und der Azimutwinkel herausgestellt. Es wurde in ein Modell mit und eins ohne den Azimutwinkel unterschieden, da die untersuchte PV-Anlage südlich ausgerichtet ist. Wie sich die Hinzunahme dieses Parameters auf eine nicht südlich ausgerichtete Anlage bezüglich des Prognosefehlers auswirken würde, kann nicht sicher gesagt werden.

Die Kombination der Eingangsparameter wurde auf die Modelle MLR, MPR, k-nächste-Nachbarn-Methode und die SVR angewendet. Den geringsten Prognosefehler bringt dabei die k-nächste-Nachbarn-Methode hervor, wobei insgesamt die Unterschiede zwischen den Modellen maximal 0,5 Prozentpunkte betragen. Die beste Variante mit dem Azimutwinkel wird bei den Parametern  $k = 6$ , der Manhattan-Distanz und der gewichteten Methode erzielt. Bei der Variante ohne den Azimutwinkel unterscheidet sich nur die Anzahl der Nachbarn mit  $k = 3$ .

Als nächsten Schritt gilt es, weitere Zeitreihen derselben PV-Anlage und danach eine nicht südlich ausgerichtete PV-Anlage zu untersuchen und die hier erzielten Ergebnisse daran zu evaluieren.

## 6 Referenzen

- [1] M. J. Mayer, G. Gróf, "Extensive comparison of physical models for photovoltaic power forecasting", *Applied Energy*, Vol. 283, S. 116239, 2021. <https://www.sciencedirect.com/science/article/pii/S0306261920316330> (02.12.2021)
- [2] V. Kushwaha, N. M. Pindoriya, "Very Short-Term Solar PV Generation Forecast Using SARIMA Model: A Case Study", 7th International Conference on Power Systems (ICPS) College of Engineering Pune, pp. 430-435, 2017.
- [3] F. O. Hocaoglu, M. Kurekci, E. Akarslan, F. Serttas, "An Experimental Study on the Modeling of the PV Output", International Conference on Photovoltaic Science and Technologies (PVCon), 2018
- [4] G. Mosaico, M. Saviozzi, "A hybrid methodology for the day-ahead PV forecasting exploiting a Clear Sky Model or Artificial Neural Networks", IEEE EUROCON 2019 -18th International Conference on Smart Technologies, pp. 1-6, 2019
- [5] M. Abuella, B. Chowdhury, "Random Forest Ensemble of Support Vector Regression Models for Solar Power Forecasting", 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017
- [6] N. Kaushika, A. Mishra, und A. K. Rai, "Solar Photovoltaics Technology, System Design, Reliability and Viability", Cham, 2018. <https://katalog.slub-dresden.de/id/0-1654952753> (02.12.2021)
- [7] B. K. Dey, I. Khan, N. Mandal, und A. Bhattacharjee, "Mathematical modelling and characteristic analysis of Solar PV Cell", in 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016, S. 1–5.
- [8] R. Krishan, Y. R. Sood, und B. Uday Kumar, "The simulation and design for analysis of photovoltaic system based on MATLAB", in 2013 International Conference on Energy Efficient Technologies for Sustainability, 2013, S. 647–651.
- [9] D.-P. Han, D.-S. Shin, J.-I. Shim, S. Kamiyama, T. Takeuchi, M. Iwaya, und I. Akasaki, "Modified Shockley Equation for GaInN-Based Light-Emitting Diodes: Origin of the Power-Efficiency Degradation Under High Current Injection", *IEEE Journal of Quantum Electronics*, Vol. 55, Nr. 4, S. 1–11, 2019.
- [10] M. Mattei, G. Notton, C. Cristofari, M. Muselli, und P. Poggi, "Calculation of the polycrystalline PV module temperature using a simple method of energy balance", *Renewable Energy*, Vol. 31, Nr. 4, S. 553–567, 2006. <https://www.sciencedirect.com/science/article/pii/S096014810500073X> (02.12.2021)
- [11] D. Faiman, "Assessing the outdoor operating temperature of photovoltaic modules", *Progress in Photovoltaics: Research and Applications*, Vol. 16, Nr. 4, S. 307–315, 2008. <https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.813> (03.12.2021)
- [12] A. Shah, "Solar Cells and Modules", 2020. <https://katalog.slub-dresden.de/id/0-1726031462> (03.12.2021)
- [13] <https://scikit-learn.org/stable/modules/classes.html#module-sklearn>. (07.12.2021)

- [14] G. James, D. Witten, T. Hastie, und R. Tibshirani, "An introduction to statistical learning with applications in R", New York, 2017. <https://katalog.slub-dresden.de/id/0-1024811107> (07.12.2021)
- [15] J. H. Hastie, T. Friedman und R. Tibshirani, "The Elements of Statistical Learning Data Mining, Inference, and Prediction", New York, 2001. <https://katalog.slub-dresden.de/id/0-1655388487> (07.12.2021)
- [16] M. Fičura, "Forecasting foreign exchange rate movements with k-Nearest-Neighbor, ridge regression and feed-forward neural networks", Prag, 2019, Kap. 1.1. <https://katalog.slub-dresden.de/id/0-1700288288> (07.12.2021)
- [17] G.-f. Fan, Y.-H. Guo, J.-M. Zheng, und W.-C. Hong, "Application of the Weighted K-Nearest Neighbor Algorithm for Short-Term Load Forecasting", *Energies*, Vol. 12, S. 916, 2019.
- [18] A. J. Smola und B. Schölkopf, "A tutorial on support vector regression", 2004. <https://katalog.slub-dresden.de/id/ai-49-aHR0cDovL2R4LmRvaS5vcmcvMTAuMTAyMy9iOnN0Y28uMDAwMDAzNTMwMS40OTU0OS44OA> (07.12.2021)
- [19] M. Awad und R. Khanna, "Support Vector Regression", Berkeley, CA: Apress, 2015, S. 67–80. [https://doi.org/10.1007/978-1-4302-5990-9\\_4](https://doi.org/10.1007/978-1-4302-5990-9_4) (07.12.2021)