

Erstellung von Langzeitprognosemodellen industrieller, elektrischer Lastgänge mittels Kurzzeitmessung und Gauß'scher Prozesse

Marta Sophia Potyka, 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: Die Anforderungen an industrielle Unternehmen bezogen auf einen effizienten Umgang mit unter anderem elektrischer Energie steigen stetig. Deren Umsetzung kann ein Energiemanagement erleichtern. Dafür ist es vorteilhaft, sich auf elektrische Messwerte des gesamten Lastganges und einzelner technischer Anlagen über mindestens ein Jahr zu stützen, um auch saisonale Komponenten berücksichtigen zu können. Doch zumeist existiert eine solche Datengrundlage nicht, da sie vor allem nicht für einzelne Anlagen gefordert wird und die wenigsten Unternehmen diese zeit- und kostenaufwendige Messung freiwillig durchführen lassen. Um diese aufwendige Messung zu ersetzen, wurde ein Langzeitprognosemodell industrieller, elektrischer Lastgänge über ein Jahr erstellt, das als Trainings- und Validierungsdatensatz nur eine Messung von vier Wochen benötigt. Das Modell wurde mittels Gauß'scher Prozesse als eine stochastische Methode erstellt.

Keywords: industrielle Langzeitprognosemodelle, elektrische Lastgänge, Kurzzeitmessung, Gauß'sche Prozesse

1 Einleitung

Mit den steigenden Anforderungen an industrielle Unternehmen hinsichtlich einer effizienten Energienutzung setzen diese vermehrt ein Energiemanagement nach DIN EN ISO 50001 oder ein Umweltmanagement nach DIN EN ISO 14001 um. Die Normen fordern unter anderem die Bildung und Bewertung von Energieleistungskennzahlen. Damit diese hinreichend validiert werden können, werden in der Regel Messwerte verwendet. Da jedoch die meisten Unternehmen maximal eine registrierende Leistungsmessung am Übergabepunkt zum Energieversorger aufweisen und selten zusätzliche Messungen einzelner technischer Anlagen durchführen, ist eine solche Datengrundlage nicht gegeben. Daher werden die Kennzahlen zumeist aus wenigen Wochen Messdaten festgelegt und stichprobenartig überprüft. [1] Damit die Bildung und Bewertung der Kennzahlen verbessert werden kann, wurden Langzeitprognosemodelle über ein Jahr erstellt. Dabei wurden elektrische Lastgänge der Wirkleistung untersucht, wobei diese auch durch andere Lastgänge ersetzt werden können. Um eine zeit- und kostenaufwendige Messung zu umgehen, dienen nur vier Wochen als Trainings- und Validierungsdaten, mit welchen die Modelle aufgestellt werden. Als Vergleichsmessung und damit Testdatensatz wurde eine Jahresmessung durchgeführt.

Zur Erstellung von Langzeitprognosemodellen wurden bisher mehrere Ansätze untersucht. Zunehmend finden *Künstliche Neuronale Netze* (KNN) und *Machine Learning* (ML) Methoden Anwendung, wobei vor allem erstere im Normalfall mit mindestens genauso vielen Daten trainiert werden müssen, wie vorhergesagt werden sollen, um eine adäquate Prognosegüte zu erreichen. Daher wurden sie für diesen Anwendungsfall nicht untersucht. Andere Ansätze sind klassische Regressionsmethoden oder Zeitreihenanalysen. Die klassischen Regressionsmethoden wie die Multiple Lineare Regression weisen in der Literatur zumeist nur mittlere Prognosegüten auf. Die Zeitreihenanalysen erzielen schon genauere Ergebnisse, wobei selten andere Eingangsparameter als Zeitparameter verwendet werden können. In der hier vorliegenden Untersuchung sollen andere jedoch nicht von vornherein ausgeschlossen werden. Als ein nicht so verbreiteter, aber im Hinblick auf die Prognosegüte bei geringer Datengrundlage vielversprechender Ansatz gelten die *Gauß'schen Prozesse* (GPs), welche daher in dieser Veröffentlichung verwendet werden. [2],[3],[4]

In der Literatur wurde keine Veröffentlichung über Langzeitprognosemodelle basierend auf einer Trainings- und Validierungsdatengrundlage von nur wenigen Wochen gefunden. Um dennoch Referenzwerte zum Evaluieren der Prognosemodelle zu erhalten, wurden solche Veröffentlichungen ausgewählt, die im Hinblick auf Prognosehorizont, Datengrundlage und untersuchten Unternehmen am ehesten diesem Anwendungsfall ähneln. Tabelle 1 zeigt die Untersuchungen mit ihren Charakteristika. ARMA steht dabei für die *Auto-Regressive Moving Average* Methode als eine Zeitreihenanalyse, SVM für *Support Vector Machine* und kNN für *k-Nearest Neighbor* als ML Methoden. Aus der Tabelle ist ersichtlich, dass die Größe des Trainingsdatensatzes mindestens so groß wie die des Testdatensatzes ist. Die besten Ergebnisse erzielen einen *Mean Absolute Percentage Error* (MAPE) von ca. 8,9 %. In Untersuchung [6] wurden als Gütekriterien stattdessen unter anderem der *Mean Absolute Error* (MAE) und der *Root Mean Square Error* (RMSE) verwendet, wobei hier zur Vergleichbarkeit mit der eigenen Untersuchung auf den maximalen Wert normalisiert wurde, weswegen das Präfix *normalized* für die beiden ergänzt wurde. Es wurde ein $nMAE \approx 9,5 \%$ und $nRMSE \approx 13,5 \%$ erzielt. Auf die Gütekriterien mit ihrer Definition und Bedeutung wird im Kapitel 2.6 tiefer eingegangen. Da in der eigenen Untersuchung ein viel kleinerer Trainings- und Validierungsdatensatz als der Testdatensatz verwendet wird, wird davon ausgegangen, dass sich deutlich höhere Prognosefehler ergeben. Wie viel höher diese ausfallen, kann nicht gesagt werden. Als Referenzwert wird 30 % für die Gütekriterien angestrebt, was ungefähr dem dreifachen Wert der besten Ergebnisse der herausgesuchten Veröffentlichungen entspricht und weniger als ein Drittel Prognosefehler bedeutet. Der $nMAE$ sollte jedoch wegen seiner Normalisierung auf den Maximalwert geringer ausfallen. In der letzten Zeile von Tabelle 1 sind die Charakteristika der eigenen Untersuchung ergänzt.

Tabelle 1: Referenzwerte aus der Literatur.

Objekt	Spitzenwert	Ausgangsparameter	Größe Trainingsdatensatz	Größe Testdatensatz	Methode	Güte	Referenz
Haus-halt	8 kW	Minütliche Leistung	2 Jahre	1 Tag	GP, ARMA	[18%,93%] (MAPE)	[5]
Ge-bäude	800 kWh	Tägliche Energie	50 Tage	50 Tage	GP, SVM	≈9,5% (nMAE), ≈13,5% (nRMSE)	[6]
Unter-nehmen	3,5 MW	Stündliche Wirkleistung	3 Jahre	1 Jahr	Curve-Fitting mit Fourier-Reihe	≈8,9% (MAPE)	[7]
Ferien-anlage	350 kW	Stündliche Wirkleistung	32 Tage	8 Tage	KNN, kNN	≈8,9% (MAPE), ≈9,2% (MAPE)	[8]
Unter-nehmen	< 120 kW	Substündliche Wirkleistung	< 4 Wochen	1 Jahr	GP	≈30% (MAPE, nRMSE)	Diese Veröffentlichung

2 Methodik

In diesem Kapitel werden zuerst die verschiedenen Datensätze, die Auswahl der Eingangsparameter sowie die Datenvorverarbeitung vorgestellt, bevor die GPs erläutert werden.

2.1 Untersuchte Datensätze

Es wurden drei Datensätze von drei unterschiedlichen industriellen Unternehmen untersucht, um das Prognosemodell breiter anwendbar aufstellen zu können. Dafür wurden Messwerte von einem Autohaus und einer Werkstatt aus Hoyerswerda in Sachsen und einer Bäckerei in der Nähe von Müncheberg in Brandenburg aufgenommen. Jedes Unternehmen weist ein anderes Muster, ein sogenanntes Pattern, im Leistungszeitverlauf auf, wie in Kap. 3.2 gezeigt wird. Charakteristika der drei Datensätze sind in Tabelle 2 aufgelistet. Die ersten zwei Messungen wurden in einer 10-minütlichen Auflösung selbst aufgenommen. Der dritte Datensatz wurde freundlicherweise durch die GICON - Großmann Ingenieur Consult GmbH in einer 15-minütlichen Auflösung zur Verfügung gestellt.

Tabelle 2: Charakteristika der drei untersuchten Datensätze.

Parameter	Autohaus	Werkstatt	Bäckerei
Zeitliche Auflösung	10 Minuten	10 Minuten	15 Minuten
Grundlast	2 – 2,5 kW	8 – 10 kW (Sommer) 10 – 20 kW (Winter)	2 – 3 kW
Spitzenlast	35 kW	111 kW	41 kW

2.2 Trainings-, Validierungs- und Testdatensätze

Mit dem Trainingsdatensatz werden die Modelle trainiert, mit dem Validierungsdatensatz die Güte von diesen evaluiert und das Modell mit der höchsten Prognosegüte aus der Validierungsphase schlussendlich am Testdatensatz angewendet. Der Trainings- und Validierungsdatensatz umfassten zusammen vier Wochen. Dabei wurde unterschieden, ob die vier Wochen am Stück in der Übergangszeit vom 16. September bis zum 14. Oktober oder einmal zwei Wochen in der Mitte des Sommers, vom 07. bis zum 21. Juli, und einmal in der Mitte des Winters, vom 07. bis zum 21. Januar, verwendet wurden. Bei der ununterbrochenen Messung wurden drei Wochen für das Training und eine für die Validierung, bei der geteilten Messung je eine Woche für das Training und eine für die Validierung genutzt. Der Testdatensatz umfasste immer ein Jahr. Bei dem Autohaus ging dieses vom 01. Mai 2021 bis zum 30. April 2022, bei der Werkstatt vom 01. April 2021 bis zum 31. März 2022 und bei der Bäckerei vom 01. Januar 2018 bis zum 31. Dezember 2018.

2.3 Eingangsparameter für das Prognosemodell

Als mögliche Eingangsparameter stehen Wetter- und Zeitparameter zur Verfügung. Die Wetterdaten werden vom Deutschen Wetterdienst [9] bezogen. Dieser stellt Daten für Deutschland von ca. 200 Wetterstationen über das Climate Data Center. Als Zeitparameter können alle genutzt werden, die aus dem Zeitstempel generiert werden können wie beispielsweise die Stunde oder der Wochentag. Mittels einer Korrelationsanalyse und einer Rastersuche wurden die Eingangsparameter wie in Tabelle 3 festgesetzt. Bei der Erstellung der Prognosemodelle wurden die Varianten nur Wetter-, nur Zeit- und alle Eingangsparameter untersucht.

Tabelle 3: Verwendete Eingangsparameter.

Wetterparameter	Zeitparameter
Außentemperatur auf 2 m Höhe [°C]	Stunde [h] [0, 1, ... 23]
Relative Feuchtigkeit auf 2 m Höhe [%]	Minute des Tages [min] [0, 10, ..., 1430]
Niederschlagsdauer [min] [0 ... 60]	Wochentag [-] [0=Montag, ..., 6=Sonntag]
Mittlere Windgeschwindigkeit [m/s]	Feiertag [-] [Wahr oder Falsch]
Taupunkt auf 2 m Höhe [°C]	
Globalstrahlung [J/cm ²]	
Sonnenscheindauer [h] [0 ... 1]	

2.4 Datenvorverarbeitung

Weil Messwerte fehlerhaft sein können, ist eine Datenvorverarbeitung unerlässlich. Daher wurden bei den vorliegenden Datensätzen doppelte sowie fehlende Zeitstempel und Leistungswerte gelöscht sowie eine untere Grenze von 0 kW für die Leistungswerte festgesetzt. Denn der Verbrauch kann im Verbraucherzählpeilsystem nicht unter 0 kW fallen. Die untersuchten Datensätze unterscheiden sich nicht nur untereinander in ihrem Leistungsverlauf, sondern es lassen sich auch in jedem von ihnen verschiedene, wiederkehrende Muster, sogenannte Tyleistungsverläufe, erkennen. Weiter wurde festgestellt, dass sich die Prognosegüte erhöht, wenn für jeden Tyleistungsverlauf ein eigenes Teilprognosemodell aufgestellt. Deshalb wurde nach einer Sortierung nach der Höhe der Werte und der Bildung der quadratischen Abweichung der k-Means Clustering Algorithmus [10] angewendet. Dieser klassifiziert die Datensätze, wie es in Tabelle 4 gegeben ist.

Tabelle 4: Tyleistungsverläufe der einzelnen Datensätze.

Autohaus	Werkstatt	Bäckerei
Arbeitsleistungsverlauf	Arbeitsleistungsverlauf	Arbeitsleistungsverlauf
Basisleistungsverlauf	Basisleistungsverlauf	
Nachtleistungsverlauf		

2.5 Gauß'sche Prozesse

Ein GP stellt einen stochastischen Prozess dar und arbeitet statt mit Regressionsgleichungen mit Wahrscheinlichkeitsverteilungen. Ein klassischer GP unterliegt einer Gauß'schen Verteilung, welche eine Normalverteilung mit einem Mittelwert $\mu = 0$ ist. Normalverteilungen vereinfachen viele Berechnungen, doch liegen Datensätze selten in dieser vor. Daher gibt es inzwischen einige Approximationen, um alle Datensätze unabhängig von der zugrundeliegenden Wahrscheinlichkeitsverteilung nutzen zu können. In dieser Arbeit wurde der Markov-Chain-Monte-Carlo-Algorithmus, speziell der Hamiltonsche Monte-Carlo-Algorithmus, von GPFlow [11] verwendet, da dieser die genaueste Approximation darstellt.

Eine Wahrscheinlichkeitsverteilung lässt sich durch ihren Mittelwert und ihre Kovarianz beschreiben. Da bei GPs $\mu = 0$ ist, bleibt nur noch die Kovarianz zum Bestimmen, sodass das Finden einer für sie geeigneten Funktion das Wesentliche bei der Bildung eines GPs ist. Bei den GPs wird die Kovarianzfunktion auch Kernel K genannt. GPFlow stellt sogenannte Basiskernel zur Verfügung, die beliebig durch Addition oder Multiplikation miteinander kombiniert werden können. Die Kernelkombination mit der höchsten Prognosegüte wurde für jeden Datensatz und jeden Tyleistungsverlauf mittels einer Rastersuche bestimmt. Die Ergebnisse dazu finden sich in Kap. 3.1. Zusätzlich zum Kernel zeichnen sich GPs durch die Wahl ihres Priors und ihrer Likelihood aus. Der Prior ist die Wahrscheinlichkeitsverteilung, die angenommen wird, ohne die Daten gesehen zu haben, und die Likelihood diejenige, die dem Modell zugeschrieben wird. Dabei sollten beide Wahrscheinlichkeitsverteilungen möglichst nahe an derjenigen der Daten liegen. Die Trainings- und Validierungsdaten wurden für die Bestimmung herangezogen, weswegen es auch in diesem Fall wichtig ist, dass die sie die Testdaten gut widerspiegeln. Zu weiteren Informationen zu den GPs, beispielsweise den Basiskerneln, sei auf [12] und [13] verwiesen.

2.6 Gütekriterien

Für eine bessere Vergleichbarkeit zu anderen Veröffentlichungen wurden zur Evaluation der Ergebnisse vier Kriterien für die Prognosegüte herangezogen: MAPE, nMAE, nRMSE und r_E . Die zugehörigen Definitionen finden sich in Gl. (1) – (4). $P_{\text{progn } t}$ steht dabei für die prognostizierte Leistung zum Zeitschritt t , $P_{\text{wahr } t}$ für die wirklich eingetretene Leistung zum Zeitschritt t , P_{max} für die maximale Leistung des Datensatzes, P_{mittel} für die durchschnittliche Leistung des Datensatzes, E_{progn} für die prognostizierte Energie und E_{wahr} für die wirklich eingetretene Energie über ein Jahr. n ist die Gesamtanzahl von Zeitschritten. Jedes Gütekriterium weist einen anderen Normierungswert auf. So bezieht sich der MAPE auf die aktuelle, der nMAE auf die maximale und der nRMSE auf die durchschnittliche Leistung. Das letzte Gütekriterium, r_E , bewertet nicht die Leistung, sondern den relativen Fehler der prognostizierten Energie.

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{P_{\text{progn } t} - P_{\text{wahr } t}}{P_{\text{wahr } t}} \right| \cdot 100 \% \quad (1) \quad \text{nMAE} = \frac{\frac{1}{n} \sum_{t=1}^n |P_{\text{progn } t} - P_{\text{wahr } t}|}{P_{\text{max}}} \cdot 100 \% \quad (2)$$

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (P_{\text{progn } t} - P_{\text{wahr } t})^2}}{P_{\text{mittel}}} \cdot 100 \% \quad (3) \quad r_E = 100 \% \cdot \left(\frac{E_{\text{progn}}}{E_{\text{wahr}}} - 1 \right) \quad (4)$$

3 Ergebnisse

3.1 Einstellungen der Gauß'schen Prozesse

Tabelle 5 zeigt die Einstellungen der GPs. Das heißt, dass die Kernelkombination, die die höchste Prognosegüte erzielte, mit ihren optimierten Hyperparametern Varianz σ^2 , Länge l und Skalenmix-Parameter α , und die Hyperparameter des Priors α_G und β_G für jeden Typleistungsverlauf aufgelistet sind. Bei der Suche nach einem geeigneten Prior und Likelihood konnten jeweils keine eindeutigen Verteilungsfunktionen gefunden werden. Die Analyse wurde mit der Fitter-Methode [14] in Python durchgeführt. Da jedoch die Verteilungen bestimmt werden müssen, wurde ausgehend von den Ergebnissen der Prior auf die Gamma- und die Likelihood auf die Exponentialverteilung festgelegt.

Für den Arbeitsleistungsverlauf ist die ausgewählte Kernelkombination für alle Datensätze die Addition aus dem *Rational Quadratischen* (RQ), dem *Kosinus* (Kos) und dem *Weißes Rauschen* (WR) Kernel. Die Definitionen zum RQ und Kos Kernel finden sich in den Gl.(5) und (6) nach [11] und [12]. X steht dabei für die Trainings- und X' für die Testdaten. Der WR Kernel wird durch einen Dirac-Impuls abgebildet. Für den Basisleistungsverlauf des Datensatzes vom Autohaus und von der Werkstatt unterscheiden sich die gewählten Kernelkombinationen, da der Basisleistungsverlauf des Datensatzes vom Autohaus mehr einer konstanten Grundlast ähnelt als derjenige von der Werkstatt, welcher eher dem Arbeitsleistungsverlauf gleicht. So wurde sowohl für den Basisleistungsverlauf als auch für den Nachtleistungsverlauf, welcher hauptsächlich einem konstanten Verlauf mit wenigen Sprüngen ähnelt, vom Datensatz des Autohauses eine Kerneladdition aus dem *Konstanten* (K) und dem WR Kernel festgelegt. Bei der Untersuchung wurde festgestellt, dass es besonders wichtig ist, die optimalen

Einstellungen, welche mit einer Rastersuche gesucht wurden, zu finden, da schon eine kleine Abweichung eine deutliche Verringerung der Prognosegüte nach sich ziehen kann.

$$RQ(X, X') = \left(1 + \frac{(X - X')^2}{2\alpha l^2}\right)^{-\alpha} \quad (5) \quad \text{Kos}(X, X') = \sigma^2 \cos\left(2\pi \frac{(X - X')}{l}\right) \quad (6)$$

Tabelle 5: Einstellungen der GPs für die jeweiligen Datensätze.

Typeleistungs- verlauf	Parameter	Wert für das Autohaus	Wert für die Werkstatt	Wert für die Bäckerei
Arbeitsleistungs- verlauf	Kernel- kombination	RQ + Kos + WR	RQ + Kos + WR	RQ + Kos + WR
	σ^2	0,1	1,0	0,001
	l	100,0	100,0	100,0
	α	0,00001	0,00001	0,001
	α_G	0,1	0,1	2,0
	β_G	0,1	0,1	2,0
Basisleistungs- verlauf	Kernel- kombination	K + WR	RQ + Kos + WR	
	σ^2	0,1	0,01	
	l		1,0	
	α		0,00001	
	α_G	0,1	0,5	
	β_G	0,1	0,5	
Nachtleistungs- verlauf	Kernel- kombination	K + WR		
	σ^2	0,1		
	α_G	0,1		
	β_G	0,1		

3.2 Ergebnisse der Prognosegüte

Tabelle 6 bis Tabelle 8 zeigen die Ergebnisse der Prognosegüte für jeden Datensatz, wobei diejenigen mit der höchsten Güte grau hinterlegt sind. Sowohl für den Datensatz des Autohauses als auch der Bäckerei wird die höchste Prognosegüte unter Verwendung von der geteilten Messung des Trainings- und Validierungsdatensatzes sowie von nur Zeitparametern als Eingangsparameter erzielt. Beim Datensatz der Werkstatt wird die höchste Güte für die Prognosevariante mit der ununterbrochenen Messung mit nur Zeitparametern erreicht. Jedoch

werden selbst bei dieser die Referenzwerte für den MAPE und den nRMSE nicht eingehalten. Auch ist r_E über dem Referenzwert von betragsmäßig 30 %.

Tabelle 6: Ergebnisse der Prognosegüte für den Datensatz des Autohauses.

Trainings- und Validierungsdatensatz	Eingangsparameter	MAPE (%)	nMAE (%)	nRMSE (%)	r_E (%)
Ununterbrochene Messung	Zeit	36,1	8,0	41,3	-12,0
	Wetter	37,8	7,0	37,6	-7,3
	Zeit und Wetter	31,2	6,1	33,3	-8,4
Geteilte Messung	Zeit	30,5	5,8	32,0	2,0
	Wetter	36,2	6,5	35,9	1,2
	Zeit und Wetter	30,6	5,7	30,9	1,2

Tabelle 7: Ergebnisse der Prognosegüte für den Datensatz der Werkstatt.

Trainings- und Validierungsdatensatz	Eingangsparameter	MAPE (%)	nMAE (%)	nRMSE (%)	r_E (%)
Ununterbrochene Messung	Zeit	40,6	10,3	48,6	-35,1
	Wetter	49,9	11,9	56,6	-37,2
	Zeit und Wetter	40,4	10,3	48,9	-36,1
Geteilte Messung	Zeit	54,9	6,2	29,9	14,1
	Wetter	66,1	7,6	36,4	14,9
	Zeit und Wetter	59,5	6,6	31,2	16,6

Tabelle 8: Ergebnisse der Prognosegüte für den Datensatz der Bäckerei.

Trainings- und Validierungsdatensatz	Eingangsparameter	MAPE (%)	nMAE (%)	nRMSE (%)	r_E (%)
Ununterbrochene Messung	Zeit	34,8	7,1	26,7	5,6
	Wetter	57,3	11,5	42,3	1,9
	Zeit und Wetter	40,8	8,1	30,4	5,0
Geteilte Messung	Zeit	30,0	6,8	26,7	-1,3
	Wetter	51,6	11,2	41,7	-4,4
	Zeit und Wetter	37,1	8,1	30,7	-0,4

Für die Prognosevarianten mit der höchsten Güte vom Datensatz des Autohauses und der Bäckerei liegen der MAPE und nRMSE ca. bei den angestrebten 30 %. Der r_E und der nMAE liegen nur zwischen -1,30 % und 6,81 %, was die Erwartungen deutlich übertrifft. Weiter kann festgehalten werden, dass die Ergebnisse mit der höchsten Prognosegüte ausschließlich unter der Verwendung von nur Zeitparametern erzielt wurden. Die Nutzung von Wetter- und Zeitparametern lieferte zum Teil ähnliche Ergebnisse, doch lässt sich kein Mehrwert aus der Nutzung von nur Wetterparametern ableiten. Hinsichtlich der prognostizierten Energie ist die geteilte Messung diejenige, welche zu genaueren Ergebnissen führt. Bezogen auf die Datensätze des Autohauses und der Bäckerei gilt dies auch für die anderen Gütekriterien. Damit scheint der Algorithmus besser den Jahreslastgang vorhersagen zu können, wenn die Extremwerte aus Sommer und Winter zur Verfügung stehen, als wenn nur die Übergangswerte erfasst werden.

Bild 1 bis Bild 4 zeigen elektrische Lastgänge vom Autohaus, der Werkstatt und der Bäckerei jeweils über eine Beispielwoche mit dem prognostizierten und tatsächlichen Leistungsverlauf der ausgewählten Prognosevariante. Es ist zu erkennen, dass der Verlauf der prognostizierten Leistung überwiegend mit demjenigen der tatsächlichen übereinstimmt. Für den Datensatz der Werkstatt wurden zwei Beispielwochen ausgewählt, um zu zeigen, dass vor allem bei diesem in manchen Wochen nur geringe und in anderen Wochen größere Abweichungen auftreten. Die Ursache liegt vermutlich in dem relativ unregelmäßigen Datensatz begründet. Denn wie in Bild 2 und Bild 3 zu erkennen ist, steigen bei dem ersten die Werte bis auf ca. 60 kW, bei dem zweiten jedoch auf über 80 kW. Je mehr Unregelmäßigkeiten in einem Datensatz auftreten, desto schwieriger ist dieser zu prognostizieren. Daher sind die Prognosegüten bei dem Datensatz der Werkstatt auch geringer als für die anderen beiden.

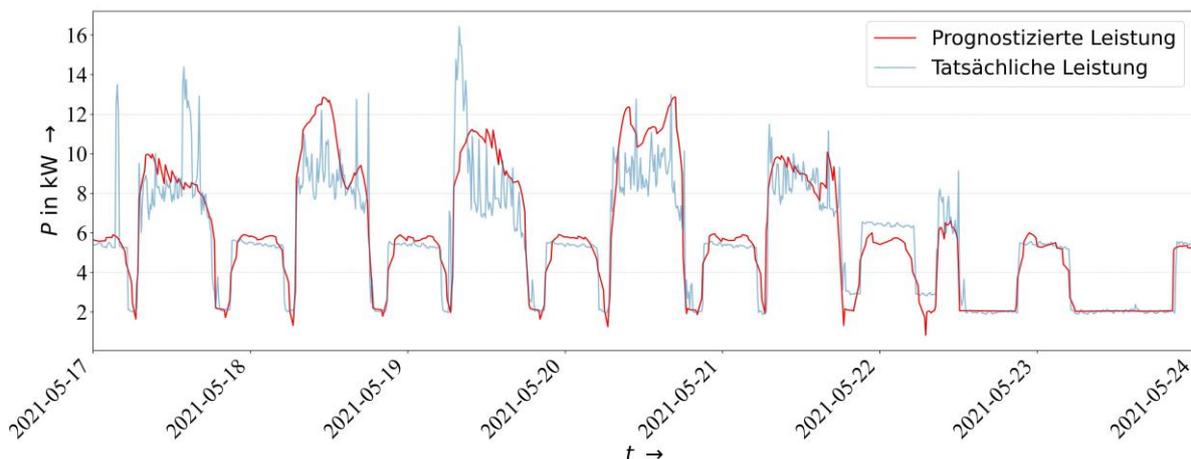


Bild 1: Lastgang vom Autohaus über eine Beispielwoche mit prognostiziertem und tatsächlichem Leistungsverlauf der Prognosevariante mit der geteilten Messung und nur Zeitparametern.

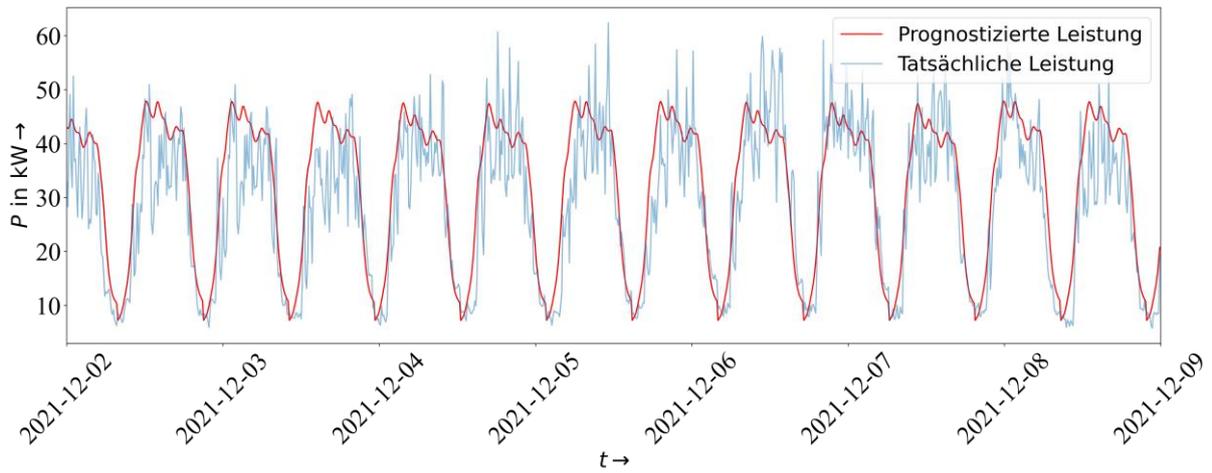


Bild 2: Lastgang von der Werkstatt über eine Beispielwoche 1 mit prognostiziertem und tatsächlichem Leistungsverlauf der Prognosevariante mit der ununterbrochenen Messung und nur Zeitparametern.

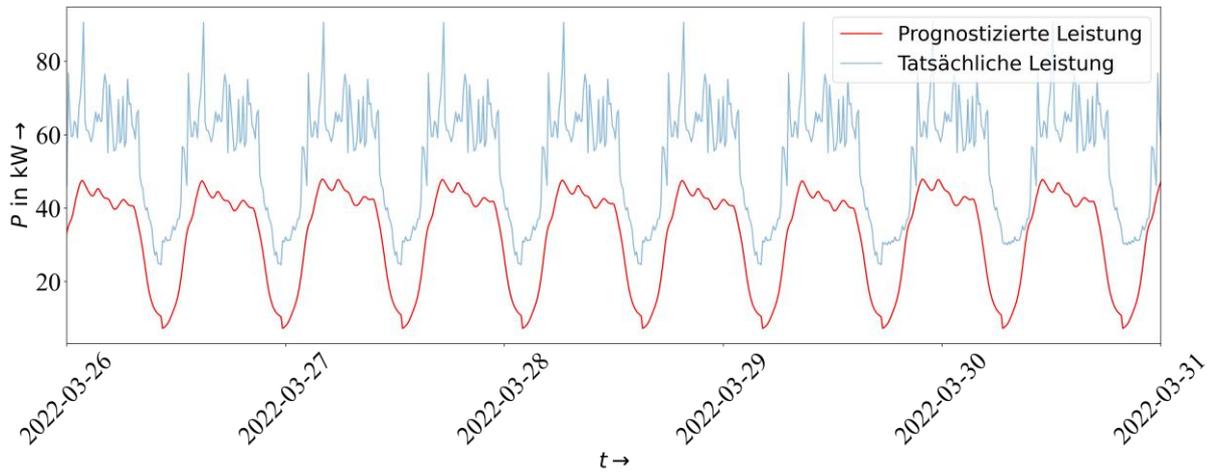


Bild 3: Lastgang von der Werkstatt über eine Beispielwoche 2 mit prognostiziertem und tatsächlichem Leistungsverlauf der Prognosevariante mit der ununterbrochenen Messung und nur Zeitparametern.

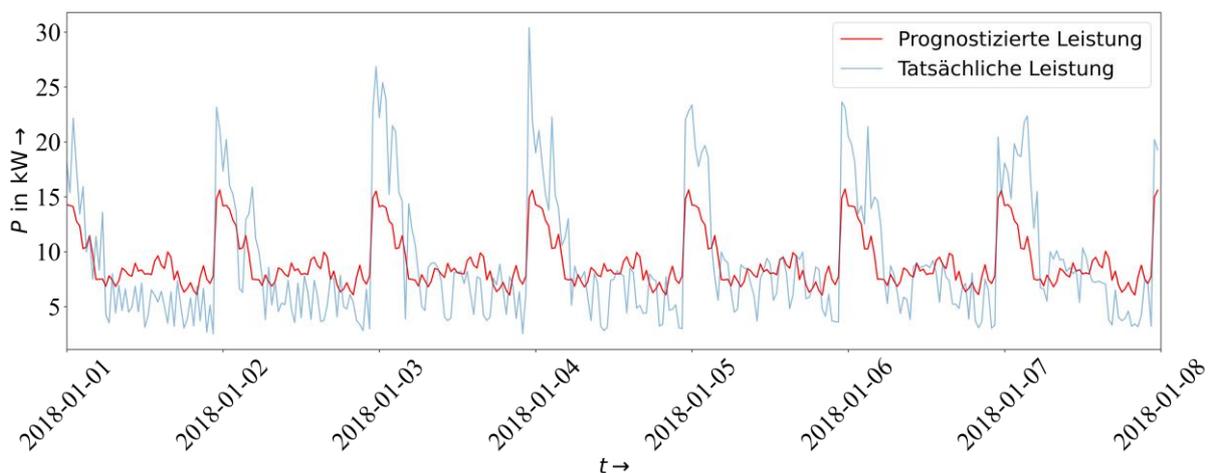


Bild 4: Lastgang von der Bäckerei über eine Beispielwoche mit prognostiziertem und tatsächlichem Leistungsverlauf der Prognosevariante mit der geteilten Messung und nur Zeitparametern.

4 Zusammenfassung

Es konnten erfolgreich Langzeitprognosemodelle für industrielle, elektrische Lastgänge mittels Kurzzeitmessung und GPs aufgestellt werden. Dabei wurden für zwei Datensätze die Referenzwerte eingehalten bzw. übertroffen, wohingegen der dritte Datensatz, derjenige der Werkstatt, höhere Prognosefehler aufgrund größerer Unregelmäßigkeiten im Datensatz aufwies. Die alleinige Verwendung von Zeitparametern erzielte bei allen und die geteilte Messung des Trainings- und Validierungsdatensatzes bei zwei von drei Datensätzen die höchste Prognosegüte.

Als nächsten Schritt gilt es, die bestehenden Prognosemodelle zu verbessern sowie weitere Datensätze für breitere Anwendungsmöglichkeiten zu untersuchen. Die Verbesserung kann beispielsweise durch eine Änderung der Zeiträume des Trainings- und Validierungsdatensatzes sowie durch das Erstellen eines eigenen Kernels für die GPs versucht werden.

5 Referenzen

- [1] H. Hilse, Interview am 21.11.2023 mit dem Geschäftsführer der GICON - Großmann Ingenieur Consult GmbH.
- [2] C. Kuster, Y. Rezgui, M. Mourshed: Electricity and power forecasting models: a critical systematic review, *Sustainable Cities and Society* (2017).
- [3] R. Patel, M. Patel, R. Patel: A review: introduction and understanding of load forecasting, *Journal of Applied Science and Computations* (2019).
- [4] S. P. V. Gupta: An overview of different types of load forecasting methods and the factors affecting the load forecasting, *International Journal for Research in Applied Science and Engineering* (2017).
- [5] M. Alamaniotis, L. H. Tsoukalas: Anticipation of minutes-ahead household active power consumption using Gaussian processes, 6th International Conference on Information, Intelligence, Systems and Applications (2015).
- [6] B. Han, S. Zhang, L. Qin, X. Liu, Y. Wangand, Z. Li: Comparison of Support Vector Machine, Gaussian Process Regression and Decision Tree Models for Energy Consumption Prediction of Campus Buildings, 8th International Conference on Hydraulic and Civil Engineering: Deep Space Intelligent Development and Utilization Forum (2022).
- [7] Ü. B. Filik, Ü. N. Gerek, M. Kurban: Hourly Forecasting of Long Term Electric Energy Demand Using a Novel Modeling Approach, Fourth International Conference on Innovative Computing, Information and Control (2009).
- [8] N. J. Johannesen, M. L. Kolhe, M. Goodwin: Load Demand Analysis of Nordic Rural Area with Holiday Resorts for Network Capacity Planning, 4th International Conference on Smart and Sustainable Technologies (2019).
- [9] https://www.dwd.de/DE/klimaumwelt/cdc/cdc_node.html, abgerufen am 04.01.2024
- [10] <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>, abgerufen am 04.01.2024.
- [11] <https://www.gpflow.org/>, abgerufen am 04.01.2024.
- [12] D. K. Duvenaud: Automatic Model Construction with Gaussian Processes, University of Cambridge (2014).
- [13] C. E. Rasmussen und C. K. I. Williams, "Gaussian Processes for Machine Learning", the MIT Press, 2006.
- [14] <https://fitter.readthedocs.io/>, abgerufen am 05.01.2024.