

Universität Kassel

Fachbereich Maschinenbau

Institut für Produktionstechnik und Logistik

Fachgebiet Umweltgerechte Produkte und Prozesse

Prof. Dr.- Ing. M. Junge

Vergleich & Anpassung zweier Strategien zur Anomalieerkennung in Lastgängen basierend auf Verfahren aus den Bereichen Machine-Learning und Statistik

Masterarbeit

zur Erlangung des akademischen Grades "Master of Science"

von Patrick Krawiec

November 2020

Universität Kassel

Masterarbeit

Vergleich & Anpassung zweier Strategien zur Anomalieerkennung in Lastgängen basierend auf Verfahren aus den Bereichen Machine-Learning und Statistik

Angefertigt von:	Patrick Krawiec
Fachgebiet:	Institut für Produktionstechnik und Logistik
Betreuer:	Prof. Dr.-Ing. Mark Junge
Zweitbetreuer:	Prof. Dr.-Ing. Jens Hesselbach
Firma:	Limón GmbH

Kurzzusammenfassung

In dieser Masterarbeit werden zwei Strategien zur Anomalieerkennung in Lastgängen ausgewertet. Dazu nutzt Strategie 1 das künstliche neuronale Netzwerk LSTM (Long Short-Term Memory) mit Datenzeitraum von einem Monat (1M) bzw. drei Monaten (3M) trainiert und Strategie 2 das Glättungsverfahren PEWMA (Probabilistic Exponential Weighted Moving Average) zur Schätzung des zu untersuchenden Lastgangmonats. Durch den Vergleich mit Originallastgangdaten werden Residuen bzw. summierte Residuen der Sequenzlängen zwei, vier, sechs und acht Stunden durch Überschreitung eines vordefinierten Schwellenwertes als Anomalien identifiziert. Die Schwellenwerte werden zum einen durch den Z-Score Test definiert, d.h. Residuenwerte größer als 2, 2,5 oder 3 Standardabweichungen gelten als anomal. Zum anderen werden mittels ESD (Extreme Studentized Deviate) Test durch drei Signifikanzniveauewerte von 0,05, 0,10 und 0,15, bei maximal $k = 40$ Iterationen, Schwellenwerte gesetzt. Es werden fünf Lastgänge untersucht, welche sich durch das Clusterverfahren k -Means als repräsentative Stichprobe aus allen verfügbaren Datensätzen der Limón GmbH ergeben haben. Die Auswertung zeigt, dass für Strategie 1 ein maximales F_T -Maß von 0,4 (1M) und für alle untersuchten Unternehmen ein durchschnittliches F_T -Maß von maximal 0,24, $SD = 0,09$ (1M) für die Untersuchung auf Einzelresiduen erreicht werden konnte. In der Variante 3M konnte mit einem durchschnittlichen F_T -Maß von 0,21, $SD = 0,06$ (3M) für summierte Residuen der Teilsequenzlänge von vier Stunden das höchste F_T -Maß erzielt werden. Die auf PEWMA basierende Strategie 2 konnte im Vergleich zu Strategie 1 in keinem untersuchten Unternehmen eine höhere Anomalieerkennungsgüte aufweisen.

Schlüsselwörter: Anomalieerkennung, LSTM, PEWMA

Abstract

In this master thesis two strategies for anomaly detection in load profiles are evaluated. To estimate the month load profile, strategy 1 uses the artificial neural network LSTM (Long Short-Term Memory), with a data period of one month (1M) or three months (3M), and strategy 2 uses the smoothing method PEWMA (Probabilistic Exponential Weighted Moving Average). By comparing with original load profile data, residuals or summed residuals of the sequence lengths of two, four, six and eight hours are identified as an anomaly by exceeding a predefined threshold. The thresholds are defined by the Z-Score test, i.e. residuals greater than 2, 2.5 or 3 standard deviations are considered anomalous. Furthermore, the ESD (Extreme Studentized Deviate) test is used to set thresholds by means of three significance level values of 0.05, 0.10 and 0.15, with a maximum of $k = 40$ iterations. Five load profiles are examined, which were obtained by the cluster procedure k -Means as a representative sample from all available data sets of the Limón GmbH. The evaluation shows that for strategy 1 a maximum F_T -value of 0.4 (1M) and for all examined companies an average F_T -value of maximum 0.24, $SD = 0.09$ (1M) could be achieved for the investigation on single residuals. In variant 3M the highest F_T -value could be achieved with an average F_T -value of 0.21, $SD = 0.06$ (3M) for summed residuals of the partial sequence length of four hours. The PEWMA-based strategy 2 did not show a higher anomaly detection quality compared to strategy 1 in any of the investigated companies.

Keywords: *Anomaly detection, LSTM, PEWMA*

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig durchgeführt und verfasst, keine anderen als die angegebenen Hilfsmittel verwendet und sämtliche Stellen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, mit Quellenangaben kenntlich gemacht habe. Desgleichen gilt für Zeichnungen, Skizzen, bildliche Darstellungen oder Gleichungen.

Kassel, den 18. November 2020

Patrick Krawiec

Inhaltsverzeichnis

1	Einleitung	1
2	Theoretische Grundlagen.....	3
2.1	Kriterien zur Systematisierung von Anomalieerkennungsmethoden	3
2.1.1	Systematisierung nach Eingabedaten	4
2.1.2	Systematisierung nach Anomalietyp.....	4
2.1.3	Systematisierung nach Anomalieerkennungsmethode	8
2.2	Verfahren zur Identifikation von Anomalien	8
2.2.1	Punktanomalien in uni- und multivariaten Zeitreihen	9
2.2.2	Teilsequenzanomalien in uni- und multivariaten Zeitreihen	13
2.3	Kriterien zum Vergleich von Anomalieerkennungsverfahren.....	15
3	Anwendungsfeld - Lastgangdaten	16
3.1	Datenmaterial & Analysestichprobe	17
3.1.1	Charakterisierung des Gesamtmaterials	17
3.1.2	Gruppierung der Daten mittels <i>k</i> -Means Algorithmus	17
3.1.3	Charakterisierung der Analysestichprobe	21
3.2	Verwendete Analysemethoden	27
3.2.1	Experteninterviews & Zählweise der Anomalien	27
3.2.2	Verfahren aus dem Bereich des Machine-Learning: LSTM	28
3.2.3	Verfahren aus der statistischen Zeitreihenanalyse: PEWMA.....	30
3.2.4	Festlegung und Prüfung von Schwellenwerten	31
4	Auswertung & Diskussion der Einzelergebnisse	34
4.1	Strategie 1: Anomalien unter Normalmodell-Schätzung mit LSTM	34

Inhaltsverzeichnis

4.1.1 Cluster 1 Unternehmen	34
4.1.2 Cluster 2 Unternehmen	40
4.1.3 Cluster 3 Unternehmen	42
4.1.4 Cluster 4 Unternehmen	45
4.1.5 Cluster 5 Unternehmen	48
4.2 Strategie 2: Anomalien unter Normalmodell-Schätzung mit PEWMA .	51
4.2.1 Cluster 1 Unternehmen	52
4.2.2 Cluster 2 Unternehmen	56
4.2.3 Cluster 3 & 4 Unternehmen	58
4.2.4 Cluster 5 Unternehmen	59
5 Strategie- und clusterübergreifende Diskussion	61
5.1 Fragen geleitetes Resümee	61
5.2 Einordnung in aktuelle und Ausblick auf zukünftige Forschung	68
Abbildungsverzeichnis	I
Tabellenverzeichnis	IV
Literaturverzeichnis	VII
Anhang	XI

1 Einleitung

Die Etablierung einer umweltverträglichen und nachhaltigen Energieversorgung ist eine der zentralen Herausforderungen unserer Zeit. Für die Umsetzung der Energiewende bis 2030 (Europäische Kommission 2014) bzw. 2050 (Europäische Kommission 2019) existieren auf europäischer Ebene vor allem zwei klimapolitische Ziele: die Erhöhung des Anteils an erneuerbarer Energie und die Steigerung der Energieeffizienz (Krewitt et al. 2009). Zur Realisierung der Ziele sind umfassende Veränderungen in verschiedenen Sektoren, wie der Energieversorgung, dem Verkehr und der Industrie nötig. Getrieben durch politische Vorgaben sowie finanzielle Förderprogramme, wird für Unternehmen somit ein effizienter Umgang mit Energie und Ressourcen bedeutsamer. Datenbasierte Energiemonitoringsysteme zur Erfassung, Analyse und Interpretation von Energiedaten können maßgebend dazu beitragen, die Energieeffizienz eines Unternehmens zu steigern (Servatius et al. 2012, Teil III)

Die computergestützte Erkennung von Anomalien in Energiezeitreihen erscheint dabei als vielversprechender Ansatz zur Erfassung von Effizienzpotenzialen. Es existieren bereits einige Methoden zur Anomalieerfassung, z.B. im Bereich des Data Mining oder auch im Feld der Statistik bzw. der Zeitreihenanalyse (Stier 2001, S. 121 ff.).

Dabei werden viele verschiedene Begriffe für Anomalien und Methoden zu ihrer Erkennung verwendet. Es wird z.B. von Ausreißern (*Outliers*), Anomalien (*Anomalies*), Neuheiten (*Novelties*) oder seltenen Ereignissen (*Rare Events*) gesprochen, wobei über die Bedeutung kein Konsens herrscht (Carreño et al. 2020). Candola et al. (2009) definieren eine Anomalie wie folgt: „*Anomalies are patterns in data that do not conform to a well defined notion of normal behavior*“ (Chandola et al. 2009, S. 2). Anomalien können daher als Auffälligkeiten in Daten angesehen werden, die nicht dem erwarteten Datenverhalten entsprechen. In der Analyse von Zeitreihen können Anomalien zwei verschiedene Bedeutungen haben (Aggarwal 2017, S. 3), die von der Perspektive auf die Daten abhängen. Zum einen können Daten als *Rauschen*, als *fehlerhaft* oder *unerwünscht*, und somit uninteressant für den Analytisten betrachtet werden. Ziel ist folglich, diese Ausreißer zu löschen oder zu korrigieren, um die Qualität des Datensatzes zu verbessern. Eine Anwendung wäre z.B. Sensorübertragungsfehler zu eliminieren, um dann im nächsten Schritt bessere Vorhersagen über den Untersuchungsgegenstand zu machen. Zum anderen lassen sich Anomalien in Zeitreihen aber auch als *unerwartete* und damit *besonders interessante* Phänomene deklarieren, die es zu erfassen und zu analysieren gilt. Als Beispiel kann hier die Erkennung von Betrugversuchen (*Fraud Detection*) genannt werden. Das Hauptziel ist die *Erkennung* und die *Analyse* der anomalen Transaktionstätigkeit. Die in dieser

Einleitung

Masterarbeit behandelten Anomalien in Lastgängen werden aus letzterer Perspektive betrachtet und gelten demnach als unerwartet, aber besonders interessante sowie relevante Auffälligkeiten.

Ziel der Arbeit ist es, Strategien zur Anomalieerkennung auf den Umgang mit Lastgangdaten zuzuschneiden sowie zu prüfen, wie gut sich die eingesetzten Strategien eigenen, um unerwartete Verläufe in Lastgängen zu erkennen.

Dafür wird im Grundlagenkapitel (Kapitel 2 Theoretische Grundlagen) zunächst eine Möglichkeit vorgestellt Anomalieerkennungsverfahren zu systematisieren (Kapitel 2.1), darauf aufbauend werden verschiedene Verfahren zur Anomalieerkennung strukturell nach Art der Anomalie präsentiert (Kapitel 2.2) und schließlich Möglichkeiten zur Eignungsprüfung der Verfahren eingeführt (Kapitel 2.4).

Im Kapitel zum Anwendungsfeld (Kapitel 3 Anwendungsfeld - Lastgangdaten) wird zunächst (Kapitel 3.1) das für diese Arbeit verwendete Datenmaterial vorgestellt, die Auswahl von prototypischen Unternehmen mithilfe von Clusteranalysen erläutert sowie die Analytestichprobe anhand verschiedener Merkmale charakterisiert. Nun werden die in dieser Masterarbeit eingesetzten Analysemethoden vorgestellt (Kapitel 3.2) angefangen von der Expertenbefragung zur analogen Identifikation von Anomalien sowie der Klärung der Zählweise von Anomalien, über den Einsatz vom Machine-Learning Algorithmus (LSTM) und Verfahren aus der Statistik (PEWMA) bis hin zu Ansätzen zur Festlegung von Schwellenwerten mittels Z-Score- und ESD-Test.

Das Auswertungskapitel (Kapitel 4 Auswertung & Diskussion der Einzelergebnisse) präsentiert und diskutiert nun die Ergebnisse der Untersuchung der Analytestichprobe. Dabei kommen zwei Strategien zum Einsatz: Im Rahmen der ersten Strategie werden Anomalien auf Grundlage der Schätzung des Lastgangverlaufs mittels LSTM Algorithmus untersucht und die Einzelergebnisse diskutiert (Kapitel 4.1). Grundlage der Analysen der zweiten Strategie ist die Schätzung von Lastgangverläufen mittels PEWMA auch hier werden folgend mögliche Anomalien untersucht und ihre Aussagekraft besprochen (Kapitel 4.2).

Im letzten Kapitel (Kapitel 5 Strategie- und clusterübergreifende Diskussion) werden die Ergebnisse noch einmal übergreifend analysiert und mit Blick auf folgende Fragen diskutiert (Kapitel 5.1): (1) Eignen sich die gewählten Ansätze der Schwellenwertbildung (hier Z-Score- und ESD-Test) zur Identifikation von Anomalien? Wie beeinflussen diese Ansätze die Anomalieerkennungsgüte? (2) Hängt die Güte der Anomalieerkennung von den spezifischen Lastgangverläufen einzelner Unternehmen ab? (3) Welche Auswirkung hat die Länge des Zeitraums der zur Schätzung eines Normalmodells herangezogenen Daten (hier Daten von einem vs. von drei Monaten) auf die Güte der Anomalieerkennung? (4) Eignet sich Strategie 1 oder Strategie 2 besser zur Anomalieerkennung? Welche Auswirkungen hat die Analyse von Einzelresiduen oder

Residuensequenzen auf die Güte der Anomalieerkennung? Die Arbeit endet schließlich mit der Einordnung der Ergebnisse in aktuelle und einem Ausblick auf zukünftige Forschungsbemühungen (Kapitel 5.2).

2 Theoretische Grundlagen

Bei der Beschäftigung mit der Erkennung von Anomalien ist es zunächst hilfreich die Vielzahl an Anomalieerkennungsverfahren zu ordnen. Für den Umgang mit Zeitreihendaten erscheint die im folgenden eingeführte Systematisierung nach Blázquez-García et al. (2020) besonders geeignet (Kapitel 2.1). Organisiert nach dem darin enthaltenen Systematisierungskriterium der *Art der Anomalie* werden darauf aufbauend schätz- und vorhersagebasierte Verfahren zur Erkennung von Punkt- und Teilsequenzanomalien in uni- sowie multivariaten Kontexten eingeführt (Kapitel 2.2). Mit Blick auf das Ziel der Arbeit Analysestrategien zu vergleichen, werden im letzten Teil dieser theoretischen Grundlagen Kriterien zur Einschätzung von Güte und Eignung der Anomalieerkennungsverfahren erarbeitet (Kapitel 2.3).

2.1 Kriterien zur Systematisierung von Anomalieerkennungsmethoden

In Anlehnung an Blázquez-García et al. (2020) lassen sich Methoden zur Anomalieerkennung in Zeitreihen nach drei Kriterien systematisieren (vgl. Abb. 2.1): (1) nach dem Eingabetyp – handelt es sich um univariate oder multivariate Zeitreihen, (2) nach dem Anomalietyp – wie gestalten sich mögliche Anomalien, ist von multi- oder univariaten Anomalien auszugehen, sowie (3) nach Art der Anomalieerkennungsmethode – zielt sie auf eine univariate oder multivariate Anomalieerkennung ab.

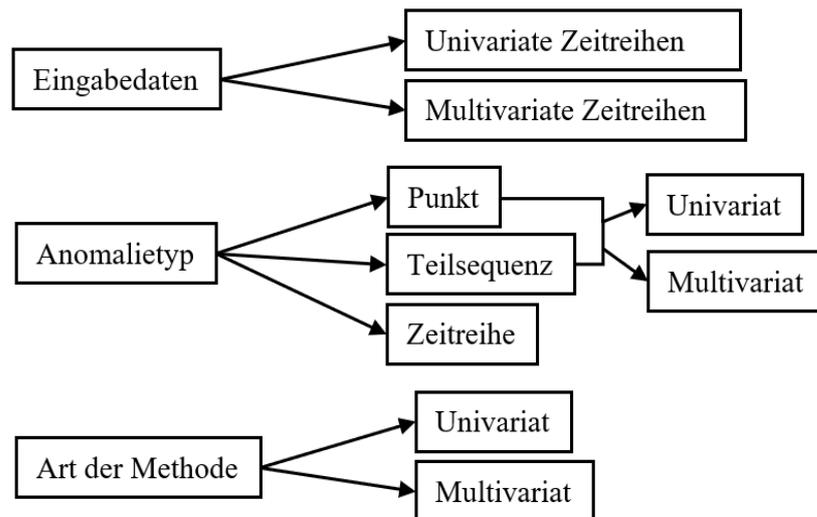


Abbildung 2.1: Systematisierung von Erkennungsmethode nach 3 Unterscheidungsmerkmalen (in Anlehnung an Blázquez-García et al. 2020, S.3)

2.1.1 Systematisierung nach Eingabedaten

Zunächst können Daten, die auf Anomalien geprüft werden sollen, danach unterschieden werden, ob es sich um univariate oder multivariate Zeitreihendaten handelt.

Eine univariate Zeitreihe repräsentiert eine durch Zeitpunkte $t \in T \subseteq \mathbb{Z}^+$ geordnete Menge reeller Zahlen $X = \{x_t\}$, wobei jede Zahl einer Beobachtung zum Zeitpunkt t entspricht (Blázquez-García et al. 2020, S. 3). Beispiele für univariate Zeitreihen finden sich in Abbildung 2.2 (a) und Abbildung 2.3 (a).

Bei einer multivariate Zeitreihe $\mathbf{X} = \{\mathbf{x}_t\}$, mit Zeitpunkten $t \in T \subseteq \mathbb{Z}^+$, liegt eine geordnete Menge von k -dimensionalen Vektoren ($k > 1$) reellwertiger Zahlen vor und damit besitzt jeder Zeitpunkt k Beobachtungen $\mathbf{x}_t = (x_{1t}, \dots, x_{kt})^T$. Daher besteht eine multivariate Zeitreihe aus $j \in \{1, \dots, k\}$ univariaten Zeitreihen $X_j = \{x_{jt}\}$. Abbildung 2.2 (b) und Abbildung 2.3 (b) stellen hierfür Beispiele dar (Blázquez-García et al. 2020, S. 3).

2.1.2 Systematisierung nach Anomalietyp

Zeitreihen lassen sich auch nach Art der existierenden Anomalien kategorisieren (Blázquez-García et al. 2020, S. 3 ff.; Wang et al. 2018, S. 1807): Punktanomalien, Teilsequenzanomalien und anomale Zeitreihen.

Punktanomalien

Ein Punktanomalie ist ein einzelner Zeitpunkt, welcher sich im Vergleich zu anderen Werten in der Zeitreihe ungewöhnlich verhält. Dabei können Punktanomalien univariat

Theoretische Grundlagen

oder multivariat sein, abhängig davon, ob sie eine oder mehrere zeitabhängige Variablen beeinflussen.

In Abbildung 2.2 (a) werden beispielsweise zwei univariate Punktanomalien O1 und O2, dargestellt. In Abbildung 2.2 (b) lassen sich in der multivariaten Zeitreihe, bestehend aus drei Variablen, univariate, O3, als auch multivariate, O1 und O2, Punktausreißer erkennen.

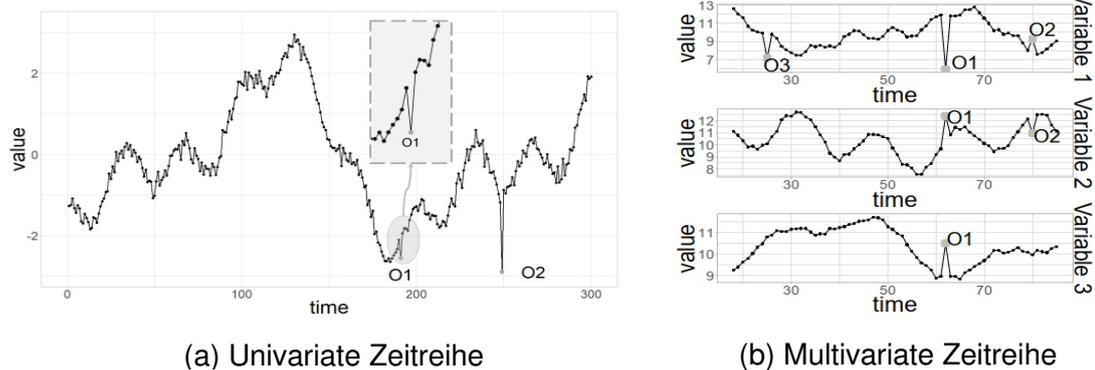


Abbildung 2.2: Punktanomalien in univariaten und multivariaten Zeitreihen (Blázquez-García et al. 2020, S. 4)

Teilsequenzanomalien

Diese Art von Anomalie bezieht sich auf das Verhalten einer Sequenz aufeinander folgender Zeitpunkte, die als ungewöhnlich erachtet werden, wobei jeder Beobachtungzeitpunkt für sich nicht notwendigerweise eine Punktanomalie sein muss (Blázquez-García et al. 2020, S. 3).

Weiterhin können Anomalieteilsequenzen zeitabhängig in einer Dimension (univariate Teilsequenz) oder in mehreren Dimensionen (multivariate Teilsequenz) auftreten.

Abbildung 2.3 zeigt Beispiele für univariate Teilsequenzanomalien mit O1 und O2 in Abbildung 2.3 (a) sowie mit O3 in Abbildung 2.3 (b), und multivariate Teilsequenzanomalien, mit O1 und O2 in Abbildung 2.3 (b). Multivariate Teilsequenzanomalien müssen nicht alle Variablen betreffen.

Theoretische Grundlagen

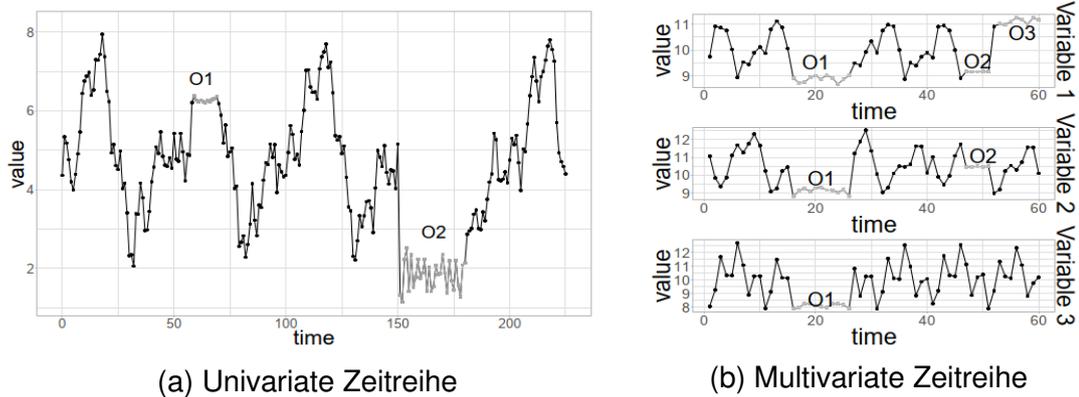


Abbildung 2.3: Teilsequenzanomalien in univariaten und multivariaten Zeitreihen (Blázquez-García et al. 2020, S. 4)

Anomale Zeitreihen

Ausschließlich im Fall von multivariaten Zeitreihen können ganze Zeitreihen anomal sein. Die Abbildung 2.4 verdeutlicht diese Art von Anomalie an der anomalen Zeitreihe „Variablen 4“, welche sich deutlich vom Rest der Verläufe unterscheidet.

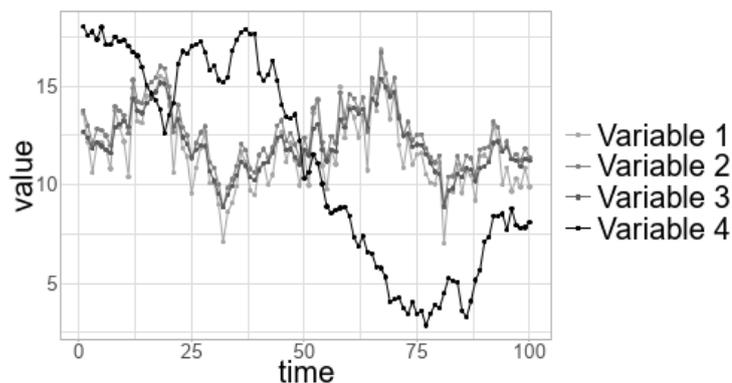


Abbildung 2.4: Anomale Zeitreihe „Variable 4“ im Kontext von multivariaten Zeitreihen (Blázquez-García et al. 2020, S. 5)

Die Einordnung von Daten als Anomalie ist stets kontextabhängig, d.h. entweder wird die gesamte Zeitreihe als Kontext gesehen oder die Methode untersucht nur gewisse Zeitfenster, sodass Ausreißer nur lokal, also im Kontext der näheren Nachbarschaft gültig sind (Blázquez-García et al. 2020, S. 4). Es gilt: Globale Ausreißer sind ebenfalls lokal, aber nicht alle lokalen Ausreißer sind global. Die Punktanomalie O1 in Abbildung 2.2 (a) zeigt einen solchen Fall auf: dieser Punkt kann global gesehen, als gewöhnlich identifiziert werden, wobei innerhalb eines kleinen Zeitfensters bzw. im Kontext der Nachbarschaft O1 als lokaler Ausreißer deklariert werden kann. Die Definition von

Theoretische Grundlagen

lokalen und globalen Ausreißern lässt sich, wenn nur die Anzahl der berücksichtigten Zeitpunkte als Kontext festgelegt ist, auf die ersten beiden Kategorien, d.h. Punktanomalien und Teilsequenzanomalien, anwenden. Jedoch können Anomalien auch in anderen Kontexten auftreten, wie z.B. bei Berücksichtigung der zeitlichen Einflüsse durch Jahreszeiten oder Werktag-Wochenend-Rhythmen. Da der Begriff der kontextualen Anomalien nicht eindeutig definiert ist, wird er, in dieser Ausarbeitung, auf anomale Zeitreihen erweitert, sodass eine anomale Zeitreihen nur im Kontext anderen Zeitreihen existieren kann. Exemplarisch, wird in Abbildung 2.5 der monatliche Temperaturzeitverlauf dargestellt. Es liegt eine kontextuale Anomalie vor, da t_2 in Bezug auf die jahreszeitlichen Schwankungen eine ungewöhnliche Temperatur aufweist, obwohl t_2 auf etwa gleichem Niveau wie t_1 liegt.

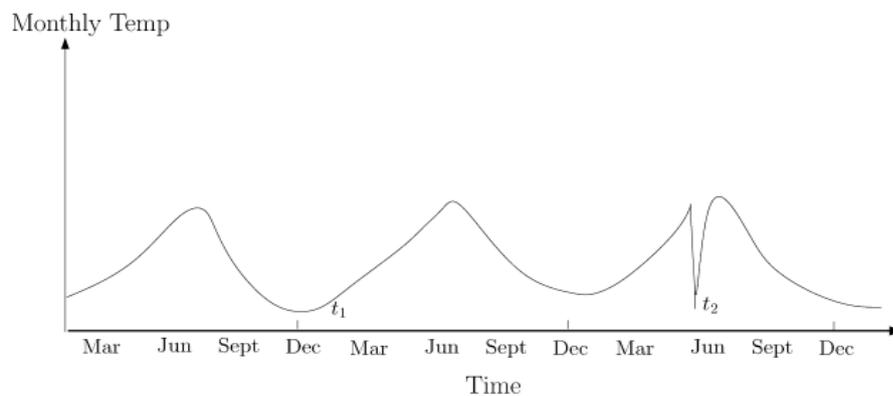


Abbildung 2.5: Kontextuale Anomalie t_2 in Bezug auf jahreszeitlichen Schwankungen (Chandola et al. 2009, S. 8)

Eine weitere und oft angewendete Möglichkeit Anomalien nach ihrer Art zu unterscheiden ist die Kategorisierung nach Punktanomalien, kontextualen Anomalien und kollektiven Anomalien (Chandola et al. 2009, S. 7 f.). Diese Aufteilung von Chandola et al. (2009) befasst sich nicht ausschließlich mit Anomalien in Zeitreihen. Die vorher aufgeführte Kategorie einer anomalen Zeitreihe im multivariaten Fall, wird daher nicht explizit als Unterscheidungsmerkmal genannt. Allerdings entsprechen die von Chandola et al. (2009, S. 7 f.) genannten kollektiven Anomalien in Zeitreihen der Definition von kontextabhängigen Punkt- und Teilsequenzanomalien aus dem Artikel von Blázquez-García et al. (2020, S. 3 f.). Deswegen werden kontextabhängige Anomalien innerhalb dieser Masterarbeit im Folgenden nicht als eigene Kategorie aufgeführt.

2.1.3 Systematisierung nach Anomalieerkennungsmethode

Bei der Systematisierung von Anomalieerkennungsverfahren, kann auch die Erkennungsmethode selbst betrachtet werden.

So lassen sich univariate von multivariaten Erkennungsmethoden unterscheiden. Eine univariate Methode berücksichtigt *eine* zeitabhängige Variable im Fall einer univariate Zeitreihe oder jede zeitabhängige Variable für sich in einer multivariaten Zeitreihe. Eine multivariate Methode eignet sich nur für multivariate Zeitreihen und ist zusätzlich in der Lage gegenseitige Abhängigkeiten der zeitabhängigen Variablen zu berücksichtigen.

Darüber hinaus lassen sich auch direkte von indirekten Anomalieerkennungsmethoden unterscheiden:

Direkte Anomalieerkennungsmethoden nutzen beispielsweise Cluster-, Klassifikation- oder Entfernungs- bzw. Dichte-basierte Verfahren. Im resultierenden Output werden Datenpunkte direkt als normal oder anomal deklariert. Ausreißer können beispielsweise in einem Dichteverfahren mit dem sogenannten Local Outlier Factor (LOF) beschrieben werden, der angibt, zu welchem Grad ein Datenpunkt anomal ist (Breunig et al. 2000). Eine differenzierte Beschreibung verschiedener direkter Methoden findet sich bei Chandola et al. (2009). Das in dieser Arbeit eingesetzte Clusterverfahren wird in Kapitel 3.1.2 näher erläutert.

Indirekte bzw. residuenbasierte Anomalieerkennungsmethoden modellieren in einem ersten Schritt das normale Verhalten eines Lastganges mit Hilfe eines mathematischen Modells. In einem zweiten Schritt werden Abweichungen (Residuen) zwischen modellierten und beobachteten Werten betrachtet. Mittels Residuenanalyse wird entschieden, ob eine Beobachtung anomal ist. Eine detaillierte Beschreibung der Residualanalyse wird in Kapitel 3.2.4 gegeben. Es existieren auch Methoden, die eine Kombination direkter und indirekter Ansätze darstellen. Zum Beispiel lassen sich durch Clusterverfahren Clusterzentren bestimmen, welche direkt dem einzelnen Datenpunkt ein Label zuweisen und damit einzelne Punkte anhand des Labels als anomal erkannt werden können. In einem weiteren Schritt könnten auch durch den Vergleich, beispielsweise mittels der euklidischen Metrik, der einzelnen Datenpunkten mit den Clusterzentren Residuen gebildet und auf anomal hohe Werte untersucht werden, um Ausreißer zu identifizieren.

2.2 Verfahren zur Identifikation von Anomalien

Wie oben beschrieben lassen sich Anomalieerkennungsverfahren nach der Analyse von Punkt- und Sequenzanomalien ordnen. Im Folgenden wird diese Ordnung aufgegriffen, weiter nach schätz- oder vorhersagebasierten Modellen differenziert sowie der Einsatz bei univariaten wie auch multivariaten Zeitreihen behandelt.

2.2.1 Punktanomalien in uni- und multivariaten Zeitreihen

Eine Punktanomalie lässt sich als eine Beobachtung definieren, in der ein beobachteter Wert einen Abstand zu einem erwartenden Wert besitzt, welcher höher als ein vordefinierter Schwellenwert τ ist. Somit gilt:

$$|x_t - \hat{x}_t| > \tau \quad (2.1)$$

mit dem beobachteten Datenpunkt x_t und dem zu erwartenden Wert \hat{x}_t .

In Abbildung 2.6 ist dieser Sachverhalt beispielhaft für eine univariate Zeitreihe dargestellt, in welchem die Punkte O1, O2, O3 und O4 außerhalb der abgeschatteten Region, also außerhalb des Schwellenwertes τ liegen und somit als Anomalien bezeichnet werden können.

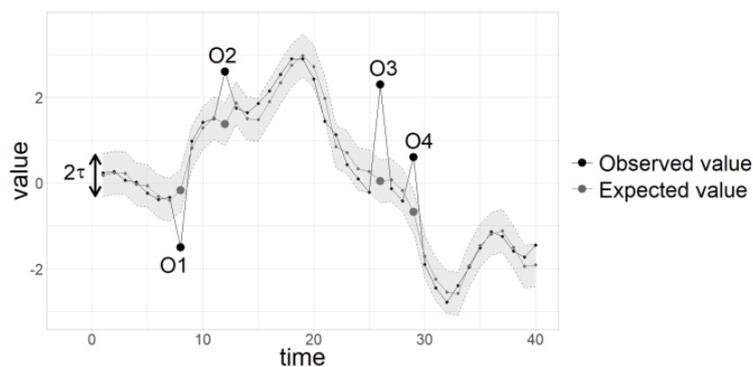


Abbildung 2.6: Punktausreißer O1, O2, O3 und O4 resultierend aus der Überschreitung des Schwellenwertes τ (Blázquez-García et al. 2020, S. 6)

Um einen Erwartungswert zu bestimmen, werden *Schätz- und Vorhersagemodelle* unterschieden. Die begriffliche Unterscheidung von Schätzung und Vorhersage orientiert sich an Blázquez-García et al. (2020). Bei Schätzmodellen werden zur Schätzung eines Erwartungswertes \hat{x}_t zurückliegende Daten (x_{t-k_1} mit $k_1 \geq 1$), aktuelle (x_t) und dem Erwartungswert zeitlich nachfolgende Daten (x_{t+k_2} mit $k_2 \geq 1$) verwendet. Im Gegensatz dazu werden bei Vorhersagemodellen nur zurückliegende Daten (von x_{t-k_1} bis x_{t-1}) genutzt, um einen zeitlich darauffolgenden (zukünftigen) Erwartungswert \hat{x}_t zu bestimmen. Bei einem Vorhersagemodell kann es sich um ein festes oder ein variables Modell handeln. Bei ersterem werden alle zukünftigen Erwartungswerte mit dem gleichen Modell bzw. den gleichen Regeln vorhergesagt, während sich variable Modelle schrittweise mithilfe neu hinzukommender Zeitreihendaten anpassen und verändern können. Beispielsweise prüfen sogenannte Streaming basierte Techniken kontinuierlich neu erhaltene Datenpunkte auf Anomalität. Hierbei können sowohl feste oder aber auch variable Vorhersagemodelle zugrunde gelegt werden. Darüber hinaus lassen sich bestimmte

Schätzmodelle, wie z.B. der Long Short-Term Memory (LSTM) Algorithmus, über spezifische Anpassungen auch als Vorhersagemodelle verwenden.

Es existiert eine ganze Reihe verschiedener Strategien zur Entwicklung von Schätzmodellen. Bei einfachsten Schätzmodellen basiert die Bestimmung des Erwartungswertes beispielsweise auf der Verwendung des Medians (Basu und Meckesheimer 2007) oder der mittleren absoluten Abweichung vom Median (engl. Median Absolute Deviation (MAD)) (Mehrang et al. 2015). Dabei können die Werte sowohl auf der vollständigen Zeitreihe als Ganzes oder auf Abschnitten gleicher Zeitdauer beruhen. Streaming basierte Anwendungen sind so allerdings nicht möglich. Deutlich komplexere Methoden nutzen z.B. ungleiche Abschnittslängen und ein von der Standardabweichung des Abschnitts abhängigen Schwellenwert (Arik et al. 2015) zur Schätzung der Erwartungswerte. Darüber hinaus existieren Glättungsmethoden wie B-Splines (Chen et al. 2010) oder gleitende Mittelwertmethoden (Carter und Streilein 2012). Auf Letztere wird im Kapitel 3.2.3 näher eingegangen.

Identifikation von Punktanomalien in *univariaten* Zeitreihen

Bei *univariaten Zeitreihendaten* können unter der Verwendung von Schätzmodellen einzelne Beobachtungswerte basierend auf Formel (2.1) direkt als Anomalien identifiziert werden. Weiter ist es möglich die Gesamtheit an erhaltenden Residuen einer Zeitreihe zu vergleichen und in Abhängigkeit davon zu entscheiden, welche Abweichungen als Anomalien gewertet werden können. Dazu werden vertiefende Erläuterungen in Kapitel 3.2.4 Festlegung und Prüfung von Schwellenwerten gegeben.

Im Gegensatz zu Schätzmodellen arbeiten Vorhersagemodelle ohne Verwendung des aktuellen Punktes x_t oder nachträglicher Beobachtungen. Auch hier gelten bei univariaten Zeitreihen Punkte als Anomalien, die von ihren vorhergesagten Werten abweichen (vgl. Formel (2.1)). Vorhersagemodelle können z.B. mithilfe des DeepAnT-Algorithmus (Munir et al. 2019) auf Grundlage von Convolutional Neural Networks (CNN) erstellt werden, oder aus statistischer Perspektive als Auto-Regressive Integrated Moving Average (ARIMA) realisiert werden (Zhou et al. 2019).

Punktanomalien lassen sich außerdem über dichtebasierte Methoden identifizieren. Punkte mit weniger als τ Nachbarn in der Entfernung R werden dabei als Anomalie definiert:

$$\text{Anomaliepunkt } x_t \Leftrightarrow |\{x \in X | d(x, x_t) \leq R\}| < \tau; R \in \mathbb{R}^+ \quad (2.2)$$

mit d dem euklidischen Abstand, x_t dem Datenpunkt zum Zeitpunkt t und X der Gesamtmenge an Datenpunkte (Blázquez-García et al. 2020, S. 9).

Theoretische Grundlagen

Ein Punkt ist demnach eine Anomalie, wenn die Summe aus vorhergehenden τ_p und nachfolgenden Nachbarn τ_s in einem Abstand kleiner gleich R unter τ liegt:

$$\tau_p + \tau_s < \tau \quad (2.3)$$

Die Anomalieerkennung mithilfe dichtebasierter Methoden ist bei Daten ohne zeitliche Ordnung eine verbreitete Herangehensweise (Blázquez-García et al. 2020, S. 9). Datenpunkte aus Zeitreihen sind voneinander abhängig, sodass sich die Verwendung des nächstliegenden Nachbarn zur Identifikation von Anomalien deutlich komplexer gestaltet. Um die temporale Anordnung zu berücksichtigen, werden dichtebasierte Methoden beispielsweise mit Hilfe von sich verschiebenden Zeitfenstern realisiert (Ishimtsev et al. 2017) und sind beispielhaft in Abbildung 2.7 dargestellt.

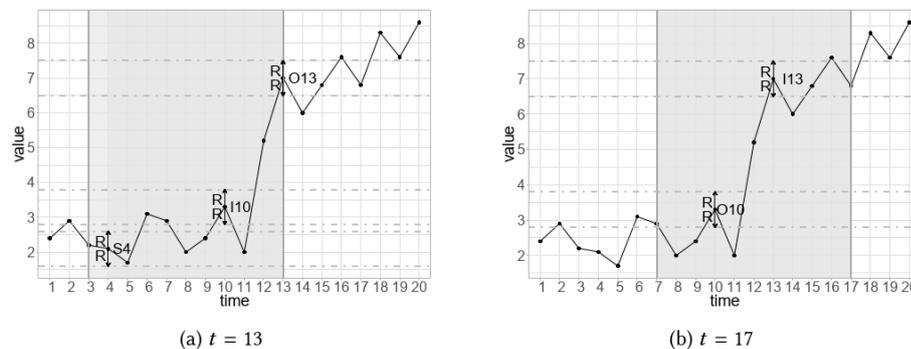


Abbildung 2.7: Dichtebasierte Erkennung von Anomalien mit gleitenden Zeitfenstern (Blázquez-García et al. 2020, S. 9)

Dieses Beispiel zeigt zwei verschiedene Zeitschritte mit $R = 0,5$, $\tau = 3$, und einer Zeitfensterlänge von 11 Schritten. In Graph (a) stellt Punkt O13 im Zeitfenster $t = 13$ eine Anomalie dar, wohingegen der gleiche Datenpunkt in Graph (b), hier I13, bei $t = 17$, nicht als Anomalie identifiziert werden kann.

Identifikation von Punktanomalien in *multivariaten* Zeitreihen

Multivariate Zeitreihen lassen sich mit univariaten oder multivariaten Methoden untersuchen, wobei letztere Korrelationen zwischen Variablen als zusätzliche Informationsquelle verwerten kann. Im Folgenden werden dazu univariaten oder multivariaten Methoden separat dargestellt.

Univariate Methoden. Univariate Analysemethoden identifizieren Punktausreißer separat in jeder der zeitabhängigen Variablen der multivariaten Zeitreihe. Daher lassen sich

im Prinzip alle Methoden zur Identifikation von Punktanomalien in univariaten Zeitreihen auch auf diese Art von Problemstellung anwenden. Hundman et al. (2018) verwendet beispielsweise ein LSTM Netzwerk zur Vorhersage von Raumfahrzeug-Telemetriedaten, wobei Formel (2.1) mit einem sich anpassenden Schwellenwert zur Identifizierung von Punktausreißern genutzt wird.

Es existieren auch Ansätze, welche multivariate Zeitreihen in ihrer Dimensionalität reduzieren, sodass ein neuer Satz an unkorrelierter Variablen erzeugt wird. Dieser Vorverarbeitungsschritt erlaubt die darauffolgende Anwendung von univariate Detektionstechniken. Durch die Berechnung von Linearkombinationen der Anfangsvariablen lassen sich beispielsweise unkorrelierte Variablen erzeugen. Papadimitriou et al. (2005) nutzen hierfür einen Hauptkomponenten Analyse Algorithmus zur Berechnung der neuen unkorrelierten Variablen. Darauf aufbauend werden univariaten Punktausreißer mit Hilfe eines autoregressiven Vorhersagemodelles (AR) identifiziert.

Multivariate Methoden. Es existieren verschiedene multivariate Methoden, welche den zuvor beschriebenen Vorverarbeitungsschritt nicht benötigen. Eine große Gruppe von Methoden basieren auf Schätzmodellen oder Vorhersagemodellen. Dafür lässt sich eine verallgemeinerte Variante von Formel (2.1) formulieren:

$$|\mathbf{x}_t - \hat{\mathbf{x}}_t| > \tau \quad (2.4)$$

mit dem k -dimensionalen beobachteten Datenpunkt \mathbf{x}_t und dem k -dimensionalen zu erwartendem Wert $\hat{\mathbf{x}}_t$.

Autoencoder sind innerhalb der auf Schätzmodellen basierenden Methoden eine der am häufigsten verwendeten Ansätze (Blázquez-García et al. 2020, S. 13). Es handelt sich dabei um eine Art von neuronalem Netz, welches versucht Eingangsinformationen zu komprimieren bzw. zu enkodieren und mit den reduzierten Daten den ursprünglichen Datensatz in einem Dekodierungsschritt zu rekonstruieren. Die komprimierten Daten enthalten die wichtigsten Merkmale eines Teildatensatzes / Trainingssatzes. Sie repräsentieren als Referenz die *Normalität* der Daten. Datenpunkte die nicht normal sind und die folglich auch die repräsentativen Merkmale des Trainingsdatensatzes nicht enthalten, werden vom Autoencoder nicht rekonstruiert und gelten als Anomalie. Aktuelle Anwendungsbeispiele sind in Publikation von Kieu et al. (2018) und Su et al. (2019) nachzulesen. Im Bereich der vorhersagebasierten Modelle lässt sich der schon in Kapitel 2.2.1 aufgeführte, DeepAnT-Algorithmus (Munir et al. 2019) als Anwendungsbeispiel nennen.

2.2.2 Teilsequenzanomalien in uni- und multivariaten Zeitreihen

Neben Punktanomalien lassen sich auch Teilsequenzanomalien identifizieren. Dabei handelt es sich um ungewöhnliche Verlaufsmuster in aufeinander folgenden Datenpunkten. Anomalieidentifikations- oder auch Nachweismethoden müssen sowohl die Länge der zu untersuchenden Teilsequenz als auch die Repräsentation der Teilsequenz berücksichtigen (Blázquez-García et al. 2020, S. 16).

Im Gegensatz zu Punktausreißern bestehen Teilsequenzen aus mehreren Punkten und besitzen somit eine *Länge*. Es gibt Methoden, welche feste oder variable Längen von Teilsequenzen detektieren können. Bei Methoden mit fester Länge muss ein Längewert vordefiniert werden. Durch Variation dieses Wertes und wiederholter Anwendung der Methode lassen sich somit auch Teilsequenzen unterschiedlicher Länge untersuchen. Neben der Länge von Teilsequenzen ist auch die Darstellung oder *Repräsentation der Daten* beim Einsatz von Teilsequenzanomalie-Erkennungsmethoden zu berücksichtigen. Da der Vergleich von Teilsequenzen ein rechen- und zeit-intensiver Prozess sein kann, wird häufig nicht mit den Rohdaten selbst, sondern mit einer Repräsentation der Daten gearbeitet. Die Diskretisierungstechnik, d.h. die Reduktion der Dimensionalität eines Datensatzes, ist eine weit verbreitete Darstellungsmethode. Wang et al. (2018) arbeitet beispielsweise bei der Detektion von Teilsequenzanomalien zur Vorverarbeitung der Daten mit dem Symbolic Aggregate approxXimation (SAX) Algorithmus (Lin et al. 2007), welcher Segmente von Zeitreihen in symbolische Repräsentationen umwandelt.

Identifikation von Sequenzanomalien in *univariaten* Zeitreihen

Auch bei der Identifikation von Sequenzanomalien können Methoden eingesetzt werden, die auf Schätz- oder Vorhersagemodellen aufbauen (vgl. Kapitel 2.2.1). Die Anomalieerkennung beruht dann auf folgendem mathematischen Zusammenhang:

$$\sum_{i=p}^{p+n+1} x_i - \hat{x}_i > \tau \quad (2.5)$$

bei welcher die Differenz der Einzelpunkte von Teilsequenz $S = x_p, \dots, x_{p+n+1}$ und erwartenden Teilsequenz \hat{S} aufsummiert werden. Die Summe lässt sich prinzipiell auch mit den Absolutwerten der Einzelresiduen bilden. Eine anomale Sequenz wird durch das Überschreiten eines Schwellenwertes τ erkannt.

Schätzungen oder Vorhersagen lassen sich prinzipiell punktweise oder sequenzweise durchführen. Daher können für die punktweise Verarbeitung der Daten auch die aus Kapitel 2.2.1 genannten Methoden eingesetzt werden. Bei Vorhersagemodellen besteht die Gefahr, dass sich Fehler anhäufen, je weiter die Prognosen in die Zukunft gemacht

werden, da jeder vorhergesagte Wert zur Vorhersage des nachfolgenden Punktes verwendet wird. Diese Fehleranfälligkeit lässt sich minimieren, indem sequenzweise Schätzungen bzw. Vorhersagen stattfinden. Auf neuronalen Netzen basierende Modelle, wie etwa CNN- (Munir et al. 2019) oder LSTM-Algorithmen (Hundman et al. 2018), lassen sich beispielsweise für punktweise oder sequenzbasierte Schätzungen und Vorhersagen bei univariaten sowie multivariaten Methoden verwenden.

Darüber hinaus existieren Methoden, die weder Schätz- noch Vorhersagemodelle zugrunde legen, sondern z.B. Algorithmen nutzen, die Teilsequenzen oder Repräsentationen von Teilsequenzen direkt miteinander vergleichen, um Unterschiede zu erkennen. Die vergleichenden Techniken nutzten unter anderem den SAX Algorithmus oder eine Abwandlung (Keogh et al. 2005) davon, welcher je nach Variante feste (Chau et al. 2018) oder variable (Senin et al. 2015) Anomaliesequenzlängen identifizieren kann. Die Repräsentationen werden beispielsweise in Wang et al. (2018) mithilfe von Clusterverfahren gruppiert, um daraus anomale Teilsequenzen zu identifizieren.

Identifikation von Sequenzanomalien in *multivariaten* Zeitreihen

Analog zu Kapitel 2.2.1, welches sich bereits mit Punktanomalie-Erkennungsmethoden in multivariaten Zeitreihen befasst hat, lassen sich hier grundsätzlich univariate oder auch multivariate Methoden unterscheiden.

Univariate Methoden. Univariate Methoden besitzen auch hier prinzipiell den Nachteil, dass Informationen über Korrelationen zwischen Variablen verloren gehen können, wenn diese bei multivariate Zeitreihen zur Anwendung kommen. Daher lassen sich Verfahren, wie sie auch für die Identifikation von Sequenzanomalien in univariaten Zeitreihen beschrieben wurden, einsetzen. Ji et al. (2014) und Wang et al. (2018) verwenden beispielsweise Techniken, welche die Dimensionalität der Daten reduzieren, um diese im nächsten Schritt auf Unähnlichkeiten zu untersuchen.

Multivariate Methoden. Im Bereich der multivariaten Methoden erweitern beispielsweise Hu et al. (2019) den Ansatz von Ji et al. (2014) auf multivariate Zeitreihen und nutzen sogenannte Rekurrenzplots (RP) um aus deren Analyse anomale Teilsequenzen zu identifizieren. Ein RP ist eine Möglichkeit zur Visualisierung bzw. erzeugt eine Repräsentation von Zeitreihendaten, indem Autokorrelationen sichtbar gemacht werden (Eckmann et al. 1987).

Analog zu Formel (2.5) lässt sich für multivariate Zeitreihen bei Überschreitung eines Schwellenwertes τ , resultierend aus der Differenzbildung der mehrdimensionalen Teilsequenz $S = \mathbf{x}_p, \dots, \mathbf{x}_{p+n+1}$ mit der zu erwartenden mehrdimensionalen Teilsequenz \hat{S} ein mathematischen Zusammenhang formulieren, der anomale Teilsequenzen identifiziert :

$$\sum_{i=p}^{p+n+1} \|x_i - \hat{x}_i\| > \tau \quad (2.6)$$

2.3 Kriterien zum Vergleich von Anomalieerkennungsverfahren

Um die Güte von Verfahren zur Identifikation von Anomalien zu quantifizieren, werden üblicherweise folgende drei Metriken verwendet: die Genauigkeit (P , engl. Precision), die Trefferquote (R , engl. Recall) und das F -Maß (Hochenbaum et al. 2017, S. 3; Aggarwal 2017, S. 27 f.). Alle Metriken geben Zahlenwert zwischen 0 und 1 aus. Um diese anwenden zu können, ist es notwendig die Anzahl an Datenpunkten zu kennen, welche von der Anomalieerkennungsmethode als anomal oder normal klassifiziert wurden. Des Weiteren muss die Grundgesamtheit an vorhandenen Anomalien im Datensatz bekannt sein. Dadurch lassen sich 4 Fälle ableiten, die bei der Analyse der Ergebnisse einer Anomalieerkennungsmethode auftreten können. Sie lassen sich in einer Konfusionsmatrix, siehe Tabelle 2-1, darstellen.

Tabelle 2-1: Konfusionsmatrix zur Analyse der Klassifikationen einer Anomalieerkennungsmethode

		Klassifikation als Anomalie durch die Erkennungsmethode	
		Ja	Nein
Tatsächlich vorhandene Anomalie	Ja	Richtig positiv	Falsch negativ
	Nein	Falsch positiv	Richtig negativ

Die *Genauigkeit* P ist definiert als die Anzahl der richtig positiven (tp , engl. true positive) Ergebnisse geteilt durch die Summe aus tp und der Anzahl an falsch positiven (fp , engl. false positive) Ergebnissen und in Formel (2.7) beschrieben.

$$P = \frac{tp}{tp + fp} \quad (2.7)$$

Diese gibt den Anteil der als korrekt positiv klassifizierten Anomalien von der Gesamtheit der durch die Erkennungsmethode als positiv klassifizierten Anomalien an. Die Genauigkeit bildet sich aus den Fällen der ersten Spalte der Konfusionsmatrix.

Die *Trefferquote* R beschreibt das Verhältnis von tp zur Summe aus tp und falsch negativer (fn , engl. false negative) Ergebnisse und ist in Formel (2.8) dargestellt.

$$R = \frac{tp}{tp + fn} \quad (2.8)$$

Das Verhältnis gibt den Anteil, der als korrekt positiv klassifizierten Anomalien, von Gesamtheit aller tatsächlich vorhandenen Anomalien an. Die Trefferquote bildet sich aus den Fällen der ersten Zeile der Konfusionsmatrix.

Das F -Maß, bzw. in Formel (2.9) das F_1 -Maß, berücksichtigt sowohl Genauigkeit als auch Trefferquote und ist wie folgt definiert:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} = \frac{2}{\frac{1}{P} + \frac{1}{R}} \quad (2.9)$$

Es beschreibt das harmonische Mittel von Genauigkeit und Trefferquote und stellt den reziproken Wert des arithmetischen Mittels der reziproken Genauigkeit und der reziproken Trefferquote dar. Dabei müssen beide Werte hoch sein, damit das harmonische Mittel einen hohen F_1 -Wert erhält.

Die allgemeine Variante des F -Maßes ist in Formel (2.10) wiedergegeben und erlaubt eine Gewichtung von Genauigkeit zu Trefferquote.

$$F_\beta = (1 + \beta^2) \cdot \frac{P \cdot R}{\beta^2 \cdot P + R} \quad (2.10)$$

Im Fall von $\beta = 1$ sind Genauigkeit und Trefferquote ausbalanciert.

3 Anwendungsfeld - Lastgangdaten

Im folgenden Kapitel wird zunächst das für diese Arbeit verwendete Datenmaterial vorgestellt, die Auswahl von prototypischen Unternehmen mit der Clusteranalyse mittels k -Means erläutert sowie die Analysestichprobe anhand verschiedener Merkmale charakterisiert (Kapitel 3.1). Anschließend werden die eingesetzten Analysemethoden vorgestellt (Kapitel 3.2), angefangen von der Expertenbefragung zur analogen Identifikation von Anomalien und der Festlegung der Zählweise von Anomalien, über den Einsatz des Machine-Learning Algorithmus (LSTM) und des statistischen Verfahrens (PEWMA) sowie abschließend der Ansätze zur Festlegung eines Schwellenwertes (Z-Score-, ESD-Test) und Informationen zur Berechnung von Residuensequenzen.

3.1 Datenmaterial & Analysestichprobe

3.1.1 Charakterisierung des Gesamtmaterials

Die Limón GmbH hat im Laufe der Firmengeschichte in einem Zeitraum der letzten 10 Jahre Daten in Form von Lastgängen von circa 200 Unternehmen auswerten können. Das gesamte Material stand für diese Ausarbeitung zur Verfügung. Die Daten liegen prinzipiell in einer Zeitauflösung von Werten im 15 Minuten-Rhythmus vor. Ein Datenpunkt setzt aus einem Leistungswert sowie dem zugehörigen Zeitstempel, bestehend aus Uhrzeit und Datum, zusammen. Es handelt sich somit bei den genutzten Eingabedaten um univariate Zeitreihendaten von Lastgängen.

Um hardwareintensive Verarbeitungsprozesse zu minimieren und den Umgang mit der Fülle der Daten im Rahmen einer Masterarbeit zu ermöglichen, gleichzeitig aber die wesentlichen Charakteristika und somit die Repräsentativität der Daten zu erhalten, wurden die Daten in drei Schritten reduziert. *Zunächst* wurden die Zeitauflösung von 15 auf 60 Minuten hochgesetzt, sodass in die weiteren Analysen der aus vier Leistungswerten gemittelte Stundenwert einging. *Zweitens* wurden die Daten eines begrenzten Zeitraumes analysiert. Da die Lastgangdaten der einzelnen Unternehmen in unterschiedlichen Zeiträumlängen sowie zu verschiedenen Zeitpunkten vorliegen, hätte eine Auswahl der Daten anhand des Kriteriums einer zeitliche Überlappung zu einer sehr begrenzten Stichprobe geführt. Deswegen wurden von jedem Unternehmen Lastgangdaten des *letzten verfügbaren* Monats verwendet. Das Datenmaterial bietet eine große Menge an unterschiedlichsten Lastgangverläufen und um dieses Spektrum an Lastgängen zu nutzen, wurden *drittens* die Lastgangdaten des letzten verfügbaren Monats jeden Unternehmens in einem Clusterverfahren gruppiert und aus jedem Cluster die Daten eines prototypischen Unternehmens ausgewählt (vgl. Kapitel 3.1.2).

3.1.2 Gruppierung der Daten mittels *k*-Means Algorithmus

Wie beschrieben wurden die Gesamtdaten u.a. mittels der Auswahl von für das Gesamt aller möglichen Lastgänge prototypischer Verläufe reduziert. Die Identifikation von Unternehmen mit charakteristischen Lastgangverläufen erfolgte über die Gruppierung der Daten unter Verwendung des *k*-Means Algorithmus.

Clusterverfahren lassen sich grundlegend in partitionierende, hierarchische, dichte basierte und kombinierte Verfahren unterteilen (Han und Kamber 2012, S. 448 ff.). Die auf dem *k*-Means Algorithmus basierende Analyse zählt zu den partitionierenden Clusterverfahren. Für den Einsatz des *k*-Means Algorithmus muss die Anzahl der zu bestimmenden Cluster, als *k* Wert, angegeben werden. Als weiterer Input werden die Daten, hier die Lastgänge der Unternehmen, benötigt. Der vorliegende Algorithmus bestimmt

zufällige initiale Clusterzentren, sogenannte Centroide. Die übrigen Datenpunkte bzw. Lastgänge werden dem Clusterzentrum zugeordnet, dem sie am ähnlichsten sind, basierend auf dem euklidischen Abstand zwischen dem Datenpunkt und dem Centroid. Nach der Zuordnung werden die Clusterzentren neu berechnet. In einer neuen Iteration werden alle Datenpunkte den aktualisierten Centroiden neu zugeordnet. Es finden so viele Iterationen statt bis die Zuordnungen stabil sind. Jeder Lastgang entspricht einem mehrdimensionalen Datenpunkt, welcher die Größe entsprechend der Stunden eines Monats besitzt. Nach dem Durchlauf der Iterationen repräsentiert ein Centroid ein gemittelten Lastgang aller Lastgänge eines Clusters.

Ein Kriterium, welches die Separierung der gebildeten Cluster bewertet, ist der mittlere Silhouetten Koeffizient sc_m aller Datenpunkte:

$$sc_m = \frac{1}{n} \sum_{i=1}^n sc_i \quad (3.1)$$

mit sc_i dem Silhouetten Koeffizienten eine Datenpunktes, bzw. Lastganges, i und n der Anzahl der Datenpunkten. Der sc_i Wert eines Datenpunktes ist definiert durch:

$$sc_i = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (3.2)$$

mit $a(i)$ den mittleren Abstand eines Datenpunktes zu allen anderen Punkten des gleichen und Clusters $b(i)$ der mittlere Abstand eines Datenpunktes zu allen anderen Punkten des nächsten Clusters (Rousseeuw 1987, S. 56). Dabei kann sc_m bzw. sc_i Werte zwischen -1 bis +1 annehmen. Tabelle 3-1 präsentiert eine Zusammenfassung entsprechender Wertebedeutungen.

Tabelle 3-1: Erläuterung zu Werterepräsentation der Silhouetten Koeffizienten

Wertebereich um	Bedeutung für den Silhouetten Koeffizienten sc_m	Bedeutung für den Silhouetten Koeffizienten sc_i
-1	Cluster sind inkorrekt gewählt	Datenpunkt i ist einem unpassenden Cluster zugeordnet
0	Cluster überlappen sich	Datenpunkt i liegt in einem Bereich zwischen Clustern
+1	Cluster sind weit voneinander entfernt und klar abgegrenzt	Datenpunkt i liegt im zugeordneten Cluster

Im vorliegenden Fall werden nicht normierte bzw. normierte Lastgänge sowie geordnete Lastgänge (Dauerlinien) untersucht. Bei der Verwendung *geordneter Lastgänge*, sind die Leistungswerte eines Lastganges absteigend beginnend mit dem Maximalwert sortiert. Die Dauerlinie zeigt an, wieviel Stunden ein bestimmter Leistungsbedarf aufgetreten ist. Zwar geht dabei die zeitliche Ordnung, d.h. die Abfolge der Tage und Stunden, verloren, jedoch lassen sich andere nutzbare Informationen, wie einzelne Betriebszustände, aus der Dauerlinie herauslesen. Mithilfe einer *Normierung* werden die z.T. sehr unterschiedlichen Leistungswerte (vgl. Abbildung 3.1) der einzelnen Lastgänge sowie Dauerlinien miteinander vergleichbar. Dazu werden alle Leistungswerte durch den Maximalwert des jeweiligen Lastganges dividiert, sodass die normierten Lastgänge in einem Wertebereich von 0 bis 1 liegen.

Insgesamt wurden für die Clusteranalyse 179 Lastgänge verwendet. Bei der Vorbereitung der Daten wurde unvollständige Monatsauszüge ausgeschlossen. Als unvollständig galten solche Datensätze, die eine Länge von weniger als drei Wochen aufwiesen. Abbildung 3.1 stellt die zur Clusteranalyse verwendeten Unternehmen dar.

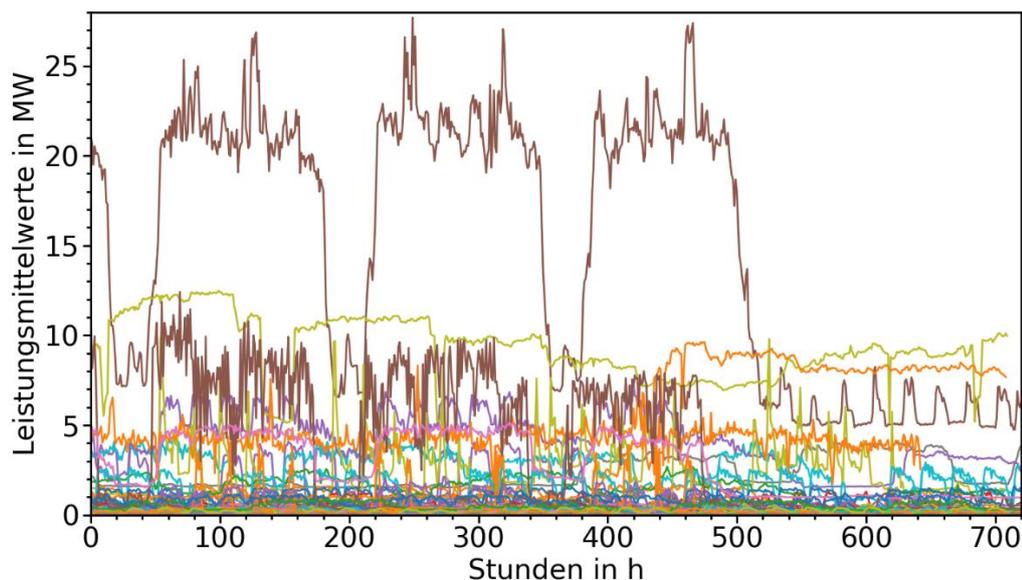


Abbildung 3.1: Darstellung aller im Clusterverfahren verwendeten nicht normierten Lastgänge

Ergebnisse. Tabelle 3-2 fasst die Ergebnisse des Clusterverfahrens für verschiedene Clustergrößen k zusammen. Der mittlere Silhouetten Koeffizient nimmt im Fall von nicht normierten (nn) Lastgängen sehr hohe Werte an. Jedoch liegt der überwiegende Teil an Unternehmen jeweils nur in einem Cluster. Dieser Umstand wird durch den hohen n_1 Wert, welcher die Anzahl der gruppierten Unternehmen im größten Cluster für ein

gewählten k Wert repräsentiert, dargestellt. Die wenigen restlichen Unternehmen werden bei fast allen k Werten den übrigen Clustern zugeordnet.

Tabelle 3-2: Ergebnisse aus der Clusterung von 179 Unternehmen für vorgegebenen Clustergrößen k , mit nicht normierten (nn) und normierten (n) Lastgängen und Angabe der Anzahl an Unternehmen im größten (n_I) und zweitgrößten (n_{II}) gebildeten Cluster

Clusteranzahl k		2	3	4	5	6	7	8	9	10
nn	sc_m	0,93	0,90	0,88	0,85	0,85	0,84	0,72	0,71	0,71
	n_I / n_{II}	175/4	171/7	168/8	165/7	165/6	165/7	148/17	146/19	146/19
n	sc_m	0,21	0,14	0,14	0,15	0,13	0,15	0,13	0,13	0,14
	n_I / n_{II}	97/82	76/56	59/49	62/43	48/36	33/26	33/28	34/31	33/29

Die normierten (n) Lastgänge zeigen insgesamt schlecht ausgebildete Cluster, da die sc_m Werte aller gewählten Clustergrößen sehr niedrig sind. Die ungenügende Separierung für nicht normierte bzw. normierte Lastgänge hat mehrere Gründe. Im Fall von nicht normierten Lastprofilen, weisen einige Unternehmen wie Stahlwerke oder Gießereien durchschnittlich sehr hohe Leistungswerte, bis 27 MW, im Monatsverlauf auf. Jedoch weist der Großteil der Unternehmen Leistungswerte unter 5 MW auf und diese werden bei den verwendeten Clustergrößen zusammen gruppiert. Die normierten Lastgänge werden eher relativ gleichmäßig auf die verfügbaren Cluster aufgeteilt und diese Cluster sind nicht scharf abgegrenzt. Das Ergebnis lässt sich damit begründen, dass hier Lastgänge aus unterschiedlichen Monaten gruppiert werden und diese im gleichen Wertebereich vorliegen. Dadurch kommt es zu unzähligen Überschneidungen der Lastverläufe, wodurch die Separierung in ungenügender Weise erfolgt.

In Tabelle 3-3 sind die Ergebnisse für das Clustern der Dauerlinien dargestellt. Für die nicht normierten geordneten Dauerlinien zeigen sich ähnliche Ergebnisse, wie für nicht normierte Lastgänge und diese lassen sich analog begründen. Die normierten Dauerlinien zeigen jedoch höhere sc_m Werte im Vergleich zu normierten Lastgänge. Die der Dauerlinie immanente Ordnungsstruktur begünstigt die Clusterbildung für die gewählten k Werte. Daher werden im weiteren Verlauf aus der Clustergröße $k = 5$, siehe dazu Abbildung 3.2, 5 Unternehmen ausgewählt, welche für den Vergleich und die Entwicklung der ausgewählten Anomaliererkennungsmethoden zur Anwendung kommen.

Tabelle 3-3: Ergebnisse aus der Clusterung von 179 Unternehmen für vorgegebenen Clustergrößen k , mit nicht normierten (nn) und normierten (n) geordneten Lastgängen und Angabe der Anzahl an Unternehmen im größten (n_I) und zweitgrößten (n_{II}) gebildeten Cluster

Clusteranzahl k		2	3	4	5	6	7	8	9	10
nn	SC_m	0,95	0,90	0,90	0,90	0,88	0,78	0,74	0,73	0,67
	n_I / n_{II}	177/ 2	169/ 8	169/ 8	158/ 11	158/ 11	146/ 20	142/ 22	135/ 22	127/ 27
n	SC_m	0,39	0,34	0,34	0,35	0,32	0,31	0,30	0,29	0,28
	n_I / n_{II}	107/ 72	86/ 63	61/ 56	57/ 47	49/ 39	39/ 38	39/ 30	38/ 31	33/ 32

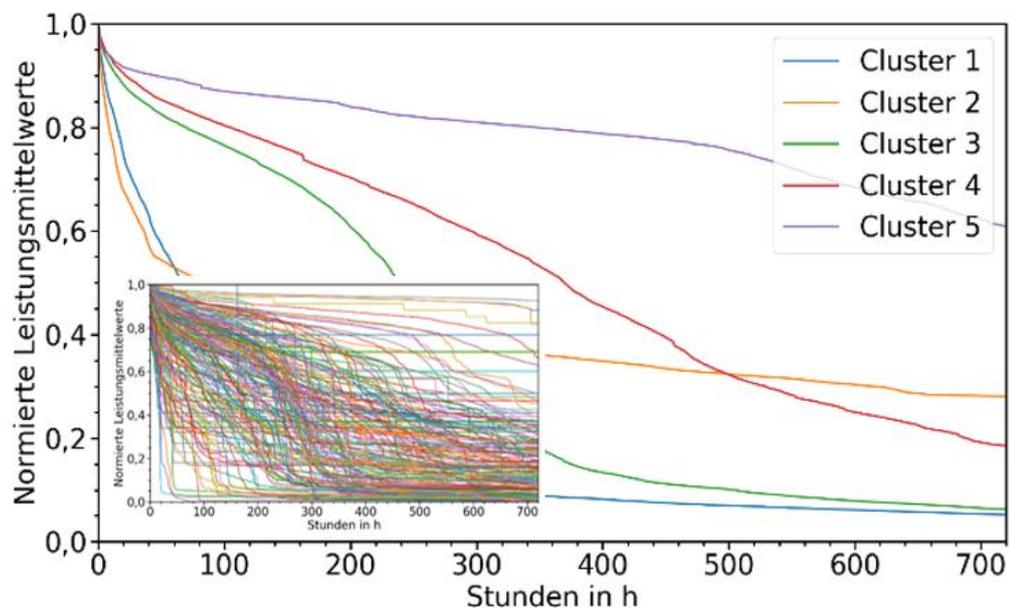


Abbildung 3.2: Darstellung der 5 gebildeten Cluster der normierten Dauerlinien (großer Rahmen) aus allen verwendeten Unternehmen (kleiner Rahmen, unten links)

Unter dem Einsatz von k Werten größer 10 bis 30 zeigten sich keine deutlichen Verbesserungen der Kennzahlen. Für die Auswahl einer möglichst geringen Anzahl repräsentativer Lastgänge, wird daher auf die Darstellung der Untersuchung von mehr als $k = 10$ Cluster verzichtet.

3.1.3 Charakterisierung der Analytestichprobe

Aus jedem Cluster ($k = 5$) wurde jeweils ein Unternehmen ausgesucht. Die Charakteristika der den Clustern entnommenen Unternehmen werden im Folgenden näher

erläutert. Der Lastgang eines ausgewählten Unternehmens entspricht nicht immer dem Verlauf des jeweiligen Dauerlinienclusterzentrums, da der jeweilige Centroid einen gemittelten Lastgang darstellt. Die Dauerlinien ausgewählter Unternehmen müssen folglich nicht zu jeder Zeit dem gemittelten Lastgang des zugehörigen Clusters ähneln.

Unternehmen aus Cluster 1 sind charakterisiert durch sehr rasch abfallenden Dauerlinien, welche im Vergleich zu anderen Clustern eine große Stundenanzahl sehr niedriger Leistungswerte, im Verhältnis zum Lastgangmaximalwert, enthalten. Die dazugehörigen Lastgänge sind dadurch gekennzeichnet, dass ein niedriges Grundlastniveau existiert, aber in relativ kurzen Zeitintervallen die Lastgänge stark ansteigende bzw. abfallende Verläufe aufweisen. Diese Art von Verläufen liegen in Unternehmen vor, in welchen eine Kernarbeitszeit, beispielsweise von 8 bis 18 Uhr, vorliegt und nachts nicht gearbeitet wird bzw. auch kein Leistungsbedarf vorhanden ist. Das aus Cluster 1 ausgewählte Unternehmen ist in Abbildung 3.3 dargestellt. Die durch das Experteninterview (vgl. Kapitel 3.2.1) vordefinierten Anomalien sind in den Diagrammen der Unternehmen auch mit abgebildet.

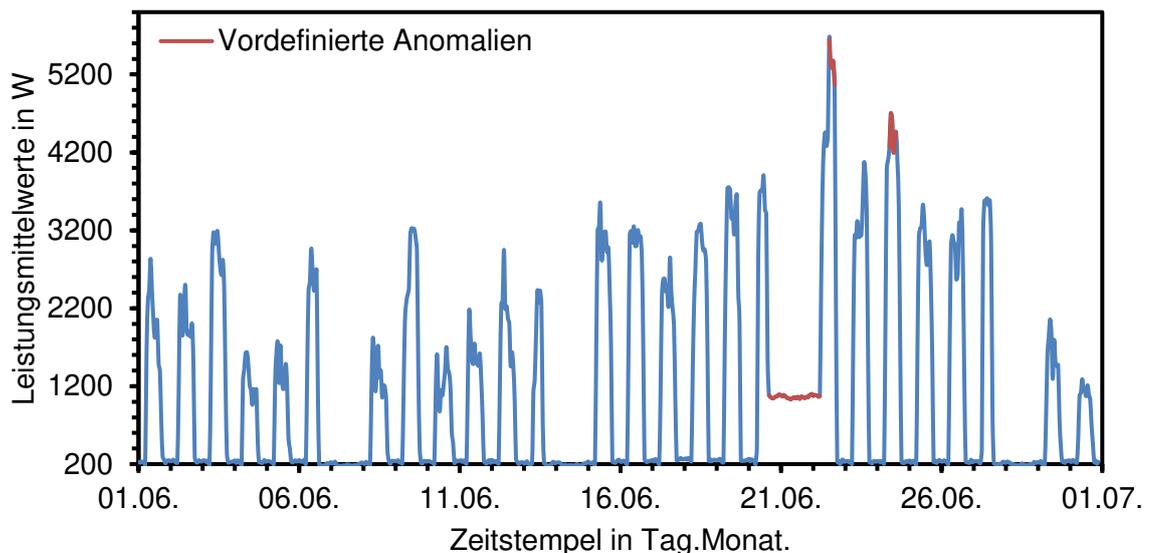


Abbildung 3.3: Cluster 1 Unternehmen mit vordefinierten Anomalien

In Tabelle 3-4 sind die vordefinierten Anomalien des Cluster 1 Unternehmens mit ihren jeweiligen Zeiträumen des Auftretens aufgelistet.

Tabelle 3-4: Übersicht der vordefinierten Anomalien des Cluster 1 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde

Anomalie	Begin	Ende
1	2015-06-20 15:00:00	2015-06-22 05:00:00
2	2015-06-22 12:00:00	2015-06-22 16:00:00
3	2015-06-24 09:00:00	2015-06-24 14:00:00

In Cluster 2 liegen Unternehmen, welche ähnlich beschrieben werden können, wie Unternehmen aus Cluster 1, jedoch liegt das Grundlastniveau im Verhältnis zu den maximalen Leistungswerten des Lastganges höher, d.h. im Durchschnitt bei circa 40 % des Maximalwertes. Das aus Cluster 2 ausgewählte Unternehmen ist in Abbildung 3.4 dargestellt.

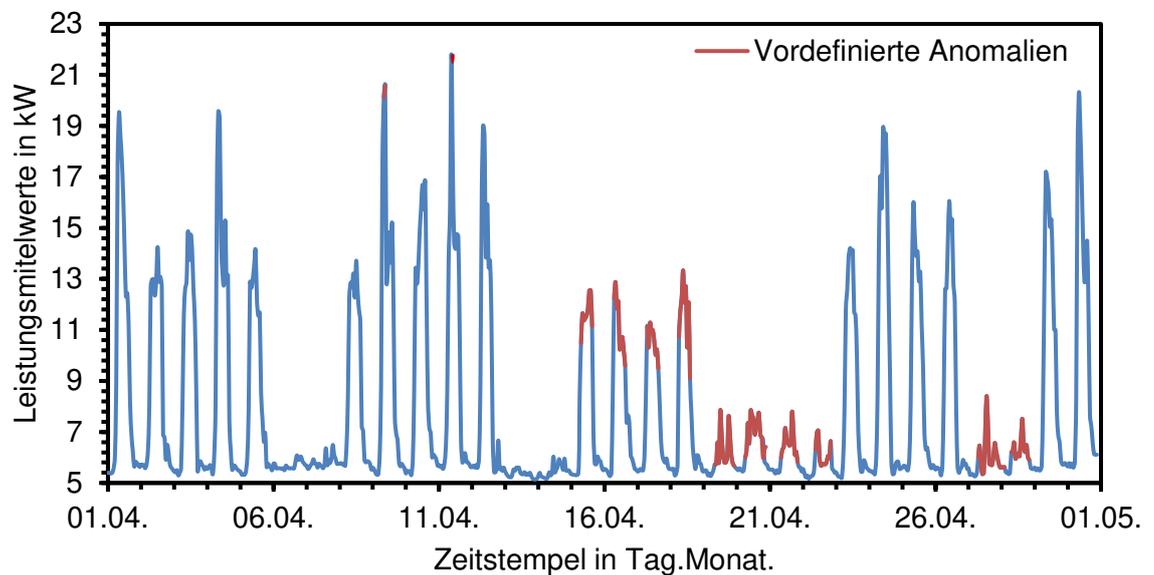


Abbildung 3.4: Cluster 2 Unternehmen mit vordefinierten Anomalien

Tabelle 3-5 gibt eine Übersicht über die vordefinierten Anomalien des Cluster 2 Unternehmens.

Tabelle 3-5: Übersicht der vordefinierten Anomalien des Cluster 2 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde

Anomalie	Begin	Ende
1	2019-04-09 08:00:00	2019-04-09 09:00:00
2	2019-04-11 09:00:00	
3	2019-04-15 07:00:00	2019-04-15 15:00:00
4	2019-04-16 07:00:00	2019-04-16 15:00:00
5	2019-04-17 07:00:00	2019-04-17 15:00:00
6	2019-04-18 06:00:00	2019-04-18 14:00:00
7	2019-04-19 08:00:00	2019-04-19 22:00:00
8	2019-04-20 06:00:00	2019-04-21 10:00:00
9	2019-04-21 08:00:00	2019-04-21 20:00:00
10	2019-04-22 09:00:00	2019-04-22 20:00:00
11	2019-04-27 07:00:00	2019-04-28 02:00:00
12	2019-04-28 07:00:00	2019-04-28 20:00:00

Unternehmen aus *Cluster 3* zeigen eine ausgeprägte Betriebsstufe in der Dauerlinie sowie ein relativ geringes Grundlastniveau im Verhältnis zu den Leistungsmaxima im betreffenden Lastgang. Die Unternehmen zeigen somit, im Vergleich zu Unternehmen aus Cluster 1 & 2, verlängerte Zeitintervalle mit einem hohem Leistungsbedarf und stark aufsteigende bzw. abfallende Verläufe am Rand dieser Zeitintervalle. Das gewählte Unternehmen aus Cluster 3 ist in Abbildung 3.5 dargestellt.

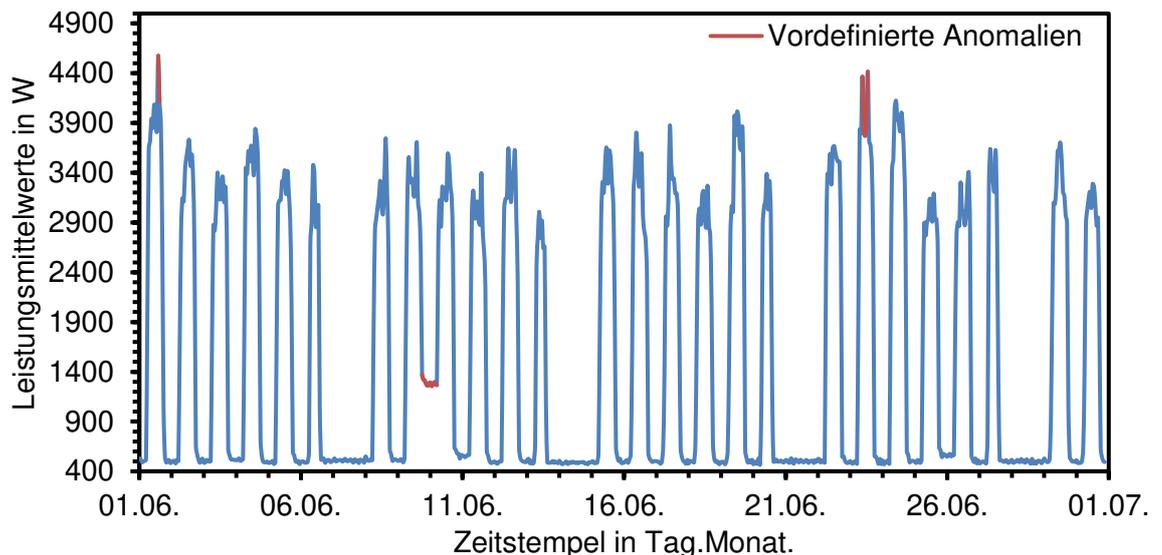


Abbildung 3.5: Cluster 3 Unternehmen mit vordefinierten Anomalien

Tabelle 3-6 listet die anomalen Lastgangabschnitte des Cluster 3 Unternehmen auf.

Tabelle 3-6: Übersicht der vordefinierten Anomalien des Cluster 3 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute: Sekunde

Anomalie	Begin	Ende
1	2015-06-01 14:00:00	2015-06-01 15:00:00
2	2015-06-09 18:00:00	2015-06-10 05:00:00
3	2015-06-23 09:00:00	2015-06-23 13:00:00

Bei Unternehmen aus *Cluster 4* liegt das Grundlastniveau im Verhältnis zu dem im Lastgang vorhandenen Leistungsmaximalwerten in der Regel höher als bei Unternehmen aus Cluster 3. Des Weiteren sind diese Unternehmen dadurch charakterisiert, dass ein höhere Stundenanzahl eines deutlich über dem Grundlastniveau gesetzten Leistungsbedarfs im Vergleich zu den vorhergehenden Unternehmen vorhanden ist. Das ausgewählte Unternehmen ist in Abbildung 3.6 dargestellt.

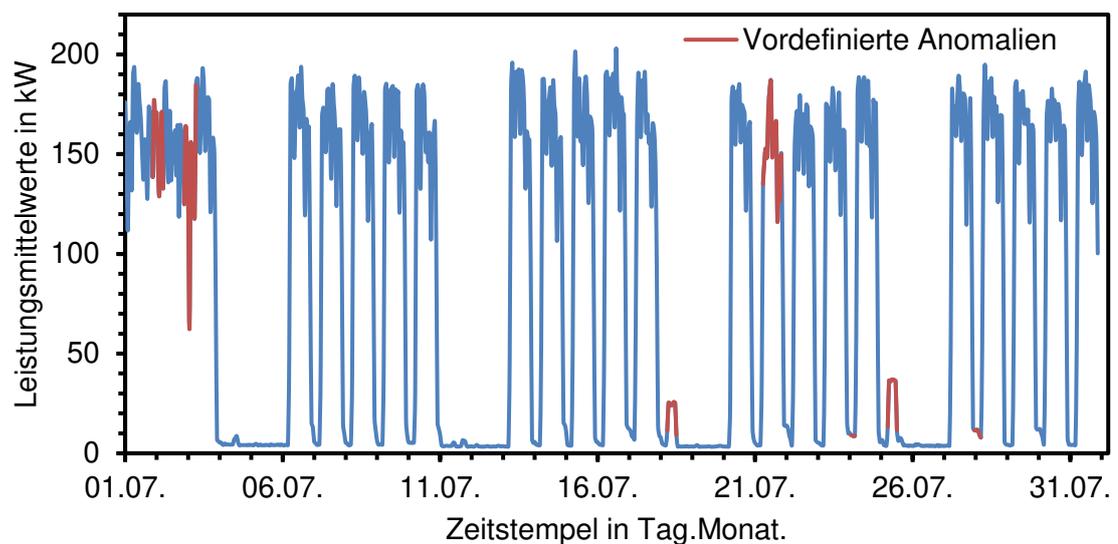


Abbildung 3.6: Cluster 4 Unternehmen mit vordefinierten Anomalien

Alle vordefinierten Anomalien des Unternehmens aus Cluster 4, sowie die Zeiträume ihres jeweiligen Auftretens, sind in Tabelle 3-7 zusammengefasst.

Tabelle 3-7: Übersicht der vordefinierten Anomalien des Cluster 4 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde

Anomalie	Begin	Ende
1	2020-07-01 21:00:00	2020-07-02 05:00:00
2	2020-07-02 21:00:00	2020-07-03 06:00:00
3	2020-07-18 05:00:00	2020-07-18 12:00:00
4	2020-07-21 06:00:00	2020-07-21 20:00:00
5	2020-07-24 01:00:00	2020-07-24 04:00:00
6	2020-07-25 05:00:00	2020-07-25 12:00:00
7	2020-07-27 23:00:00	2020-07-28 04:00:00

Unternehmen aus *Cluster 5* zeigen vergleichsweise flache Dauerlinienverläufe. Der Leistungsbedarf ändert sich in Relation zum Maximalwert des jeweiligen Lastgangs im Vergleich zu Unternehmen anderer Cluster am wenigsten. Diese Art von Lastgangverläufen zeigt in den meisten Fällen sehr geringe Abhängigkeiten von Tag-Nacht-Wechsel oder Werktag-Wochenende- bzw. Feiertag-Wechsel, anders als Unternehmen aus den Clustern 1 bis 4. In Abbildung 3.7 ist das ausgewählte Unternehmen aus Cluster 5 präsentiert.

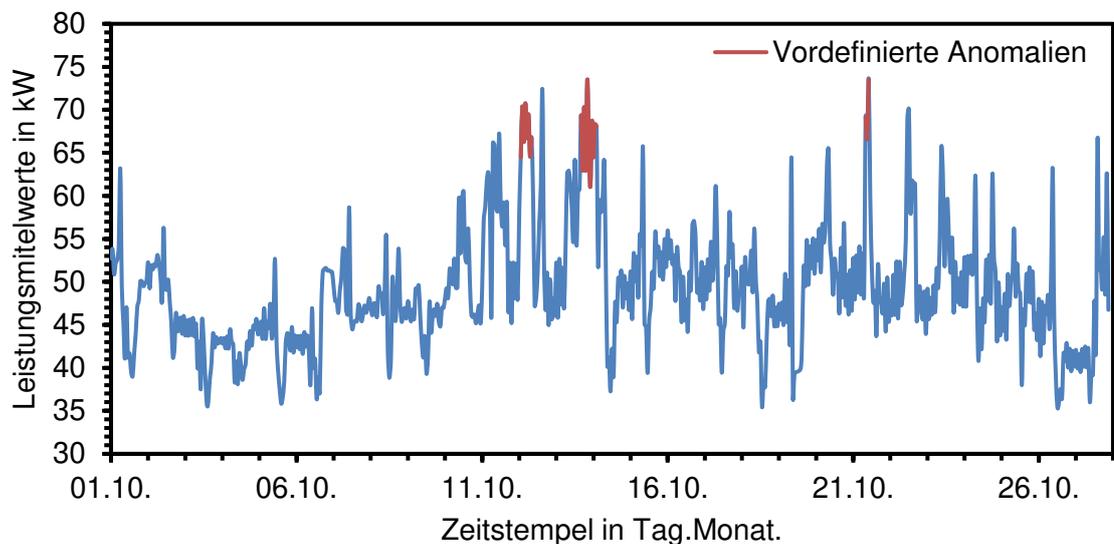


Abbildung 3.7: Cluster 5 Unternehmen vordefinierten Anomalien

Tabelle 3-8 zeigt wieder zeitliche Übersicht der vordefinierten Anomalien, diesmal von dem ausgewählten Unternehmen aus Cluster 5.

Tabelle 3-8: Übersicht der vordefinierten Anomalien des Cluster 5 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde

Anomalie	Begin	Ende
1	2017-10-12 01:00:00	2017-10-12 08:00:00
2	2017-10-13 16:00:00	2017-10-14 02:00:00
3	2017-10-21 08:00:00	2017-10-21 10:00:00

Die vordefinierten Anomalien der Unternehmen aus Cluster 1 bis 4 lassen sich als zeitlich kontextuale Anomalien beschreiben, da die Lastgänge einem zeitlichen Rhythmus folgen, d.h. Tag-Nacht- und Werktag-Wochenende-Abhängigkeiten vorhanden sind, und ungewöhnliche Verläufe im Kontext des betreffenden Zeitpunktes existieren. Das Unternehmen aus Cluster 5 unterliegt keiner dieser Abhängigkeiten. Die hier vordefinierten Anomalien lassen sich nur im Kontext des Gesamtverlaufes sehen, sodass ungewöhnliche maximale Leistungswerte als Ausreißer definiert werden. Da außer dem Zeitstempel und dem zugehörigen Leistungswert keine weiteren Einflussgrößen gegeben sind, können keine weiteren Kontexte untersucht werden, obwohl diese wahrscheinlich existieren. Im Fall von den Unternehmen aus Cluster 5 könnte beispielsweise die Produktionsmenge oder die Auftragslage ein Parameter sein, der den Lastgangverlauf beeinflusst.

3.2 Verwendete Analysemethoden

3.2.1 Experteninterviews & Zählweise der Anomalien

Experteninterviews. Um Anomalien vorzudefinieren wurden im Vorfeld der Auswertung die zu untersuchenden Unternehmen in einem *Experteninterview* auf anomale Verläufe hin untersucht. Prof. Junge sowie ein erfahrener Mitarbeiter der Limón GmbH im Bereich der Lastganganalyse beurteilten die Lastgänge. Die von den Experten definierten Anomalien gelten als Referenzen zur Bestimmung von Güte und Eignung (vgl. Kapitel 2.3) der im weiteren Verlauf der Arbeit eingesetzten Strategien zur Anomalieerkennung.

Zählweise der Anomalien. In Kapitel 3.1.3 sind die zu untersuchenden Unternehmen dargestellt. Es zeigt sich, dass der überwiegende Teil der Unternehmen mehr vordefinierte anomale Sequenzen als Punktanomalien aufweist. Punktanomalien werden dadurch identifiziert, dass die Residuenwerte einzelner Zeitpunkte ausgewertet werden. Es ist dabei möglich, dass Punktanomalieuntersuchungen auch anomale Teilsequenzen erkennen. Denn die Schätzung im Bereich der anomalen Sequenzen unterscheidet sich von den gegebenen Lastgangdaten z.T. ebenfalls durch nur *einzelne* hohe Residuenwerte. Allerdings kann auch die Untersuchung von Teilsequenzanomalien mittels

summierter Residuenwerte einzelne Punktanomalien identifizieren. Um dieser Schwierigkeit zu begegnen wird folgende Zählweise für Anomalien festgelegt:

- Anomale Sequenzbereiche werden bei mehrmaliger Identifizierung als nur einmal erkannt gewertet.
- Treten Punktanomalien und anomale Sequenzen im einem Lastgang gemeinsam auf, dann zählt jede anomale Sequenz jeweils als eine Anomalie.

Unter diesem Vorgehen ergibt sich somit folgende Anzahl an Referenzanomalien bei den aus den Clustern gewählten Unternehmen:

Tabelle 3-9: Auflistung der Anzahl an Anomalien je Unternehmen eines Clusters

	Anzahl an Punktanomalien	Anzahl an Teilsequenzanomalien	Gesamtanzahl an Anomalien
Cluster 1	0	3	3
Cluster 2	1	11	12
Cluster 3	0	3	3
Cluster 4	0	7	7
Cluster 5	0	3	3

3.2.2 Verfahren aus dem Bereich des Machine-Learning: LSTM

Als Grundlage zur Identifikation von Anomalien ist es erforderlich, ein Normalmodell zu definieren. Dies kann als Schätzung oder Vorhersage erfolgen (vgl. Kapitel 2.2). Der Vergleich von statistischen und machine-learning basierenden Verfahren zur Schätzung bzw. Vorhersage von Zeitreihen und die resultierende Frage, welches Verfahren bessere Ergebnisse zeigt, ist bisher nicht geklärt und Gegenstand aktueller Forschung (Makridakis et al. 2018). Festhalten lässt sich jedoch, dass Long Short-Term Memory (LSTM, dt.: langes Kurzzeitgedächtnis) Netzwerke erfolgreich im Bereich der Schätzung bzw. Vorhersage von Zeitreihen eingesetzt werden können (Siarni-Namini et al. 2018).

LSTM Netzwerke gehören zur Gruppe der Recurrent neural networks (RNNs, dt.: rekurrente bzw. rückgekoppelte neuronale Netze). Dabei handelt es sich um einen Sonderfall künstlicher neuronaler Netze. Im Gegensatz zu „vorwärtsbetriebenen“ künstlichen neuronalen Netzen (engl. feed forward networks), bei welchen die Informationsweitergabe vom Eingabeneuron bis hin zum Ausgabeneuron nur in eine Richtung realisiert ist, werden in RNNs Informationen, durch Verknüpfungen von Neuronen mit Neuronen der gleichen oder einer vorangegangenen Neuronenschicht je nach Verknüpfungsstruktur zum Teil rückgekoppelt (Kruse et al. 2015, S. 35). In Abbildung 3.8 ist zur Verdeutlichung ein einziges nicht rekurrentes (a) und rekurrentes Neuron (b) dargestellt (Elsworth und

Güttel 2020, S. 2 f.). Im Fall des nicht rekurrenten Neurons werden alle Daten einer Sequenz, bestehend aus Datenpunkten $x^{(1)}$ bis $x^{(\ell)}$, in der Trainingsphase in einem Schritt verwendet. Hiermit geht die sequenzielle bzw. zeitliche Ordnung der Daten verloren. Dagegen kann das rekurrente Neuron, in Abbildung 3.8 (b, links) in verallgemeinerter Form darstellt, die Datenpunkte $x^{(1)}$ bis $x^{(\ell)}$ zeitlich nacheinander verarbeiten, siehe dazu Abbildung 3.8 (b, rechts).

$$\sigma(\mathbf{w}_1^T \mathbf{x}_1 + \mathbf{w}_2^T \mathbf{x}_2 + \dots + \mathbf{w}_\ell^T \mathbf{x}_\ell + b) \quad \sigma(w_h(\dots \sigma(w_h \sigma(\mathbf{w}^T \mathbf{x}^{(1)} + b) + \mathbf{w}^T \mathbf{x}^{(2)} + b)) \dots) + \mathbf{w}^T \mathbf{x}^{(\ell)}$$

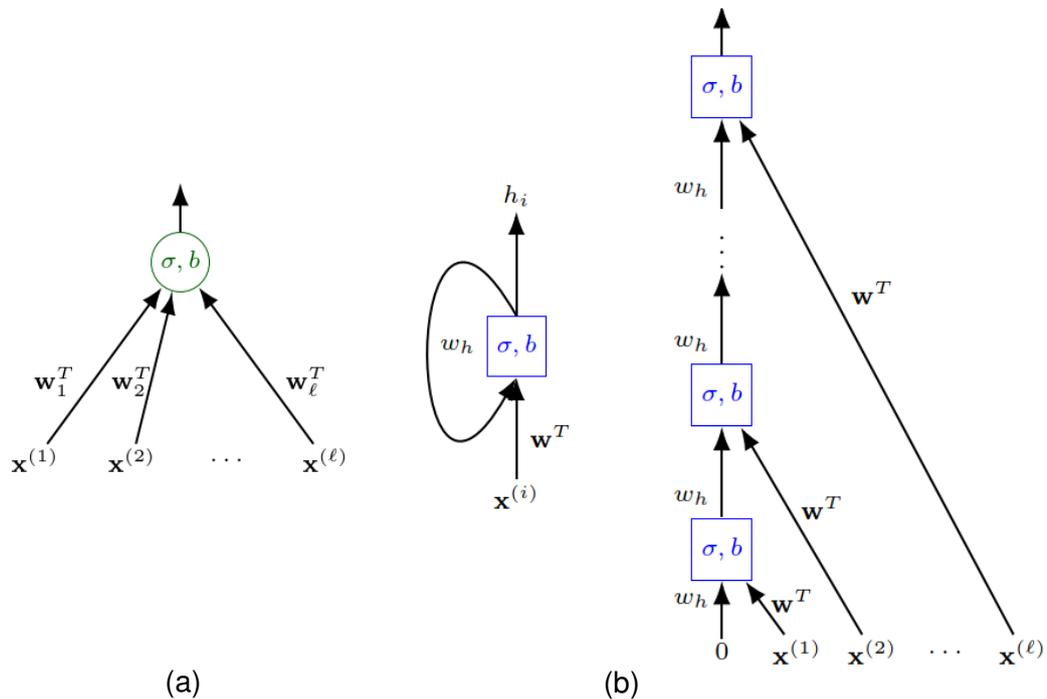


Abbildung 3.8: Darstellung eines einzigen (a) nicht rekurrenten und (b) rekurrenten Neurons. Mit σ der Aktivierungsfunktion, b dem Biaswert und w^1 bis w^ℓ den Gewichten (Elsworth und Güttel 2020, S. 2 f.).

Das Training der RNN geschieht durch den Backpropagation Through Time (BPTT) Algorithmus, d.h. die Gewichte werden, in umgekehrter zeitlicher Reihenfolge neu justiert, sodass die Gewichte der ersten Datenpunkte zuletzt korrigiert werden (Elsworth und Güttel 2020, S. 3). Bei RNN und auch anderen künstlichen neuronalen Netzen, die insbesondere lange Sequenzen verarbeiten sollen, entsteht die Problematik, dass die Gewichte in ungenügender Weise neu justiert werden, je weiter die Daten in der Sequenz zurückliegen (Hochreiter und Schmidhuber 1997, S. 1). LSTM Netzwerke lösen dieses Problem und stellen daher eine Weiterentwicklung von RNNs dar. Diese besitzen komplexer aufgebaute Neuronen, welche als Speicherblöcke bezeichnet werden, sodass LSTM Netzwerke in sequentiellen Daten effektiver langfristige Abhängigkeiten berücksichtigen und erlernen können (Hochreiter und Schmidhuber 1997, S. 7).

Das LSTM Netzwerk lässt sich in Python durch die Keras Bibliothek verwenden. (Keras Google Group). Die in der Masterarbeit verwendete Realisierung der Codestruktur orientiert sich an Brownlee (2020).

3.2.3 Verfahren aus der statistischen Zeitreihenanalyse: PEWMA

Neben Verfahren aus dem Bereich des Machine-Learning lassen sich auch in der Statistik Verfahren zum Umgang mit und zur Schätzung bzw. Vorhersage von Zeitreihendaten finden. So gibt es bspw. s.g. Glättungsmethoden, die Zeitreihen durch neu berechnete Werte fortführen (Schlittgen und Sattarhoff 2020, S. 61 f.) oder Unregelmäßigkeiten im Verlauf minimieren (Schlittgen 2001, S. 35). Durch das Beseitigen von irregulären Schwankungen im Verlauf der Daten wird eine neue Zeitreihe erzeugt. Diese neuen Zeitreihenpunkten lassen sich mit den Originaldaten vergleichen, sodass durch die resultierenden Residuen auf Anomalien geschlossen werden kann. Es existiert eine Reihe unterschiedlicher Glättungsverfahren, die sich in ihrer Komplexität unterscheiden.

Der einfache gleitende Mittelwert (SMA, engl. Simple Moving Average) bildet sich beispielsweise als arithmetische Mittelwerte von n Datenpunkten. Dazu wird ein Zeitfenster der Länge n schrittweise über die Zeitreihendaten verschoben und die Daten innerhalb des Fensters gemittelt, siehe dazu Formel (3.3).

$$y_{t,SMA} = \frac{1}{n} \sum_{i=0}^{n-1} y_{t-i} \quad (3.3)$$

Das SMA Verfahren gewichtet jeden vorherigen Datenpunkt gleich, sodass z.B. bei einem ansteigenden Kurvenverlauf der gebildete Mittelwert immer unterhalb des zuletzt verwendeten Datenpunktes liegt. Daraus kann allerdings eine ungenügende Schätzung resultieren.

Eine Methode, die versucht, dieser Problematik zu begegnen, indem Zeitreihenwerte gewichtet werden, ist das exponentiell gewichtete gleitende Mittelwert Verfahren (EWMA, engl. Exponential Weighted Moving Average) (Zwerenz 2012, S. 259 (Teil III); Roberts 1959). Die allgemeine Gleichung lautet wie folgt:

$$y_{t,EWMA} = \sum_{i=1}^t \alpha^{t-i} (1 - \alpha) \cdot y_i \quad (3.4)$$

mit $y_{t,EWMA}$ dem geglätteten Wert an Zeitpunkt t und α der Glättungskonstante bzw. Gewichtungskonstante im Wertebereich von 0 bis 1. Gleichung (3.4) lässt sich in rekursiver Form darstellen und ist beschreibbar durch:

$$y_{t,EWMA} = \alpha \cdot y_{t-1,EWMA} + (1 - \alpha) \cdot y_t \quad (3.5)$$

Je kleiner α , desto mehr Gewicht besitzt der aktuelle Wert y_t und desto eher passt sich $y_{t,EWMA}$ dem aktuellen Verlauf an und wirkt dadurch weniger geglättet. Laut Carter und Streilein (2012) besitzt EWMA den Nachteil, dass sprunghafte Verläufe den geglätteten Wert von der Originalkurve weit entfernen und damit die Identifizierung von Anomalien durch den Vergleich der beiden Werte stark beeinträchtigen. PEWMA (engl. Probabilistic EWMA) ist eine Weiterentwicklung von EWMA (Carter und Streilein 2012) und ist folgendermaßen definiert:

$$y_{t,PEWMA} = \alpha_p \cdot (1 - \beta \cdot P_t) \cdot y_{t-1,PEWMA} + (1 - \alpha_p \cdot (1 - \beta \cdot P_t)) \cdot y_t \quad (3.6)$$

mit α_p der Gewichtungskonstante im Wertebereich zwischen 0 und 1, β einen gewichtenden Faktor für Wahrscheinlichkeitsfunktion P_t und $y_{t,PEWMA}$ den durch das PEWMA Verfahren geglätteten Wert zum Zeitpunkt t .

Im Weiteren berechnet sich P_t durch Formel (3.7) mit σ_t der Standardabweichung.

$$P_t = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_t - y_{t-1,PEWMA})^2}{2 \cdot \sigma_t^2}\right) \quad (3.7)$$

Durch Anpassung von α lässt sich, analog zum EWMA Verfahren, bestimmen, wieviel Gewicht dem vorangegangenen Wert $y_{t-1,PEWMA}$ beigemessen werden soll. Je höher β , desto mehr wird verhindert, dass anomale Datenpunkte den Verlauf von $y_{t,PEWMA}$ beeinflussen. Angelehnt an Renshaw (2016) wird das PEWMA Verfahren in dieser Arbeit mit $\beta = 0.5$ verwendet. Des Weiteren werden die drei α_p Werte 0,3, 0,6 und 0,9 hinsichtlich der Möglichkeit zur gelingenden Identifikation von Anomalien geprüft.

Für diese Arbeit wurde ein Python Code für das PEWMA Verfahren verwendet, welcher im Internet zugänglich ist (PEWMA Python 2019).

3.2.4 Festlegung und Prüfung von Schwellenwerten

Mittels der in Kapitel 2.2 erläuterten Verfahren zur Identifikation von Anomalien werden Residuen – also Abweichungen von z.B. mithilfe von PEWMA oder LSTM geschätzten bzw. vorhergesagten Erwartungswerten sichtbar. Um zuverlässig *echte* Anomalien erkennen zu können, die weder zufällig entstanden sind noch zur normalen Variabilität gezählt werden müssen, ist es möglich s.g. Residuenanalysestrategien einzusetzen, welche im Folgenden beschrieben werden.

Z-Score Test

Residuenwerten lassen durch eine Transformation standardisieren (Wooldridge 2013, S. 736). Die Originalmaßeinheit geht dabei verloren und die resultierenden Z-Werte werden in Vielfachen der Standardabweichung s dargestellt. Formel (3.8) stellt die mathematische Formulierung der Z-Transformation dar:

$$z_i = \frac{x_i - \bar{x}}{s} \quad (3.8)$$

mit x_i den Originalwerten der Stichprobe, \bar{x} den Mittelwert der Stichprobe und s der Standardabweichung der Stichprobe. Die Z-Transformation verändert nicht die Form der Stichprobenverteilung, aber deren Mittelwert und Standardabweichung. Es lässt sich nun ein Schwellenwert definieren ab wieviel des Vielfachen der Standardabweichung ein Datenpunkt als Ausreißer deklariert werden soll. In dieser Ausarbeitung werden die drei Schwellenwerte 2, 2,5, und 3 gesetzt und die Eignung der gewählten Schwellen zur Identifikation von richtig-positiven Anomalien im Abgleich mit den durch Experten gesetzten Referenzanomalien bewertet.

ESD-Test

Der generalisierte Extreme Studentized Deviate (ESD) Test (Rosner 1983) wird angewandt, um statistische Ausreißer in einer Menge von Werten, bzw. in dem Kontext der Masterarbeit aus der Menge der Residuen, zu erkennen. Der ESD Test ist für folgende Hypothesen definiert:

H_0 : Es existieren keine Anomalien im Datensatz.

H_1 : Es existieren bis zu k Anomalien im Datensatz.

Im ESD Test ist die Angabe des Signifikanzniveaus α und eine obere Grenze k der vermuteten Anzahl an Ausreißern erforderlich. Der Test wird k mal iterativ durchgeführt und in jedem Schritt die Null Hypothese überprüft. Dazu wird folgender Prüfwert berechnet:

$$R_i = \frac{\max_i |x_i - \bar{x}|}{s} \quad i = 1, \dots, k \quad (3.9)$$

mit x_i dem i -ten Extremwert und s der Standardabweichung. Der errechnete Prüfwert wird mit dem kritischen Wert λ_i , resultierend aus Formel (3.10), verglichen:

$$\lambda_i = \frac{(n - i)t_{p, n-i-1}}{\sqrt{(n - i - 1 + t_{p, n-i-1}^2)(n - i + 1)}} \quad (3.10)$$

mit $t_{p, v}$ den $100 \cdot p$ Prozentsatz Punkt der t-Verteilung mit v Freiheitsgraden, n der

$$p = 1 - \frac{\alpha}{2(n - i + 1)} \quad (3.11)$$

Gesamtanzahl der Datenpunkte und α dem Signifikanzniveau, welcher die Wahrscheinlichkeit angibt, dass H_0 abgelehnt wird, obwohl die Null Hypothese richtig ist. Im Fall von $R_i > \lambda_i$ wird die Null Hypothese H_0 abgelehnt. Unabhängig von der Entscheidung über die Null Hypothese, wird bei jeder Iteration der Wert x_i , basierend auf der Berechnung von R_i , aus dem Datensatz entfernt und in der nächsten Iteration $n-1$ Datenpunkten verwendet. Anomalien sind somit alle von den Iterationen zuvor entfernten Datenpunkten bis einschließlich dem Datenpunkt, bei welchem die Null Hypothese verworfen wurde. Rosner (1983) erläutert die Bestimmung von Anomalien mittels ESD Test anhand eines Beispieldatensatzes und die publizierten Testergebnisse sind in Tabelle 3-10 für die weitere Erläuterung dargestellt.

Tabelle 3-10: Zusammenfassung der Ergebnisse des ESD Tests (Rosner 1983)

Nummer des Ausreißers i	Test Statistik Wert R_i	Kritischer Wert λ_i
1	3,118	3,158
2	2,942	3,151
3 *	3,179	3,143
4	2,810	3,136
5	2,815	3,128
6	2,848	3,120
7	2,279	3,111
8	2,310	3,103
9	2,101	3,094
10	2,067	3,085

In diesem beispielhaften ESD-Test ($k = 10$, $\alpha = 0.05$) ist der dritte Datenpunkt anomal, da $R > \lambda$ erfüllt wird. Die vorherliegenden Punkte $i = 1$ und $i = 2$ gelten ebenfalls als Anomalien. Es existieren somit insgesamt drei Anomalien in diesem Datensatz. In den Analysen dieser Masterarbeit werden die drei Signifikanzniveauwerte $\alpha = 0.05$, $\alpha = 0.10$ und $\alpha = 0.15$ als Schwellenwerte zur Identifikation von Anomalien verwendet.

Teilsequenzanalyse

Die Residuenanalyse mittels Z-Score oder ESD lässt sich sowohl auf einzelne Datenpunkte als auch auf Teilsequenzen anwenden. Formel (2.5) addiert z.B. Werte von einzelnen Residuen auf. Die Summenwerte lassen sich wiederum mit den zwei zuvor genannten Analysemethoden untersuchen, um so ungewöhnliche hohe summierte Residuenwerte identifizieren zu können. Es werden die Teilsequenzen der Länge zwei, vier,

sechs, acht Stunden in der Auswertung untersucht. Dabei handelt es sich jeweils um *gleitende Sequenzfenster*, d.h. bspw. für Sequenzen der Länge vier Stunden werden zunächst die Werte von Stunde eins, zwei, drei und vier als erste Sequenz summiert, dann die von Stunde zwei, drei, vier und fünf als zweite Sequenz usw..

4 Auswertung & Diskussion der Einzelergebnisse

Ziel der Arbeit ist es, zu prüfen, welches Verfahren sich eignet, um Anomalien in den Lastgängen von fünf Unternehmen, die stellvertretend für fünf Cluster der Gesamtzahl aller möglichen Unternehmen stehen, aufzudecken. Unter Berücksichtigung theoretischer Vorüberlegungen und der Datencharakteristiken wurde sich für zwei Strategien zur Erkennung von Anomalien in Lastgangdaten entschieden. *Strategie 1* nutzt den Machine-learning Algorithmus LSTM um ein Normalmodell ohne Anomalien zu schätzen. *Strategie 2* schätzt das Normalmodell mithilfe eines glättenden Verfahrens aus der statistischen Zeitreihenanalyse, dem PEWMA. Auf die Schätzung folgt in beiden Strategien zunächst die Untersuchung der Residuen auf Punkt- und dann auf Teilsequenzanomalien mittels Z-Score- und ESD-Test. Die Ergebnisse werden einzeln für die Lastgänge jedes zu analysierenden Unternehmens berichtet und die Eignung der jeweiligen Strategien zur Anomalieerkennung für die Einzelergebnisse diskutiert (vgl. die in Kapitel 2.3 spezifizierten Kriterien).

4.1 Strategie 1: Anomalien unter Normalmodell-Schätzung mit LSTM

Die Schätzung des Normalmodells für Anomalieerkennungsstrategie 1 erfolgt mittels LSTM Algorithmus. Dabei wird das Schätzmodell in zwei Varianten genutzt. Die erste Variante verwendet zur Bildung des Modells einen Zeitraum von einem Monat, im Weiteren mit 1M abgekürzt. Dieser entspricht dem Zeitraum, in dem Anomalien identifiziert werden sollen. In der zweiten Variante wird das Modell mit Daten aus drei Monaten (3M), d.h. zwei Vormonate und den für das Unternehmen aktuellen Monat, gebildet und daraus eine Schätzung für den zu untersuchenden Zeitraum von einem Monat erstellt. Für die zusätzliche Analyse unter Verwendung eines 3M-Schätzmodells wurde sich entschieden, da damit die Gefahr gegenüber dem 1M-Schätzmodell sinkt, mögliche Anomalien im Normalmodell zu schätzen und damit zu verschleiern.

4.1.1 Cluster 1 Unternehmen

Untersuchung von Einzelresiduen

Für die Analyse nach auffälligen Einzelresiduen des Unternehmenslastgangs aus Cluster 1 wurde zunächst das Normalmodell (1M) mittels LSTM Algorithmus geschätzt. Mit

Auswertung & Diskussion der Einzelergebnisse

einer Anpassungsgüte von $R^2 = 0.96$ erweist sich die Schätzung als geeignet, die beobachteten Werte vorherzusagen (zur Veranschaulichung s. Abbildung 4.1, oben). Um aus den Abweichungen der beobachteten von den geschätzten Werten (Residuendiagramm, s. Abbildung 4.1, unten) Anomalien zu identifizieren wurden im nächsten Schritt sowohl Z-Score Test als auch ESD Test ($k = 40$) durchgeführt. Die so identifizierten Anomalien wurden mit den Referenzanomalien der Experteneinschätzungen verglichen und den Zellen der Konfusionsmatrix (vgl. Tabelle 2-1) zugeordnet. Dabei zeigte sich, dass maximal 2 der 3 Referenzanomalien als richtig positiv identifiziert werden konnten, bei allerdings zugleich sehr hohen falsch positiven Zuordnungen (s. Tabelle 4-1).

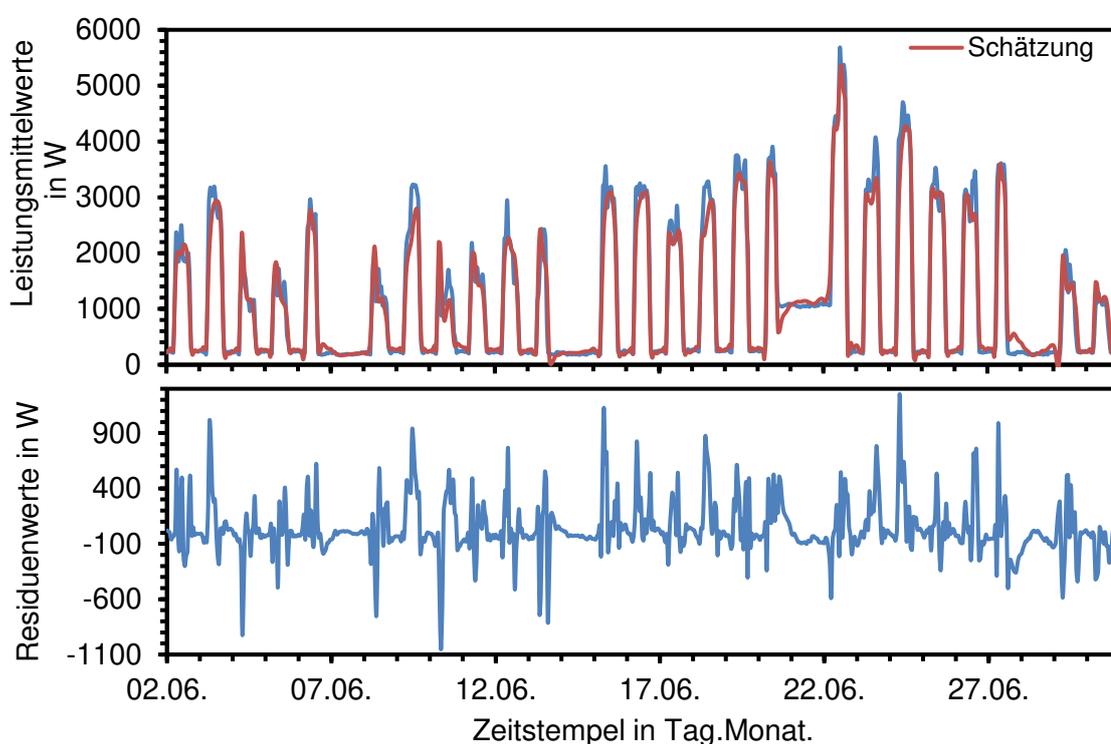


Abbildung 4.1: Darstellung (oben) der gegebenen Lastgang Daten, des Cluster 1 Unternehmens, sowie der Schätzung (1M) aus Daten eines Zeitraumes von einem Monat, und (unten) den resultierenden Residuenwerten

Die Analyse mit dem ESD Test bei $\alpha = 0,05$ zeigt mit $P = 0,14$ die höchste Genauigkeit, eine Trefferquote von $R = 0,33$ und mit $F_1 = 0,22$ das höchste F_1 -Maß, als das harmonische Mittel von Genauigkeit und Trefferquote. Für weniger strenge Signifikanzniveaus ($\alpha = 0,10$ und $0,15$) verschlechtern sich die Werte. Für die Analyse nach Z-Score-Schwellenwerten können zwar unter den beiden Schwellen von $Z = 2,0$ und $Z = 2,5$ zwei der drei Referenzanomalien als richtig positiv identifiziert werden ($R = 0,67$), gleichzeitig steigt allerdings die Zahl falsch positiver Zuordnungen erheblich auf 19 bzw. 34 an. Werden für die LSTM Schätzung 3 Monate einbezogen (Anpassungsgüte $R^2 = 0,95$),

ändern sich die Werte in ihrer Zusammensetzung, allerdings führt dies zu keiner Verbesserung der Erkennungsleistung richtig positiver Anomalien. Die Ergebnisse aller Schwellenwerte für Z-Score- und ESD-Test sowie die LSTM Schätzung 1M und 3M lassen sich Tabelle 4-1 entnehmen, sowie graphisch aufbereitet Abbildung 4.2.

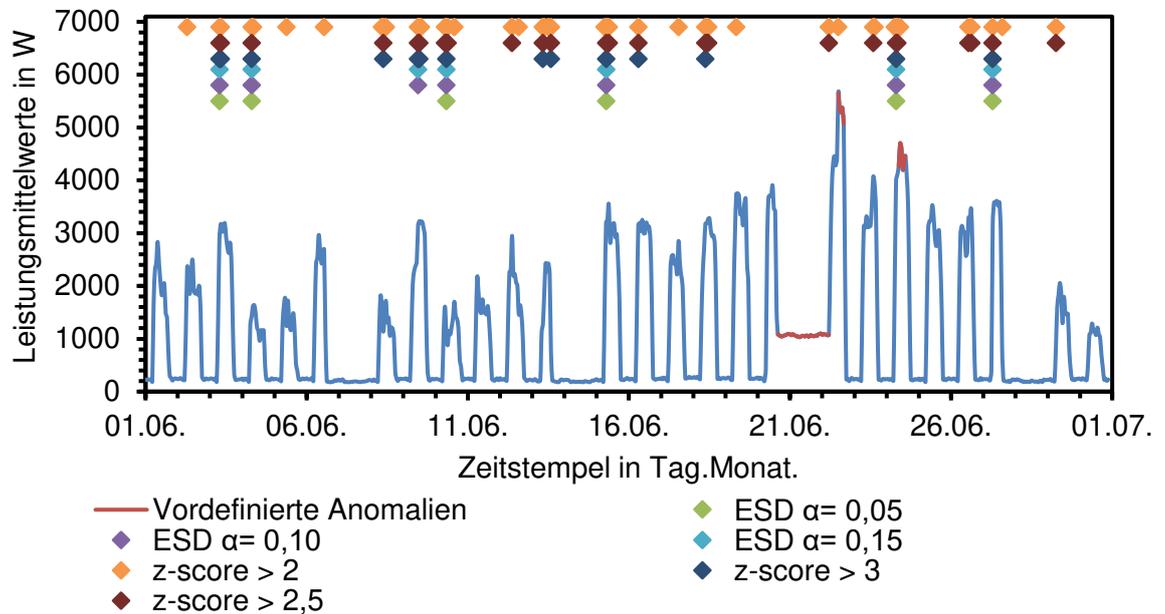


Abbildung 4.2: Darstellung der identifizierten Anomalien des Cluster 1 Unternehmens, resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$)

Tabelle 4-1: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 1 Unternehmens (1M) für drei Z-Score Schwellenwerte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$)

C1 (1M, $R^2 = 0,96$)	Genauigkeit (P)	Trefferquote (R)	F_1 -Maß	tp	fp	fn
ESD $\alpha = 0,05$	0,17	0,33	0,22	1	5	2
ESD $\alpha = 0,10$	0,14	0,33	0,20	1	6	2
ESD $\alpha = 0,15$	0,14	0,33	0,20	1	6	2
Z-Score > 3	0,08	0,33	0,13	1	11	2
Z-Score > 2,5	0,10	0,67	0,17	2	19	1
Z-Score > 2	0,06	0,67	0,10	2	34	1
C1 (3M, $R^2 = 0,95$)						
ESD $\alpha = 0,05$	0,09	0,33	0,14	1	10	2
ESD $\alpha = 0,10$	0,08	0,33	0,13	1	12	2
ESD $\alpha = 0,15$	0,13	0,67	0,21	2	14	1
Z-Score > 3	0,13	0,67	0,21	2	14	1
Z-Score > 2,5	0,08	0,67	0,14	2	23	1
Z-Score > 2	0,05	0,67	0,10	2	35	1

Auswertung & Diskussion der Einzelergebnisse

Übergreifend lässt sich erkennen, dass das Schätzmodell (1M & 3M) nicht in der Lage ist jede Spitze im Lastgangverlauf zu modellieren, sodass hohe Residuenwerte auch außerhalb der vordefinierten anomalen Bereiche existieren, siehe dazu für Variante 1M Abbildung 4.1 (unten) und dadurch auch viele falsch positive Ergebnisse existieren.

Untersuchung von summierten Residuen

Für die Analyse Unternehmenslastgangs aus Cluster 1 nach *Teilsequenzanomalien*, mittels summierter Residuen, wurden die gleichen Normalmodell-Schätzungen (1M & 3M) mithilfe des LSTM Algorithmus wie bei der Analyse der Einzelresiduen zugrunde gelegt. Diesmal wurden die Abweichungen der geschätzten von den beobachteten Werten (Residuen) zu Teilsequenzen aufsummiert. Dabei wurden zwei Stunden (sum2), vier Stunden (sum4), sechs Stunden (sum6) oder acht Stunden (sum8) als Teilsequenzlängen gesetzt. Tabelle 4-2 zeigt die Ergebnisse der Analyse nach Anomalien für das 1M- und das 3M-Schätzmodell.

Tabelle 4-2: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 1 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang A Tabelle 1 & 2 für eine Auflistung der Werte für tp , fp und fn .

	C1 (1M, $R^2 = 0,96$)			C1 (3M, $R^2 = 0,95$)		
	<i>P</i>	<i>R</i>	F_1	<i>P</i>	<i>R</i>	F_1
ESD $\alpha = 0,05$ sum4	0,25	0,33	0,29	0	0	0
ESD $\alpha = 0,05$ sum2	0,10	0,33	0,15	0	0	0
ESD $\alpha = 0,10$ sum6	1,00	0,33	0,50	1,00	0,33	0,50
ESD $\alpha = 0,10$ sum4	0,20	0,33	0,25	0	0	0
ESD $\alpha = 0,10$ sum2	0,08	0,33	0,13	0	0	0
ESD $\alpha = 0,15$ sum8	1,00	0,33	0,50	0	0	0
ESD $\alpha = 0,15$ sum6	1,00	0,33	0,50	1,00	0,33	0,50
ESD $\alpha = 0,15$ sum4	0,20	0,33	0,25	0,17	0,33	0,22
ESD $\alpha = 0,15$ sum2	0,08	0,33	0,13	0	0	0
Z-Score > 3 sum8	0,20	0,33	0,25	0,08	0,33	0,13
Z-Score > 3 sum6	0,11	0,33	0,17	0,09	0,33	0,14
Z-Score > 3 sum4	0,08	0,33	0,13	0,08	0,33	0,13
Z-Score > 3 sum2	0,08	0,33	0,13	0,07	0,33	0,12
Z-Score > 2,5 sum8	0,06	0,33	0,11	0,11	0,67	0,19
Z-Score > 2,5 sum6	0,05	0,33	0,09	0,11	0,67	0,19
Z-Score > 2,5 sum4	0,04	0,33	0,08	0,10	0,67	0,17
Z-Score > 2,5 sum2	0,04	0,33	0,07	0,05	0,33	0,08
Z-Score > 2 sum8	0,05	0,67	0,09	0,05	0,67	0,10
Z-Score > 2 sum6	0,05	0,67	0,09	0,04	0,67	0,08
Z-Score > 2 sum4	0,03	0,33	0,05	0,07	1,00	0,14
Z-Score > 2 sum2	0,07	1,00	0,13	0,06	1,00	0,12

Auswertung & Diskussion der Einzelergebnisse

Es zeigt sich, dass für alle getesteten Fälle mindestens eine Referenzanomalie als richtig positiv erkannt wurde ($R > 0,33$). Diesmal sind auch Trefferquoten von $R = 1,0$ zu beobachten (M1-Schätzung mit $Z > 2$ und sum2 bzw. M3-Schätzung mit $Z > 2$ und sum2 bzw. sum4) – d.h. alle drei Referenzanomalien konnten unter Verwendung bestimmter Analyseparameter richtig positiv erkannt werden. Allerdings geht die hohe Trefferquote auch hier wieder mit einer hohen Anzahl falsch positiver Zuordnungen einher, sodass die Genauigkeit dabei sehr niedrig ausfällt ($P = 0,03$ bis $0,07$). Der ESD-Test zeigt zwar z.T. höhere Genauigkeitswerte mit $P = 1,0$ als die Tests mithilfe des Z-Scores, gleichzeitig wird aber nur maximal eine der drei Referenzanomalien richtig positiv erkannt ($R = 0,33$). Beim Vergleich der F_I -Werte von Einzelresiduen und summierter Residuenwerte lässt sich eine verbesserte Identifikation der Anomalien im Vergleich von $Z > 2,5$ ($F_I = 0,14$, 3M, Einzelresiduen) zu $Z > 2,5$ ($F_I \geq 0,17$, 3M, sum4, sum6, sum8) sowie $Z > 2$ ($F_I = 0,10$, 3M, Einzelresiduen) zu $Z > 2$ ($F_I \geq 0,12$, 3M, sum2, sum4) erkennen.

Bei den durchgeführten ESD Tests zeigt sich, dass Summierungslängen von 8 Stunden, auch im Vergleich mit den Z-Score Test, in der Regel null richtig positive Ergebnisse ergeben, außer beim Schwellenwert ESD $\alpha = 0,15$ (sum8, 1M). Daher werden die Einzelresiduen und die vier summierten Teilsequenzlängen des Cluster 1 Unternehmens der Variante 3M exemplarisch untersucht, siehe dazu Abbildung 4.3.

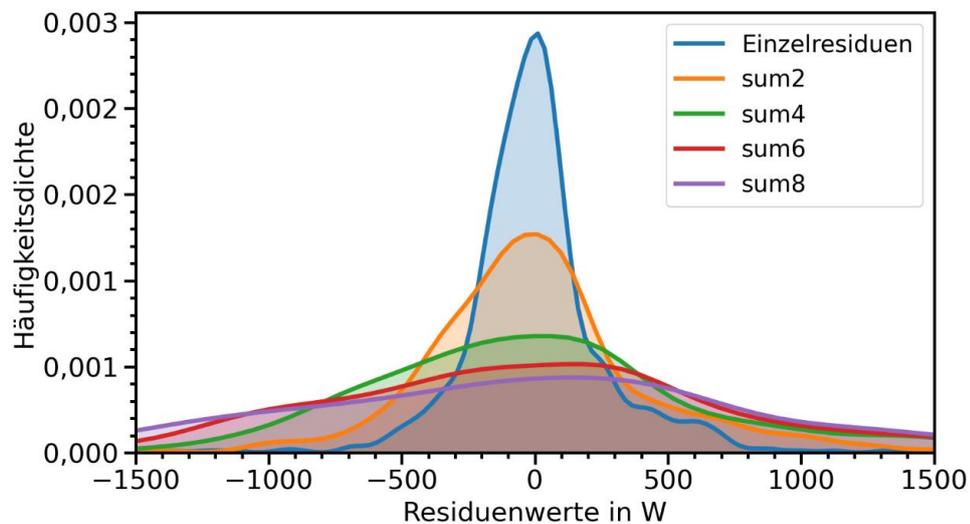


Abbildung 4.3: Häufigkeitsdichte der Einzelresiduen und der vier summierten Teilsequenzlängen des Cluster 1 Unternehmens (3M)

Die Verteilung der Residuenwerte ändert sich mit der Länge der summierten Teilsequenzlängen, d.h. je länger die Summationslänge, desto flacher, asymmetrischer und schwach bimodal (sum8) wird Verteilung. In einer übergreifenden Diskussion wird die Anwendbarkeit des ESD Test noch einmal erörtert, da Verteilungen die zu stark von der

Auswertung & Diskussion der Einzelergebnisse

Normalverteilung abweichen, kein aussagekräftiges Ergebnis eines ESD Test zulassen (siehe Kapitel 5.1 *Fragestellung (1)*). Die Häufigkeitsverteilung der summierten Teilsequenzlängen, und auch der Einzelresiduen, hängt generell stark von der erhaltenden Schätzung ab, sodass der ESD Test im weiteren Verlauf der Auswertung angewendet wird, um einen Überblick über alle getesteten Unternehmen und auch Strategieansätze zu erhalten.

Neben der Häufigkeitsverteilung, die im besonderen Maße die ESD Test beeinflusst, wird nun die Auswirkung der verschiedene Summationslängen auf die Ergebnisse von Tabelle 4-2 untersucht, siehe dazu Abbildung 4.4.

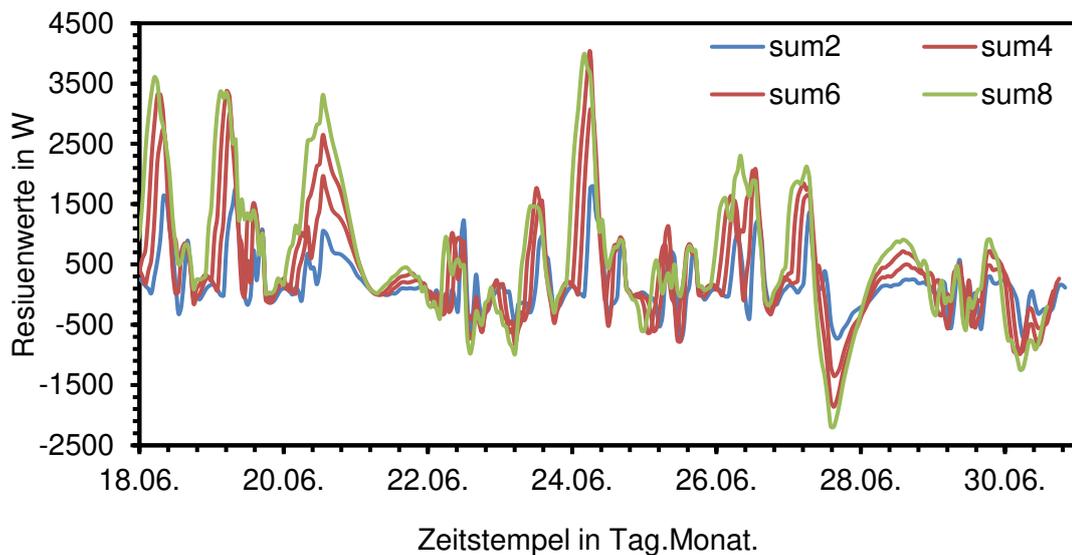


Abbildung 4.4: Darstellung der vier aufsummierten Residuenlängen für einen ausgewählten Zeitraum des Monats resultierend aus der Schätzung (3M)

Zwei Effekte lassen sich erkennen, welche durch die Länge der aufsummierten Residuen beeinflusst wird. Zum einen werden, je nach Länge der Summation Residuenverläufe in bestimmten Zeitabschnitten im Vergleich zu anderen Residuenverläufen verstärkt oder abgeschwächt werden, wie beispielsweise der Residuenverlauf vom 20.06 bis zum 20.06 (dritter Teilstrich) eine Verstärkung erfährt, d.h. die Residuenwerte in diesem Bereich erhöhen sich im größeren Maße als in anderen zeitlichen Abschnitte. Dadurch kann dieser Bereich beim Z-Score Test oder ESD Test eher als anomal eingestuft werden und die Beurteilungsmetriken zur Anomalieerkennung beeinflussen. Zum anderen verschiebt sich der Maximalpunkt eines Residuenverlauf eines gewissen Zeitabschnittes, je nach Summationslänge um wenige Stunden, siehe dazu bspw. den Maximalwert der summierten Residuen am 24.06 (erster Teilstrich). Daher wird im folgenden Verlauf der Auswertung nicht jedes Ergebnis eines Z-Score- oder ESD-Test

analysiert, da sie hauptsächlich durch die beiden angesprochenen Effekte die Anomalieerkennung beeinflussen.

4.1.2 Cluster 2 Unternehmen

Untersuchung von Einzelresiduen

Für das untersuchte Unternehmen aus Cluster 2 identifizierten die konsultierten Experten 12 Referenzanomalien. Zwei der Referenzanomalien fallen durch besonders hohe Leistungswerte auf (09.04. & 11.04.). Sowohl die Analysen mittels Z-Score- und ESD-Test zeigen sich in der Lage diese beiden Referenzanomalien unter allen möglichen Bedingungsvariationen richtig positiv ($R = 0,17$) zu erkennen (siehe dazu Abbildung 4.5 und Tabelle 4-3). Im Fall der LSTM-Schätzung aus Daten von einem Monat ($R^2 = 0.87$) können maximal 5 der 12 Referenzanomalien richtig positiv erkannt werden (1M $Z > 2$, $P = 0,15$, $R = 0,42$, $F_1 = 0,22$), bei dem Abgleich mit der LSTM-Schätzung aus Daten von drei Monaten ($R^2 = 0,92$) werden maximal 6 der 12 Referenzanomalien richtig positiv erkannt (3M, $Z > 2$, $P = 0,16$, $R = 0,50$, $F_1 = 0,24$).

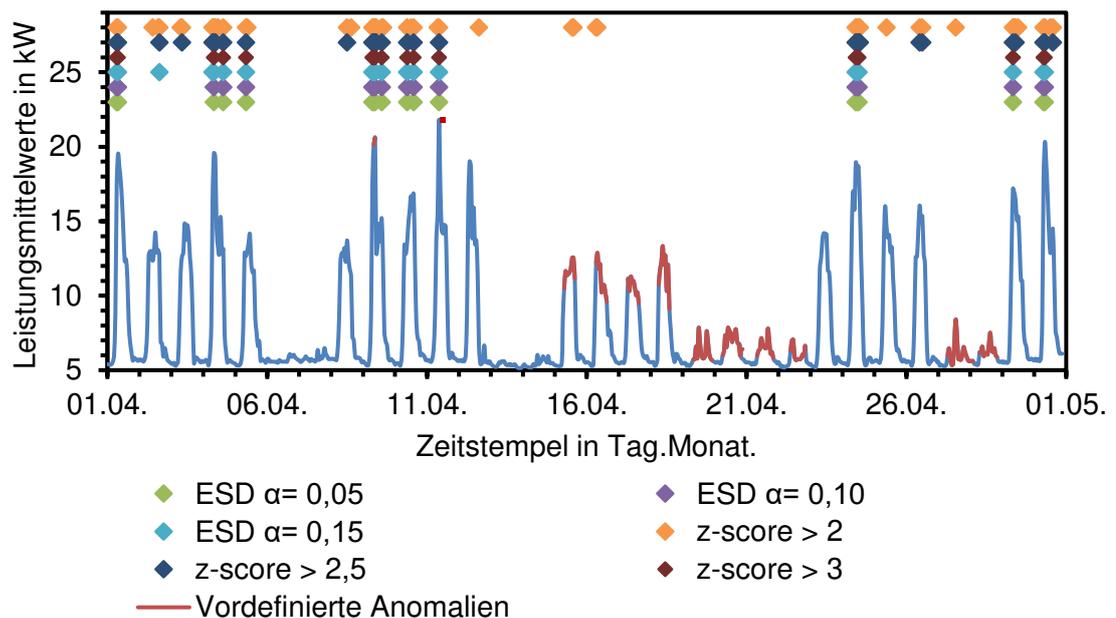


Abbildung 4.5: Darstellung der identifizierten Anomalien des Cluster 2 Unternehmens in der Variante 1M, resultierend aus den drei Z-Score Werten und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$)

Auswertung & Diskussion der Einzelergebnisse

In der Variante 3M erhöhen sich die F_1 -Maße für alle ausgewerteten Alpha- und Z-Score-Werte. Da das Modell über den Zeitraum von drei Monaten trainiert wurde, entstehen bei der Schätzung des zu untersuchenden Monats, in den Tagen vom 15.04 bis 18.04 und dem Wochenende des 27. und 28. Aprils, bei dem Vergleich zu den Originaldaten größere Residuenwerte, als in der Variante 1M. Dadurch erhöht sich die Anomalieerkennungsgüte.

Tabelle 4-3: Berechnete Genauigkeiten (P), Trefferquoten (R) und F_1 -Maße des Cluster 2 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang B Tabelle 1 für eine Auflistung der Werte für tp , fp und fn .

	C2 (1M, $R^2 = 0,87$)			C2 (3M, $R^2 = 0,92$)		
	<i>P</i>	<i>R</i>	F_1	<i>P</i>	<i>R</i>	F_1
ESD $\alpha = 0,05$	0,13	0,17	0,15	0,33	0,25	0,29
ESD $\alpha = 0,10$	0,13	0,17	0,15	0,25	0,25	0,25
ESD $\alpha = 0,15$	0,13	0,17	0,14	0,21	0,25	0,23
Z-Score > 3	0,15	0,17	0,16	0,23	0,25	0,24
Z-Score > 2,5	0,10	0,17	0,12	0,13	0,25	0,17
Z-Score > 2	0,15	0,42	0,22	0,16	0,50	0,24

Untersuchung von summierten Residuen

Tabelle 4-4 stellt die Ergebnisse der Auswertung verschiedener Teilsequenzlängen für die Variante 1M und Variante 3M für den Lastgang des Unternehmens aus Cluster 2 dar. Es fällt auf, dass die F_1 -Maße in allen Fällen bei 1M und 3M im Vergleich zu den F_1 Werten bei der Untersuchung von Punktanomalien niedrigere Werte annehmen. Diese Ergebnisse resultieren aus dem Umstand, dass das Schätzmodell nicht in jedem Fall im Leistungsmaximalbereich des Tagesverlaufs die Lastgangspitzen nachbilden kann. Wenn der LSTM Algorithmus relativ lange Verläufe im Bereich der Maximalwerte eines Tages nicht ausreichend gut modellieren kann, kommt es bei der Summierung der Residuenwerte zu höheren Werten, als bei schmalen und abrupten Lastgangspitzen. Somit resultieren an vielen weiteren Tagen prominente Residuenwerte und diese erschweren die Identifikation der vordefinierten Anomalien.

Tabelle 4-4: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 2 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang B Tabelle 2 & 3 für eine Auflistung der Werte für tp , fp und fn .

	C2 (1M, $R^2 = 0,87$)			C2 (3M, $R^2 = 0,92$)		
	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$ sum4	0,06	0,08	0,07	0,09	0,08	0,09
ESD $\alpha = 0,05$ sum2	0	0	0	0,10	0,17	0,12
ESD $\alpha = 0,10$ sum4	0,06	0,08	0,07	0,08	0,08	0,08
ESD $\alpha = 0,10$ sum2	0	0	0	0,10	0,17	0,12
ESD $\alpha = 0,15$ sum6	0,05	0,08	0,06	0	0	0
ESD $\alpha = 0,15$ sum4	0,06	0,08	0,07	0,08	0,08	0,08
ESD $\alpha = 0,15$ sum2	0	0	0	0,14	0,25	0,18
Z-Score > 3 sum8	0	0	0	0,17	0,08	0,11
Z-Score > 3 sum6	0,10	0,08	0,09	0,08	0,08	0,08
Z-Score > 3 sum4	0,06	0,08	0,07	0,07	0,08	0,08
Z-Score > 3 sum2	0,07	0,08	0,08	0,07	0,08	0,07
Z-Score > 2,5 sum8	0,05	0,08	0,06	0,05	0,08	0,06
Z-Score > 2,5 sum6	0,08	0,17	0,11	0,09	0,17	0,12
Z-Score > 2,5 sum4	0,07	0,17	0,10	0,11	0,25	0,15
Z-Score > 2,5 sum2	0,09	0,25	0,13	0,11	0,33	0,17
Z-Score > 2 sum8	0,05	0,20	0,08	0,10	0,33	0,15
Z-Score > 2 sum6	0,05	0,22	0,09	0,11	0,42	0,18
Z-Score > 2 sum4	0,08	0,27	0,12	0,10	0,33	0,15
Z-Score > 2 sum2	0,12	0,42	0,18	0,11	0,42	0,18

4.1.3 Cluster 3 Unternehmen

Untersuchung von Einzelresiduen

Für den Lastgang des Unternehmens aus Cluster 3 identifizierten die Experten drei Referenzanomalien. Die Z-Score Tests unter allen Bedingungsvariationen sind in der Lage 2 der 3 Referenzanomalien als richtig positiv zu identifizieren. Der letzte im Monat vordefinierte anomale Bereich lässt sich bspw. für die Variante 1M & 3M zuverlässig identifizieren ($R > 0,33$), siehe dazu Abbildung 4.6 am 24.06 zw. dem zweiten und dritten Teilstrich. Bei dieser Anomalie treten 2 Lastgangspitzen im Tagesverlauf in sehr kurzer Zeit auf und können durch das Schätzmodell nur unzureichend nachgebildet werden, sodass hier hohe Residuenwerte entstehen. Bei einem Schwellenwert von $Z > 3$ zeigen sich hier die besten Ergebnisse. So liegt für das Schätzmodell aus einem Monat eine Genauigkeit von $P = 0,20$, eine Trefferquote von $R = 0,67$ und ein F_1 Wert von 0,31 und für das 3M Modell eine Genauigkeit von $P = 0,29$, eine Trefferquote von $R = 0,67$ und ein F_1 Wert von 0,40 vor. Weiter zeigt sich, dass der Abgleich der beobachteten Daten

Auswertung & Diskussion der Einzelergebnisse

mit dem Normalmodell aus 3 Monaten insgesamt höhere F_1 -Werte erzielt. Das 3M-Normalmodell scheint also besser geeignet richtig positive Anomalien sichtbar zu machen als das 1M-Normalmodell, vermutlich, weil das 1M-Modell Abweichungen als „normal“ nachbildet, die über die Hinzunahme weiterer Monate zur Schätzung nicht mehr als normal gelten können. Tabelle 4-5 fasst die Ergebnisse der Untersuchung von Punktanomalien für die Variante 1M & 3M zusammen, Abbildung 4.6 veranschaulicht sie graphisch.

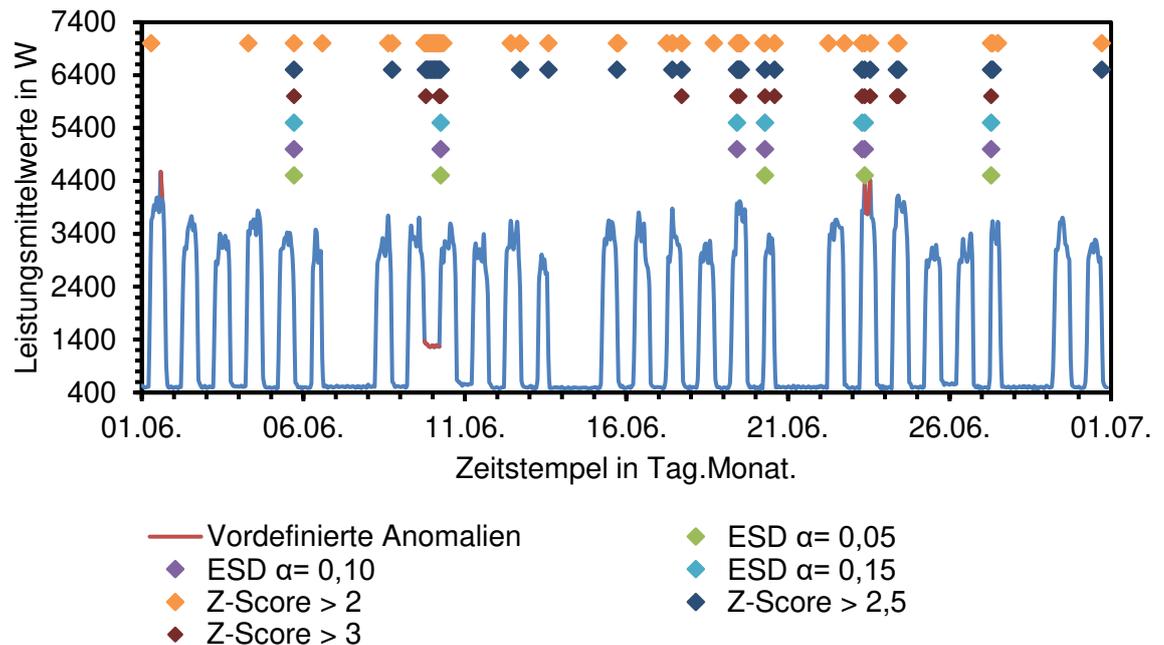


Abbildung 4.6: Darstellung der identifizierten Anomalien des Cluster 3 Unternehmens in der Variante 1M, resultierend aus den drei Z-Score Werten und den drei Signifikanzniveauewerten des ESD Tests ($k = 40$)

Tabelle 4-5: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 3 Unternehmens (1M) für drei Z-Score Werten und den drei Signifikanzniveauewerten des ESD Tests ($k = 40$). Siehe Anhang C Tabelle 1 für eine Auflistung der Werte für tp , fp und fn .

	C3 (1M, $R^2 = 0,92$)			C3 (3M, $R^2 = 0,98$)		
	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$	0,17	0,33	0,22	1,00	0,33	0,50
ESD $\alpha = 0,10$	0,17	0,33	0,22	1,00	0,33	0,50
ESD $\alpha = 0,15$	0,20	0,33	0,25	0,50	0,33	0,40
Z-Score > 3	0,20	0,67	0,31	0,29	0,67	0,40
Z-Score > 2,5	0,10	0,67	0,17	0,11	0,67	0,19
Z-Score > 2	0,06	0,67	0,11	0,07	0,67	0,12

Untersuchung von summierten Residuen

Tabelle 4-6 zeigt die Ergebnisse der Analyse des Unternehmenslastgangs von Cluster 3 auf Teilsequenzanomalien. Es werden, analog zur Untersuchung der Einzelresiduen, maximal 2 der 3 Referenzanomalien ($R = 0,67$) richtig positiv identifiziert. Bei der Analyse basierend auf dem 1M-Modell und unter Verwendung des Z-Score-Test trifft dies auf alle Bedingungsvariationen zu. Hier werden Genauigkeitswerte von $P = 0,08$ bis $P = 0,22$ erreicht, die Trefferquote liegt in allen Fällen bei $R = 0,67$ und der F_1 Wert variiert von 0,14 bis 0,33. Abgesehen von $Z > 3$ sum6 und sum8 gilt dies ($R = 0,67$) auch für alle Bedingungsvariationen des Z-Score-Tests basierend auf dem 3M-Modell, wobei hier die Genauigkeit zwischen $P = 0,07$ und $P = 0,50$ liegt und die F_1 Werte von 0,13 bis 0,50 variieren. Für die Analysen unter Verwendung des ESD-Tests zeigen sich deutlich schlechtere Trefferquoten von maximal $R = 0,33$ (ein Drittel der Referenzanomalien, richtig positiv) unter Verwendung des 1M Modell mit $\alpha = 0,15$ und sum4 bzw. sum2. In beiden Fällen liegt eine Genauigkeit von $P = 0,25$ und eine F_1 Wert von 0,36 vor. Weiterhin zeigt sich im Vergleich von der Einzelresiduenanalyse zur Untersuchung der summierten Residuen, für $Z > 2$ ($F_1 = 0,12$, Einzelresiduen; $F_1 \geq 0,13$, alle Sequenzlängen), $Z > 2,5$ ($F_1 = 0,19$, Einzelresiduen; $F_1 \geq 0,25$, sum2, sum4, sum6) sowie $Z > 3$ ($F_1 = 0,40$, Einzelresiduen; $F_1 \geq 0,50$, sum2, sum4), dass bei Variante 3M eine erhöhte Anomalieerkennungsgüte für ausgewählte summierte Residuen vorliegt.

Bei genauerer Betrachtung der Referenzanomalien fällt auf, dass die erste Referenzanomalie des Monats in keinem Testfall identifiziert wird. Diese erste Referenzanomalie zeigt sich als vergleichsweise kurze Zeitsequenz. Dadurch entstehen bei der Summierung zu Residuensequenzen kleinere Werte als bei der Summierung von Residuen über längere Zeiträume, wie z.B. am 05.06, 19.06 und 20.06. Somit kann in der Variante 3M mit kleinen Summationslängen und einem Schwellenwerte von $Z > 3$ das höchste F_1 -Maß von 0,57 (sum2) und mit $Z > 2,5$, das höchste F_1 -Maß für den gewählten Schwellenwert, von 0,27 (sum4) erreicht werden.

Tabelle 4-6: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 3 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang C Tabelle 2 & 3 für eine Auflistung der Werte für tp , fp und fn .

	C3 (1M, $R^2 = 0,92$)			C3 (3M, $R^2 = 0,98$)		
	P	R	F_1	P	R	F_1
ESD $\alpha = 0,10$ sum4	0	0	0	0,50	0,33	0,40
ESD $\alpha = 0,10$ sum2	0,14	0,33	0,20	1,00	0,33	0,50
ESD $\alpha = 0,15$ sum4	0,25	0,67	0,36	0,50	0,33	0,40
ESD $\alpha = 0,15$ sum2	0,25	0,67	0,36	0	0	0
Z-Score > 3 sum8	0,22	0,67	0,33	0,50	0,33	0,40
Z-Score > 3 sum6	0,20	0,67	0,31	0,20	0,33	0,25
Z-Score > 3 sum4	0,22	0,67	0,33	0,40	0,67	0,50
Z-Score > 3 sum2	0,20	0,67	0,31	0,50	0,67	0,57
Z-Score > 2,5 sum8	0,13	0,67	0,21	0,06	0,33	0,11
Z-Score > 2,5 sum6	0,14	0,67	0,24	0,15	0,67	0,25
Z-Score > 2,5 sum4	0,17	0,67	0,27	0,17	0,67	0,27
Z-Score > 2,5 sum2	0,14	0,67	0,24	0,15	0,67	0,25
Z-Score > 2 sum8	0,09	0,67	0,15	0,09	0,67	0,15
Z-Score > 2 sum6	0,09	0,67	0,16	0,10	0,67	0,17
Z-Score > 2 sum4	0,11	0,67	0,18	0,10	0,67	0,17
Z-Score > 2 sum2	0,08	0,67	0,14	0,07	0,67	0,13

4.1.4 Cluster 4 Unternehmen

Untersuchung von Einzelresiduen

Der Lastgang des Unternehmens aus Cluster 4 weist 7 Referenzanomalien auf. Die Ergebnisse der Analysen finden sich in Abbildung 4.7 und Tabelle 4-8. Es zeigt sich, dass maximal drei Referenzanomalien als richtig positiv erkannt werden ($R = 0,43$). Dies gilt unter Verwendung des 1M-Modell im Z-Score Test mit den Schwellenwerten $Z > 2,5$ und $Z > 2$ bei einer Genauigkeit von $P = 0,17$ bzw. $P = 0,08$ und einem F_1 -Wert von 0,24 bzw. 0,13. Der ESD-Test schafft es unter allen Bedingungsvariationen maximal 1 von 7 Referenzanomalien richtig positiv einzuordnen ($R = 0,14$, mit $P = 0,33$ bis 0,50 und $F_1 = 0,2$ bis 0,22).

Vier der sieben Referenzanomalien konnten in keinem Fall als solche erkannt werden, weder unter Verwendung der 1M-Schätzung noch der 3M-Schätzung. Die Ursache für diesen Umstand lässt sich mithilfe von Abbildung 4.8 erklären: Die im Zeitintervall vom 18.07. bis zum 31.07. gezeigte Schätzung (1M) bildet den Lastgangverlauf ausreichend gut ab, so dass die resultierenden kleinen Residuen entgegen der

Auswertung & Diskussion der Einzelergebnisse

Referenzeinschätzungen nicht mehr als Anomalien erkannt werden können. Gleiches gilt unter Verwendung der 3M-Schätzung.

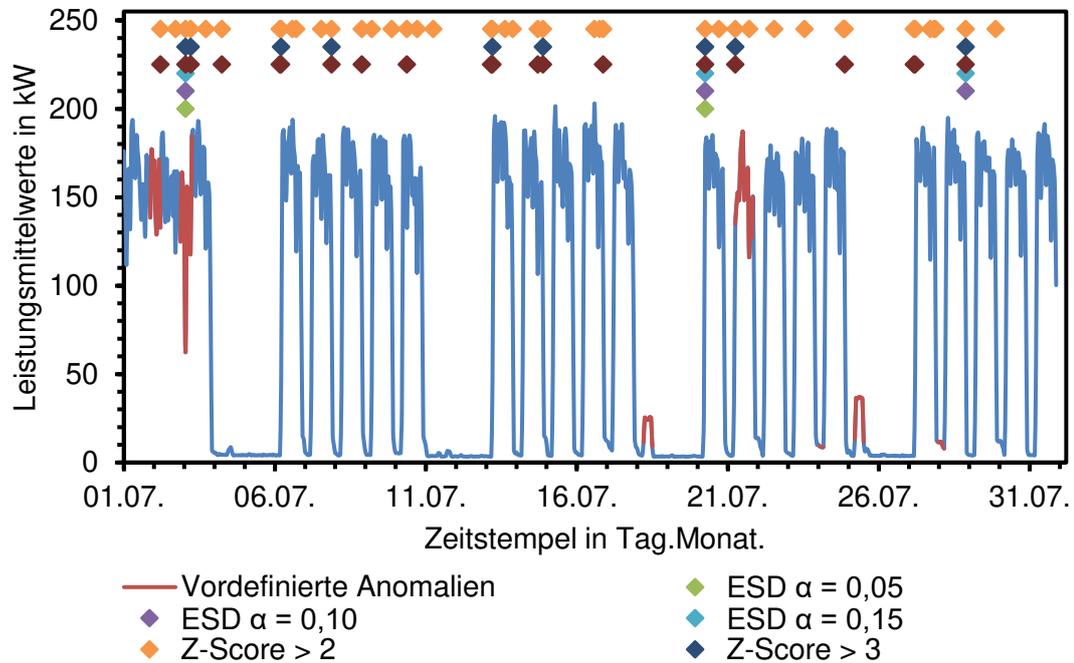


Abbildung 4.7: Darstellung der identifizierten Anomalien des Cluster 4 Unternehmens in der Variante 1M, resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$)

Tabelle 4-7: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 4 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang D Tabelle 1 für eine Auflistung der Werte für tp , fp und fn .

C4 (1M, $R^2 = 0,93$)	P	R	F_1
ESD $\alpha = 0,05$	0,50	0,14	0,22
ESD $\alpha = 0,10$	0,33	0,14	0,20
ESD $\alpha = 0,15$	0,33	0,14	0,20
Z-Score > 3	0,25	0,29	0,27
Z-Score > 2,5	0,17	0,43	0,24
Z-Score > 2	0,08	0,43	0,13
C4 (3M, $R^2 = 0,83$)			
Z-Score > 2	0,02	0,14	0,03

Auswertung & Diskussion der Einzelergebnisse

Bei der Detailbetrachtung der 3M-Schätzung zeigt sich darüber hinaus, dass Zeitintervalle zwischen den Werktagen schlecht nachgebildet werden können, vermutlich wegen des Lastgangverlaufs zwischen dem 01.07. und dem 04.07., der in den zwei Vormonaten mehrmals auftritt. Die Anpassungsgüte liegt folglich bei einem vergleichsweise geringen Wert von $R^2 = 0.83$. Es entstehen somit in der Zeit zwischen den Werktagen relativ hohe Residuenwerte und die Referenzanomalien werden schwer als anomal identifiziert. Abbildung 4.8 (unten) veranschaulicht weiter, dass relativ große Residuen im Bereich der Tagesmaximalwerte entstehen, viele geschätzten Werte weichen relativ weit von den beobachteten Werten ab, wodurch eine Identifikation der Referenzanomalien erschwert wird. Weiter zeigen sich große Residuen zu Zeiten, beispielsweise in Abbildung 4.8 am 20.07. zw. dem ersten und zweiten Teilstrich bzw. eine Woche später in einer schwächer ausgeprägten Form, in denen der Lastgangverlauf sehr stark ansteigt. Vermutlich beginnt hier die Kernarbeitszeit in den Unternehmen. Das Schätzmodell kann sich diesem Verlauf nur unzureichend anpassen – es entstehen große Residuen und damit vermeintliche Anomalien, obwohl in Wahrheit keine vorliegen. Auch in Abbildung 4.7 lässt sich erkennen, dass an Zeitpunkten bei welchem der Lastgang stark ansteigt und abfällt, also am Anfang und Ende der Kernarbeitszeiten, und insbesondere an Montagen, gehäuft Anomalien identifiziert werden.

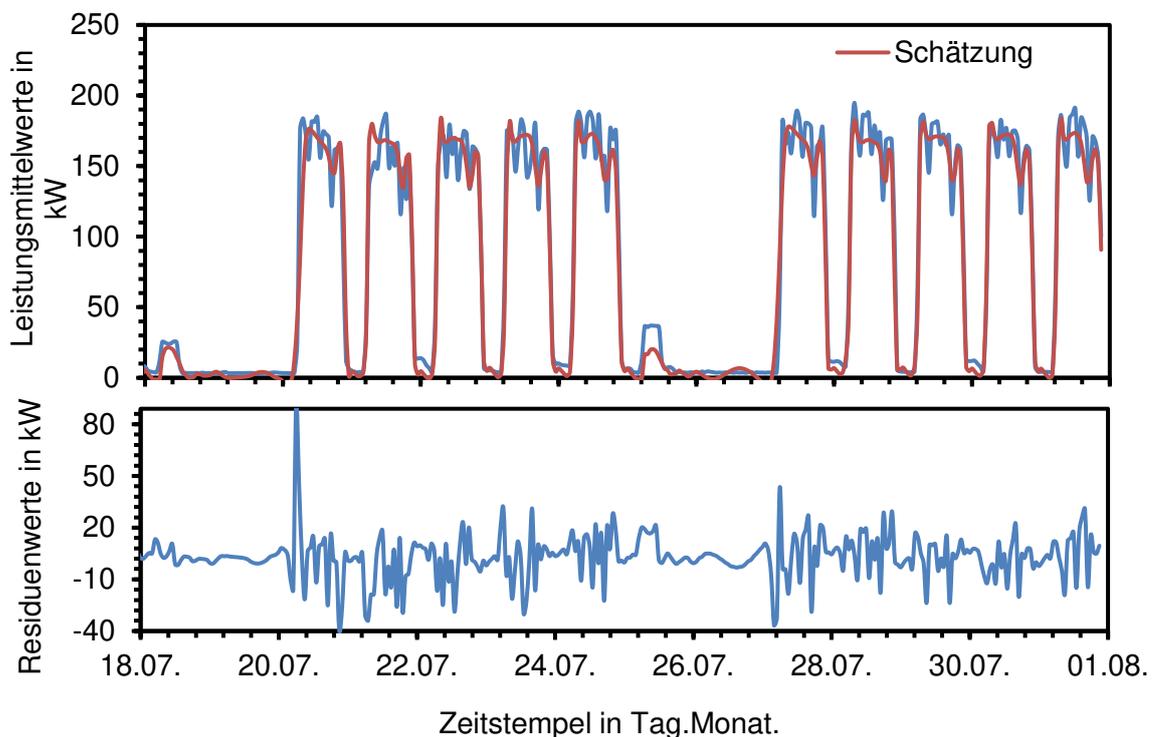


Abbildung 4.8: Darstellung (oben) der Schätzung und gegebenen Daten des Cluster 4 Unternehmens in der Variante 1M sowie (unten) die resultierenden Residuenwerte

Untersuchung von summierten Residuen

Die Auswertungsergebnisse sind in Tabelle 4-8 zusammengefasst. In Variante 1M liegen alle F_1 -Maße der summierten Residuen im Bereich $0,11 < F_1 < 0,25$ und zeigen vergleichbare Ergebnisse zur Einzelresiduenanalyse ($0,13 < F_1 < 0,27$). In Variante 3M liegen die F_1 -Maße der summierten Residuen im Wertebereich $0,05 < F_1 < 0,10$ und in der Untersuchung der Einzelresiduen bei $F_1 = 0,03$ ($Z > 2$). Die, bei der Untersuchung der Einzelresiduen, erläuterten Schwierigkeiten beeinflussen in gleicher Weise die Ergebnisse der Untersuchung verschiedener Teilsequenzlängen. Da hohe Residuenwerte an Zeitpunkten entstehen, die nicht im Bereich der vordefinierten Anomalien existieren, kann eine Summation auch zu keiner verbesserten Wiedererkennung der vordefinierten Anomalien führen.

Tabelle 4-8: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 4 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang D Tabelle 2 für eine Auflistung der Werte für tp , fp und fn .

	C4 (1M, R ² = 0,93)			C4 (3M, R ² = 0,83)		
	P	R	F ₁	P	R	F ₁
Z-Score > 3 sum2	0,22	0,29	0,25	0	0	0
Z-Score > 2,5 sum6	0,33	0,14	0,20	0	0	0
Z-Score > 2,5 sum4	0,09	0,14	0,11	0,07	0,14	0,10
Z-Score > 2,5 sum2	0,15	0,29	0,20	0,05	0,14	0,07
Z-Score > 2 sum8	0,14	0,14	0,14	0,04	0,14	0,06
Z-Score > 2 sum6	0,09	0,14	0,11	0,04	0,14	0,06
Z-Score > 2 sum4	0,12	0,29	0,17	0,06	0,14	0,09
Z-Score > 2 sum2	0,10	0,29	0,15	0,03	0,14	0,05

4.1.5 Cluster 5 Unternehmen

Untersuchung von Einzelresiduen

Das Unternehmen aus Cluster 5 besitzt drei vordefinierte Anomalien. Im Vergleich zu den Unternehmen der anderen Cluster ist kein zeitlich kontextualer Rhythmus im Lastgang erkennbar, d.h. es existieren keine Werktag-Wochenende- oder Tag-Nacht-Abhängigkeit. Daher kann das Schätzmodell in der Variante 1M und 3M nur Anpassungsgütwerte von $R^2 = 0,58$ (1M) und $R^2 = 0,62$ (3M) erreichen.

Auswertung & Diskussion der Einzelergebnisse

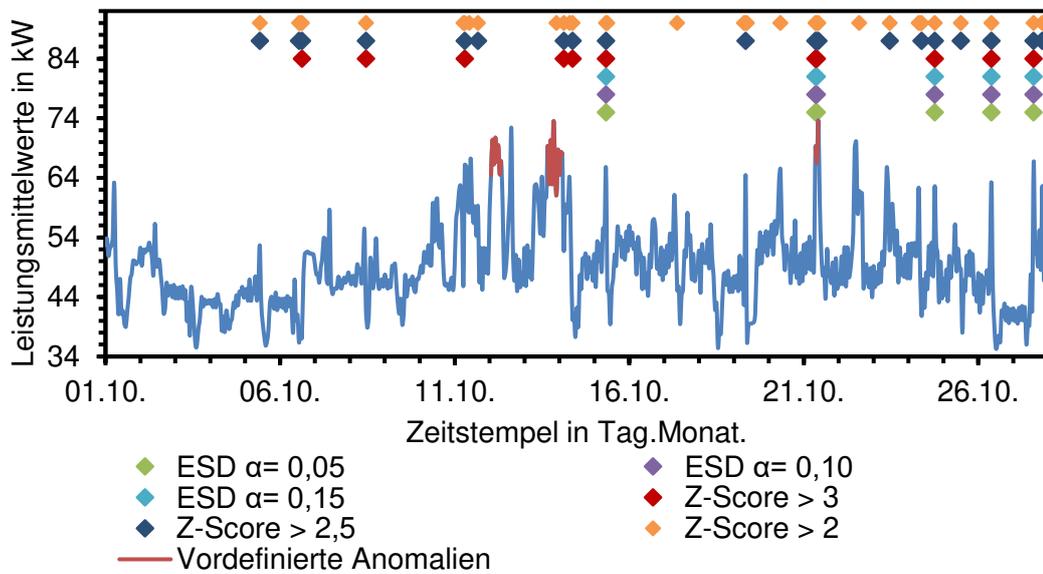


Abbildung 4.9: Darstellung der identifizierten Anomalien des Cluster 5 Unternehmens in der Variante 1M, resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauewerten des ESD Tests ($k = 40$)

Die Ergebnisse, für Variante 1M in Abbildung 4.9 graphisch dargestellt, und in Tabelle 4-9 zeigen, obwohl die R^2 -Werte relativ niedrig sind, können die F_1 -Maße in allen Bedingungsvariationen des ESD- und Z-Score-Tests ($0,14 \leq F_1 \leq 0,50$) vergleichbar hohe Werte, wie Unternehmen aus Cluster 1 bis 4 ($0,10 \leq F_1 \leq 0,50$), erreichen. Zudem liegt die Genauigkeit bei den ESD-Testvariationen (1M: $P = 0,20$; 3M: $0,50 \leq P < 1,00$) über den der Z-Score Tests (1M: $0,08 \leq P < 0,20$; 3M: $0,15 \leq P < 0,17$) jedoch werden maximal ein Drittel der Referenzanomalien wiederentdeckt.

Tabelle 4-9: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 5 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauewerten des ESD Tests ($k = 40$). Siehe Anhang E Tabelle 1 für eine Auflistung der Werte für tp , fp und fn .

	C5 (1M, $R^2 = 0,58$)			C5 (3M, $R^2 = 0,62$)		
	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$	0,20	0,33	0,25	0,50	0,33	0,40
ESD $\alpha = 0,10$	0,20	0,33	0,25	0,50	0,33	0,40
ESD $\alpha = 0,15$	0,20	0,33	0,25	1,00	0,33	0,50
Z-Score > 3	0,20	0,67	0,31	0,17	0,33	0,22
Z-Score $> 2,5$	0,12	0,67	0,20	0,17	0,67	0,27
Z-Score > 2	0,08	0,67	0,14	0,15	1,00	0,26

Auswertung & Diskussion der Einzelergebnisse

In der Variante 3M zeigen die Ergebnisse, dass sich die F_1 -Maße deutlich erhöhen, außer beim Schwellenwert $Z > 3$ (3M). Durch den längeren Trainingszeitraum kann der LSTM Algorithmus eine vorteilhaftere Schätzung durchführen, sodass die Anzahl an falsch positiven Ergebnissen niedriger ausfällt und sich damit die Genauigkeit, außer bei $Z > 3$ (3M), erhöht. Die im Vergleich zu Cluster 1 bis 4 ähnlich hohen F_1 -Maße resultieren wahrscheinlich aus dem Umstand, da die vordefinierten Anomalien die Bereiche mit den Maximalwerten im Kontext des Lastganggesamtverlauf darstellen. Das Schätzmodell muss daher die Lastgangspitzen nicht genau nachbilden können, sodass eine Art geschätzter mittlerer Lastgangverlauf ausreicht um Maximalwerte als Anomalien zu identifizieren, siehe dazu Schätzung 3M in Abbildung 4.10.

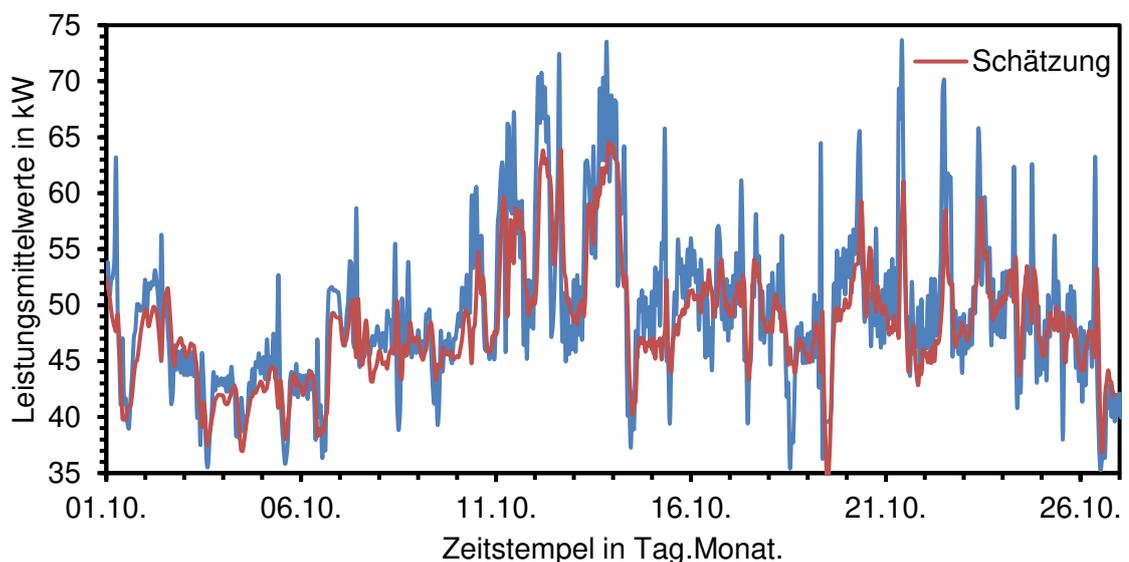


Abbildung 4.10: Darstellung der Schätzung (3M, $R^2 = 0.65$) und der Originaldaten des Clusters 5 Unternehmens

Untersuchung von summierten Residuen

Die Ergebnisse der Auswertung sind für Variante 1M und 3M in Tabelle 4-10 dargestellt. Der größere Bezugszeitraum von 3 Monaten zur Erstellung des Schätzmodells verbessert analog zu den Ergebnissen bei der Untersuchung der Einzelresiduen die F_1 -Maße für alle Sequenzlängen, wenn diese unterhalb von acht Stunden liegen.

Für den Schwellenwert $Z > 3$ (3M) ergibt sich ein F_1 -Maß von 0,22 und dieser kann durch Summierung von zwei, vier und acht Stunden die Anomalieerkennungsgüte mit Werten von $F_1 \geq 0,29$ verbessern. Für einen Z-Schwellenwert größer als 2,5 (3M) ergibt sich für summierte Einzelresiduen nur bei sum2 ein leicht erhöhtes F_1 -Maß ($F_1 = 0,27$, Einzelresiduen; $F_1 = 0,31$, sum2). Je kleiner der Z-Wert desto höher die Anzahl an falsch positiven Ergebnisse, jedoch steigt diese Anzahl bei summierten Residuen deutlich

Auswertung & Diskussion der Einzelergebnisse

stärker an, sodass die Genauigkeitswerte unter denen der Einzelresiduenanalyse liegen ($Z > 2$ (3M): $0,15 \leq P \leq 0,17$, Einzelresiduen; $0,09 \leq P \leq 0,13$, alle Sequenzlängen).

Tabelle 4-10: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 5 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauewerten des ESD Tests ($k = 40$). Siehe Anhang E Tabelle 2 & 3 für eine Auflistung der Werte für tp , fp und fn .

	C5 (1M, $R^2 = 0,58$)			C5 (3M, $R^2 = 0,65$)		
	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$ sum2	0,08	0,33	0,13	0,50	0,33	0,40
ESD $\alpha = 0,10$ sum4	0	0	0	0,20	0,33	0,25
ESD $\alpha = 0,10$ sum2	0,06	0,33	0,11	0,50	0,33	0,40
ESD $\alpha = 0,15$ sum4	0	0	0	0,20	0,33	0,25
ESD $\alpha = 0,15$ sum2	0,06	0,33	0,10	0,50	0,33	0,40
Z-Score > 3 sum8	0,17	0,33	0,22	0	0	0
Z-Score > 3 sum6	0,25	0,33	0,29	0,25	0,33	0,29
Z-Score > 3 sum4	0,10	0,33	0,15	0,29	0,67	0,40
Z-Score > 3 sum2	0,08	0,33	0,13	0,25	0,33	0,29
Z-Score $> 2,5$ sum8	0,20	0,33	0,25	0	0	0
Z-Score $> 2,5$ sum6	0,05	0,33	0,09	0,13	0,33	0,18
Z-Score $> 2,5$ sum4	0,06	0,33	0,11	0,14	0,67	0,24
Z-Score $> 2,5$ sum2	0,05	0,33	0,09	0,20	0,67	0,31
Z-Score > 2 sum8	0,03	0,33	0,06	0,10	1,00	0,18
Z-Score > 2 sum6	0,03	0,33	0,06	0,12	1,00	0,21
Z-Score > 2 sum4	0,04	0,33	0,07	0,13	1,00	0,22
Z-Score > 2 sum2	0,08	0,67	0,15	0,09	1,00	0,17

4.2 Strategie 2: Anomalien unter Normalmodell-Schätzung mit PEWMA

Für jedes Unternehmen eines Clusters werden die drei PEWMA Alphawerte $\alpha_P = 0,3$, $\alpha_P = 0,6$ sowie $\alpha_P = 0,9$ untersucht. Analog zu Strategie 1 werden die aus dem Vergleich von geschätzten und beobachteten Werten entstehenden Residuen (einzeln oder summiert) mittels Z-Score- oder ESD Test analysiert und wieder Genauigkeit, Trefferquote sowie das F_1 -Maß bestimmt.

4.2.1 Cluster 1 Unternehmen

Untersuchung von Einzelresiduen

In Abbildung 4.11 zeigt sich für $\alpha_P = 0,3$, dass eine große Anzahl an anomalen Punkten durch die verschiedenen Testvarianten bestimmt werden. Für den kleinsten untersuchten PEWMA Alphawert, bei dem die geglättete Kurve den Originaldaten am ehesten folgt, zeigt sich, dass steile Leistungsanstiege bzw. -abfälle die größten Residuenwerte erzeugen. Diese werden vom Z-Score- und ESD-Test als Anomalien detektiert (vgl. Abbildung 4.11).

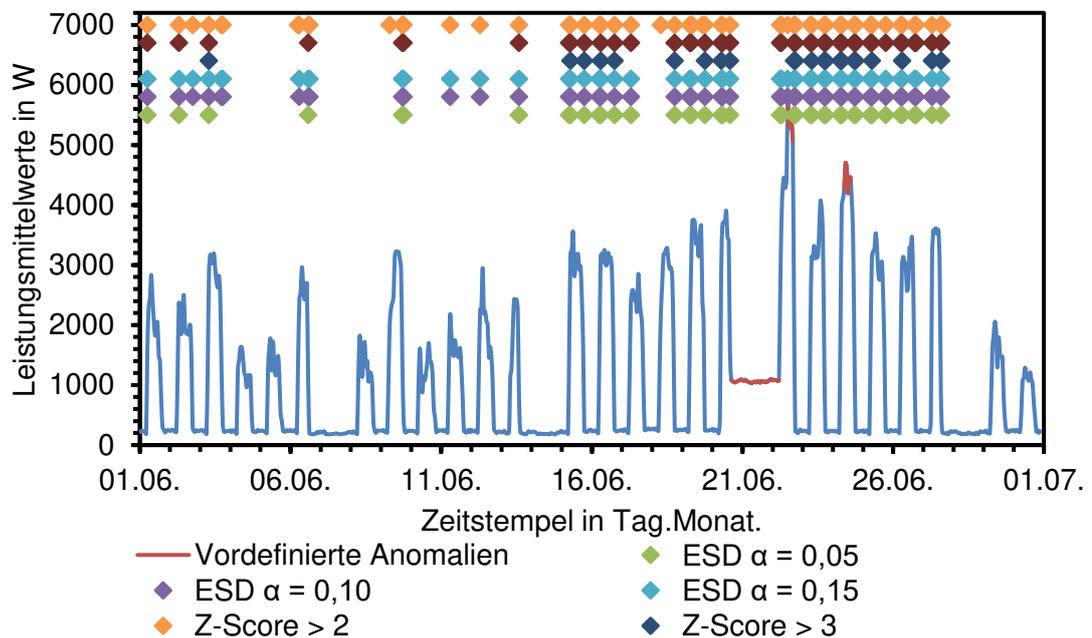


Abbildung 4.11: Darstellung der identifizierten Anomalien des Cluster 1 Unternehmens für PEWMA $\alpha_P = 0,3$ resultierend aus den drei Z-Score Werten und den drei Signifikanzniveauwerten des ESD Tests ($k = 60$)

Tabelle 4-11 fasst die Auswertungsergebnisse für die drei Alphawerte des PEWMA Verfahrens zusammen. Exemplarisch werden die Einordnungen von Anomalien in richtig positiv (tp), falsch positiv (fp) und falsch negativ (fn) für das Unternehmen aus Cluster 1 dargestellt. So lässt sich deutlicher erkennen, dass die Wahl von α_P einen großen Einfluss auf die Ergebnisse des Z-Score Tests und insbesondere auf den ESD Test haben. Die durch Abbildung 4.11 dargestellten vielen falsch positiven Ergebnisse, ergeben für $\alpha_P = 0,3$ in Tabelle 4-11 maximal Genauigkeitswerte von $P = 0,02$. Für größere α_P -Werte resultieren für den ESD Test keine richtig positiven Ergebnisse. $Z > 2$ mit $\alpha_P = 0,9$ zeigt die höchste Anomalieerkennungsgüte mit $F_1 = 0,21$ bei zugleich maximal zwei von drei erkannten Referenzanomalien ($R = 0,67$).

Auswertung & Diskussion der Einzelergebnisse

Die Ergebnisse des ESD Test und die Auswirkung auf die geglättete Kurve, bei Variation des PEWMA Alphawertes, lassen sich mithilfe von Abbildung 4.12 erklären. Sie zeigt die geglätteten Lastgangverläufe für verschieden PEWMA Alphawerte sowie die resultierenden Residuen. Für $\alpha_P = 0,3$ folgt die geglättete Kurve dem Lastgangverlauf der gegebenen Daten relativ stark im Vergleich zu größeren Alphawerten, sodass sich insgesamt die kleinsten Residuenwerte ergeben. Je größer der PEWMA Alphawert wird, desto eher tendieren die Residuen dazu sich von einer unimodalen Verteilung zu entfernen. Dadurch werden beim ESD Test wenige ($\alpha_P = 0,6$) bis gar keine anomalen Punkte ($\alpha_P = 0,9$) und damit verbunden auch keine falsch positiven Ergebnisse identifiziert (vgl. Tabelle 4-11).

Tabelle 4-11: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 1 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 60$) bei drei PEWMA Alphawerten

	Genauigkeit	Trefferquote	F_1 -Maß	tp	fp	fn
C1 ($\alpha_P = 0,3$)						
ESD $\alpha = 0,05$	0	0	0	0	42	3
ESD $\alpha = 0,10$	0,02	0,33	0,04	1	49	2
ESD $\alpha = 0,15$	0,02	0,33	0,04	1	42	2
Z-Score > 3	0	0	0	0	22	3
Z-Score > 2,5	0	0	0	0	41	3
Z-Score > 2	0,02	0,33	0,03	1	55	3
C1 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,05$	0	0	0	0	5	3
ESD $\alpha = 0,10$	0	0	0	0	7	3
ESD $\alpha = 0,15$	0	0	0	0	8	3
Z-Score > 3	0	0	0	0	18	3
Z-Score > 2,5	0	0	0	0	31	3
Z-Score > 2	0,02	0,33	0,03	1	57	2
C1 ($\alpha_P = 0,9$)						
Z-Score > 3	0	0	0	0	3	3
Z-Score > 2,5	0,13	0,67	0,21	2	14	1
Z-Score > 2	0,04	0,67	0,08	2	44	1

Im Fall $\alpha_P = 0,9$ wird der Verlauf sehr stark geglättet. Für die zwei der drei Referenzanomalien mit ungewöhnlichen Maximalwerten entstehen sehr große Residuen. Ihre richtig positive Identifikation ist damit erleichtert. Dies lässt sich in Tabelle 4-11 auch an den hohen Trefferquoten bei Z-Score größer als 2 und 2,5 für $\alpha_P = 0,9$ erkennen. Die Anwendung der auf PEWMA basierenden Strategie 2 bei der Analyse von Einzelresiduen

Auswertung & Diskussion der Einzelergebnisse

des Unternehmens aus Cluster 1 erscheint wenig geeignet für Anomalien, die bspw. durch Werktag-Wochenende-Abhängigkeiten hervorgerufen werden.

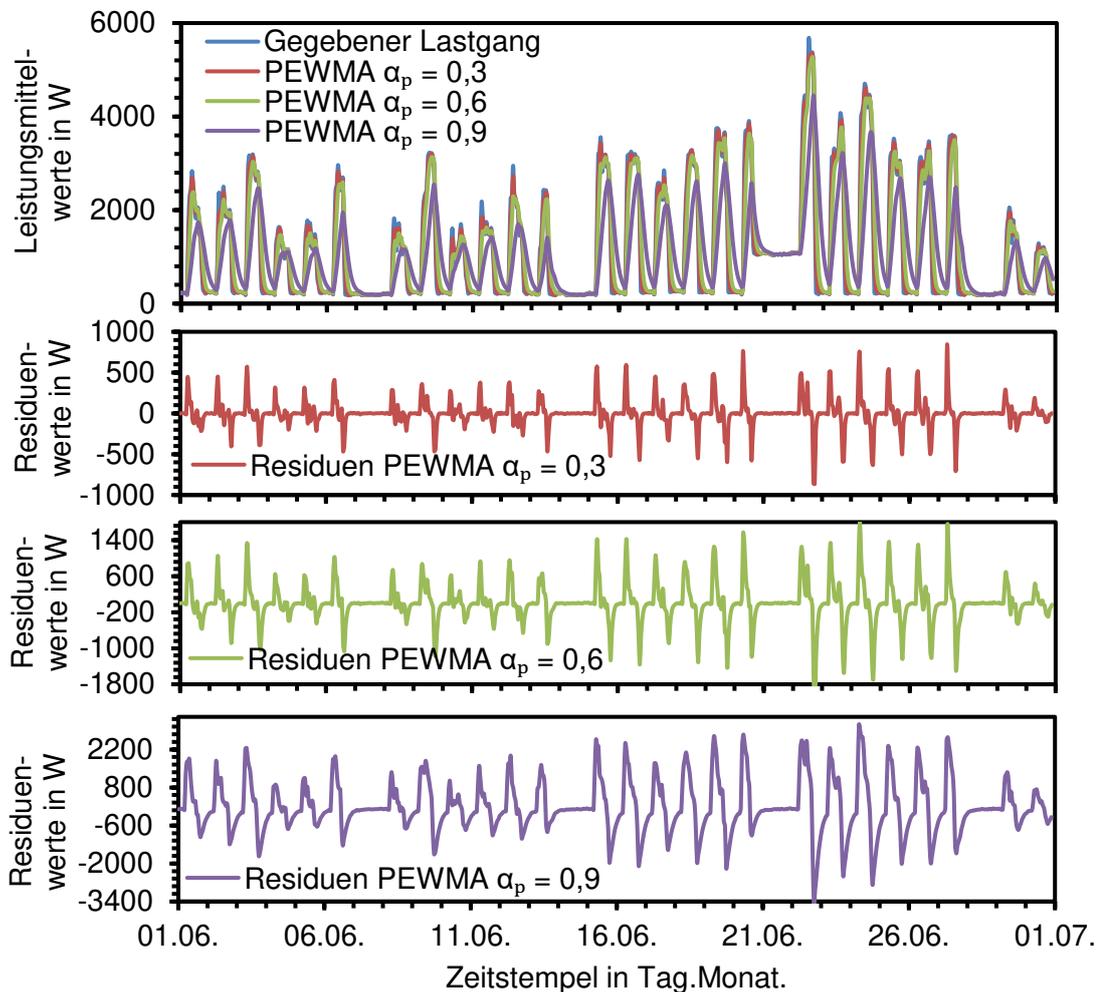


Abbildung 4.12: Darstellung der glättenden Verläufe für verschieden PEWMA Alphawerte und den jeweils resultierenden Residuen des Clusters 1 Unternehmens

Untersuchung von summierten Residuen

In Tabelle 4-13 sind die Ergebnisse der Auswertung für verschiedene Teilsequenzlängen tabelliert. Analog zur Untersuchung von Einzelresiduen lässt sich auch hier der Trend erkennen, dass für größere Alphawerte die F_T -Maße für die Z-Schwellenwerte steigen. Auch bei summierten Residuen werden maximal zwei der drei Referenzanomalien richtig positiv identifiziert ($R \leq 0,67$).

Weiterhin ergibt sich, dass überwiegend beim Test auf anomale Werte bei Z-Score größer als 2 und 2,5 längere Teilsequenzsumationen höhere F_T -Maße bedingen, wenn $Z > 2,5$ sum8 aus diesem Vergleich herausgenommen wird. Dieser Effekt hängt damit zusammen, wie unterschiedliche Summierungslängen die resultierenden Verteilungen

Auswertung & Diskussion der Einzelergebnisse

summierter Residuenwerte beeinflussen. Residuen längerer Sequenzen verschieben die Maxima des Residuenverlaufs auf der Zeitachse. Darüber hinaus erhöhen oder senken gleitende Residuensummen den Leistungsverlauf in verschiedenen Abschnitten unterschiedlich stark (s. Abbildung 4.13). Ein ähnliches Verhalten wurde für Strategie 1, exemplarisch für das Cluster 1 Unternehmen, beschrieben. Dadurch kann z.B. die Referenzanomalie vom 21.06. bis 22.06. bei $Z > 2$ und Residuensequenzen sum6 und sum8 sowie $\alpha_P = 0,3$ identifiziert werden.

Tabelle 4-12: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 1 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang F Tabelle 1, 2 & 3 für eine Auflistung der Werte für tp , fp und fn .

	C1 ($\alpha_P = 0,3$)			C1 ($\alpha_P = 0,6$)			C1 ($\alpha_P = 0,9$)		
	P	R	F_1	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$ sum2	0,06	0,33	0,11	0	0	0	0	0	0
ESD $\alpha = 0,10$ sum2	0,10	0,33	0,15	0	0	0	0	0	0
ESD $\alpha = 0,15$ sum2	0,13	0,33	0,18	0	0	0	0	0	0
Z-Score > 3 sum4	0,14	0,33	0,20	0,17	0,33	0,22	0,33	0,33	0,33
Z-Score > 2,5 sum8	0	0	0	0	0	0	0,40	0,67	0,50
Z-Score > 2,5 sum6	0,11	0,33	0,17	0,13	0,33	0,18	0,33	0,67	0,44
Z-Score > 2,5 sum4	0,04	0,33	0,08	0,05	0,33	0,09	0,10	0,33	0,15
Z-Score > 2,5 sum2	0,03	0,33	0,05	0,03	0,33	0,06	0,07	0,33	0,12
Z-Score > 2 sum8	0,12	0,67	0,20	0,13	0,67	0,22	0,10	0,67	0,17
Z-Score > 2 sum6	0,05	0,67	0,09	0,06	0,67	0,11	0,07	0,67	0,13
Z-Score > 2 sum4	0,02	0,33	0,03	0,02	0,33	0,03	0,05	0,67	0,10
Z-Score > 2 sum2	0,02	0,33	0,03	0,03	0,67	0,06	0,05	0,67	0,09

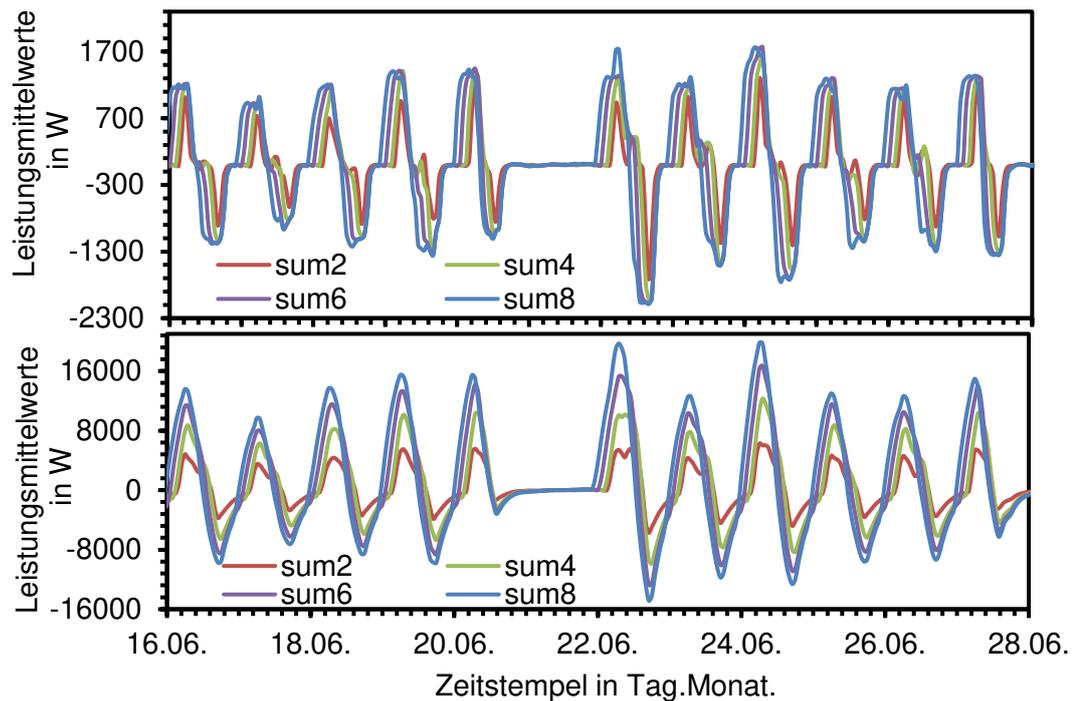


Abbildung 4.13: Darstellung der summierten Residuen verschiedener Teilsequenzlängen für (oben) $\alpha_P = 0,3$ und (unten) $\alpha_P = 0,9$ für das Cluster 1 Unternehmen

4.2.2 Cluster 2 Unternehmen

Untersuchung von Einzelresiduen

In Tabelle 4-13 sind Ergebnisse der Auswertung aufgelistet. Für das Unternehmen aus Cluster 2 liegen zwei Referenzanomalien mit hohen Leistungswerten an den Tagen 9.04. & 11.04. vor. Die restlichen acht Referenzanomalien sind im Kontext der sonst erreichten Leistungswerte im Tages- oder Wochenend-Verlauf ungewöhnlich und daher durch eine auf Glättung basierende Anomalieerkennungsmethode schwer zu identifizieren. Analog zum Unternehmen aus Cluster 1 werden fast an jedem Tag durch den steilen Auf- und Abfall des Lastgangverlaufes bei kleineren Alphawerten falsch positive Ergebnisse identifiziert. Mit steigenden Alphawerten erhöht sich die Wahrscheinlichkeit, dass die zwei angesprochenen Referenzanomalien ($R = 0,17$) richtig positiv als solche erkannt werden können. Entsprechend steigt die Genauigkeit mit größerem Alpha an. Die F_T -Werte zeigen jedoch keine systematischen Änderungen in Abhängigkeit vom gewählten α_P -Wert. Dies liegt daran, dass bei kleineren PEWMA Alphawerten die Trefferquote im Mittel höher liegt. Bei $\alpha_P = 0,3$ und $\alpha_P = 0,6$ werden Referenzanomalien im Zeitraum vom 15.04. bis einschließlich 18.04. zum Teil wiedererkannt (vgl. Abbildung 3.4). Dadurch ergeben sich R -Werte von 0,33 oder 0,25. Das Wiedererkennen der Referenzanomalien im genannten Zeitraum resultiert daraus, dass bei kleinen PEWMA Alphawerten die steilen

Auswertung & Diskussion der Einzelergebnisse

Lastgangflanken im Tagesverlauf große Residuenwerte erzeugen. Zusätzlich sind die vordefinierten Anomalien im genannten Zeitraum über einen längere Zeitspanne definiert, d.h. der vordefinierte anomale Bereich schließt das Ende der aufsteigenden Lastgangflanke am Tagesbeginn und den Anfang der abfallenden Lastgangflanke am Tagesende mit ein. Somit kommt es zu einem zufälligen Wiedererkennen der vordefinierten Anomalien.

Tabelle 4-13: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 2 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang G Tabelle 1 für eine Auflistung der Werte für tp , fp und fn .

	C2 ($\alpha_P = 0,3$)			C2 ($\alpha_P = 0,6$)			C2 ($\alpha_P = 0,9$)		
	P	R	F_1	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$	0,09	0,33	0,14	0,15	0,17	0,16	0,40	0,17	0,24
ESD $\alpha = 0,10$	0,09	0,33	0,14	0,09	0,17	0,12	0,22	0,17	0,19
ESD $\alpha = 0,15$	0,09	0,33	0,14	0,09	0,25	0,14	0,30	0,25	0,27
Z-Score > 3	0,05	0,08	0,06	0,11	0,17	0,13	0,12	0,17	0,14
Z-Score > 2,5	0,11	0,33	0,16	0,09	0,25	0,13	0,09	0,17	0,12
Z-Score > 2	0,09	0,33	0,14	0,07	0,25	0,10	0,05	0,17	0,08

Untersuchung von summierten Residuen

Die Ergebnisse der Untersuchung von summierten Residuen auf Anomalien finden sich in Tabelle 4-14. I.d.R. können hier zwei der zwölf Referenzanomalien richtig positiv identifiziert werden. Dabei handelt es sich um die im Kontext des Gesamtverlaufes höchsten Leistungswerte. Der im Vorabschnitt diskutierte vermutete Zufallsfund richtig positiver Referenzanomalien, zeigt sich im Falle der Analyse von summierten Residuen nicht.

Tabelle 4-14: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 2 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 60$). Siehe Anhang G Tabelle 2, 3 & 4 für eine Auflistung der Werte für tp , fp und fn .

	C2 ($\alpha_P = 0,3$)			C2 ($\alpha_P = 0,6$)			C2 ($\alpha_P = 0,9$)		
	P	R	F_1	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$ sum2	0,13	0,17	0,15	0,29	0,17	0,21	0	0	0
ESD $\alpha = 0,10$ sum2	0,13	0,17	0,14	0,17	0,17	0,17	0,18	0,17	0,17
ESD $\alpha = 0,15$ sum2	0,11	0,17	0,13	0,15	0,17	0,16	0,17	0,17	0,17
Z-Score > 3 sum8	0	0	0	0	0	0	0	0	0
Z-Score > 3 sum6	1,00	0,17	0,29	0	0	0	0	0	0
Z-Score > 3 sum4	0,18	0,17	0,17	0,17	0,17	0,17	0,10	0,08	0,09
Z-Score > 3 sum2	0,13	0,17	0,14	0,14	0,17	0,15	0,17	0,17	0,17
Z-Score > 2,5 sum8	1,00	0,17	0,29	0	0	0	0,00	0,00	0
Z-Score > 2,5 sum6	0,12	0,17	0,14	0,12	0,17	0,14	0,00	0,00	0
Z-Score > 2,5 sum4	0,08	0,17	0,11	0,09	0,17	0,11	0,10	0,17	0,13
Z-Score > 2,5 sum2	0,06	0,17	0,09	0,06	0,17	0,09	0,10	0,17	0,12
Z-Score > 2 sum8	0,06	0,17	0,09	0,06	0,17	0,09	0	0	0
Z-Score > 2 sum6	0,05	0,17	0,07	0,05	0,17	0,07	0,02	0,08	0,04
Z-Score > 2 sum4	0,04	0,17	0,07	0,04	0,17	0,07	0,05	0,17	0,07
Z-Score > 2 sum2	0,09	0,42	0,14	0,08	0,42	0,14	0,06	0,17	0,09

4.2.3 Cluster 3 & 4 Unternehmen

Die Auswertung mittels PEWMA hat bei den Unternehmen aus Cluster 3 und 4 gezeigt, dass ähnliche Ergebnisse und Effekte auftreten, wie die bei den zuvor ausgewerteten Unternehmen. Zwei der drei Referenzanomalien im Lastgang des Unternehmens aus Cluster 3 sind Maximalwerte im Kontext des gesamten Verlaufs. Bei einem hohen PEWMA Alphawert lassen sich diese wiedererkennen. Die Unternehmen aus Cluster 3 und 4 besitzen beide sehr steile An- und Abstiege im Tagesverlauf, wahrscheinlich bedingt durch eine Kernarbeitszeit. Diese werden, wie bei Unternehmen aus Cluster 1 und 2, bei geringen α_P Werten vornehmlich als (falsch positive) Anomalien identifiziert. Das Unternehmen aus Cluster 4 besitzt überwiegend Referenzanomalien, welche nicht durch ungewöhnliche Maximalwerte auffallen, daher kann die auf PEWMA basierende Methode hier nur sehr selten die Referenzanomalien wiedererkennen. Die ausgewerteten Genauigkeiten, Trefferquoten und F_1 -Maße für die Unternehmen aus Cluster 3 & 4 und die dazugehörigen Werte für tp , fp und fn finden sich im Anhang H und I.

4.2.4 Cluster 5 Unternehmen

Untersuchung von Einzelresiduen

In Abbildung 4.14 sind die Ergebnisse aus der Auswertung mit $\alpha_P = 0.9$ graphisch dokumentiert. Es zeigt sich, dass analog zu den Unternehmen der Cluster 1 bis 4 Referenzanomalien, welche Maximalwerte im Lastganggesamtverlauf darstellen, mit einem hohen PEWMA Alphawert wiedererkannt werden können (siehe dazu auch Tabelle 4-15). Jedoch vermindert sich die Genauigkeit, je mehr Datenpunkte als Ausreißer gewertet werden, wie beispielsweise im Test auf Residuenwerte mit einem Z-Score größer als 2.

Tabelle 4-15: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 5 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauewerten des ESD Tests ($k = 40$). Siehe Anhang J Tabelle 1 für eine Auflistung der Werte für tp , fp und fn .

	C5 ($\alpha_P = 0,3$)			C5 ($\alpha_P = 0,6$)			C5 ($\alpha_P = 0,9$)		
	P	R	F_1	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$	0,25	0,33	0,29	0,33	0,33	0,33	0	0	0
ESD $\alpha = 0,10$	0,25	0,33	0,29	0,20	0,33	0,25	0,17	0,33	0,22
ESD $\alpha = 0,15$	0,20	0,33	0,25	0,14	0,33	0,20	0,14	0,33	0,20
Z-Score > 3	0,09	0,33	0,14	0,09	0,33	0,14	0,18	0,67	0,29
Z-Score > 2,5	0,05	0,33	0,09	0,11	0,67	0,19	0,12	0,67	0,20
Z-Score > 2	0,03	0,33	0,05	0,06	0,67	0,10	0,08	1,00	0,14

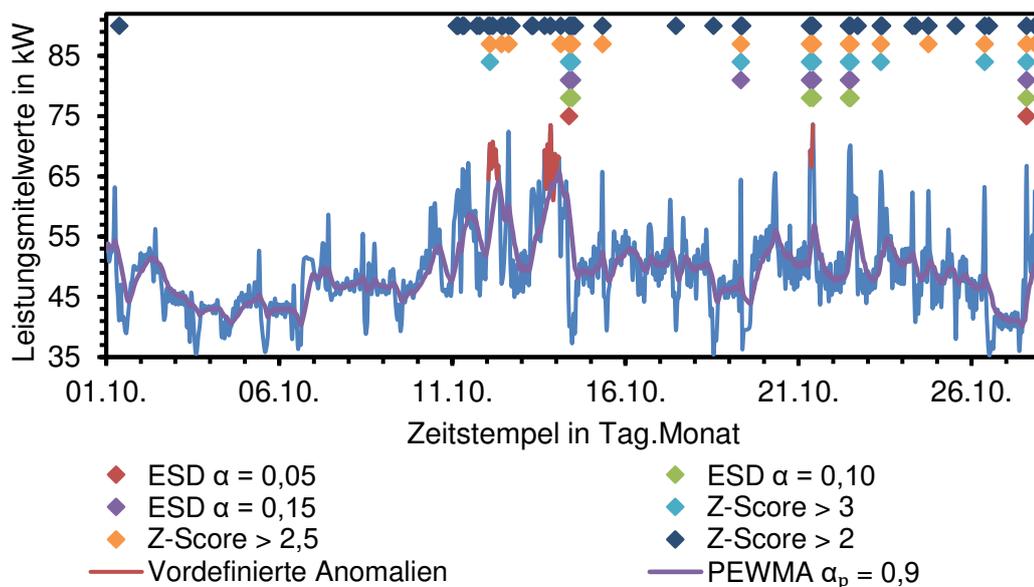


Abbildung 4.14: Darstellung der identifizierten Anomalien des Cluster 5 Unternehmens für PEWMA $\alpha_P = 0,9$ resultierend aus dem Z-Score Test und dem ESD Test ($k = 60$)

Untersuchung von summierten Residuen

Tabelle 4-16 listet die Ergebnisse für die Summierung verschiedener Teilsequenzlängen auf. Die Identifikation richtig positiver Anomalien gelingt bei der Analyse von summierten Residuen besser als bei der Analyse von Einzelresiduen. Allerdings steigt auch die Anzahl falsch positiver Zuordnungen, so dass sich das F_1 -Maß nur unwesentlich verbessert. Insgesamt zeigen sich bei der Analyse der Residuensequenzen ähnliche Tendenzen, wie bei den Analysen unter Verwendung von Strategie 1 und der Schätzung des Normalmodells mit LSTM Algorithmus: Das höchste harmonische Mittel aus Genauigkeit und Trefferquote zeigt sich mit $F_1 = 0,40$ bei $Z > 3$ und Residuensequenzen von 6 Stunden (sum6) – allerdings können auch hier nur 2 der 3 Referenzanomalien richtig positiv identifiziert werden.

Tabelle 4-16: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 5 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 60$). (Siehe Anhang 9.) Tabelle 2 für eine Auflistung der Werte für tp , fp und fn .

	C5 ($\alpha_P = 0,3$)			C5 ($\alpha_P = 0,6$)			C5 ($\alpha_P = 0,9$)		
	P	R	F_1	P	R	F_1	P	R	F_1
ESD $\alpha = 0,05$ sum2	0	0	0	0	0	0	0	0	0
ESD $\alpha = 0,10$ sum2	0,10	0,33	0,15	0,20	0,33	0,25	0,14	0,33	0,20
ESD $\alpha = 0,15$ sum2	0,09	0,33	0,14	0,17	0,33	0,22	0,13	0,33	0,18
Z-Score > 3 sum8	0	0	0	0	0	0	0,13	0,33	0,18
Z-Score > 3 sum6	0	0	0	0,25	0,33	0,29	0,29	0,67	0,40
Z-Score > 3 sum4	0,11	0,33	0,17	0,18	0,67	0,29	0,20	0,67	0,31
Z-Score > 3 sum2	0,07	0,33	0,11	0,14	0,67	0,24	0,20	0,67	0,31
Z-Score > 2,5 sum8	0,15	0,67	0,25	0,17	0,67	0,27	0,07	0,33	0,12
Z-Score > 2,5 sum6	0,11	0,67	0,19	0,17	0,67	0,27	0,15	0,67	0,25
Z-Score > 2,5 sum4	0,09	0,67	0,16	0,10	0,67	0,17	0,13	0,67	0,21
Z-Score > 2,5 sum2	0,07	0,67	0,13	0,08	0,67	0,14	0,11	0,67	0,18
Z-Score > 2 sum8	0,06	0,67	0,11	0,05	0,67	0,10	0,08	1,00	0,15
Z-Score > 2 sum6	0,06	0,67	0,11	0,05	0,67	0,09	0,08	1,00	0,15
Z-Score > 2 sum4	0,05	0,67	0,09	0,05	0,67	0,09	0,05	0,67	0,10
Z-Score > 2 sum2	0,05	0,67	0,10	0,05	0,67	0,09	0,06	0,67	0,10

5 Strategie- und clusterübergreifende Diskussion

Ziel der Arbeit war es, Strategien zur Anomalieerkennung auf den Umgang mit Lastgangdaten zuzuscheiden, sowie die Güte der Strategien zu prüfen. Aufbauend auf theoretischen Vorüberlegungen zur Anomalieerkennung wurden zwei Strategien entwickelt. Der wesentliche Unterschied der beiden Strategien besteht in der Art der Schätzung eines *normalen* monatlichen Lastgangverlaufs. Strategie 1 schätzt die Lastgänge mithilfe des LSTM Algorithmus aus dem Bereich des Machine Learnings. Strategie 2 hingegen nutzt das Schätzverfahren PEWMA aus der statistischen Zeitreihenanalyse. Die Schätzungen erfolgten bei Strategie 1 mit Daten von einem Monat (1M) sowie mit Daten von 3 drei exemplarischen Monaten (3M). Die geschätzten Lastgänge (auch Normalmodelle genannt) wurden mit den beobachteten Lastgängen eines Monats verglichen und die Abweichungen analysiert (Residuenanalyse). Diese Analyse erfolgt entweder auf Ebene der Einzelresiduen (Punktanomalien) oder auf Basis der Summierung von aufeinanderfolgenden Residuen als gleitende Sequenzen (Teilsequenzanomalien). Residuen und Residuensequenzen, deren Ausprägung eine definierte Schwelle überschritt, wurden als Anomalien gewertet. Die Schwellen wurden systematisch in einem ESD-Testverfahren oder als Z-Score-Test variiert (je drei Prüfschwellen). Die so identifizierten Anomalien wurden mit von Experten im Vorhinein identifizierten Referenzanomalien verglichen. Dadurch wurde es möglich die Eignung der Strategien zur Anomalieerkennung in den vorliegenden Daten hinsichtlich Genauigkeit (P), Trefferquote (R) und harmonischem Mittel daraus (F_1) zu prüfen.

Mit Hilfe der so gewonnenen Erkenntnisse wird nun ein Fragen geleitetes Resümee gezogen. Anschließend und die Arbeit abschließend, werden die Befunde in aktuelle Forschungsbeiträge eingeordnet, sowie ein Ausblick auf interessante Fragen für zukünftige Forschungsbemühungen gegeben.

5.1 Fragen geleitetes Resümee

(1) Eignen sich die gewählten Ansätze zur Schwellenwertbildung (hier Z-Score und ESD Test) für die Identifikation von Anomalien? Wie beeinflussen die Ansätze die Anomalie-Erkennungsgüte?

In Abbildung 5.1 ist die durchschnittliche Trefferquote des Z-Score- und ESD-Tests je Schwellenwert der Strategie 1 von Variante 1M und 3M dargestellt. Es zeigt sich, dass die durchschnittliche Trefferquote jedes ESD Tests unterhalb des Z-Schwellenwertes $Z > 3$ liegt. Das ist insofern bemerkenswert, als dass die für den ESD Test gewählten Signifikanzniveaus überdurchschnittlich hoch angesetzt waren. Üblicherweise werden Signifikanzniveauewerte bei $\alpha = 0,5$ oder kleiner gesetzt (Rosner 1983). Bei der Voranalyse der vorliegenden Daten zeigte sich jedoch, dass unter einem

α -Wert kleiner 0,05 keine Anomalien aufgedeckt werden konnte. Deshalb wurden drei deutlich höhere Signifikanzniveaus als Schwellenwerte gesetzt ($\alpha = 0,05; 0,10; 0,15$). Ein Alphawert von 0,15 bedeutet in dem vorliegenden Kontext, dass mit 15 %iger Wahrscheinlichkeit der ESD-Test anomale Punkte identifiziert, obwohl in Wahrheit keine Anomalien vorhanden sind (Alphafehler-Wahrscheinlichkeit). Je höher α , desto weniger Aussagekraft hat der ESD Test. Darüber hinaus identifiziert der ESD-Test Anomalien in Abhängigkeit von der Menge der eingehenden Datenpunkte. Es werden keine Anomalien erkannt, wenn die in Formel (3.9) gebildeten R_i -Werte unterhalb des kritischen Wertes von λ_i liegen. Die exemplarische Auswertung (siehe Anhang K Tabelle 1) für C1 (1M, ESD $\alpha = 0,15$) der gebildeten R_i -Wert zeigt, dass die kritischen Werte je Iteration im Vergleich zu den R_i -Werten sehr langsam kleiner werden. Die hohe Zahl eingehender Daten von $n = 720$ zu Beginn führt dazu, dass je Iteration in Formel (3.11) der errechnete p -Wert und damit auch der λ_i -Wert relativ stabil bleiben. Der Test scheint für Anomalien in großen Datenmengen folglich zu wenig sensibel. Weiter zeigte sich in der Auswertung, dass sich mit steigender Sequenzlänge der summierten Residuen (vgl. Abbildung 4.3) sowie mit steigenden PEWMA Alphawerten (vgl. Abbildung 4.12) die Häufigkeitsverteilung der erhaltenen Residuen immer weiter von einer Normalverteilung entfernt. Die Normalverteilungsannahme zählt jedoch zu den Voraussetzungen eines ESD-Tests. Ist sie nicht gegeben sind die Ergebnisse unter Vorbehalt zu interpretieren. Auch dies spricht gegen die Eignung des ESD-Tests als Anomalieerkennungsverfahren für die vorliegenden Daten. All diese Einschränkungen gelten allerdings nicht für den Einsatz des Z-Score Tests.

Der Z-Score Test kann demnach im Gegensatz zum ESD-Test als geeignet angesehen werden, die gegebenen Daten zu analysieren.

Für die Diskussion um die Interpretation der Beurteilungskriterien Genauigkeit, Trefferquote und dem harmonischen Mittel werden im Folgenden nur noch Daten aus den Analysen mit Z-Score Verfahren herangezogen.

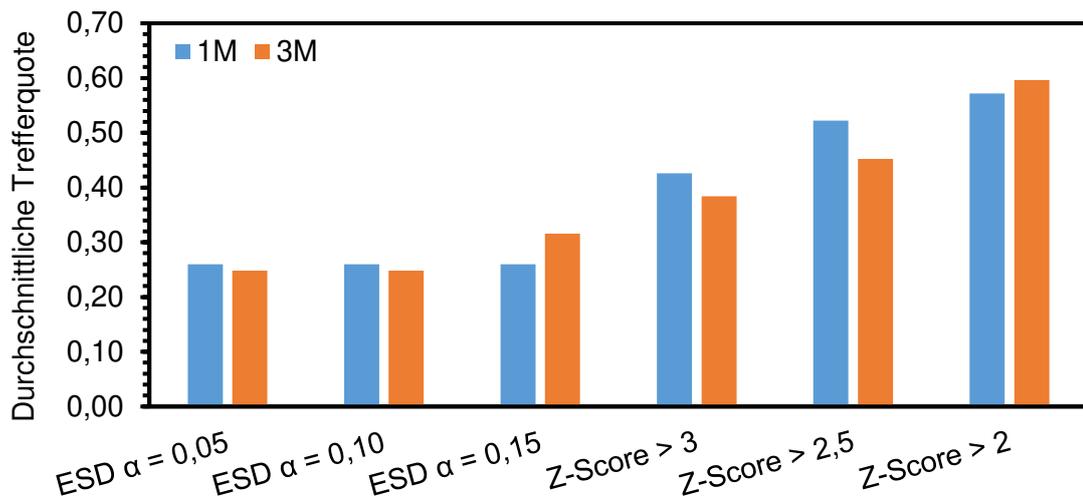


Abbildung 5.1: Darstellung der durchschnittlichen Trefferquote von Unternehmen C1 bis C5 je Schwellenwert für den Z-Score- und ESD-Test von Strategie 1 (1M & 3M)

(2) Hängt die Güte der Anomalieerkennung von den spezifischen Lastgangverläufen einzelner Unternehmen ab?

Um die dieser Arbeit zugrundeliegenden Daten handhabbar zu machen, wurden sie in mehreren Schritten reduziert. Einer dieser Schritte bestand darin, die vorliegenden Lastgänge aller Unternehmen so zu gruppieren, dass sich die den einzelnen Gruppen zugeordneten Lastgänge maximal voneinander unterscheiden sollten (*k*-Means Clusteranalyse). Es ergaben sich 5 Gruppen oder auch Cluster voneinander unterscheidbarer Unternehmen. Aus jedem Cluster wurde ein prototypisches Unternehmen, insgesamt also 5 Unternehmen, für die weitere Verarbeitung und Analyse ausgewählt. Dabei zeigte sich, dass die Unternehmen aus Cluster 1 bis 4 relativ ähnliche Grundvoraussetzungen für eine Normalmodell-Schätzung mitbrachten: Es existieren nachvollziehbare zeitliche Rhythmen, die davon abhängig sind, ob ein Werktag, Wochenende oder Feiertag vorlag oder auch Kernarbeitszeiten anzeigten. Die Experten identifizierten u.a. Abweichungen als Referenzanomalien, die diesen Rhythmen nicht folgten.

Das der Strategie 1 zugrunde liegende LSTM Netzwerk kann dank der Informationen aus dem Zeitstempel prinzipiell auf viele dieser Input-Variablen zurückgreifen. Folgend wäre anzunehmen, dass der LSTM Algorithmus in der Lage sein sollte, die zyklischen Veränderungen der Leistungsdaten zu schätzen. In der Ausführung zeigen sich allerdings Schwierigkeiten einzelne Lastgangspitzen im Leistungsmaximalbereich des Tagesverlaufes oder die Flanken plötzlich ansteigender oder abfallender Leistungswerte zu schätzen (vgl. dazu

Abbildung 4.1 oder Abbildung 4.8). In diesen Bereichen entstehen beim Vergleich des mit LSTM geschätzten Modells mit den beobachteten Daten hohe Residuenwerte. Diese werden in der Residuennalyse als Anomalien erkannt, entsprechen allerdings nicht den Referenzanomalien, so dass niedrige F_T -Maße resultieren. Dem entgegen zeigt der Lastgang des Unternehmens aus Cluster 5 keine zeitlichen Abhängigkeiten wie sie bei den Lastgängen aus Cluster 1 bis 4 auftreten. Die nach Strategie 1 analysierten Anomalien stimmen dabei mit den Referenzanomalien überdurchschnittlich gut überein, da sie sich deutlich vom Kontext des gesamten Lastgangverlaufes abheben.

Auch bei der auf PEWMA basierenden Strategie 2 lässt sich beobachten, dass nicht die Zugehörigkeit von Lastgängen zu einem Cluster die Güte der Anomalieerkennung beeinflusst, sondern eher in welchem Kontext Anomalien stehen. Dabei können Residuen anomal für einen zeitlichen Rhythmus sein oder aber als ungewöhnlich im Vergleich zum Gesamtverlauf auffallen.

Sowohl unter Verwendung von Strategie 1 als auch von Strategie 2 gilt, dass die Güte der Erkennung von Anomalien nicht von der Zugehörigkeit der analysierten Lastgänge zu einem Cluster abhängt.

(3) Welche Auswirkung hat die Länge des Zeitraums der zur Schätzung eines Normalmodells herangezogenen Daten (hier Daten von einem vs. von drei Monaten) auf die Güte der Anomalieerkennung?

Die den beiden Anomalieerkennungsstrategien zugrundeliegenden Schätzmodelle, schätzen das Normalmodell entweder aus Daten von einem Monat (1M), oder aus Daten von drei Monaten (3M). In Abbildung 5.2 sind die F_T -Maße der Z-Score Tests jedes Unternehmens aus Clusters 1 bis 5 für beide Varianten dargestellt (auf eine Abbildung der ESD-Maße wird verzichtet, s. Überlegungen zu Frage 1 dieser Resümeees). Die Abbildung zeigt außerdem das Verhältnis von hohen F_T -Maßen bei der 1M-Schätzung im Vergleich zu denen aus der 3M-Schätzung bei jeweils gleichem Schwellenwert. Es ist ersichtlich, dass sich unter der Berücksichtigung der 3M-Schätzung in den meisten Fällen bessere F_T -Maße ergeben.

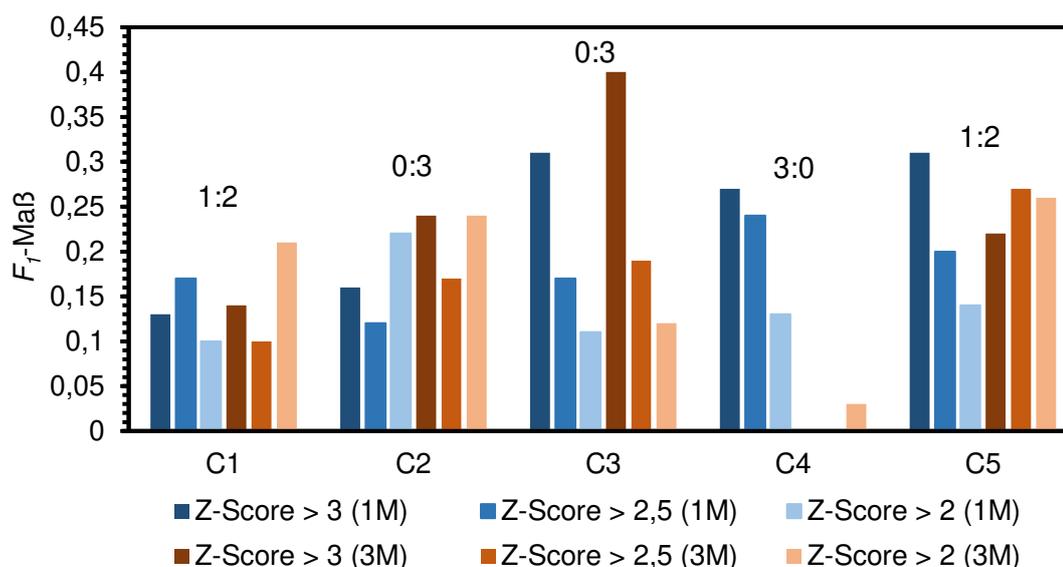


Abbildung 5.2: Darstellung F_1 -Maße der Z-Score Tests für jedes Unternehmen des Clusters 1 bis 5 in den Varianten 1M & 3M und Angabe des Verhältnisses von A:B mit A: Anzahl der höheren F_1 -Maße im Vergleich zu 3M und B: Anzahl der höheren F_1 -Maße im Vergleich zu 1M

Bei genauerer Betrachtung der Datengrundlage des Unternehmens aus Cluster 4 wird deutlich, dass zwei der zur 3M-Schätzung herangezogenen Vormonate sehr unterschiedliche Verläufe aufweisen. Die resultierende Schätzung bildet folglich keine gute Grundlage zur Identifikation der Referenzanomalien.

Es zeigen sich allerdings auch insgesamt vergleichsweise niedrige F_1 -Werte, da je nach Schwellenwert eine hohe Anzahl an falsch positiven Zuordnungen vorliegt. Vermutlich könnte die Verwendung deutlich längerer Zeiträume für das Trainieren des LSTM Netzwerkes dazu dienen, die Anomalieerkennungsgüte zu verbessern. Da die Limón GmbH über viele Datensätze verfügt, deren Informationen sich über einen Zeitraum von einem halben bis zu einem Jahr erstrecken, wäre es sinnvoll mit längere Zeiträume als Grundlage für weitere Untersuchungen zu nutzen.

(4) Eignet sich Strategie 1 oder Strategie 2 besser zur Anomalieerkennung? Welche Auswirkungen hat die Analyse von Einzelresiduen oder Residuensequenzen auf die Güte der Anomalieerkennung?

Abbildung 5.3 stellt die durchschnittlichen F_1 -Maße (1M: bläuliche Farbtöne, 3M: rötliche Farbtöne) sowie Standardabweichungen für die Unternehmen aus Cluster C1 bis C5 für summierte Residuen (gleitende Sequenzen) und Einzelresiduen in einer Vergleichsübersicht dar (basierend auf Z-Score-Test, vgl. Frage 1). Das Unternehmen aus Cluster

4 ist bei diesem Vergleich in Variante 3M nicht inkludiert (vgl. Überlegungen zur Schätzung des Lastganges von Cluster 4 unter Frage 3). Für die beiden Schätzvarianten 1M und 3M zeigen sich unterschiedliche Tendenzen. In Variante 1M sinken die durchschnittlichen F_T -Werte der Einzelresiduen je kleiner der Schwellenwert wird. Für summierte Residuen zeigt sich ebenfalls eine leichte Tendenz, dass die F_T -Werte mit kleiner werdenden Schwellenwerten sinken.

Weiter ist in Abbildung 5.3 zu sehen, dass unter der 1M-Schätzung der Z-Score Test nach Einzelresiduen in allen Bedingungsvariationen zuverlässig höhere F_T -Werte aufweist als die gleichen Analysen mit summierten Residuen (Teilsequenzen). Ein differenzierteres Bild zeigt sich, wenn die F_T -Werte betrachtet werden, die auf Grundlage der 3M-Schätzung entstanden sind. Die Kombination aus Schwellenwerten von $Z > 2,5$ bzw. $Z > 3$ und Residuensequenzen von vier aufsummierten Stunden (sum4) unter 3M-Schätzung zeigt im Gesamtvergleich die höchsten F_T -Werte. Die kleinsten Standardabweichungen zeigen sich für ein $Z > 2,5$ unter der 3M-Schätzung. Im Durchschnitt erzielt unter Strategie 1, also der Schätzung des Normalmodells mittels LSTM Algorithmus, die Bedingungskombination 3M, $Z > 2,5$ und sum4 die besten Anomalieerkennungswerte.

In der Analyse der Daten mittels Strategie 2 zeigte sich, dass sich dafür nur der Lastgangverlauf des Unternehmens aus Cluster 5 eignet (s. Kapitel 4.2.4). Die höchsten harmonischen Mittel aus Genauigkeit und Trefferquote mit F_T -Maß bis zu 0,40 ergaben sich bei der Residuensummierung von vier und sechs Stunden sowie bei der Z-Schwelle von > 3 . Es stellte sich weiterhin heraus, dass der größte untersuchte PEWMA Alpha-wert $\alpha_P = 0.9$ die höchsten F_T -Maße erzielen konnte (s. Kapitel 4.2). Vergleicht man diese Werte mit der Strategie-1-Analyse der Daten des Unternehmens aus Cluster 5 zeigen sich identisch hohe F_T -Werte.

Insgesamt lässt sich unter Berücksichtigung aller analysierten Daten und Verwendung der Z-Score Schwellen festhalten, dass der Einsatz der auf dem LSTM-Netzwerk basierenden Strategie 1 dem Einsatz von Strategie 2 vorzuziehen ist. Strategie 1 erzielt eine bessere Anomalieerkennungsleistung.

Strategie- und clusterübergreifende Diskussion

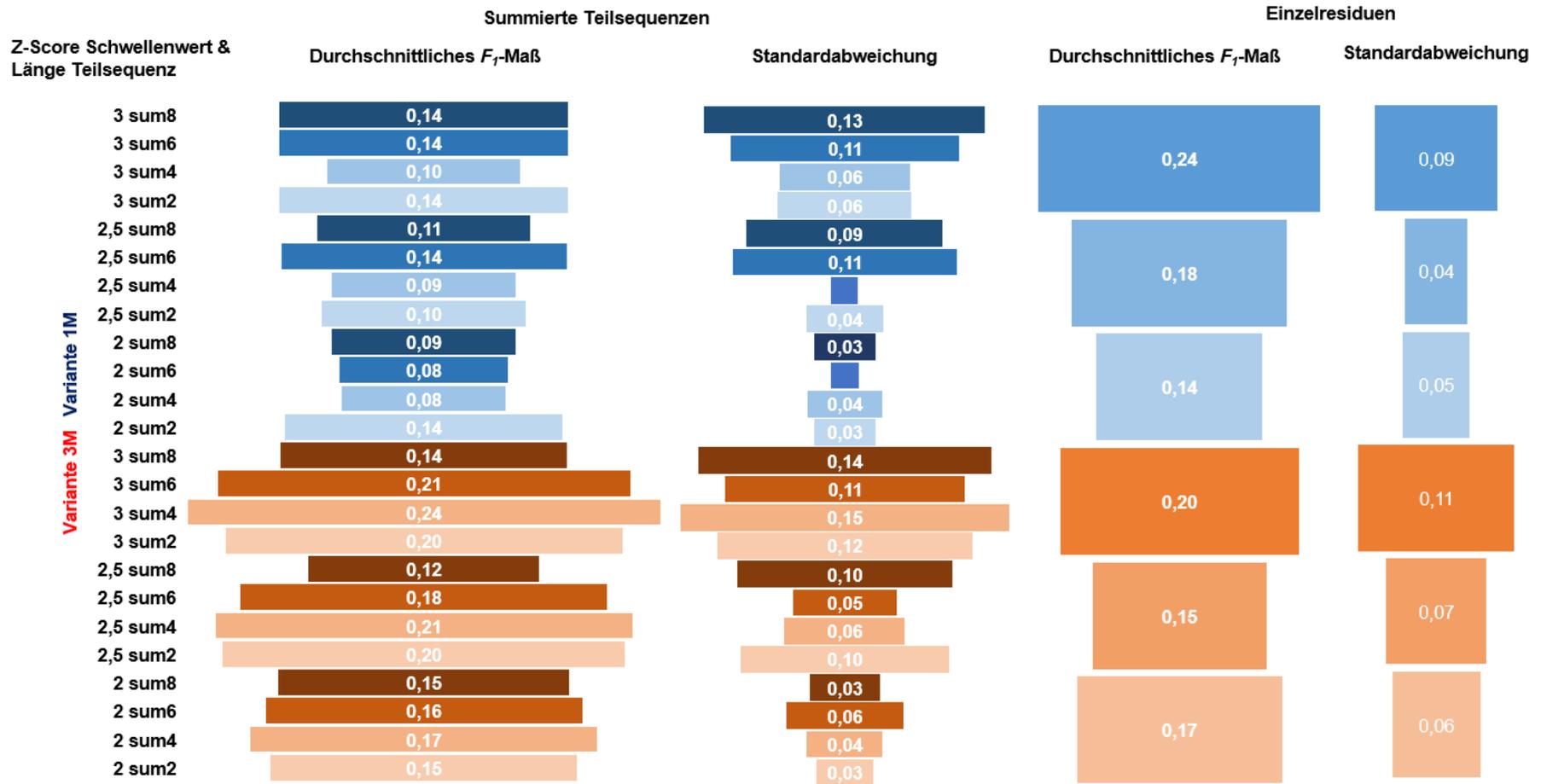


Abbildung 5.3: Darstellung der durchschnittlichen F_1 -Maße und Standardabweichung (1M: bläuliche Farbtöne, 3M: rötliche Farbtöne) für die Unternehmen aus C1 bis C5 für aufsummierte Teilsequenzen und Einzelresiduen im Vergleich

5.2 Einordnung in aktuelle und Ausblick auf zukünftige Forschung

Der Ansatz, allein mittels Informationen aus Zeitstempel und den zugehörigen Leistungswerten Anomalien identifizieren zu können, ist mindestens ambitioniert. Tatsächlich wird ein Lastgangverlauf potenziell von vielen Faktoren beeinflusst. Khan et al. (2014) verwenden einen zur Strategie 1 ähnlichen Ansatz. Sie nutzen ebenfalls ein künstliches neuronales Netzwerk (hier das s.g. Multilayer Perceptron, MLP) zur Vorhersage eines Normalmodells. Das Modell dient als Grundlage für die Analyse von Anomalien im Leistungsbedarf der Beleuchtung eines einzelnen Stockwerkes eines bekannten Gebäudes. Den Forschern sind dabei detaillierte und weitreichende Informationen über den bloßen Leistungsbedarf hinaus bekannt, z.B. die Anzahl der Lichtquellen und Mitarbeiterzahlen. Auf Grundlage weitreichender Informationen wird ein realistisches und genaues „anomaliefreies“ Normalmodell geschätzt und schließlich mit den beobachteten Daten verglichen. Auftretende Anomalien können mit großer Genauigkeit erkannt werden. Diese Ergebnisse weisen darauf hin, dass es die Integration detaillierter Informationen bedarf, um ein möglichst genaues Normalmodell schätzen zu können. Bezogen auf die vorliegende Arbeit könnte demnach eine höhere Anomalieerkennungsgüte mit höheren F_1 -Werten erreicht werden, indem weitere auf die Leistungswerte wirkende Faktoren in die Schätzung des Normalmodells integriert werden.

Hundman et al. (2018) verwenden ein LSTM Netzwerk um Anomalien in multivariaten Zeitreihen von Telemetriedaten eines Raumfahrzeuges zu identifizieren, bei welchen verschiedenartige Verfahren zur Schwellenwertbildung ausgetestet werden. Die Autoren berichten Genauigkeitswerte von bis zu $P = 0,92$ in einer Kombination mit einer Trefferquote von $R = 0,63$. Die Publikation lässt offen, welche Charakteristika die Verläufe der beschriebenen Telemetriedaten aufweisen, entsprechend lassen sich die Ergebnisse schwer substanziell mit denen dieser Arbeit vergleichen. Allerdings sprechen sie dafür, dass der auch hier zur Anwendung gekommene LSTM Algorithmus prinzipiell durchaus eine geeignete Methode zur Schätzung eines Normalmodells darstellen kann. Darüber hinaus regt die Publikation von Hundman et al. (2018) an, weitere Verfahren zur Bildung von Schwellenwerten im Kontext der in dieser Arbeit analysierten Daten auf ihre Nützlichkeit hin zu prüfen.

Die Schätzung eines Normalmodells mittels PEWMA-Verfahren hat sich für den Anwendungsfall dieser Arbeit als weniger geeignet erwiesen. In der Literatur finden sich allerdings Befunde, die eindeutig auf eine Güte des Ansatzes im Kontext der Identifikation von Anomalien hinweisen. So wurde das PEWMA Verfahren beispielsweise bei Streamingdaten von Twitter (Patel et al. 2015) mit resultierende F_1 -Werten von bis zu 0,80 oder bei Temperaturzeitreihen (Novacic 2019) mit F_1 -Werten von bis zu 0,84 erfolgreich angewendet. Die in diesen Fällen analysierten Zeitreihendaten weisen

Strategie- und clusterübergreifende Diskussion

allerdings deutlich andere Charakteristika auf, als die hier analysierten Leistungszeitreihen. Temperaturzeitreihen verlaufen bspw. im Vergleich zu Stromlastgängen deutlich träger. Stromlastgänge hingegen können in kurzer Zeit steil an- bzw. absteigen. Da die Limón GmbH zukünftig voraussichtlich vermehrt Gaslastgänge von Kunden analysieren wird, die im Vergleich zu Stromlastgängen ebenfalls trägere Verläufe aufweisen, könnte die in dieser Arbeit entwickelte Strategie 2 zur Anomalieerkennung auf Grundlage der Normalmodellsschätzung mittels PEWMA an dieser Stelle erfolgreich zum Einsatz kommen.

Abbildungsverzeichnis

Abbildung 2.1: Systematisierung von Erkennungsmethode nach Unterscheidungsmerkmalen (in Anlehnung an Blázquez-García et al. 2020, S.3).....	3 4
Abbildung 2.2: Punktanomalien in univariaten und multivariaten Zeitreihen (Blázquez-García et al. 2020, S. 4).....	5
Abbildung 2.3: Teilsequenzanomalien in univariaten und multivariaten Zeitreihen (Blázquez-García et al. 2020, S. 4).....	6
Abbildung 2.4: Anomale Zeitreihe „Variable 4“ im Kontext von multivariaten Zeitreihen (Blázquez-García et al. 2020, S. 5).....	6
Abbildung 2.5: Kontextuale Anomalie t_2 in Bezug auf jahreszeitlichen Schwankungen (Chandola et al. 2009, S. 8).....	7
Abbildung 2.6: Punktausreißer O1, O2, O3 und O4 resultierend aus der Überschreitung des Schwellenwertes τ (Blázquez-García et al. 2020, S. 6).....	9
Abbildung 2.7: Dichtebasierte Erkennung von Anomalien mit gleitenden Zeitfenstern (Blázquez-García et al. 2020, S. 9).....	11
Abbildung 3.1: Darstellung aller im Clusterverfahren verwendeten nicht normierten Lastgänge.....	19
Abbildung 3.2: Darstellung der 5 gebildeten Cluster der normierten Dauerlinien (großer Rahmen) aus allen verwendeten Unternehmen (kleiner Rahmen, unten links).....	21
Abbildung 3.3: Cluster 1 Unternehmen mit vordefinierten Anomalien.....	22
Abbildung 3.4: Cluster 2 Unternehmen mit vordefinierten Anomalien.....	23
Abbildung 3.5: Cluster 3 Unternehmen mit vordefinierten Anomalien.....	24
Abbildung 3.6: Cluster 4 Unternehmen mit vordefinierten Anomalien.....	25
Abbildung 3.7: Cluster 5 Unternehmen vordefinierten Anomalien	26
Abbildung 3.8: Darstellung eines einzigen (a) nicht rekurrenten und (b) rekurrenten Neurons. Mit σ der Aktivierungsfunktion, b dem Biaswert und w^1 bis w^l den Gewichten (Elsworth und Güttel 2020, S. 2 f.)	29

Abbildungsverzeichnis

Abbildung 4.1: Darstellung (oben) der gegebenen Lastgang Daten, des Cluster 1 Unternehmens, sowie der Schätzung (1M) aus Daten eines Zeitraumes von einem Monat, und (unten) den resultierenden Residuenwerten	35
Abbildung 4.2: Darstellung der identifizierten Anomalien des Cluster 1 Unternehmens, resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$).....	36
Abbildung 4.3: Häufigkeitsdichte der Einzelresiduen und der vier summierten Teilsequenzlängen des Cluster 1 Unternehmens (3M).....	38
Abbildung 4.4: Darstellung der vier aufsummierten Residuenlängen für einen ausgewählten Zeitraum des Monats resultierend aus der Schätzung (3M)	39
Abbildung 4.5: Darstellung der identifizierten Anomalien des Cluster 2 Unternehmens in der Variante 1M, resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$).....	40
Abbildung 4.6: Darstellung der identifizierten Anomalien des Cluster 3 Unternehmens in der Variante 1M, resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$).....	43
Abbildung 4.7: Darstellung der identifizierten Anomalien des Cluster 4 Unternehmens in der Variante 1M, resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$).....	46
Abbildung 4.8: Darstellung (oben) der Schätzung und gegebenen Daten des Cluster 4 Unternehmens in der Variante 1M sowie (unten) die resultierenden Residuenwerte	47
Abbildung 4.9: Darstellung der identifizierten Anomalien des Cluster 5 Unternehmens in der Variante 1M, resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$).....	49
Abbildung 4.10: Darstellung der Schätzung (3M, $R^2 = 0.65$) und der Originaldaten des Clusters 5 Unternehmens	50
Abbildung 4.11: Darstellung der identifizierten Anomalien des Cluster 1 Unternehmens für PEWMA $\alpha_P = 0,3$ resultierend aus den drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 60$).....	52
Abbildung 4.12: Darstellung der glättenden Verläufe für verschieden PEWMA Alphawerte und den jeweils resultierenden Residuen des Clusters 1 Unternehmens.....	54

Abbildungsverzeichnis

Abbildung 4.13: Darstellung der summierten Residuen verschiedener Teilsequenzlängen für (oben) $\alpha_P = 0,3$ und (unten) $\alpha_P = 0,9$ für das Cluster 1 Unternehmen.....	56
Abbildung 4.14: Darstellung der identifizierten Anomalien des Cluster 5 Unternehmens für PEWMA $\alpha_P = 0,9$ resultierend aus dem Z-Score Test und dem ESD Test ($k = 60$).....	59
Abbildung 5.1: Darstellung der durchschnittlichen Trefferquote von Unternehmen C1 bis C5 je Schwellenwert für den Z-Score- und ESD-Test von Strategie 1 (1M & 3M).....	63
Abbildung 5.2: Darstellung F_1 -Maße der Z-Score Tests für jedes Unternehmen des Clusters 1 bis 5 in den Varianten 1M & 3M und Angabe des Verhältnisses von A:B mit A: Anzahl der höheren F_1 -Maße im Vergleich zu 3M und B: Anzahl der höheren F_1 -Maße im Vergleich zu 1M.....	65
Abbildung 5.3: Darstellung der durchschnittlichen F_1 -Maße und Standardabweichung (1M: bläuliche Farbtöne, 3M: rötliche Farbtöne) für die Unternehmen aus C1 bis C5 für aufsummierte Teilsequenzen und Einzelresiduen im Vergleich	67

Tabellenverzeichnis

Tabelle 2-1: Konfusionsmatrix zur Analyse der Klassifikationen einer Anomalieerkennungsmethode	15
Tabelle 3-1: Erläuterung zu Werterepräsentation der Silhouetten Koeffizienten	18
Tabelle 3-2: Ergebnisse aus der Clusterung von 179 Unternehmen für vorgegebenen Clustergrößen k , mit nicht normierten (nn) und normierten (n) Lastgängen und Angabe der Anzahl an Unternehmen im größten (n_I) und zweitgrößten (n_{II}) gebildeten Cluster	20
Tabelle 3-3: Ergebnisse aus der Clusterung von 179 Unternehmen für vorgegebenen Clustergrößen k , mit nicht normierten (nn) und normierten (n) geordneten Lastgängen und Angabe der Anzahl an Unternehmen im größten (n_I) und zweitgrößten (n_{II}) gebildeten Cluster	21
Tabelle 3-4: Übersicht der vordefinierten Anomalien des Cluster 1 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde.....	23
Tabelle 3-5: Übersicht der vordefinierten Anomalien des Cluster 2 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde.....	24
Tabelle 3-6: Übersicht der vordefinierten Anomalien des Cluster 3 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde.....	25
Tabelle 3-7: Übersicht der vordefinierten Anomalien des Cluster 4 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde.....	26
Tabelle 3-8: Übersicht der vordefinierten Anomalien des Cluster 5 Unternehmens mit dem Zeitstempelformat Jahr-Monat-Tag Stunde:Minute:Sekunde.....	27
Tabelle 3-9: Auflistung der Anzahl an Anomalien je Unternehmen eines Clusters	28
Tabelle 3-10: Zusammenfassung der Ergebnisse des ESD Tests (Rosner 1983).....	33
Tabelle 4-1: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 1 Unternehmens (1M) für drei Z-Score Schwellwerte und den drei Signifikanzniveauewerten des ESD Tests ($k = 40$).....	36
Tabelle 4-2: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 1 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauewerten des ESD Tests ($k = 40$). Siehe Anhang A Tabelle 1 & 2 für eine Auflistung der Werte für tp , fp und fn	37

Tabellenverzeichnis

Tabelle 4-3: Berechnete Genauigkeiten (P), Trefferquoten (R) und F_1 -Maße des Cluster 2 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang B Tabelle 1 für eine Auflistung der Werte für tp , fp und fn	41
Tabelle 4-4: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 2 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang B Tabelle 2 & 3 für eine Auflistung der Werte für tp , fp und fn	42
Tabelle 4-5: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 3 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang C Tabelle 1 für eine Auflistung der Werte für tp , fp und fn	43
Tabelle 4-6: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 3 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang C Tabelle 2 & 3 für eine Auflistung der Werte für tp , fp und fn	45
Tabelle 4-7: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 4 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang D Tabelle 1 für eine Auflistung der Werte für tp , fp und fn	46
Tabelle 4-8: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 4 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang D Tabelle 2 für eine Auflistung der Werte für tp , fp und fn	48
Tabelle 4-9: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 5 Unternehmens (1M) für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang E Tabelle 1 für eine Auflistung der Werte für tp , fp und fn	49
Tabelle 4-10: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 5 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang E Tabelle 2 & 3 für eine Auflistung der Werte für tp , fp und fn	51
Tabelle 4-11: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 1 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 60$) bei drei PEWMA Alphawerten	53

Tabellenverzeichnis

Tabelle 4-12: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 1 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang F Tabelle 1, 2 & 3 für eine Auflistung der Werte für tp , fp und fn	55
Tabelle 4-13: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 2 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang G Tabelle 1 für eine Auflistung der Werte für tp , fp und fn	57
Tabelle 4-14: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 2 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 60$). Siehe Anhang G Tabelle 2, 3 & 4 für eine Auflistung der Werte für tp , fp und fn	58
Tabelle 4-15: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 5 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 40$). Siehe Anhang J Tabelle 1 für eine Auflistung der Werte für tp , fp und fn	59
Tabelle 4-16: Berechnete Genauigkeiten, Trefferquoten und F_1 -Maße des Cluster 5 Unternehmens für drei Z-Score Werte und den drei Signifikanzniveauwerten des ESD Tests ($k = 60$). Siehe Anhang 9.) Tabelle 2 für eine Auflistung der Werte für tp , fp und fn	60

Literaturverzeichnis

Aggarwal, Charu C. (2017): *Outlier Analysis*. 2nd ed. 2017. Cham, s.l.: Springer International Publishing.

Arik, Sabri; Huang, Tingwen; Lai, Weng Kin; Liu, Qingshan (Hg.) (2015): *Neural Information Processing*. Cham: Springer International Publishing (Lecture Notes in Computer Science).

Basu, Sabyasachi; Meckesheimer, Martin (2007): Automatic outlier detection for time series: an application to sensor data. In: *Knowl Inf Syst* 11 (2), S. 137–154. DOI: 10.1007/s10115-006-0026-6.

Blázquez-García, Ane; Conde, Angel; Mori, Usue; Lozano, Jose A. (2020): A review on outlier/anomaly detection in time series data. In: *arXiv preprint arXiv:2002.04236v1*.

Breunig, Markus M.; Kriegel, Hans-Peter; Ng, Raymond T.; Sander, Jörg (2000): LOF. In: *SIGMOD Rec.* 29 (2), S. 93–104. DOI: 10.1145/335191.335388.

Brownlee, Jason (2020): *Multivariate Time Series Forecasting with LSTMs in Keras*. Online verfügbar unter <https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>, zuletzt geprüft am 09.11.2020.

Carter, Kevin M.; Streilein, William W. (Hg.) (2012): *2012 IEEE Statistical Signal Processing Workshop (SSP)*. 2012 IEEE Statistical Signal Processing Workshop (SSP). Ann Arbor, MI, USA, 2012: IEEE.

Chandola, Varun; Banerjee, Arindam; Kumar, Vipin (2009): Anomaly detection. In: *ACM Comput. Surv.* 41 (3), S. 1–58. DOI: 10.1145/1541880.1541882.

Chau, Pham Minh; Duc, Bui Minh; Anh, Duong Tuan (2018): *Discord Discovery in Streaming Time Series based on an Improved HOT SAX Algorithm*. In: Unknown (Hg.): *Proceedings of the Ninth International Symposium on Information and Communication Technology - SolCT 2018*. Danang City, Viet Nam, 06.12.2018 - 07.12.2018. New York, New York, USA: ACM Press, S. 24–30.

Chen, Jiyi; Li, Wenyan; Lau, Adriel; Cao, Jiguo; Wang, Ke (2010): Automated Load Curve Data Cleansing in Power Systems. In: *IEEE Trans. Smart Grid* 1 (2), S. 213–221. DOI: 10.1109/TSG.2010.2053052.

Eckmann, J.-P.; Kamphorst, S. Oliffson; Ruelle, D. (1987): Recurrence Plots of Dynamical Systems. In: *Europhys. Lett.* 4 (9), S. 973–977. DOI: 10.1209/0295-5075/4/9/004.

Elsworth, Steven; Güttel, Stefan (2020): Time Series Forecasting Using LSTM Networks: A Symbolic Approach. Online verfügbar unter <http://arxiv.org/pdf/2003.05672v1>.

Han, Jiawei; Kamber, Micheline (2012): Data mining. Concepts and techniques. [Nachdr.]. San Francisco: Morgan Kaufmann (The Morgan Kaufmann series in data management systems).

Hochenbaum, J.; Vallis, O.; Kejariwal, A. (2017): Automatic Anomaly Detection in the Cloud Via Statistical Learning. In: *arXiv preprint arXiv:1704.07706v1*.

Hochreiter, S.; Schmidhuber, J. (1997): Long short-term memory. In: *Neural computation* 9 (8), S. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

Hu, Min; Feng, Xiaowei; Ji, Zhiwei; Yan, Ke; Zhou, Shengchen (2019): A novel computational approach for discord search with local recurrence rates in multivariate time series. In: *Information Sciences* 477, S. 220–233. DOI: 10.1016/j.ins.2018.10.047.

Hundman, Kyle; Constantinou, Valentino; Laporte, Christopher; Colwell, Ian; Soderstrom, Tom (2018): Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. In: *In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, S. 387–395.

Ishimtsev, V.; Bernstein, A.; Burnaev, E.; Nazarov, I. (2017): Conformal k-NN Anomaly Detector for Univariate Data Streams. In: *Proceedings of Machine Learning Research on Conformal and Probabilistic Prediction and Applications* 60, S. 1–15.

Ji, Zhiwei; Wang, Bing; Deng, SuPing; You, Zhuhong (2014): Predicting dynamic deformation of retaining structure by LSSVR-based time series method. In: *Neurocomputing* 137, S. 165–172. DOI: 10.1016/j.neucom.2013.03.073.

Keogh, E.; Lin, J.; Fu, A. (2005): HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence. In: Fifth IEEE International Conference on Data Mining (ICDM'05). Houston, TX, USA, 27-30 Nov. 2005: IEEE, S. 226–233.

Keras Google Group (Hg.): LSTM Keras Bibliothek für Python. Online verfügbar unter https://keras.io/api/layers/recurrent_layers/lstm/, zuletzt geprüft am 09.11.2020.

Kieu, Tung; Yang, Bin; Jensen, Christian S. (2018): Outlier Detection for Multidimensional Time Series Using Deep Neural Networks. In: 2018 19th IEEE International Conference on Mobile Data Management (MDM). Aalborg, Denmark, 2018: IEEE, S. 125–134.

Kruse, Rudolf; Borgelt, Christian; Braune, Christian; Klawonn, Frank; Moewes, Christian; Steinbrecher, Matthias (2015): Computational Intelligence. Eine methodische Einführung in künstliche neuronale Netze, evolutionäre Algorithmen, Fuzzy-Systeme und

Bayes-Netze. 2., überarbeitete und erweiterte Auflage. Wiesbaden: Springer Vieweg (Computational Intelligence).

Lin, Jessica; Keogh, Eamonn; Wei, Li; Lonardi, Stefano (2007): Experiencing SAX: a novel symbolic representation of time series. In: *Data Min Knowl Disc* 15 (2), S. 107–144. DOI: 10.1007/s10618-007-0064-z.

Makridakis, Spyros; Spiliotis, Evangelos; Assimakopoulos, Vassilios (2018): Statistical and Machine Learning forecasting methods: Concerns and ways forward. In: *PloS one* 13 (3), e0194889. DOI: 10.1371/journal.pone.0194889.

Mehrang, Saeed; Helander, Elina; Pavel, Misha; Chieh, Angela; Korhonen, Ilkka: Outlier detection in weight time series of connected scales. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Washington, DC, USA: IEEE, S. 1489–1496.

Munir, Mohsin; Siddiqui, Shoaib Ahmed; Dengel, Andreas; Ahmed, Sheraz (2019): DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series. In: *IEEE Access* 7, S. 1991–2005. DOI: 10.1109/ACCESS.2018.2886457.

Papadimitriou, Spiros; Sun, Jieng; Faloutsos, Christos (2005): Streaming Pattern Discovery in Multiple Time-Series. In: *In Proceedings of the 31st International Conference on Very Large Data Bases* ACM, Trondheim, Norway, S. 697–708.

PEWMA Python (2019). Online verfügbar unter <https://codereview.stackexchange.com/questions/221962/optimising-probabilistic-weighted-moving-average-pewma-df-iterrows-loop-in-pan>, zuletzt geprüft am 14.11.2020.

Renshaw, John (2016): Anomaly Detection Using AWS IoT and AWS Lambda. Online verfügbar unter <https://aws.amazon.com/de/blogs/iot/anomaly-detection-using-aws-iot-and-aws-lambda/>, zuletzt geprüft am 27.10.2020.

Roberts, S. W. (1959): Control Chart Tests Based on Geometric Moving Averages. In: *Technometrics* 1 (3), S. 239. DOI: 10.2307/1266443.

Rosner, Bernard (1983): Percentage Points for a Generalized ESD Many-Outlier Procedure. In: *Technometrics* 25 (2), S. 165. DOI: 10.2307/1268549.

Rousseeuw, Peter J. (1987): Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. In: *Journal of Computational and Applied Mathematics* 20, S. 53–65. DOI: 10.1016/0377-0427(87)90125-7.

Schlittgen, Rainer (2001): Zeitreihenanalyse. s.l.: Oldenbourg Wissenschaftsverlag (Lehr- und Handbücher der Statistik). Online verfügbar unter <http://lib.mylibrary.com/detail.asp?id=609353>.

Literaturverzeichnis

Schlittgen, Rainer; Sattarhoff, Cristina (2020): Angewandte Zeitreihenanalyse mit R. 4., erweiterte und aktualisierte Auflage (Lehr- und Handbücher der Statistik).

Senin, Pavel; Lin, Jessica; Wang, Xing; Oates, Tim; Gandhi, Sunil; Boedihardjo, Arnold P. et al. (2015): Time series anomaly discovery with grammar-based compression.

Siami-Namini, Sima; Tavakoli, Neda; Siami Namin, Akbar (2018): A Comparison of ARIMA and LSTM in Forecasting Time Series. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). Orlando, FL, 2018: IEEE, S. 1394–1401.

Su, Ya; Zhao, Youjian; Niu, Chenhao; Liu, Rong; Sun, Wei; Pei, Dan (2019): Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. In: Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi und George Karypis (Hg.): Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage AK USA, 2019. New York, NY, USA: ACM, S. 2828–2837.

Wang, Xing; Lin, Jessica; Patel, Nital; Braun, Martin (2018): Exact variable-length anomaly detection algorithm for univariate and multivariate time series. In: *Data Min Knowl Disc* 32 (6), S. 1806–1844. DOI: 10.1007/s10618-018-0569-7.

Wooldridge, Jeffrey M. (2013): Introductory econometrics. A modern approach. 5. ed., internat. ed.

Zhou, Yuxun; Arghandeh, Reza; Zou, Han; Spanos, Costas J. (2019): Nonparametric Event Detection in Multiple Time Series for Power Distribution Networks. In: *IEEE Trans. Ind. Electron.* 66 (2), S. 1619–1628. DOI: 10.1109/TIE.2018.2840508.

Zwerenz, Karlheinz (2012): Statistik. Einführung in die computergestützte Datenanalyse. 5., überarb. Aufl. München: Oldenbourg (Managementwissen für Studium und Praxis). Online verfügbar unter <http://www.oldenbourg-link.com/isbn/9783486707694>.

Anhang

A

Tabelle 1

C1 (1M, R² = 0,96)	Genauigkeit (R)	Trefferquote (P)	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,05$ sum4	0,25	0,33	0,29	1	3	2
ESD $\alpha = 0,05$ sum2	0,10	0,33	0,15	1	9	2
ESD $\alpha = 0,10$ sum8	0,00	0,00	0,00	0	0	3
ESD $\alpha = 0,10$ sum4	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,10$ sum2	0,08	0,33	0,13	1	11	2
ESD $\alpha = 0,15$ sum6	1,00	0,33	0,50	1	0	2
ESD $\alpha = 0,15$ sum4	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,15$ sum2	0,08	0,33	0,13	1	12	2
Z-Score > 3 sum8	0,20	0,33	0,25	1	4	2
Z-Score > 3 sum6	0,11	0,33	0,17	1	8	2
Z-Score > 3 sum4	0,08	0,33	0,13	1	11	2
Z-Score > 3 sum2	0,08	0,33	0,13	1	11	2
Z-Score > 2,5 sum8	0,06	0,33	0,11	1	15	2
Z-Score > 2,5 sum6	0,05	0,33	0,09	1	18	2
Z-Score > 2,5 sum4	0,04	0,33	0,08	1	22	2
Z-Score > 2,5 sum2	0,04	0,33	0,07	1	26	2
Z-Score > 2 sum8	0,05	0,67	0,09	2	39	1
Z-Score > 2 sum6	0,05	0,67	0,09	2	38	1
Z-Score > 2 sum4	0,03	0,33	0,05	1	38	2
Z-Score > 2 sum2	0,07	1,00	0,13	3	40	0

Tabelle 2

C1 (3M, R² = 0,95)	Genauigkeit (R)	Trefferquote (P)	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,10$ sum4	0	0	0	0	1	3
ESD $\alpha = 0,10$ sum2	0,00	0,00	0,00	0	10	3
ESD $\alpha = 0,15$ sum6	0,00	0,00	0,00	1	0	2
ESD $\alpha = 0,15$ sum4	0,17	0,33	0,22	1	5	2
ESD $\alpha = 0,15$ sum2	1,00	0,33	0,50	1	0	2
Z-Score > 3 sum8	0,08	0,33	0,13	1	11	2
Z-Score > 3 sum6	0,09	0,33	0,14	1	10	2
Z-Score > 3 sum4	0,08	0,33	0,13	1	11	2
Z-Score > 3 sum2	0,07	0,33	0,12	1	13	2
Z-Score > 2,5 sum8	0,11	0,67	0,19	2	16	1
Z-Score > 2,5 sum6	0,11	0,67	0,19	2	16	1
Z-Score > 2,5 sum4	0,10	0,67	0,17	2	18	1
Z-Score > 2,5 sum2	0,05	0,33	0,08	1	20	2
Z-Score > 2 sum8	0,05	0,67	0,10	2	35	1
Z-Score > 2 sum6	0,04	0,67	0,08	2	47	1
Z-Score > 2 sum4	0,07	1,00	0,14	3	38	0
Z-Score > 2 sum2	0,06	1,00	0,12	3	45	0

B

Tabelle 1

C2 (1M, R² = 0,87)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,15$	0,13	0,17	0,14	2	14	10
ESD $\alpha = 0,10$	0,13	0,17	0,15	2	13	10
ESD $\alpha = 0,05$	0,13	0,17	0,15	2	13	10
Z-Score > 3	0,15	0,17	0,16	2	11	10
Z-Score > 2,5	0,10	0,17	0,12	2	19	10
Z-Score > 2	0,15	0,42	0,22	5	29	7
C2 (3M, R² = 0,92)						
ESD $\alpha = 0,15$	0,21	0,25	0,23	3	11	9
ESD $\alpha = 0,10$	0,25	0,25	0,25	3	9	9
ESD $\alpha = 0,05$	0,33	0,25	0,29	3	6	9
Z-Score > 3	0,23	0,25	0,24	3	10	9
Z-Score > 2,5	0,13	0,25	0,17	3	20	9
Z-Score > 2	0,16	0,50	0,24	6	32	6

Tabelle 2

C2 (1M, R² = 0,87)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,05$ sum4	0,07	0,08	0,08	1	13	11
ESD $\alpha = 0,10$ sum4	0,06	0,08	0,07	1	16	11
ESD $\alpha = 0,15$ sum6	0,05	0,08	0,06	1	20	11
ESD $\alpha = 0,15$ sum4	0,06	0,08	0,07	1	16	11
Z-Score > 3 sum6	0,10	0,08	0,09	1	9	11
Z-Score > 3 sum4	0,06	0,08	0,07	1	15	11
Z-Score > 3 sum2	0,07	0,08	0,08	1	13	11
Z-Score > 2,5 sum8	0,05	0,08	0,06	1	21	11
Z-Score > 2,5 sum6	0,08	0,17	0,11	2	23	10
Z-Score > 2,5 sum4	0,07	0,17	0,10	2	26	10
Z-Score > 2,5 sum2	0,09	0,25	0,13	3	30	9
Z-Score > 2 sum8	0,05	0,20	0,08	2	36	8
Z-Score > 2 sum6	0,05	0,22	0,09	2	36	7
Z-Score > 2 sum4	0,08	0,27	0,12	3	37	8
Z-Score > 2 sum2	0,12	0,42	0,18	5	38	7

Tabelle 3

C2 (3M, R² = 0,92)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,05$ sum4	0,09	0,08	0,09	1	10	11
ESD $\alpha = 0,05$ sum2	0,10	0,17	0,12	2	19	10
ESD $\alpha = 0,10$ sum4	0,08	0,08	0,08	1	11	11
ESD $\alpha = 0,10$ sum2	0,10	0,17	0,12	2	19	10
ESD $\alpha = 0,15$ sum4	0,08	0,08	0,08	1	11	11
ESD $\alpha = 0,15$ sum2	0,14	0,25	0,18	3	19	9
Z-Score > 3 sum8	0,17	0,08	0,11	1	5	11
Z-Score > 3 sum6	0,08	0,08	0,08	1	12	11
Z-Score > 3 sum4	0,07	0,08	0,08	1	13	11
Z-Score > 3 sum2	0,07	0,08	0,07	1	14	11
Z-Score > 2,5 sum8	0,05	0,08	0,06	1	18	11
Z-Score > 2,5 sum6	0,09	0,17	0,12	2	20	10
Z-Score > 2,5 sum4	0,11	0,25	0,15	3	25	9
Z-Score > 2,5 sum2	0,11	0,33	0,17	4	31	8
Z-Score > 2 sum8	0,10	0,33	0,15	4	38	8
Z-Score > 2 sum6	0,11	0,42	0,18	5	39	7
Z-Score > 2 sum4	0,10	0,33	0,15	4	37	8
Z-Score > 2 sum2	0,11	0,42	0,18	5	39	7

C

Tabelle 1

C3 (1M, R² = 0,92)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,05$	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,10$	0,17	0,33	0,22	1	5	2
ESD $\alpha = 0,15$	0,17	0,33	0,22	1	5	2
Z-Score > 3	0,20	0,67	0,31	2	8	1
Z-Score > 2,5	0,10	0,67	0,17	2	18	1
Z-Score > 2	0,06	0,67	0,11	2	32	1
C3 (3M, R² = 0,98)						
ESD $\alpha = 0,05$	1,00	0,33	0,50	1	0	2
ESD $\alpha = 0,10$	1,00	0,33	0,50	1	0	2
ESD $\alpha = 0,15$	0,50	0,33	0,40	1	1	2
Z-Score > 3	0,29	0,67	0,40	2	5	1
Z-Score > 2,5	0,11	0,67	0,19	2	16	1
Z-Score > 2	0,07	0,67	0,12	2	28	1

Tabelle 2

C3 (1M, R² = 0,92)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,10$ sum2	0,14	0,33	0,20	1	6	2
ESD $\alpha = 0,15$ sum4	0,25	0,67	0,36	2	6	1
ESD $\alpha = 0,15$ sum2	0,25	0,67	0,36	2	6	1
Z-Score > 3 sum8	0,22	0,67	0,33	2	7	1
Z-Score > 3 sum6	0,20	0,67	0,31	2	8	1
Z-Score > 3 sum4	0,22	0,67	0,33	2	7	1
Z-Score > 3 sum2	0,20	0,67	0,31	2	8	1
Z-Score > 2,5 sum8	0,13	0,50	0,20	2	14	2
Z-Score > 2,5 sum6	0,14	0,67	0,24	2	12	1
Z-Score > 2,5 sum4	0,17	0,67	0,27	2	10	1
Z-Score > 2,5 sum2	0,14	0,67	0,24	2	12	1
Z-Score > 2 sum8	0,09	0,67	0,15	2	21	1
Z-Score > 2 sum6	0,09	0,67	0,16	2	20	1
Z-Score > 2 sum4	0,11	0,67	0,18	2	17	1
Z-Score > 2 sum2	0,08	0,67	0,14	2	24	1

Tabelle 3

C3 (3M, R² = 0,98)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,10$ sum4	0,50	0,33	0,40	1	1	2
ESD $\alpha = 0,10$ sum2	1,00	0,33	0,50	1	0	2
ESD $\alpha = 0,15$ sum4	0,50	0,33	0,40	1	1	2
ESD $\alpha = 0,15$ sum2	1,00	0,33	0,50	1	0	2
Z-Score > 3 sum8	0,50	0,33	0,40	1	1	2
Z-Score > 3 sum6	0,20	0,33	0,25	1	4	2
Z-Score > 3 sum4	0,40	0,67	0,50	2	3	1
Z-Score > 3 sum2	0,50	0,67	0,57	2	2	1
Z-Score > 2,5 sum8	0,06	0,33	0,11	1	15	2
Z-Score > 2,5 sum6	0,15	0,67	0,25	2	11	1
Z-Score > 2,5 sum4	0,17	0,67	0,27	2	10	1
Z-Score > 2,5 sum2	0,15	0,67	0,25	2	11	1
Z-Score > 2 sum8	0,09	0,67	0,15	2	21	1
Z-Score > 2 sum6	0,10	0,67	0,17	2	19	1
Z-Score > 2 sum4	0,10	0,67	0,17	2	18	1
Z-Score > 2 sum2	0,07	0,67	0,13	2	25	1

D

Tabelle 1

C4 (1M, R² = 0,93)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,15$	0,33	0,14	0,20	1	2	6
ESD $\alpha = 0,10$	0,33	0,14	0,20	1	2	6
ESD $\alpha = 0,05$	0,50	0,14	0,22	1	1	6
Z-Score > 3	0,25	0,29	0,27	2	6	5
Z-Score > 2,5	0,17	0,43	0,24	3	15	4
Z-Score > 2	0,08	0,43	0,13	3	35	4
C4 (3M, R² = 0,83)						
Z-Score > 2	0,02	0,14	0,03	1	59	6

Tabelle 2

C4 (1M, R² = 0,93)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
Z-Score > 3 sum2	0,22	0,29	0,25	2	7	5
Z-Score > 2,5 sum6	0,33	0,14	0,20	1	2	6
Z-Score > 2,5 sum4	0,09	0,14	0,11	1	10	6
Z-Score > 2,5 sum2	0,15	0,29	0,20	2	11	5
Z-Score > 2 sum8	0,14	0,14	0,14	1	6	6
Z-Score > 2 sum6	0,09	0,14	0,11	1	10	6
Z-Score > 2 sum4	0,12	0,29	0,17	2	15	5
Z-Score > 2 sum2	0,10	0,29	0,15	2	18	5
C4 (3M, R² = 0,83)						
Z-Score > 2,5 sum4	0,07	0,14	0,10	1	13	6
Z-Score > 2,5 sum2	0,05	0,14	0,07	1	20	6
Z-Score > 2 sum8	0,04	0,14	0,06	1	25	6
Z-Score > 2 sum6	0,04	0,14	0,06	1	24	6
Z-Score > 2 sum4	0,06	0,14	0,09	1	15	6
Z-Score > 2 sum2	0,03	0,14	0,05	1	34	6

E

Tabelle 1

C5 (1M, R² = 0,58)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,15$	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,10$	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,05$	0,20	0,33	0,25	1	4	2
Z-Score > 3	0,20	0,67	0,31	2	8	1
Z-Score > 2,5	0,12	0,67	0,20	2	15	1
Z-Score > 2	0,08	0,67	0,14	2	24	1
C5 (3M, R² = 0,62)						
ESD $\alpha = 0,15$	0,50	0,33	0,40	1	1	2
ESD $\alpha = 0,10$	0,50	0,33	0,40	1	1	2
ESD $\alpha = 0,05$	1,00	0,33	0,50	1	0	2
Z-Score > 3	0,17	0,33	0,22	1	5	2
Z-Score > 2,5	0,17	0,67	0,27	2	10	1
Z-Score > 2	0,15	1,00	0,26	3	17	0

Tabelle 2

C5 (1M, R² = 0,58)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,05$ sum2	0,08	0,33	0,13	1	11	2
ESD $\alpha = 0,10$ sum2	0,06	0,33	0,11	1	15	2
ESD $\alpha = 0,15$ sum2	0,06	0,33	0,10	1	16	2
Z-Score > 3 sum8	0,17	0,33	0,22	1	5	2
Z-Score > 3 sum6	0,25	0,33	0,29	1	3	2
Z-Score > 3 sum4	0,10	0,33	0,15	1	9	2
Z-Score > 3 sum2	0,08	0,33	0,13	1	11	2
Z-Score > 2,5 sum8	0,20	0,33	0,25	1	4	2
Z-Score > 2,5 sum6	0,05	0,33	0,09	1	18	2
Z-Score > 2,5 sum4	0,06	0,33	0,11	1	15	2
Z-Score > 2,5 sum2	0,05	0,33	0,09	1	18	2
Z-Score > 2 sum8	0,03	0,33	0,06	1	32	2
Z-Score > 2 sum6	0,03	0,33	0,06	1	28	2
Z-Score > 2 sum4	0,04	0,33	0,07	1	24	2
Z-Score > 2 sum2	0,08	0,67	0,15	2	22	1

Tabelle 3

C5 (3M, R² = 0,65)	Genauigkeit	Trefferquote	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,05$ sum2	0,50	0,33	0,40	1	1	2
ESD $\alpha = 0,10$ sum4	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,10$ sum2	0,50	0,33	0,40	1	1	2
ESD $\alpha = 0,15$ sum4	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,15$ sum2	0,33	0,33	0,33	1	2	2
Z-Score > 3 sum6	0,25	0,33	0,29	1	3	2
Z-Score > 3 sum4	0,29	0,67	0,40	2	5	1
Z-Score > 3 sum2	0,25	0,33	0,29	1	3	2
Z-Score > 2,5 sum6	0,13	0,33	0,18	1	7	2
Z-Score > 2,5 sum4	0,14	0,67	0,24	2	12	1
Z-Score > 2,5 sum2	0,20	0,67	0,31	2	8	1
Z-Score > 2 sum8	0,10	1,00	0,18	3	27	0
Z-Score > 2 sum6	0,12	1,00	0,21	3	22	0
Z-Score > 2 sum4	0,13	1,00	0,22	3	21	0
Z-Score > 2 sum2	0,09	1,00	0,17	3	30	0

F

Tabelle 1

C1 ($\alpha_P = 0,3$)	Genauigkeit (<i>R</i>)	Trefferquote (<i>P</i>)	<i>F₁</i> -Maß	<i>tp</i>	<i>fp</i>	<i>fn</i>
ESD $\alpha = 0,05$ sum4	0	0	0	0	2	3
ESD $\alpha = 0,05$ sum2	0	0	0	0	15	2
ESD $\alpha = 0,10$ sum4	0	0	0	0	2	3
ESD $\alpha = 0,10$ sum2	0	0	0	0	9	2
ESD $\alpha = 0,15$ sum4	0	0	0	0	2	3
ESD $\alpha = 0,15$ sum2	0	0	0	0	7	2
Z-Score > 3 sum8	0	0	0	0	0	0
Z-Score > 3 sum6	0	0	0	0	4	3
Z-Score > 3 sum4	0,14	0,33	0,20	1	6	2
Z-Score > 3 sum2	0	0	0	0	18	3
Z-Score > 2,5 sum8	0	0	0	0	5	3
Z-Score > 2,5 sum6	0,11	0,33	0,17	1	8	2
Z-Score > 2,5 sum4	0,04	0,33	0,08	1	22	2
Z-Score > 2,5 sum2	0,03	0,33	0,05	1	36	2
Z-Score > 2 sum8	0,12	0,67	0,20	2	15	1
Z-Score > 2 sum6	0,05	0,67	0,09	2	38	1
Z-Score > 2 sum4	0,02	0,33	0,03	1	57	2
Z-Score > 2 sum2	0,02	0,33	0,03	1	60	2

Tabelle 2

C1 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,05$ sum4	0	0	0	0	7	3
ESD $\alpha = 0,05$ sum2	0	0	0	0	3	2
ESD $\alpha = 0,10$ sum4	0	0	0	0	2	3
ESD $\alpha = 0,10$ sum2	0	0	0	0	2	2
ESD $\alpha = 0,15$ sum4	0	0	0	0	2	3
ESD $\alpha = 0,15$ sum2	0	0	0	0	1	2
Z-Score > 3 sum8	0	0	0	0	0	3
Z-Score > 3 sum6	0	0	0	0	3	3
Z-Score > 3 sum4	0,17	0,33	0,22	1	5	2
Z-Score > 3 sum2	0	0	0	0	10	3
Z-Score > 2,5 sum8	0	0	0	0	5	3
Z-Score > 2,5 sum6	0,13	0,33	0,18	1	7	2
Z-Score > 2,5 sum4	0,05	0,33	0,09	1	19	2
Z-Score > 2,5 sum2	0,03	0,33	0,06	1	32	2
Z-Score > 2 sum8	0,13	0,67	0,22	2	13	1
Z-Score > 2 sum6	0,06	0,67	0,11	2	32	1
Z-Score > 2 sum4	0,02	0,33	0,03	1	56	2
Z-Score > 2 sum2	0,03	0,67	0,06	2	61	1

Tabelle 3

C1 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$ sum2	0	0	0	0	8	3
ESD $\alpha = 0,10$ sum2	0	0	0	0	4	3
ESD $\alpha = 0,05$ sum2	0	0	0	0	3	3
Z-Score > 3 sum8	0	0	0	0	0	3
Z-Score > 3 sum6	0	0	0	0	1	3
Z-Score > 3 sum4	0,33	0,33	0,33	1	2	2
Z-Score > 3 sum2	0,00	0,00	0	0	3	3
Z-Score > 2,5 sum8	0,40	0,67	0,50	2	3	1
Z-Score > 2,5 sum6	0,33	0,67	0,44	2	4	1
Z-Score > 2,5 sum4	0,10	0,33	0,15	1	9	2
Z-Score > 2,5 sum2	0,07	0,33	0,12	1	13	2
Z-Score > 2 sum8	0,10	0,67	0,17	2	19	1
Z-Score > 2 sum6	0,07	0,67	0,13	2	25	1
Z-Score > 2 sum4	0,05	0,67	0,10	2	35	1
Z-Score > 2 sum2	0,05	0,67	0,09	2	40	1

G

Tabelle 1

C2 ($\alpha_P = 0,3$)	Genauigkeit	Trefferquote	F_1-Maß	tp	fp	fn
ESD $\alpha = 0,15$	0,09	0,33	0,14	4	41	8
ESD $\alpha = 0,10$	0,09	0,33	0,14	4	41	8
ESD $\alpha = 0,05$	0,09	0,33	0,14	4	41	8
Z-Score > 3	0,05	0,08	0,06	1	18	11
Z-Score > 2,5	0,11	0,33	0,16	4	34	8
Z-Score > 2	0,09	0,33	0,14	4	43	8
C2 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,15$	0,09	0,25	0,14	3	29	9
ESD $\alpha = 0,10$	0,09	0,17	0,12	2	20	10
ESD $\alpha = 0,05$	0,15	0,17	0,16	2	11	10
Z-Score > 3	0,11	0,17	0,13	2	16	10
Z-Score > 2,5	0,09	0,25	0,13	3	30	9
Z-Score > 2	0,07	0,25	0,10	3	43	9
C2 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$	0,30	0,25	0,27	3	7	9
ESD $\alpha = 0,10$	0,22	0,17	0,19	2	7	10
ESD $\alpha = 0,05$	0,40	0,17	0,24	2	3	10
Z-Score > 3	0,12	0,17	0,14	2	15	10
Z-Score > 2,5	0,09	0,17	0,12	2	20	10
Z-Score > 2	0,05	0,17	0,08	2	35	10

Anhang

Tabelle 2

C2 ($\alpha_P = 0,3$)	Genauigkeit (R)	Trefferquote (P)	F₁-Maß	tp	fp	fn
ESD $\alpha = 0,05$ sum2	0,13	0,17	0,15	2	13	10
ESD $\alpha = 0,10$ sum2	0,13	0,17	0,14	2	14	10
ESD $\alpha = 0,15$ sum4	0,25	0,17	0,20	2	6	10
ESD $\alpha = 0,15$ sum2	0,11	0,17	0,13	2	16	10
Z-Score > 3 sum8	0	0	0	0	0	12
Z-Score > 3 sum6	0	0	0	0	0	12
Z-Score > 3 sum4	0	0	0	0	0	12
Z-Score > 3 sum2	0	0	0	0	0	12
Z-Score > 2,5 sum8	0	0	0	0	0	12
Z-Score > 2,5 sum6	1,00	0,17	0,29	2	0	10
Z-Score > 2,5 sum4	0,18	0,17	0,17	2	9	10
Z-Score > 2,5 sum2	0,13	0,17	0,14	2	14	10
Z-Score > 2 sum8	1,00	0,17	0,29	2	0	10
Z-Score > 2 sum6	0,12	0,17	0,14	2	15	10
Z-Score > 2 sum4	0,08	0,17	0,11	2	22	10
Z-Score > 2 sum2	0,06	0,17	0,09	2	29	10

Tabelle 3

C2 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,15$ sum2	0,15	0,17	0,16	2	11	10
ESD $\alpha = 0,10$ sum2	0,17	0,17	0,17	2	10	10
ESD $\alpha = 0,05$ sum2	0,29	0,17	0,21	2	5	10
Z-Score > 3 sum8	0	0	0	0	0	12
Z-Score > 3 sum6	0	0	0	0	0	12
Z-Score > 3 sum4	0,17	0,17	0,17	2	10	10
Z-Score > 3 sum2	0,14	0,17	0,15	2	12	10
Z-Score > 2,5 sum8	0	0	0	0	0	12
Z-Score > 2,5 sum6	0,12	0,17	0,14	2	15	10
Z-Score > 2,5 sum4	0,09	0,17	0,11	2	21	10
Z-Score > 2,5 sum2	0,06	0,17	0,09	2	29	10
Z-Score > 2 sum8	0,06	0,17	0,09	2	32	10
Z-Score > 2 sum6	0,05	0,17	0,07	2	42	10
Z-Score > 2 sum4	0,04	0,17	0,07	2	45	10
Z-Score > 2 sum2	0,08	0,42	0,14	5	54	7

Tabelle 4

C2 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$ sum2	0,17	0,17	0,17	2	10	10
ESD $\alpha = 0,10$ sum2	0,18	0,17	0,17	2	9	10
ESD $\alpha = 0,05$ sum2	0	0	0	0	0	12
Z-Score > 3 sum8	0	0	0	0	0	12
Z-Score > 3 sum6	0	0	0	0	2	12
Z-Score > 3 sum4	0,10	0,08	0,09	1	9	12
Z-Score > 3 sum2	0,17	0,17	0,17	2	10	12
Z-Score > 2,5 sum8	0	0	0	0	6	12
Z-Score > 2,5 sum6	0	0	0	0	19	12
Z-Score > 2,5 sum4	0,10	0,17	0,13	2	18	12
Z-Score > 2,5 sum2	0,10	0,17	0,12	2	19	12
Z-Score > 2 sum8	0,00	0,00	0	0	41	12
Z-Score > 2 sum6	0,02	0,08	0,04	1	42	12
Z-Score > 2 sum4	0,05	0,17	0,07	2	40	12
Z-Score > 2 sum2	0,06	0,17	0,09	2	32	11

H

Tabelle 1

C3 ($\alpha_P = 0,3$)	Genauigkeit	Trefferquote	F_1-Maß	tp	fp	fn
ESD $\alpha = 0,15$	0	0	0	0	50	3
ESD $\alpha = 0,10$	0	0	0	0	50	3
ESD $\alpha = 0,05$	0	0	0	0	31	3
Z-Score > 3	0	0	0	0	24	3
Z-Score > 2,5	0	0	0	0	42	3
Z-Score > 2	0,01	0,33	0,03	1	67	2
C3 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,15$	0	0	0	0	0	3
ESD $\alpha = 0,10$	0	0	0	0	0	3
ESD $\alpha = 0,05$	0	0	0	0	0	3
Z-Score > 3	0	0	0	0	14	3
Z-Score > 2,5	0	0	0	0	33	3
Z-Score > 2	0	0	0	0	59	3
C3 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$	0	0	0	0	0	3
ESD $\alpha = 0,10$	0	0	0	0	0	3
ESD $\alpha = 0,05$	0	0	0	0	0	3
Z-Score > 3	0	0	0	0	0	3
Z-Score > 2,5	0	0	0	0	0	3
Z-Score > 2	0,04	0,33	0,07	1	24	2

Anhang

Tabelle 2

C3 ($\alpha_P = 0,3$)	Genauigkeit (<i>R</i>)	Trefferquote (<i>P</i>)	<i>F</i> ₁ -Maß	<i>tp</i>	<i>fp</i>	<i>fn</i>
ESD $\alpha = 0,15$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,10$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,05$ sum2	0	0	0	0	0	0
Z-Score > 3 sum8	0	0	0	0	0	0
Z-Score > 3 sum6	0	0	0	0	0	0
Z-Score > 3 sum4	0	0	0	0	0	0
Z-Score > 3 sum2	0	0	0	0	0	0
Z-Score > 2,5 sum8	0	0	0	0	0	0
Z-Score > 2,5 sum6	0	0	0	0	0	0
Z-Score > 2,5 sum4	0,25	0,33	0,29	1	3	2
Z-Score > 2,5 sum2	0	0	0	0	43	3
Z-Score > 2 sum8	0	0	0	0	0	3
Z-Score > 2 sum6	0,22	0,67	0,33	2	7	1
Z-Score > 2 sum4	0,02	0,33	0,04	1	42	2
Z-Score > 2 sum2	0	0	0	0	81	3

Tabelle 3

C3 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,15$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,10$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,05$ sum2	0	0	0	0	0	0
Z-Score > 3 sum8	0	0	0	0	0	0
Z-Score > 3 sum6	0	0	0	0	0	0
Z-Score > 3 sum4	0	0	0	0	0	0
Z-Score > 3 sum2	0	0	0	0	2	3
Z-Score > 2,5 sum8	0	0	0	0	0	3
Z-Score > 2,5 sum6	0	0	0	0	0	3
Z-Score > 2,5 sum4	0,50	0,33	0,40	1	1	2
Z-Score > 2,5 sum2	0	0	0	0	21	3
Z-Score > 2 sum8	0	0	0	0	0	3
Z-Score > 2 sum6	0,13	0,33	0,18	1	7	2
Z-Score > 2 sum4	0,03	0,33	0,05	1	38	2
Z-Score > 2 sum2	0	0	0	0	77	3

Tabelle 4

C3 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,10$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,05$ sum2	0	0	0	0	0	0
Z-Score > 3 sum8	0	0	0	0	0	0
Z-Score > 3 sum6	0	0	0	0	0	0
Z-Score > 3 sum4	0	0	0	0	0	0
Z-Score > 3 sum2	0	0	0	0	0	0
Z-Score > 2,5 sum8	0	0	0	0	0	0
Z-Score > 2,5 sum6	0	0	0	0	0	0
Z-Score > 2,5 sum4	0	0	0	0	0	0
Z-Score > 2,5 sum2	0	0	0	0	0	0
Z-Score > 2 sum8	0,40	0,67	0,50	2	3	1
Z-Score > 2 sum6	0,20	0,33	0,25	1	4	2
Z-Score > 2 sum4	0,08	0,33	0,13	1	11	2
Z-Score > 2 sum2	0,06	0,33	0,10	1	17	2

I

Tabelle 1

C4 ($\alpha_P = 0,3$)	Genauigkeit	Trefferquote	F_1-Maß	tp	fp	fn
ESD $\alpha = 0,15$	0,03	0,14	0,04	1	39	6
ESD $\alpha = 0,10$	0,03	0,14	0,04	1	39	6
ESD $\alpha = 0,05$	0,03	0,14	0,04	1	39	6
Z-Score > 3	0,04	0,14	0,06	1	24	6
Z-Score > 2,5	0,03	0,14	0,04	1	39	6
Z-Score > 2	0,05	0,29	0,08	2	42	5
C4 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,15$	0,02	0,14	0,04	1	41	6
ESD $\alpha = 0,10$	0,03	0,14	0,04	1	38	6
ESD $\alpha = 0,05$	0	0	0	0	15	7
Z-Score > 3	0	0	0	0	22	7
Z-Score > 2,5	0,02	0,14	0,04	1	40	6
Z-Score > 2	0,02	0,14	0,03	1	55	6
C4 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$	0	0	0	0	0	7
ESD $\alpha = 0,10$	0	0	0	0	0	7
ESD $\alpha = 0,05$	0	0	0	0	0	7
Z-Score > 3	0	0	0	0	0	7
Z-Score > 2,5	0	0	0	0	7	7
Z-Score > 2	0,02	0,14	0,03	1	64	6

Tabelle 2

C4 ($\alpha_P = 0,3$)	Genauigkeit (<i>R</i>)	Trefferquote (<i>P</i>)	<i>F</i>₁-Maß	<i>tp</i>	<i>fp</i>	<i>fn</i>
ESD $\alpha = 0,15$ sum2	0,03	0,14	0,04	1	39	6
ESD $\alpha = 0,10$ sum2	0,03	0,14	0,04	1	39	6
ESD $\alpha = 0,05$ sum2	0,03	0,14	0,04	1	39	6
Z-Score > 3 sum8	0	0	0	0	0	0
Z-Score > 3 sum6	0	0	0	0	0	0
Z-Score > 3 sum4	0	0	0	0	0	0
Z-Score > 3 sum2	0	0	0	0	0	0
Z-Score > 2,5 sum8	0	0	0	0	0	0
Z-Score > 2,5 sum6	0	0	0	0	0	0
Z-Score > 2,5 sum4	0	0	0	0	0	0
Z-Score > 2,5 sum2	0	0	0	0	0	0
Z-Score > 2 sum8	0	0	0	0	0	0
Z-Score > 2 sum6	0	0	0	0	16	7
Z-Score > 2 sum4	0,02	0,29	0,04	2	86	5
Z-Score > 2 sum2	0,03	0,29	0,05	2	74	5

Tabelle 3

C4 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,15$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,10$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,05$ sum2	0	0	0	0	0	0
Z-Score > 3 sum8	0	0	0	0	0	0
Z-Score > 3 sum6	0	0	0	0	0	0
Z-Score > 3 sum4	0	0	0	0	0	0
Z-Score > 3 sum2	0	0	0	0	15	7
Z-Score > 2,5 sum8	0	0	0	0	0	3
Z-Score > 2,5 sum6	0	0	0	0	0	3
Z-Score > 2,5 sum4	0	0	0	0	11	7
Z-Score > 2,5 sum2	0,04	0,29	0,08	2	44	5
Z-Score > 2 sum8	0	0	0	0	0	7
Z-Score > 2 sum6	0	0	0	0	20	7
Z-Score > 2 sum4	0,02	0,29	0,04	2	83	5
Z-Score > 2 sum2	0,03	0,29	0,05	2	77	5

Tabelle 4

C4 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,10$ sum2	0	0	0	0	0	0
ESD $\alpha = 0,05$ sum2	0	0	0	0	0	0
Z-Score > 3 sum8	0	0	0	0	0	0
Z-Score > 3 sum6	0	0	0	0	0	0
Z-Score > 3 sum4	0	0	0	0	0	0
Z-Score > 3 sum2	0	0	0	0	1	7
Z-Score > 2,5 sum8	0	0	0	0	0	7
Z-Score > 2,5 sum6	0	0	0	0	3	7
Z-Score > 2,5 sum4	0	0	0	0	4	7
Z-Score > 2,5 sum2	0	0	0	0	6	7
Z-Score > 2 sum8	0	0	0	0	22	7
Z-Score > 2 sum6	0	0	0	0	38	7
Z-Score > 2 sum4	0,05	0,29	0,09	2	38	5
Z-Score > 2 sum2	0,02	0,14	0,03	1	50	6

J

Tabelle 1

C5 ($\alpha_P = 0,3$)	Genauigkeit	Trefferquote	F_1-Maß	tp	fp	fn
ESD $\alpha = 0,15$	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,10$	0,25	0,33	0,29	1	3	2
ESD $\alpha = 0,05$	0,25	0,33	0,29	1	3	2
Z-Score > 3	0,09	0,33	0,14	1	10	2
Z-Score > 2,5	0,05	0,33	0,09	1	18	2
Z-Score > 2	0,03	0,33	0,05	1	38	2
C5 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,15$	0,14	0,33	0,20	1	6	2
ESD $\alpha = 0,10$	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,05$	0,33	0,33	0,33	1	2	2
Z-Score > 3	0,09	0,33	0,14	1	10	2
Z-Score > 2,5	0,11	0,67	0,19	2	16	1
Z-Score > 2	0,06	0,67	0,10	2	34	1
C5 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$	0,14	0,33	0,20	1	6	2
ESD $\alpha = 0,10$	0,17	0,33	0,22	1	5	2
ESD $\alpha = 0,05$	0	0	0	0	2	3
Z-Score > 3	0,18	0,67	0,29	2	9	1
Z-Score > 2,5	0,12	0,67	0,20	2	15	1
Z-Score > 2	0,08	1,00	0,14	3	37	0

Tabelle 2

C5 ($\alpha_P = 0,3$)	Genauigkeit (<i>R</i>)	Trefferquote (<i>P</i>)	<i>F₁</i>-Maß	<i>tp</i>	<i>fp</i>	<i>fn</i>
ESD $\alpha = 0,15$ sum2	0,09	0,33	0,14	1	10	2
ESD $\alpha = 0,10$ sum2	0,10	0,33	0,15	1	9	2
ESD $\alpha = 0,05$ sum2	0	0	0	0	5	3
Z-Score > 3 sum8	0	0	0	0	0	0
Z-Score > 3 sum6	0	0	0	0	0	0
Z-Score > 3 sum4	0,11	0,33	0,17	1	8	2
Z-Score > 3 sum2	0,07	0,33	0,11	1	14	2
Z-Score > 2,5 sum8	0,15	0,67	0,25	2	11	1
Z-Score > 2,5 sum6	0,11	0,67	0,19	2	16	1
Z-Score > 2,5 sum4	0,09	0,67	0,16	2	20	1
Z-Score > 2,5 sum2	0,07	0,67	0,13	2	26	1
Z-Score > 2 sum8	0,06	0,67	0,11	2	33	1
Z-Score > 2 sum6	0,06	0,67	0,11	2	33	1
Z-Score > 2 sum4	0,05	0,67	0,09	2	39	1
Z-Score > 2 sum2	0,05	0,67	0,10	2	37	1

Tabelle 3

C5 ($\alpha_P = 0,6$)						
ESD $\alpha = 0,15$ sum2	0,17	0,33	0,22	1	5	2
ESD $\alpha = 0,10$ sum2	0,20	0,33	0,25	1	4	2
ESD $\alpha = 0,05$ sum2	0	0	0	0	0	0
Z-Score > 3 sum8	0	0	0	0	3	3
Z-Score > 3 sum6	0,25	0,33	0,29	1	3	2
Z-Score > 3 sum4	0,18	0,67	0,29	2	9	1
Z-Score > 3 sum2	0,14	0,67	0,24	2	12	1
Z-Score > 2,5 sum8	0,17	0,67	0,27	2	10	1
Z-Score > 2,5 sum6	0,17	0,67	0,27	2	10	1
Z-Score > 2,5 sum4	0,10	0,67	0,17	2	18	1
Z-Score > 2,5 sum2	0,08	0,67	0,14	2	23	1
Z-Score > 2 sum8	0,05	0,67	0,10	2	37	1
Z-Score > 2 sum6	0,05	0,67	0,09	2	40	1
Z-Score > 2 sum4	0,05	0,67	0,09	2	40	1
Z-Score > 2 sum2	0,05	0,67	0,09	2	39	1

Tabelle 4

C5 ($\alpha_P = 0,9$)						
ESD $\alpha = 0,15$ sum2	0,13	0,33	0,18	1	7	2
ESD $\alpha = 0,10$ sum2	0,14	0,33	0,20	1	6	2
ESD $\alpha = 0,05$ sum2	0	0	0	0	4	3
Z-Score > 3 sum8	0,13	0,33	0,18	1	7	2
Z-Score > 3 sum6	0,29	0,67	0,40	2	5	1
Z-Score > 3 sum4	0,20	0,67	0,31	2	8	1
Z-Score > 3 sum2	0,20	0,67	0,31	2	8	1
Z-Score > 2,5 sum8	0,07	0,33	0,12	1	13	2
Z-Score > 2,5 sum6	0,15	0,67	0,25	2	11	1
Z-Score > 2,5 sum4	0,13	0,67	0,21	2	14	1
Z-Score > 2,5 sum2	0,11	0,67	0,18	2	17	1
Z-Score > 2 sum8	0,08	1,00	0,15	3	33	0
Z-Score > 2 sum6	0,08	1,00	0,15	3	33	0
Z-Score > 2 sum4	0,05	0,67	0,10	2	36	1
Z-Score > 2 sum2	0,06	0,67	0,10	2	34	1

K

Tabelle 1

Iteration i	R_i -Wert	λ_i -Wert	Iteration i	R_i -Wert	λ_i -Wert
1	4,85	3,678	21	3,39	3,670
2	4,49	3,677	22	3,29	3,669
3	4,51	3,677	23	3,26	3,669
4	4,55	3,677	24	3,27	3,669
5	4,59	3,676	25	3,27	3,668
6	3,99	3,676	26	3,23	3,668
7	3,84	3,675	27	3,22	3,667
8	3,83	3,675	28	3,19	3,667
9	3,81	3,675	29	3,19	3,667
10	3,71	3,674	30	3,21	3,666
11	3,70	3,674	31	3,22	3,666
12	3,70	3,673	32	3,22	3,665
13*	3,57	3,673	33	3,24	3,665
14	3,52	3,673	34	3,25	3,665
15	3,53	3,672	35	3,23	3,664
16	3,41	3,672	36	3,16	3,664
17	3,41	3,671	37	3,13	3,663
18	3,43	3,671	38	3,14	3,663
19	3,45	3,671	39	3,15	3,662
20	3,48	3,670	40	3,10	3,662